Goblet cell differentiation in colorectal cancer

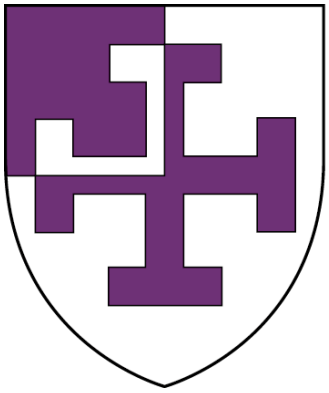
Gulnar Abdullayeva

St Cross College

A thesis submitted to the Division of Medical Sciences, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Oncology

Michaelmas Term, 2023

MRC Weatherall Institute of Molecular Medicine
University of Oxford



DEDICATION

I dedicate this thesis to my amazing parents, my sister, and her family.

Thank you very much for your unconditional love and endless support. This milestone would not have been possible without your guidance, motivation, and belief in me. I want to dedicate this thesis especially to my dear grandparents who unfortunately passed away. I deeply wish that they could be here with me, and I am sure that they would be immensely proud to know about my success in earning my DPhil from the University of Oxford, the most prestigious institution in the world.

ACKNOWLEDGEMENTS

First, I would like to express my endless gratitude to my supervisor, Professor Sir Walter Bodmer for giving me an amazing opportunity to study at the University of Oxford and do my research in the Cancer and Immunogenetics laboratory. I am extremely grateful for his continuous support, help, and guidance throughout my DPhil project. Working under Professor Sir Walter Bodmer's supervision is an immense pride, as well as enriching experience for me. I am deeply inspired his dedication, relentless motivation, discipline, amazing ideas, and expertise to solve difficult problems, all of which shaped me and will continue to forever influence me.

I am extremely thankful to Academician Professor Irada Huseynova and all the colleagues at the Institute of Molecular Biology and Biotechnologies, Ministry of Science and Education of the Republic of Azerbaijan. My scientific achievements are attributed to the support and guidance provided by Academician Professor Irada Huseynova. Her inspiration and unwavering dedication to the science have consistently motivated me. I also express my deepest gratitude to the Islamic Development Bank for providing me with the scholarship and Azerbaijan National Academy of Sciences for covering the bench fees during my DPhil studies.

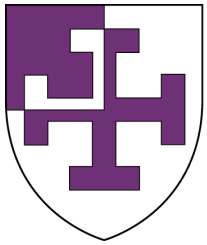
I extend my thanks to all members of Cancer and Immunogenetics laboratory. I am extremely thankful to Dr Viorica Liebe for her outstanding guidance and tremendous support. Thank you very much for your patience in providing detailed answers to all of my questions and always helping me through various situations. I want to sincerely thank to Dr Johanna Michl for guiding me throughout my studies and meticulously reviewing my thesis. Huge thanks to Professor Pawel Swietach, Dr Stefania Monterisi, Dr Shazia Irshad, Dr Jeff (Ta-Chun) Liu, Norah Alrishedan, Dr Luyao Wang, Dr Yang Hu, Wenliang Bai, Yining Sun, Dr Laura Colling, Leticia Campo, Tomas Goncalves, Shamir Montazid, May Sin Ke, Felix Dingler, Kaima Tsukada, Malitha Ratnaweera, and Robert Horsley. Especial thanks to Dr Jeff Liu and James Ashford for providing invaluable assistance with data analysis and for teaching me the basics of R and Phyton programming languages. You are all great scientists and I feel incredibly lucky to have had a privilege to working with each one of you.

I want to acknowledge the honour of being a student at St Cross College. Throughout my time as a Student Vice-President, I had a privilege of working closely with an incredible St Cross team. I owe heart-felt acknowledgements to Carole Souter, Sharon Durno, John Tranter, Dr Joanna Ashbourn, Professor Kevin Marsh, Joanne Russell, and Penny Berry for their tremendous support and help. I would like to sincerely thank Professor Alison Simmons and Professor Ahmed Ahmed for their valuable recommendations and guidance during my transfer and confirmation of status. I extend my deepest gratitude to Professor Marco Novelli for his invaluable help with histological analysis of tissue sections, as well as, for kindly agreeing to be my examiner. I am very grateful to Christoffer Lagerholm, Jana Koth, Lior Pytowski, Cyril Lai, and Rhodri Wilson for their expertise in microscopy and image analysis, Kevin Clark, Craig Waugh and Sally Clark for the expertise in cell sorting. Huge thanks to Dr Philip Hublitz for his guidance and support on site-directed mutagenesis and cloning experiments. I am grateful the opportunity to work with Dr Neil Ashley on single-cell sequencing analysis. My sincere appreciation goes out to the Oncology Facilities team and the Health Safety office for their invaluable contributions.

Special thanks to my dear parents, a sister and her family for their unconditional love, encouragement, and support. Last, but not least, I extend my appreciation to all my friends. I am proud of the dedication, efforts, and hard work I have made for my DPhil journey at the University of Oxford, and I am confident that it will lead me to a bright future.

FUNDING STATEMENT

This study was supported by the Islamic Development Bank (IsDB), Azerbaijan National Academy of Sciences, and the Institute of Molecular Biology and Biotechnologies, Ministry of Science and Education of the Republic of Azerbaijan. I am also grateful for the support from St Cross College.



AUTHOR DECLARATION

I, Gulnar Abdullayeva, hereby declare that the whole work of this thesis is based is my original work at the MRC Weatherall Institute of Molecular Medicine, Department of Oncology, University of Oxford, and that neither the whole work nor any part of it has been, is being, or will be submitted for any other degree at this or any other university. The work is wholly original except as indicated by reference in the text and listed as follows: computational and statistical analysis of single-cell RNA sequencing data (Chapter 5) by Dr Ta-Chun Liu (Bodmer group) and James Ashford (Sahakyan group).

Signed: ..... Date: .....

ABSTRACT

Disruption of goblet cell differentiation, resulting in reduced numbers of goblet cells is one of the major features of colorectal cancer (CRC). Therefore, investigating the regulatory mechanisms associated with goblet cell differentiation is important for a better understanding of CRC and finding new approaches to its prevention and treatment.

The Cancer and Immunogenetics laboratory own a large collection of CRC cell lines, which have been thoroughly examined regarding gene expression and mutation analyses. We assessed MUC2 and TFF3 protein expression levels in 100 CRC cell lines and categorised them into 5 distinct groups. These groups included high, medium, and low MUC2/TFF3-expressing cell lines, cell lines with no/very low MUC2 but high TFF3 expression, and cell lines completely lacking these proteins. MUC2-negative and TFF3-positive cell lines are novel CRC types that have not been described before.

While tracking goblet cell differentiation, we detected that TFF3 is expressed earlier than MUC2, possibly due to its faster folding and low molecular weight. It is also needed for holding mucin together during secretion. The proportion of MUC2/TFF3-positive cells was much higher in single-derived LS180 colonies compared to bulk culture.

We also investigated the effects of the main transcription factors on goblet cell differentiation. Downregulation of ATOH1, SPDEF, CDX1, and CDX2 led to a significant reduction in MUC2/TFF3-positive cells in multiple CRC cell lines. Among these factors, ATOH1 had a more substantial effect compared to others. But it is still not enough to induce MUC2/TFF3 expression in completely negative CRC cell lines. Suppression of methylation had variable effects on different CRC cell lines. We suggest that in the group 4 cell lines, MUC2 is most likely methylated as opposed to other group cell lines that might be regulated by demethylation of transcription factors, such as ATOH1 and HES1.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

5-AZA-CdR	5-Aza-2′-deoxycytidine
5-FU	5-fluorouracil
AGR2	Anterior gradient protein 2, protein disulphide isomerase family member
AKT	Protein kinase B
ALPI	Intestinal alkaline phosphatase
ALPP	Placental alkaline phosphatase
AMPC	2-amino-4-(4-(6-fluoro-5-methylpyridin-3-yl)phenyl)-5-oxo-4H,5H-pyrano[3,2-c]chromene-3-carbonitrile
APC	Adenomatous polyposis coli
AQP8	Aquaporin 8
ASCL2	Achaete-scute family BHLH transcription factor 2
ATCC	American Tissue Culture Collection
ATOH1	Atonal BHLH transcription factor 1
ATOH1	Atonal BHLH transcription factor 1
AXIN2	Axin 2
BCA	Bicinchoninic acid
BEST2	Bestrophin 2
BEST4	Bestrophin 4
BMP	Bone morphogenetic protein
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
c-Myc	MYC Proto-oncogene, BHLH transcription factor
C19orf33	Chromosome 19 open reading frame 33
CA1	Carbonic Anhydrase 1
CBFA2T3	CBFA2/RUNX1 partner transcriptional co-repressor 3
CCNA2	Cyclin A2
CD147	Cluster of differentiation 147
CD52	CD52 molecule
CDC27	Cell division cycle
CDCA7	Cell division cycle associated 7
cDNA	Complementary DNA
CDX1	Caudal type homeobox 1
CDX2	Caudal type homeobox 2
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1
CEACAM6	Carcinoembryonic antigen-related cell adhesion molecule 6
CEACAM7	Carcinoembryonic antigen-related cell adhesion molecule 7
CFTR	Cystic fibrosis transmembrane conductance regulator
CHGA	Chromogranin A
CIMP	CpG island methylation phenotype
CIN	Chromosomal instability
CK	Cysteine-knot

CLCA1	Calcium-activated chloride channel regulator 1
CLCA1	Calcium-activated chloride channel regulator 1
CLCA1	Calcium-activated chloride channel regulator 1
CLRN3	Clarin 3
CMS	Consensus molecular subtype
CRC	Colorectal cancer
CREB3L1	CAMP responsive element binding protein 3 like 1
Ct	Cycle threshold
CXCR4	C-X-C chemokine receptor type 4
CXCR7	C-X-C chemokine receptor type 7
DAB	3,3'-diaminobenzidine
DAM	Donkey anti-mouse
DAPI	4′,6-diamidino-2-phenylindole
DAPT	N-[N-(3,5-difluorophenacetyl)-l-alanyl]-S-phenylglycine t-butyl ester
DAR	Donkey anti-rabbit
DBZ	Dibenzazepine
DKFZ	Deutsches Krebsforschungszentrum
DLL1	Delta-like 1
DLL3	Delta-like 3
DLL4	Delta-like 4
DMBT1	Deleted in malignant brain tumour 1
DMSO	Dimethyl sulfoxide
DSMZ	German Collection of Microorganisms and Cell Cultures GmbH
DTT	Dithiothreitol
E-cadherin	Epithelial cadherin
ECACC	European Collection of Authenticated Cell Cultures
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
EpCAM	Epithelial cell adhesion molecule
EPHB2	EPH receptor B2
ERK	Extracellular signal-regulated kinase
ETS2	ETS proto-oncogene 2, transcription factor
FAM3D	FAM3 metabolism regulating signaling molecule D
FAP	Familial adenomatous polyposis
FCGBP	Fc Gamma Binding Protein
FCGBP	Fc gamma binding protein
FEV	FEV transcription factor ETS family member
FFAR4	Free fatty acid receptor 4
FFPE	Formalin-fixed paraffin-embedded
FOXM1	Forkhead box M1
Fu	Fucose

g	Gram
Gal	Galactose
GalNAc	N-acetylgalactosamine
GALNT8	Polypeptide N-acetylgalactosaminyltransferase 8
GAM	Goat-anti mouse
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GAPs	Goblet cell-associated antigen passages
GAR	Goat-anti rabbit
GATA6	GATA binding protein 6
GD/PH	Gly-Asp-Pro-His
GlcNAc	N-acetylglucosamine
GMM	Gaussian mixture model
GSK3β	Glycogen synthase kinase 3 beta
GSTP1	Glutathione S-transferase Pi 1
GUCA2A	Guanylate cyclase activator 2A
GUCA2B	Guanylate cyclase activator 2B
H&E	Haematoxylin and Eosin
HEPH	Hephaestin
HES1	Hairy and enhancer of split 1
HES5	Hairy and enhancer of split 5
HEY1	Hes related family BHLH transcription factor with YRPW motif 1
HEY2	Hes related family BHLH transcription factor with YRPW motif 2
HOPX	HOP homeobox
HRP	Horseradish peroxidase
IFABP	Intestinal fatty acid binding protein
IHC	Immunohistochemistry
IL33	Interleukin 33
IMS	Industrial methylated spirit
INSM1	INSM transcriptional repressor 1
ITLN1	Intelectin 1
JAG1	Jagged 1
JAG2	Jagged 2
JCRB	Japanese Collection of Research Bioresources
KCLB	Korean Cell Line Bank
KLK1	Kallikrein 1
KRAS	Kirsten rat sarcoma viral oncogene homolog
KRT20	Cytokeratin 20
LB	Lysogeny broth
LEF1	Lymphoid enhancer-binding factor 1
LGALS4	Galectin 4
LGALS9C	Galectin 9C
LGR5	Leucine rich repeat containing G protein-coupled receptor 5

LINGO2	Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 2
LRIG1	Leucine rich repeats and immunoglobulin like domains 1
LSP1	Lymphocyte specific protein 1
LYPD8	LY6/PLAUR Domain Containing 8
M	Molar
MAML-1	Mastermind Like Transcriptional Coactivator 1
MAPK	Mitogen-activated protein kinase
MCM5	Minichromosome maintenance complex component 5
MECOM	MDS1 and EVI1 complex locus
MEX3A	Mex-3 RNA binding family member A
mg	Milligram
min	Minutes
MKI67	Marker of proliferation Ki-67
ml	millilitre
MS4A12	Membrane spanning 4-domains A12
MSI	Microsatellite instability
MSI1	Musashi RNA binding protein 1
MT-ND4L	Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4L
MUC2	Mucin 2
MYBL2	MYB proto-oncogene like 2
NEB	New England Biolabs
NEUROD1	Neuronal differentiation 1
NEUROG3	Neurogenin 3
ng	Nanogram
NICD	Notch intracellular domain
nm	Nanometre
nM	Nanomolar
nmol	Nanomole
NUSAP1	Nucleolar and spindle associated protein 1
OLFM4	Olfactomedin 4
ORF	Open reading frame
OTOP2	Otopetrin 2
PABPC1	Poly(A) binding protein cytoplasmic 1
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PFA	Paraformaldehyde
PHGR1	Proline, histidine and glycine rich 1
PI3K	Phosphatidylinositol-3-kinase
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha

PTK7	Protein tyrosine kinase 7
PTPRC	Protein Tyrosine Phosphatase Receptor Type C
PVDF	Polyvinylidene difluoride
qRT-PCR	Quantitative real-time polymerase chain reaction
RCB	Research Cell Bank
REG4	Regenerating family member 4
REG4	Regenerating islet-derived family member 4
RELMβ	Resistin-like molecule β
RETNLβ	Resistin like beta
RGMB	Repulsive guidance molecule BMP co-receptor B
RPBJ (CSL or CBF1)	Recombination signal binding protein for immunoglobulin kappa J region
RQ	Relative quantification
RT	Room temperature
SA	Sialic acid
SCNN1B	Sodium channel epithelial 1 subunit beta
scRNA-seq	Single-cell RNA sequencing
SDS	Sodium dodecyl sulphate
sec	Seconds
SLC12A8	Solute carrier family 12 member 8
SLC26A3	Solute carrier family 26 member 3
SLC9A2	Solute carrier family 9 member A2
SMAD4	SMAD family member 4
SMOC2	SPARC related modular calcium binding 2
SOC	Super optimal broth with catabolite repression
SPDEF	SAM pointed domain containing ETS transcription factor
SPINK4	Serine peptidase inhibitor Kazal type 4
SRB	Sulforhodamine B
ST3GAL4	ST3 beta-galactoside alpha-2,3-sialyltransferase 4
ST6GALNAC1	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 1
STAT3	Signal transducer and activator of transcription 3
TAE	Tris-Acetate-EDTA
TCGA	The Cancer Genome Atlas
TFF	Trefoil factor
TFF3	Trefoil factor 3
TGF-β	Transforming growth factor-beta
TMEM37	Transmembrane protein 37
TOP2A	DNA topoisomerase II alpha
TOX3	TOX high mobility group box family member 3
TP53	Tumour protein p53
TSNE	T-distributed stochastic neighbour embedding
TYMP	Thymidine phosphorylase
UMAP	Uniform manifold approximation and projection

UMIs	Unique molecular identifier
VWD	von Willebrand D
WDHD1	WD repeat and HMG-box DNA binding protein 1
ZG16	Zymogen granule protein 16
µg	Microgram
µl	Microliter
µm	Micrometre
µM	Micromolar

CHAPTER 1

INTRODUCTION

CHAPTER 1: INTRODUCTION

1.1 Large intestine: anatomy and function

The large intestine consists of the colon and rectum and has a length of approximately 150 to 180 cm in adults (Baran et al., 2018; Siri et al., 2020). Its segments include the cecum, appendix, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, S-shaped sigmoidal colon, rectum, and anus (Kahai. et al., 2022). The right-sided colon refers to the first four parts and the upper two-thirds of the transverse colon, whereas the remaining sections belong to the left-sided colon (Baran et al., 2018). The large intestine plays a key role in the absorption of water, electrolytes, and vitamins from waste products, as well as storing faeces prior to their elimination. It also acts as a host for commensal microbiota which metabolises partially absorbed energy substrates and produces short-chain fatty acids and vitamins (Milla, 2009; Okumura & Takeda, 2017).

The large intestinal wall is composed of four layers, namely mucosa, submucosa, muscularis propria, and serosa (Siri et al., 2020). Mucosa is the inner layer of the intestinal wall and comprises the epithelium, lamina propria, and muscularis mucosa (a muscle layer). It is lined with epithelial cells that form a tubular configuration known as a crypt. In contrast to the small intestine, the colon lacks villi structures which are small finger-shaped projections which increase the surface area of the small intestine (Kong et al., 2018). The submucosa is a layer of connective tissue surrounding the mucosa. The large intestine is mainly supplied with blood and lymph vessels through this section (Kachlik et al., 2010; Siri et al., 2020). The muscularis propria is made up of two types of muscle layers, including longitudinal and circular muscle layers interacted via nerves (Washabau & Day, 2013). These muscle layers enable the

passage of food from the proximal to the distal direction by peristaltic movements. The outer surface of the intestine is covered by a protective layer known as the serosa which includes squamous epithelial cells and the mesothelium (Siri et al., 2020).

1.2 Colonic crypt structure and stem cell niche

The large intestine, or colon, contains lamina propria, connective tissue, and multiple crypts. A colonic crypt is a tubular invagination that forms the main functional unit of the epithelium. It consists of distinct types of epithelial cells, surrounded by myofibroblasts and an extracellular matrix (Figure 1.1).

Stem cells control intestinal integrity by self-renewal and differentiate into mature cells. They are located at the bottom of the crypts and can give rise to progenitors. Following several cycles of division, progenitors are able to differentiate into either absorptive or secretory lineages (Figure 1.1) (Noah et al., 2011). The enterocytes are absorptive cells, whereas the mucin-secreting goblet and hormone-producing enteroendocrine cells are considered the secretory lineages. As the cells migrate to the top of the crypts, they undergo terminal differentiation and are shed into the lumen. The colonic epithelium is the most rapidly regenerating tissue in mammals with a turnover time of 3-5 days (Dame et al., 2014).

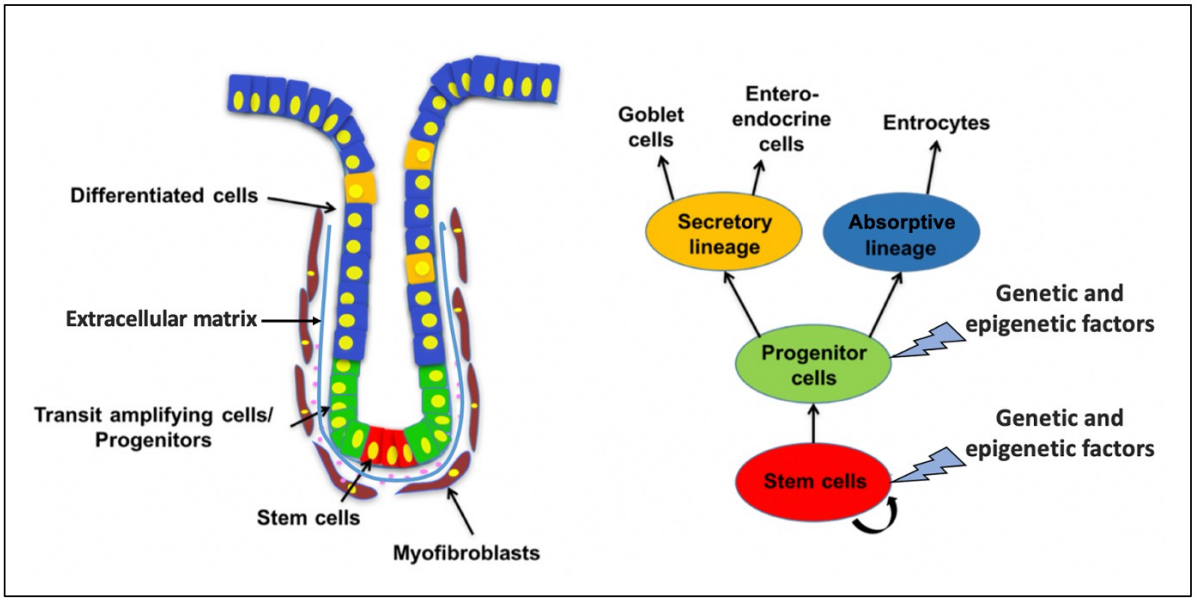


Figure 1.1 Colonic crypt microarchitecture and stem cell differentiation. Colonic stem cells are located at the bottom of the crypt. Progenitors give rise to absorptive and secretory lineages. The figure of the colonic crypt structure was modified from Haoyu Liu's DPhil thesis (2017) in the Bodmer lab.

In the small intestine, stem cells are able to differentiate into five various cell types, namely enterocytes, goblet cells, tuft cells, enteroendocrine cells, and Paneth cells. Paneth cells are the only specialised cell type in the intestine that move to the basal part of the crypt after differentiating from stem cells (Lueschow & McElroy, 2020). Their main function is to provide antimicrobial peptides and proteins, including α-defensins, lysozymes, angiogenin-4, and cryptdin-related sequence peptides. These proteins, alongside mucus, form a protective barrier to prevent bacteria from accessing the intestinal epithelium (Wallaeys et al., 2023). Paneth cells are normally present throughout the right colon into but not beyond the transverse colon but may be seen in the left colon following longstanding mucosal injury (e.g. IBD) (Elphick & Mahida, 2005; Simmonds et al., 2014). This could be explained by the protective role of

metaplastic Paneth cells against bacterial invasion in damaged colonic epithelium (Elphick & Mahida, 2005).

Intestinal stem cells are protected by the surrounding cellular and physical niches (Meran et al., 2017). The physical niche comprises the extracellular matrix, a fibrous network of structural proteins that preserve the three-dimensional organisation of the intestine. The cellular niche is formed by various types of stromal cells, such as fibroblasts, pericryptal myofibroblasts, immune cells, endothelial cells, and neural cells which are embedded in the extracellular matrix. Intestinal stem cells and niches interact with each other to maintain a balance between cell proliferation and differentiation. This communication is mediated by several proteins and signalling pathways, such as Wnt/β-catenin, Notch, Hedgehog, Phosphatidylinositol-3-kinase (PI3K)/Protein kinase B (Akt), and Transforming growth factor-beta (TGF-β)/bone morphogenetic protein (BMP) pathways (Gjorevski & Ordonez-Moran, 2017; Kim & Khan, 2013; Meran et al., 2017).

1.3 Intestinal goblet cells

1.3.1 Goblet cells and mucus organisation

Goblet cells are one of the three differentiated epithelial cell types within the large intestine and are named for their typical goblet- or cup-like appearance. The nucleus, Golgi apparatus, endoplasmic reticulum, and mitochondria are localised at the base of the cell, while the remaining part of the cytoplasm is filled with mucin granules. The proportion of goblet cells increases from the duodenum (4%) to the distal colon (16%), corresponding to a rise in microorganisms in the same direction (Paone & Cani, 2020; Zhang & Wu, 2020).

Goblet cells are responsible for synthesizing and secreting mucins which are crucial components of the mucus layer. Mucus is found in different parts of the body, such as the ocular surface, upper and lower respiratory tracts, gastrointestinal system, and reproductive organs (Bansil & Turner, 2018). Mucus is made of water, mucins, electrolytes, lipids, ions, components of the immune system, anti-microbial peptides, and other proteins (Kim & Khan, 2013; Paone & Cani, 2020). Additional to mucins, goblet cells produce and release a variety of other compounds, such as trefoil factors (*TFFs*), Fc gamma binding protein (*FCGBP*), Calcium-activated chloride channel regulator 1 (*CLCA1*), and Resistin-like molecule β (*RELMβ*) which contribute to the composition of intestinal mucus (Gustafsson & Johansson, 2022; Yang & Yu, 2021).

The mucus layer coats epithelial cells and protects them against pathogens, toxins, and allergens. Mucus has lubricating and moisturizing properties that prevent epithelial cells from dehydrating and being mechanically stressed during peristalsis. It also allows nutrients, ions, and other small molecules to penetrate through the layer and access the enterocytes (Paone & Cani, 2020). In the human body, a thin layer of mucus also covers the stool, which facilitates its removal and minimizes mechanical stress (Gustafsson & Johansson, 2022; Kamphuis et al., 2017).

The colon has one inner and one outer mucus layer, whereas the small intestine has only one mucus layer. The inner layer is attached to the colonic epithelium and impenetrable to bacteria, whereas the outer mucus layer contains commensal microbiota (Hansson, 2020). The inner mucus layer undergoes volume expansion by endogenous proteases, causing its transformation into the outer mucus layer.

Goblet cells were also described to have an immunological role in antigen sampling and delivery. A distinct subtype of goblet cell creates goblet cell-associated antigen

passages (GAPs) (Gustafsson et al., 2021) and presents luminal substances to $CD103^{+}CD11c^{+}$ dendritic cells in the lamina propria (Johansson & Hansson, 2016). These dendritic cells interact with T cells, triggering an immune response. According to the study by Gustafsson et al., GAP formation is not compulsory for the maintenance of mucus secretion or mucus barrier integrity (Gustafsson et al., 2021). Instead, GAPs serve as enhanced transport machinery that allows the delivery of luminal antigens via endocytosis.

Another distinct subset of goblet cells, called sentinel goblet cells, are able to internalise microbe-associated molecular patterns that penetrate the inner layer (Birchenough et al., 2016; Gustafsson & Johansson, 2022). They trigger a cascade of events, leading to mucus secretion from surrounding goblet cells via intercellular gap junctions. This coordinated response washes away pathogenic bacteria and prevents severe damage to colonic crypts. Once the sentinel goblet cells release their contents, they are expelled into the lumen.

1.3.2 Mucins and MUC2

Mucins are large, complex, and viscoelastic glycoproteins that are the major building components of mucus. Various mucin types were distinguished according to their structure and localization. Mucins can be classified into transmembrane or secreted, with the latter being further subdivided as gel-forming or non-gel-forming mucins. Transmembrane or membrane-bound mucins include MUC1, MUC3A/3B, MUC4, MUC12, MUC13, MUC14, MUC15, MUC16, MUC17, MUC20, MUC21 and MUC22 (Dhanisha et al., 2018; Song et al., 2023). Secreted gel-forming mucins are MUC2,

MUC5AC/5B, MUC6, and MUC19, whereas MUC7, MUC8, and MUC9 belong to secreted non-gel-forming mucins (Zhang & Wu, 2020).

Transmembrane mucins are typically found in the apical part of epithelial cells and contain an extracellular N-terminal domain, mucin domains, a transmembrane domain, and a short cytoplasmic C-terminal tail. They play a significant role in building a protective glycocalyx coat across the intestinal epithelium. Additionally, transmembrane mucins are actively involved in cell signalling and sensing processes (Grondin et al., 2020; Johansson et al., 2013).

Gel-forming mucins are produced and secreted by goblet cells. Mucin 2 (MUC2) is the first secretory mucin identified in humans and is the most abundant mucin in the intestine (Kim & Ho, 2010). It has a crucial role in gut homeostasis by contributing to mucus production and preventing pathogen invasion. MUC2 deletion or deficiency in mice results in tumour development and chronic inflammation (Gundamaraju & Chong, 2021).

MUC2 contains a protein core coated with linear and branched oligosaccharide chains. A MUC2 monomer has over 5000 amino acids and its central part is enriched with tandem repeats, predominantly composed of proline, threonine, and serine amino acids (known as PTS domains) (Figure 1.2). O-glycosylation occurs in the hydroxyl groups of threonine and serine residues, whereas proline holds the unfolded structure of MUC2, facilitating the initial glycosyltransferase to access its correct sites during the O-glycosylation process (Liu et al., 2023). There are five sugars, including N-acetylgalactosamine (GalNAc), galactose (Gal), N-acetylglucosamine (GlcNAc), fucose (Fu), and sialic acid (SA), commonly observed in mucin oligosaccharides and linked to the protein core though covalent bonds (Kim & Khan, 2013).

Figure 1.2 Domain structure of MUC2. MUC2 contains cysteine-rich N- and C-terminal regions as well as three complete (D1-D3) and one partial (D') von Willebrand D (VWD) domains located at the N-terminus and one VWD4 domain at the C-terminus. The central PTS domains are separated by the CysD domains. CK: cysteine-knot; GD/PH: Gly-Asp-Pro-His. The figure was adapted from Johansson et al. (2011).

MUC2 undergoes different modifications prior to being packed into secretory mucin granules (Figure 1.3). It is co-translationally translocated to the endoplasmic reticulum where it acquires N-glycosylation and C-terminal dimerisation in Cys-knot domains via disulphide bonds (Dhanisha et al., 2018; Hansson, 2019). MUC2 is then exported to the *cis*-Golgi network where it is O-glycosylated and the N-glycans are further modified. O-linked glycosylation of the PTS or mucin repeat domains starts with the addition of GalNAc to the hydroxyl groups of threonine or serine residues via the peptidyl-GalNAc transferases, resulting in the formation of the Tn antigen (Bergstrom & Xia, 2013; Hansson, 2020). Once GalNAc is covalently attached to the protein core, it serves as a substrate for other glycosyltransferases. Various combinations of glycans give MUC2 its bottle brush-like appearance and water-binding capacity, enabling it to form a gel. In the *trans*-Golgi network, a large and net-like MUC2 structure is arranged via disulphide-bonded trimers in the VWD3 domain of the N-terminal regions and localised inside the secretory granules (Ambort et al., 2012; Liu et al., 2023).

MUC2 is located within the secretory vesicles prior to secretion, and this process requires an increase in Ca^{2+} concentrations and a reduction in pH. When MUC2 is

released by goblet cells, it undergoes significant volume expansion (100-1000 times), which entails lower Ca^{2+} concentrations and higher pH (Figure 1.3) (Birchenough et al., 2015; Schroeder, 2019).

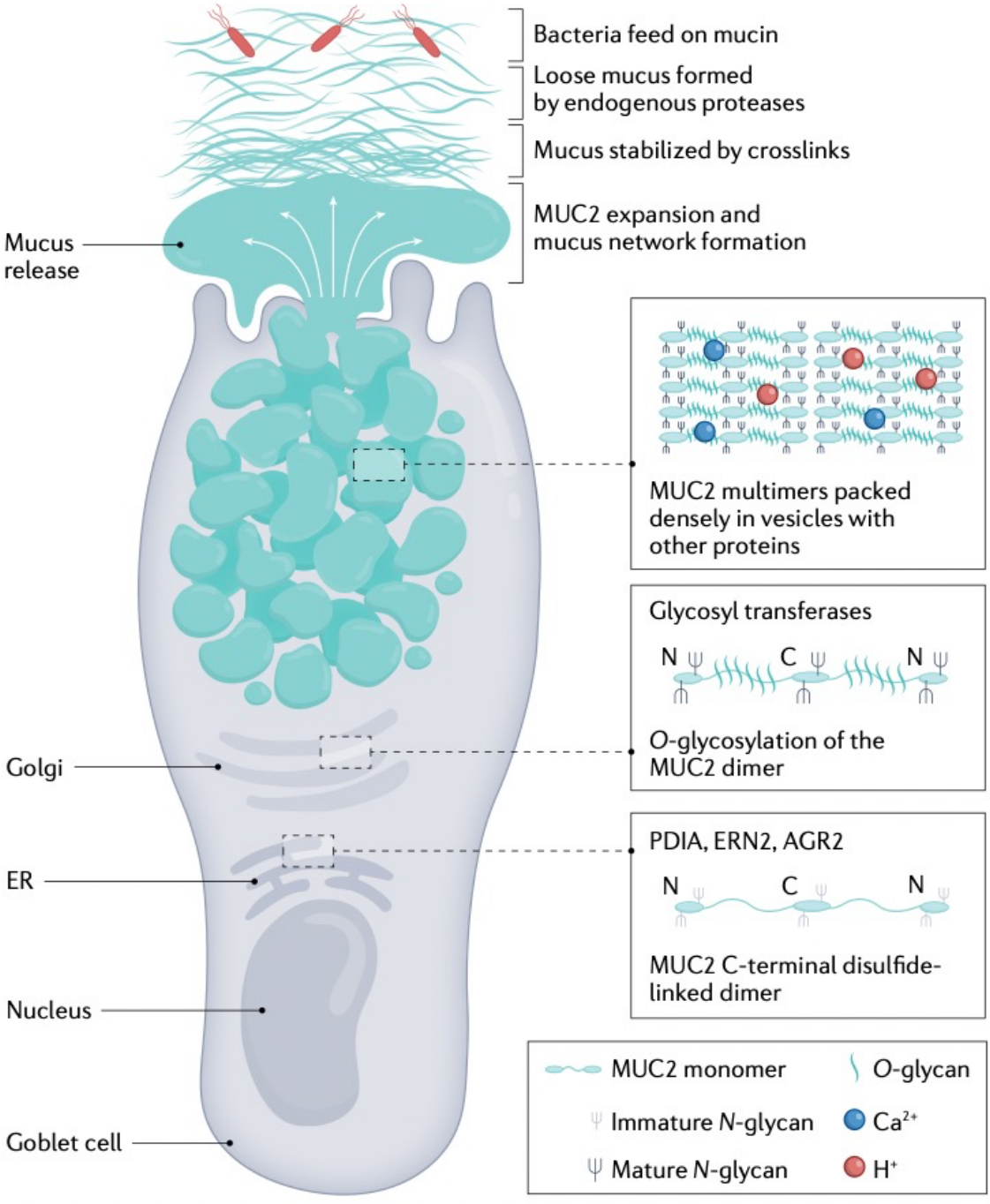


Figure 1.3 Biosynthesis and secretion of MUC2. MUC2 involves C-C dimerization and N glycosylation in the endoplasmic reticulum. Then, it is heavily O-glycosylated in the Golgi

apparatus before being packed in the secretory vesicles. Following secretion, the volume expands and MUC2 is stabilised by cross-links. The figure was adapted from Gustafsson and Johansson (2022) with permission from Springer Nature under the licence number: 5557120891474.

Mucins are secreted into the lumen either through basal or compound exocytosis. Basal or constitutive mucin secretion leads to the continuous release of individual mucin granules during normal physiological states. It allows the thickness of the mucus layer to remain constant. During compound or stimulated mucin secretion, most of the secretory vesicles release their contents into the lumen under exposure to various stimuli, such as carbachol, histamine, and acetylcholine (Birchenough et al., 2015; Yang & Yu, 2021).

1.3.3 TFF3

The trefoil factor (*TFF*) peptides comprise six cysteine residues connected via disulphide bridges (Aihara et al., 2017). This structure forms a clover-leaf-shaped trefoil or P-domain, which is conserved among the various types of TFFs, enabling them to be more resistant to degradation. Three types of trefoil factors have been identified in mammals, namely TFF1, TFF2, and TFF3 (Aamann et al., 2014; Aihara et al., 2017). TFF1 and TFF3 each contain one trefoil domain, while TFF2 is characterised by having two trefoil domains. Both TFF1 and TFF3 carry an additional cysteine residue in the C-terminal regions which allows them to interact with TFF peptides or various other proteins. TFF1 and TFF2 are primarily detected in the stomach, whereas TFF3 is highly expressed in the human small intestine and colon (Madsen et al., 2007; Podolsky et al., 1993; Ribieras et al., 1998).

TFFs have a vital role in promoting epithelial regeneration and mucosal repair within the gastrointestinal tract. The TFF peptides are connected to mucins in various regions of the gastrointestinal tract. TFF1, TFF2, and TFF3 colocalize with MUC5AC, MUC6, and MUC2, respectively (Aihara et al., 2017). It has been shown that TFFs exhibit lectin-like activity by recognizing the GlcNAc-α-1,4-Gal disaccharide which enables them to non-covalently cross-link with mucin glycoproteins (Jarva et al., 2020). Additional to its non-covalent interactions, TFF3 also forms a heterodimer with FCGBP, which is then covalently linked to the MUC2 mucin (Albert et al., 2010).

TFF3 protein contains 59 amino acids and its molecular weight is around 6.6 kD for a monomer and 13 kD for a dimer (Kjellev, 2009). It is produced by intestinal goblet cells and released into the luminal surface together with mucin glycoproteins (Podolsky et al., 1993). TFF3, as a secreted protein, was demonstrated to bind to specific receptors and stimulate intracellular signalling pathways, including signal transducer and activator of transcription 3 (*STAT3*) (Rivat et al., 2005), the mitogen-activated protein kinase (*MAPK*)/Extracellular signal-regulated kinase (*ERK*) (Chen et al., 2019), and PI3K/AKT (Taupin et al., 2000).

Several binding partners have been determined for TFF3 in different organs. In human ocular surface tissue and conjunctival epithelial cell lines, C-X-C chemokine receptor type 4 (*CXCR4*) and CXCR7 dimers regulate cell migration in a TFF3-dependent manner, suggesting they might be potential TFF3 binding partners (Dieckow et al., 2016). Moreover, TFF3 was described to interact with the cluster of differentiation 147 (*CD147*) receptor and promote cell invasion, proliferation, and migration in several CRC cell lines (Cui et al., 2021).

Upon injury, TFF3 was shown to bind to leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 2 (*LINGO2*), thereby separating LINGO2 from the epidermal growth factor receptor (*EGFR*) (Belle et al., 2019). In turn, it leads to upregulated EGFR activity, which promotes mucosal barrier function, enhances immune responses against gastrointestinal parasites and reduces colitis. LINGO2 is also needed for TFF3-induced STAT3 activation in intestinal cells which plays an anti-apoptotic role (Belle et al., 2019). Moreover, human deleted in malignant brain tumour 1 (*DMBT1*), an upregulated protein during inflammatory bowel diseases, was shown to also interact with TFF3 dimers (Madsen et al., 2013). It indicates the role of TFF3 in mucosal defence and innate immunity.

TFF3 affects epithelial integrity, cell migration, and mucosal healing during injury. An elevated level of TFF3 was detected in ulcerative colitis patients, and it was associated with the disease's biochemical and clinical characteristics (Gronbaek et al., 2006). When mice were treated with dextran sulphate sodium, a substance that mimics the condition of human ulcerative colitis, TFF3 knockout mice displayed more severe colonic damage as compared to wild-type mice (Mashimo et al., 1996). In TFF3-deficient mice, although the colon showed a normal crypt structure, epithelial cell proliferation (Mashimo et al., 1996) and apoptosis were raised (Taupin et al., 2000). Furthermore, the migration of intestinal crypt cells towards the epithelium lining, as a part of cell turnover was significantly delayed, suggesting a potential role of TFF3 in the gut homeostasis (Mashimo et al., 1996).

TFF3 expression was upregulated in various cancer types, including colon, gastric, prostate, breast, hepatocellular, and pancreatic carcinomas (Busch & Dünker, 2015) and its overexpression was associated with the promotion of cancer growth (Yang et

al., 2022). Administration of recombinant TFF3 in HT29 and HCT116 CRC cell lines resulted in cell detachment and reduced epithelial (E)-cadherin-mediated intercellular adhesion (Efstathiou et al., 1998). TFF3 overexpression in HT29 cells led to cell growth and migration, upregulated Vimentin, Twist1, and Snail, and downregulated E-cadherin expression, indicating its involvement in cancer progression through epithelial-mesenchymal transition (EMT) (Yusufu et al., 2019). It should be noted that there are several contradictory reports, leading to uncertainties regarding the function of TFF3 in cancer development (Uchino et al., 2000; Uchino et al., 1999). Another study with HCT116 CRC cell line and a rat intestinal epithelial cell line (IEC-6) showed that exogenous TFF3 prevented cells from apoptosis (Taupin et al., 2000). Blocking of TFF3 activity by 2-amino-4-(4-(6-fluoro-5-methylpyridin-3-yl)phenyl)-5-oxo-4H,5H-pyrano[3,2-c]chromene-3-carbonitrile (AMPC) or a siRNA-mediated approach lowered TFF3-driven oncogenic behaviour and stimulated apoptosis in several CRC and mammary carcinoma cell lines (Chen et al., 2019; Pandey et al., 2022).

1.3.4 Other mucus components

Apart from MUC2 and TFF3, goblet cells secrete other mucus components. Here we provide a brief overview of FCGBP, ZG16, and CLCA1 mucus components.

Human Fc gamma binding protein (*FCGBP*) consists of 5405 amino acids with a molecular weight of around 500 kDa (Q. Liu et al., 2022). It is considered a mucin-like protein due to its typical mucin characteristics, such as a high molecular weight, cysteine-rich regions, extensive glycosylation, intramolecular disulphide bonds, production, and secretion by goblet cells. FCGBP forms heterodimers with MUC2 (Johansson et al., 2009) and TFF3 (Albert et al., 2010) and prevents bacterial

attachment to mucosal surfaces and alleviates their removal. However, Ehrencrona et al. were unable to verify the covalent interaction between FCGBP and MUC2 (Ehrencrona et al., 2021). The predominant existence of FCGBP in the soluble fraction of the intestinal mucus layer, along with MUC2 being mainly found in the insoluble component, implies that the majority of secreted FCGBP does not covalently bind to MUC2 (Ehrencrona et al., 2021).

According to several studies, FCGBP and MUC2 levels diminish along with the progression of colorectal cancer (Al-Khayal et al., 2016; Mizoshita et al., 2007; Yuan et al., 2021), thereby leading to an impairment of epithelial integrity and facilitating metastasis. Downregulation of FCGBP correlated with shorter survival time in the primary CRC (Zhuang et al., 2021) and CRC with liver metastases (Yuan et al., 2021). MUC2, FCGBP, and other mucin components also declined in patients with ulcerative colitis (van der Post et al., 2019).

Zymogen granule protein 16 (*ZG16*) is a 16 kDa lectin-like protein excreted by goblet cells and is one of the core components of the mucus layer (Bergstrom et al., 2016). It functions as a bacterial aggregator by binding to the peptidoglycan of Gram-positive bacteria's cell walls (Bergstrom et al., 2016). This cooperation along with the inner mucus layer holds bacteria at a far distance from colonic epithelial cells, thus retaining the integrity of the mucosal barrier. Moreover, ZG16 expression was reduced in CRC patients, which was associated with poor overall and progression-free survivals (Chen et al., 2016; Meng et al., 2018).

Calcium-activated chloride channel regulator 1 (*CLCA1*) is another secreted protein that is found in the intestinal mucosa. Regardless of its ion current activity, CLCA1 protein acts as a metalloprotease by possibly cleaving the N-terminal part of MUC2,

leading to the expansion of mucus structures (Nyström et al., 2019). This process promotes the transition from the inner mucus layer to the outer mucus layer (Nyström et al., 2018). A study by Li et al. revealed that CLCA1 levels significantly declined in CRC patients compared to healthy individuals (Li et al., 2017). Additionally, the study found that CLCA1 expression showed a negative correlation with CRC progression and metastasis (Li et al., 2017).

1.4 Study of factors involved in goblet cell differentiation

1.4.1 Notch pathway

Notch signalling plays a substantial role in the regulation of cell proliferation and differentiation in the intestine. The Notch signalling pathway begins via interactions between Notch ligands and Notch receptors in neighbouring cells (Figure 1.4). In humans, there are four transmembrane Notch receptors (*Notch 1*, *2*, *3*, and *4*) and five Notch ligands (Jagged (*JAG*) 1, 2 and Delta-like (*DLL*) 1, 3, 4) (Totaro et al., 2018; Zhou et al., 2022). The Notch intracellular domain (*NICD*) Is released from the cell membrane by two consecutive proteolytic events mediated by metalloprotease-disintegrins and γ-secretase (complex of presenilin-1 and nicastrin). The NICD translocates into the nucleus, where it interacts with recombination signal binding protein for immunoglobulin kappa J region (*RPBJ*) (also known as *CSL* or *CBF1*) and transcriptional coactivator, Mastermind like transcriptional coactivator 1 (*MAML-1*) (Liang et al., 2019). This complex activates the transcription of Notch target genes, including the hairy and enhancer of split (*HES*) 1, HES5, Hes related family BHLH transcription factor with YRPW motif (*HEY*) 1, and HEY2 (Shaik et al., 2020).

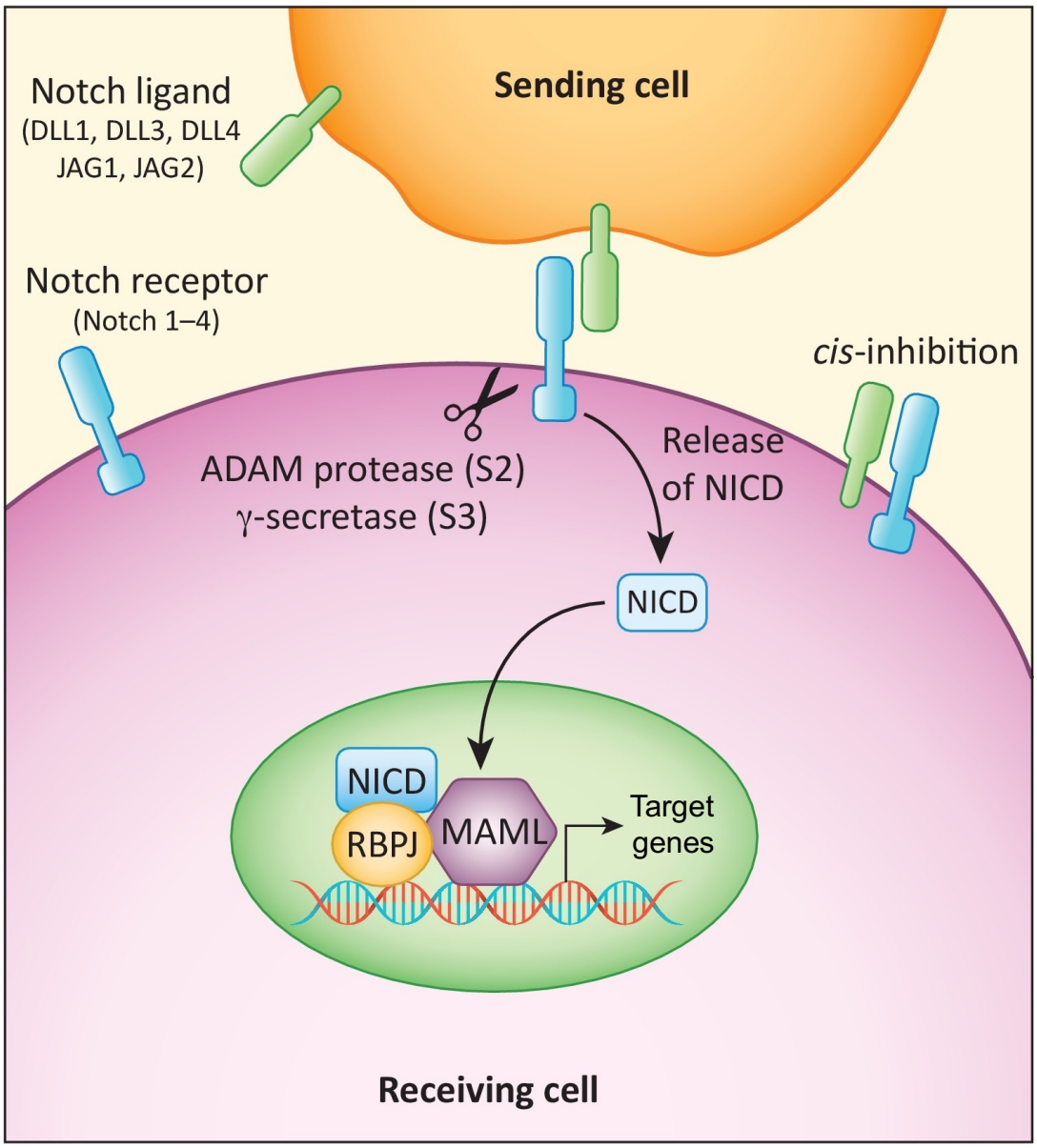


Figure 1.4 Notch signalling pathway. Following interaction with Notch ligands, Notch receptors undergo cleavages by metalloprotease (S2) and γ-secretase (S3), leading to a release of Notch intracellular domain (*NICD*). NICD transfers to the nucleus where it recruits transcriptional co-activators to activate Notch target genes. The figure was adapted from Totaro et al. (2018) from Trends in Cell Biology under the licence number: 5584531433459.

It was described that the Notch target gene, Hairy and enhancer of split 1 (HES1), is overexpressed in both Notch-dependent and Notch-independent manners during

murine and human intestinal carcinogenesis (Peignon et al., 2011). The Notch-dependent way leads to high expression of Notch ligands and receptors, whereas the independent process occurs via direct activation of HES1 promoter through β-catenin/TCF signalling pathway, suggesting a complex crosstalk between Wnt/β-catenin and Notch pathways (Peignon et al., 2011). Another study also reported Notch as a downstream pathway of Wnt signalling through an interaction between β-catenin and JAG1, as well as defining two TCF-binding binding sites in the JAG1 promoter region (Rodilla et al., 2009).

The blocking of Notch activity in mice causes a reduction of crypt base columnar stem cells, decreased cell proliferation and programmed cell death in the small intestine (VanDussen et al., 2012). Mice treated with γ-secretase inhibitor (Dibenzazepine, DBZ) exhibited upregulation of markers associated with enteroendocrine, tuft, goblet, and Paneth cells, implying that inactivation of Notch signalling induces the overall expansion of secretory cell populations within the intestine (VanDussen et al., 2012). Similar results were obtained in the LS174T cell line, as blocking of the Notch pathway via (N-[N-(3,5-difluorophenacetyl)-l-alanyl]-S-phenylglycine t-butyl ester (DAPT)) resulted in elevated expression levels of Atonal BHLH transcription factor 1 (*ATOH1*) and secretory lineage markers, such as MUC2, TFF3, Resistin like beta (*RETNLβ*), Serine peptidase inhibitor Kazal type 4 (*SPINK4*), SAM pointed domain containing ETS transcription factor (*SPDEF*), Neurogenin 3 (*NEUROG3*) (Kazanjian et al., 2010). In addition, upon administration of a gamma-secretase inhibitor, cell viability was reduced in LOVO, LS174T, SW480, HT29, and HCT116 CRC cell lines, suggesting the growth inhibitory role of the gamma-secretase inhibitor (Kazanjian et al., 2010).

Mice carrying deletions of both Notch1 and Notch2 showed enhanced secretory cell hyperplasia, substantially reduced cell division, and poor crypt regeneration following irradiation (Carulli et al., 2015). Inactivation of DLL1 and DLL4 simultaneously resulted in a loss of crypt base columnar stem cells and promoted a complete transformation of transit-amplifying cells into secretory goblet cells in mice, suggesting cooperation between DLL1 and DLL4 (Pellegrinet et al., 2011). However, disruption of DLL4 or JAG1 alone did not impact goblet cell differentiation in the small intestine and colon (Pellegrinet et al., 2011). Transgenic mice with Notch1 and Notch2 knock-ins demonstrated higher Notch activity in intestinal stem cells and absorptive progenitors, whereas cells differentiated into secretory lineages did not exhibit detectable Notch activity (Fre et al., 2011).

Upregulation of Notch-related factors, such as Notch1, JAG ligands, and HES1 was observed in human colorectal cancers (Reedijk et al., 2008; Vinson et al., 2016) . A cohort covering mRNA expression of five ligands (JAG1, JAG2, DLL1, DLL3, and DLL4), four Notch receptors (Notch1-4), and four Notch target genes (HES1, HES5, HEY1, and HEY2) in human colorectal adenomas and cancers showed that these components are highly expressed in colorectal tumours compared to normal samples (Shaik et al., 2020). Moreover, the overall survival rate was lower in patients with overexpression of Notch3, Notch4, and HEY1 genes (Shaik et al., 2020). On the other hand, there are some studies that suggest low levels of Notch2 expression in colorectal cancerous tissues and its downregulation during the transition from well to poorly differentiated tumours, leading to poor survival (Chu et al., 2011; Chu et al., 2009).

1.4.2 ATOH1 and SPDEF

ATOH1 (also called Math1 or Hath) is a basic helix-loop-helix transcriptional factor that plays a primary role in cell fate specification. It is a positive regulator of secretory cell lineages in the small and large intestines. Depletion of ATOH1 results in a significant reduction of secretory lineages without influencing absorptive cells in the mouse small intestine (Yang et al., 2001), whereas induced expression of ATOH1 leads to the conversion of progenitors into secretory cells (VanDussen & Samuelson, 2010).

The interaction between ATOH1 and Notch pathways controls cell fate between absorptive and secretory lineages: HES1, a Notch target gene, downregulates the transcription of ATOH1 (Fu et al., 2020; Zheng et al., 2011). Besides, in the absence of ATOH1, the Notch pathway loses its function to regulate stem cell proliferation and absorptive lineage regulation, indicating that Notch activity is strongly correlated with ATOH1 (Kazanjian et al., 2010).

It was identified that the Wnt/β-catenin pathway can also regulate ATOH1 activity through Glycogen synthase kinase 3 beta (*GSK3β*) (Peignon et al., 2011; Tsuchiya et al., 2007). When the Wnt pathway is active, ATOH1 appears to undergo proteasomal degradation by GSK3β in CRC cell lines (Aragaki et al., 2008; Tsuchiya et al., 2007). A study by Tsuchiya showed that the 54th and 58th serine residues of ATOH1 protein are vital for protein degradation (Tsuchiya et al., 2007). Conditional activation of ATOH1 by modifying these serine residues or utilising GSK3β inhibitors in CRC cell lines with an active Wnt pathway led to upregulation of MUC2 expression, thereby promoting cancer cell differentiation (Aragaki et al., 2008; Peignon et al., 2011).

ATOH1 also plays a regenerative role during normal conditions and tissue damage in the small intestine and colon (Ishibashi et al., 2018; Tomic et al., 2018). ATOH1-positive progenitors were described to revert to the stem cell population upon injury in mice, thereby providing self-renewal and epithelial regeneration (Tomic et al., 2018). The plasticity of secretory progenitors is controlled by ATOH1's multiple phosphorylation sites as mice with phospho-mutant ATOH1 demonstrated poor tissue regeneration and decreased clonogenicity (Tomic et al., 2018).

ATOH1 expression was dramatically downregulated in CRC by either deletion of its allele/alleles or CpG methylation (Bossuyt et al., 2009). Overexpression of ATOH1 in HT29 CRC cell line blocked cell proliferation and growth in both xenografts and soft agar conditions (Leow et al., 2004). According to a study of HCT116 and LOVO CRC cell lines, upregulation of ATOH1 through anti-micRNA-613 significantly suppressed cell division, migration, and invasion (Yang et al., 2018).

SPDEF is another factor shown to promote terminal differentiation and maturation of goblet cells in the intestine. Intestine-specific SPDEF knockout results in a significant loss of mature goblet and Paneth cells in mice, accompanied by an accumulation of immature secretory transit amplifying cells (Gregorieff et al., 2009). Immature goblet cells exhibit poorly organised vacuoles and a distinct brush border (Gregorieff et al., 2009). On the other hand, enterocyte differentiation within the intestinal tract remained normal in SPDEF-deficient mice (Gregorieff et al., 2009).

SPDEF has been reported to be a downstream target of ATOH1, as well as to have positive feedback on the amplification of ATOH1 target genes (Lo et al., 2017). Several genes including ATOH1, SPDEF, Regenerating family member 4 (*REG4*), MUC2, CAMP responsive element binding protein 3 like 1 (*CREB3L1*), Kallikrein 1 (*KLK1*),

and Solute carrier family 12 member 8 (*SLC12A8*) were identified to carry SPDEF motifs and associate with goblet cells (Lo et al., 2017). Interestingly, in ATOH1 deleted cells of the mouse model, SPDEF, even after its transgenic activation, was not capable of inducing transcription of KLK1, REG4, CBFA2/RUNX1 partner transcriptional co-repressor 3 (*CBFA2T3*), MUC2, SLC12A8, and CREB3L1 genes (Lo et al., 2017). Furthermore, SPDEF deletion significantly reduced transcription of several genes, such as MUC2, CREB3L1, REG4, and SLC12A8 during overexpression of ATOH1 (Lo et al., 2017). These findings indicate a mutual interaction between ATOH1 and SPDEF.

1.4.3 CDX1 and CDX2

Caudal Type Homeobox 1 (CDX1) and 2 (CDX2) are transcription factors involved in the regulation of intestinal epithelial development. In contrast to CDX1, CDX2-null mice are lethal during embryogenesis, which could be explained by their initial expression in extraembryonic tissue (Silberg et al., 2000). Their expression pattern varies during postnatal development with both displaying an increased trend from the proximal to the distal small intestine while CDX1 being the highest and CDX2 being lower or absent in the distal colon (Silberg et al., 2000). Mice with CDX1 knockout scaled up CDX2 expression levels in the distal colon where its existence is limited, suggesting a compensatory interaction between CDX1 and CDX2 (Bonhomme et al., 2008; Grainger et al., 2013). Conditional ablation of CDX2 in adult mice showed that endodermal cells in the intestine were incapable of differentiating into enterocytes, Paneth or goblet cells despite maintaining long-term viability, highlighting the importance of CDX2 in intestinal stem cell differentiation and homeostasis (Stringer et al., 2012).

CDX1 was identified to modulate the expression of cytokeratin 20 (KRT20), a marker of differentiated intestinal cells, as well as the expression of glycoprotein A33 cell surface protein by interacting directly with their promoter regions (Chan et al., 2009; Johnstone et al., 2002). CDX2 controls the expression of intestinal specific proteins, such as MUC2, sucrase-isomaltase, and carbonic anhydrase I (Drummond et al., 1998; Suh et al., 1994; Yamamoto et al., 2003). It was also described to regulate cell proliferation in CRC cell lines through binding β-catenin and interrupting its interaction with TCF complex, causing downregulation of Wnt target genes, such as MYC Proto-Oncogene, BHLH Transcription Factor (*c-Myc*), Cyclin D1, and Lymphoid enhancer-binding factor 1 (*LEF1*) (Guo et al., 2010).

CDX1 expression was reduced in colonic adenocarcinomas when compared with normal crypts (Silberg et al., 1997). A study of 37 CRC cell lines conducted by our lab revealed that a substantial proportion of the cell lines exhibit a lack of CDX1 expression due to promoter methylation, presumably resulting in decreased differentiation and increased growth rate (Wong et al., 2004). Based on research on HT29, HTC116, and SW1222 CRC cell lines, HCT116, which does not express CDX1, represents a more aggressive phenotype due to its lack of differentiation ability and primarily stem cell characteristics (Yeung et al., 2010). Induced expression of CDX1 in HCT116 caused fewer colonies and promoted lumen formation in Matrigel (Yeung et al., 2010).

The role of CDX2 in colon cancer is controversial. Some reports suggest that CDX2 expression is lower in colonic carcinomas compared to adenomas and tumours lacking CDX2 expression display poor differentiation, advanced TNM stage, and a lower survival rate (Bae et al., 2015; Bakaris et al., 2008; Hinoi et al., 2001). However,

a study by Witek et al. (2005) pointed out the overexpression of CDX2 mRNA and protein levels in colorectal tumours in comparison with adjacent mucosa that was histologically normal (Witek et al., 2005). Moreover, the disruption of CDX2 prevented the anchorage-independent cell growth in LOVO and cell death in SW48 CRC cell lines (Dang et al., 2006). According to a population-based study covering 796 metastatic CRC patients, CDX2 expression in BRAF-mutated patients is associated with a better outcome as opposed to controls, whereas the absence of CDX2 in a small proportion of KRAS-mutated patients had a poor prognosis (Aasebø et al., 2020). CDX1 and CDX2 expression in the COLO205 CRC cell line promoted a pattern of gene expression associated with differentiation, activated cell adhesion, and reduced cell proliferation (Keller et al., 2004).

1.4.4 GATA6 and ETS2

GATA binding protein 6 (GATA6) is a zinc finger transcription factor expressed along the gastrointestinal tract (Haveri et al., 2008). It plays an important role in the earliest stages of embryogenesis and continues its expression during adulthood. GATA6-null mice embryos died before the heart formed, displaying atypical characteristics at 5.5 days postcoitum (Koutsourakis et al., 1999). Mice lacking GATA4 and GATA6 in the intestine were lethal during 24 hours of birth (Walker et al., 2014).

GATA6 occupancy in promoter or enhancer regions of several genes, such as SPINK4 and solute carrier family 9 member A2 (SLC9A2) was detected by chromatin immunoprecipitation (ChIP) analysis in the LS174T cell line (Beuling et al., 2012). Another gene, intestinal fatty acid binding protein (IFABP), a marker of intestinal enterocytes, contains GATA-binding sites in the promoter region, suggesting it being

a potential target for GATA family proteins, including GATA4, 5, and 6 (Gao et al., 1998).

GATA6 was described to be expressed in dividing crypt cells and all differentiated cell types in the mouse colonic epithelium (Beuling et al., 2012). Deficiency of GATA6 caused morphological changes, a reduction in cell division and migration from the crypt base to the surface in the colon (Beuling et al., 2012). Furthermore, it downregulated the expression of goblet- and enterocyte-specific genes, as well as hormone production by enteroendocrine cells in the colon, implying its dual role in cell proliferation and differentiation (Beuling et al., 2012). Similar results were obtained in the LS174T cell line in terms of goblet- and colonocyte-associated genes during the knockdown of GATA6 (Beuling et al., 2012). A recent study by Laudisi et al. revealed that mice with GATA6 deletion exhibited disrupted mucosal barrier and intestinal homeostasis, as well as enhanced dysbiosis (Laudisi et al., 2022). GATA6 gene expression was observed to diminish during colon cancer progression from adenoma to carcinoma (Haveri et al., 2008) and in inflammatory bowel diseases (Laudisi et al., 2022). Study with GATA6 Antisense RNA 1 also supports its reduction during Crohn's disease, ulcerative colitis, and coeliac disease and this dropdown correlates with more severe ulcerative colitis conditions by affecting mitochondrial respiration (Sosnovski et al., 2023).

ETS proto-oncogene 2, transcription factor (ETS2) is considered a Wnt target gene and its disruption led to higher levels of cell proliferation at the base of the intestinal crypts and increases colonic crypt fission (Múnera et al., 2011). During the administration of dextran sulfate sodium and azoxymethane, ETS2-deficient mice were more susceptible to developing adenomas than wild-type animals (Múnera et al.,

2011). A recent study by Chen et al. revealed that the transcriptional activity of ETS2 is controlled by distal super-enhancer regulatory elements, which recruit the MDS1 and EVI1 complex locus (*MECOM*) transcription factor. This complex was remarkably increased in colorectal cancer, as well as in inflammatory bowel disease (Chen et al., 2023).

Overall, the transcription factors involved in goblet cell differentiation in the colon can be visualised in Figure 1.5.

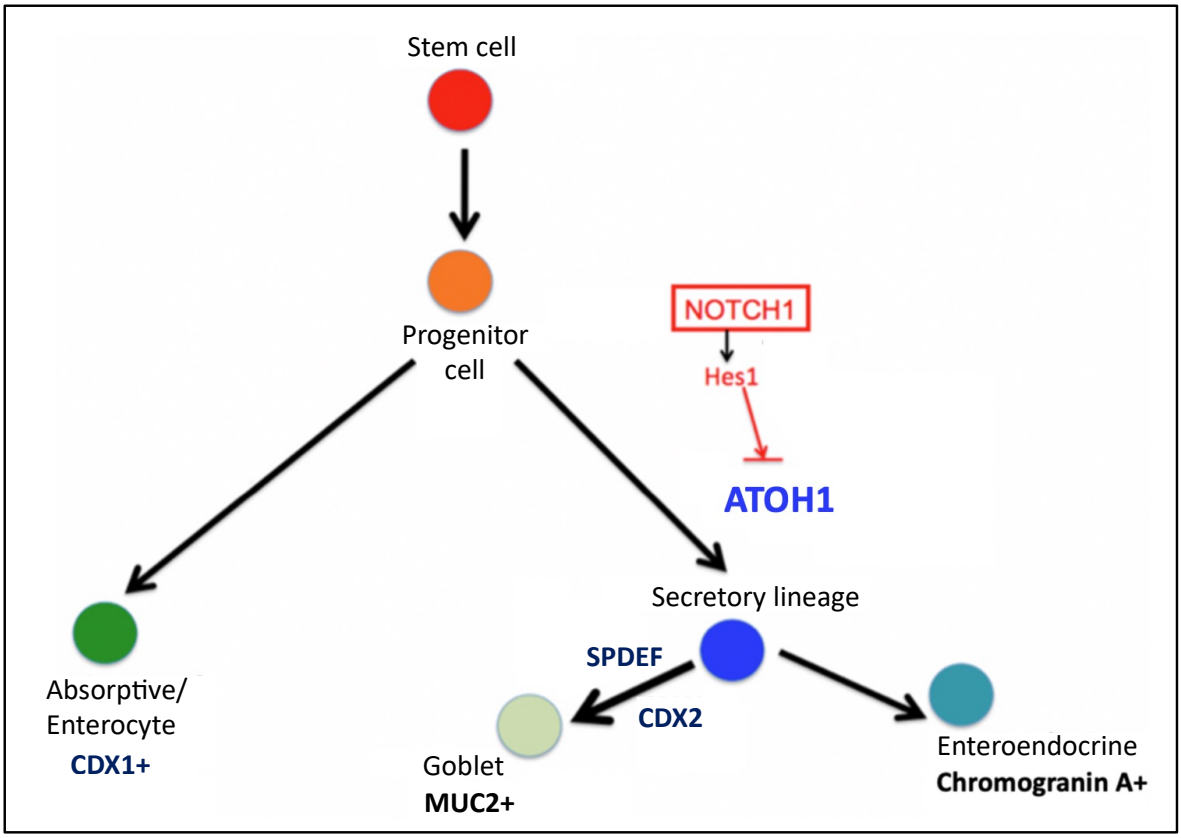


Figure 1.5 Differentiation pathway in normal colon. ATOH1 is the key transcription factor for secretory lineage differentiation and is negatively regulated by Notch target gene, HES1. SPDEF is involved in terminal differentiation of goblet cells. The figure was adapted from Professor Walter Bodmer.

1.5 Colorectal cancer: epidemiology, causes, and genetics

Colorectal cancer (CRC) ranks third in the world for the number of diagnoses and second in terms of mortality (Hossain et al., 2022; Sawicki et al., 2021). CRC can be categorised into three types: sporadic, hereditary, and colitis-associated (Hossain et al., 2022). The majority of colorectal cancer cases are sporadic, while others are familial or hereditary and are associated with inflammatory bowel diseases (Hisamuddin & Yang, 2006; Yamagishi et al., 2016). Genetic and environmental factors that provide survival advantages for stem or progenitor cells contribute to CRC progression. Suggested environmental factors include a sedentary lifestyle, high consumption of red meat and preserved foods, obesity, smoking, and alcohol consumption.

Familial CRC can be divided into two groups: hereditary polyposis CRC and hereditary non-polyposis CRC (Rebuzzi et al., 2023). The most frequent and well-defined hereditary non-polyposis CRC condition is Lynch syndrome, caused by mutations in key DNA mismatch repair genes, including MLH1, PMS2, MSH2, MSH6 or epithelial cell adhesion molecule (EpCAM) (Fearnhead et al., 2002; Hossain et al., 2022; Rebuzzi et al., 2023). Patients with hereditary non-polyposis CRC have high levels of microsatellite instability because of DNA replication errors (Hisamuddin & Yang, 2006; Rebuzzi et al., 2023). Microsatellite instability in sporadic CRC is generally associated with hypermethylation of the MLH1 promoter rather than gene mutation as in inherited cancers (Bodmer, 2006; Li et al., 2021).

Familial adenomatous polyposis (FAP) is the most typical cause of hereditary polyposis CRC, accounting for less than 1% of CRCs (Fearnhead et al., 2002; Rebuzzi et al., 2023). FAP syndrome is due to germline mutations in the *Adenomatous*

polyposis coli (*APC*) tumour suppressor gene which plays an essential role in the Wnt/β-catenin signalling pathway, cell adhesion and apoptosis (Hisamuddin & Yang, 2006; Rebuzzi et al., 2023). In addition to germline mutations, somatic mutations and/or loss of heterozygosity of the APC gene have been identified in the sporadic CRCs (Bodmer, 2006; Hull et al., 2020).

Somatic APC mutations occur in nearly 80% of colorectal cancers and traditional adenomas (Bodmer, 2006), whereas BRAF mutations are commonly found in serrated polyps (Kuipers et al., 2015; Nguyen et al., 2020). Other driver genes mutations, such as tumour protein p53 (*TP53*), Kirsten rat sarcoma viral oncogene homolog (*KRAS*), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*), and SMAD family member 4 (*SMAD4*) were detected at the late stages of the transition from adenoma to carcinoma, well before the development of metastasis (Figure 1.6) (Bodmer, 2006; Hull et al., 2020; Li et al., 2021).

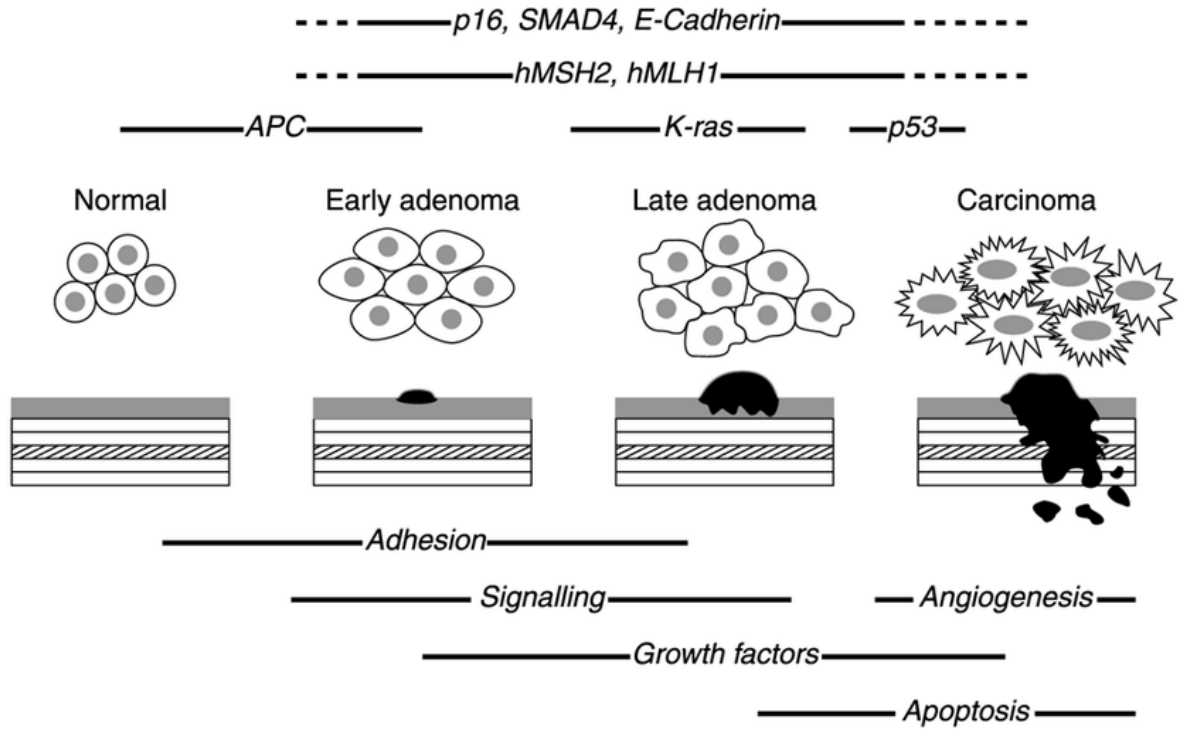


Figure 1.6 Adenoma to carcinoma transition. The schematic describes the stages and factors involved in the adenoma to carcinoma sequence in colorectal cancer. The figure was adapted from Fearnhead et al. (2002).

Tumours in the right- and left-sided colon vary in terms of their histological, molecular, and genetic characteristics. A higher proportion of right-sided colon cancer cases was observed among women and older individuals, whereas left-sided colon cancer tends to be more common among men (Baran et al., 2018). In terms of signet-ring cell carcinomas, sessile serrated adenomas, and mucinous adenocarcinomas, there are larger increases in the prevalence of these types in right-sided colon cancer than in left-sided colon cancer, making right-sided colon cancer difficult to detect at early stages (Baran et al., 2018; Nawa et al., 2008).

Right-sided colon cancer patients are usually considered large, poorly differentiated, and advanced tumours (Baran et al., 2018; Benedix et al., 2010; Bustamante-Lopez et al., 2019). In terms of genetic composition, right-sided cancers exhibit higher microsatellite instability (MSI) and CpG island methylation phenotype (CIMP), whereas left-sided cancers present chromosomal instability (CIN) related mutations, such as APC, KRAS, and p53 mutations (Iacopetta, 2002; Sugai et al., 2006). Based on a cohort that included 17,641 CRC patients (8297 right-sided and 9344 left-sided), the metastatic spread was different between the two types and right-sided carcinomas had a considerably poorer outcome as compared to left-sided colon cancers (Benedix et al., 2010). Furthermore, there was a higher incidence of hepatic and pulmonary metastases in left-sided carcinomas, while peritoneal metastases were often detected in right-sided colon cancer patients (Benedix et al., 2010). Another retrospective study of 206 right-sided and 214 left-sided colon cancer patients in The Cancer Genome

Atlas (TCGA) revealed that patients with right-sided colon cancers have a lower 5-year overall survival rate than those with left-sided cancers (Ward et al., 2022).

1.6 Goblet cell in colorectal cancer

Colorectal cancers can be classified as mucinous with abundant MUC2 secretion and non-mucinous with reduced or absent MUC2 secretion, emphasising dysregulation of goblet cell differentiation. Mucinous adenocarcinoma is defined as a tumour in which mucin covers more than 50% of the extracellular space (Huang et al., 2021; Luo et al., 2019). This type of cancer is more frequent in females and younger individuals, and it is usually determined at a later stage of its progression (Luo et al., 2019).

The incidence of mucinous carcinomas, in general of cancers in the small intestine is much lower than in the large intestine (Kim & Ho, 2010), representing around 10-20% of colorectal cancers (Gundamaraju & Chong, 2021; Luo et al., 2019). Cancers with elevated MUC2 levels appear to have lymph node metastasis, microsatellite instability, an increased recurrence rate, and a diverse mutation pattern (Gundamaraju & Chong, 2021). The frequency of B-Raf proto-oncogene, serine/threonine kinase (BRAF), and KRAS mutations was greater in mucinous carcinomas, while the frequency of TP53 mutations was lower (Huang et al., 2021).

It has been suggested that in colorectal cancer, there are two main alterations in mucins, including abnormal glycosylation and gene expression. MUC2 is highly expressed in mucinous carcinomas due to aberrant genetic and epigenetic regulation. MUC5AC, despite its primary expression in the respiratory tract and stomach, is also overexpressed in mucinous carcinomas (Luo et al., 2019). In addition, truncated glycan structures and a raised level of neo-glycans, such as sialylLeA and sialylLeX

antigens were defined in both mucinous and non-mucinous carcinomas (Hollingsworth & Swanson, 2004; Kim & Deng, 2008; Kim & Ho, 2010).

Compared to their non-mucinous counterparts, mucinous colorectal adenocarcinomas are mostly diagnosed at a more advanced stage (Kim & Ho, 2010) and have poorer response rates to chemotherapy (Luo et al., 2019). The study by Reynolds et al. revealed that the mucinous adenocarcinomas showed upregulation of various genes, such as Thymidine phosphorylase (TYMP) and Glutathione S-transferase Pi 1 (GSTP1) involved in 5-fluorouracil (5-FU) and oxaliplatin drug resistance, respectively (Reynolds et al., 2020). These modifications might cause a poor therapeutic response owing to accelerated drug metabolism and detoxification. Furthermore, mucus released during mucinous adenocarcinomas could potentially act as a physical obstacle for drugs, prohibiting their effective delivery to cancer cells.

Signet-ring cell carcinoma is a very rare CRC subtype which is identified by having more than 50% intracellular mucin content with signet-ring cell appearance (An et al., 2021; Kim & Ho, 2010). Similar to mucinous carcinoma, signet-ring cell carcinoma exhibits high BRAF mutations, an MSI-high status and a CIMP-positive phenotype (Luo et al., 2019). Patients with signet-ring cell carcinoma have a poorer survival rate and increased risk of developing metastases at distinct sites (An et al., 2021; Luo et al., 2019).

1.7 Aims of the project

Colorectal cancer is primarily characterised by the low frequency of goblet cells and an impaired mucus barrier. This was our main reason for focusing on goblet cell differentiation. Investigation of the mechanisms responsible for goblet cell differentiation is essential to improve current efforts in the prevention and treatment of CRC.

Our laboratory has a large collection of CRC cell lines, well characterised and truly representative of the tumours from which they are derived (Mouradov et al., 2014; Wilding & Bodmer, 2014). We used in vitro cell line models to study goblet cell differentiation and aimed to address following objectives.

- Reclassification of primary CRC cell lines based on MUC2 and TFF3 protein expression profile. Validation of our classification in human tissue sections, as well as in single-cell RNA sequencing data.

- To identify what triggers the goblet cell differentiation process in the colon using our cell lines as the models for CRC? Why do most CRC cell lines not produce differentiated goblet cells and when they are produced, why goblet cells present at such a low frequency except for the rare signet ring tumours or mucinous carcinomas?

- Analyse the roles of known, and possible new factors involved in goblet cell differentiation in the CRC cell lines, for example, by knocking them out or in, by studying the timing of differentiation in clonal outgrowth, and by single-cell mRNA analysis of cell lines under various conditions of growth.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2: MATERIALS AND METHODS

2.1 Cell culture methods

2.1.1 Cell lines

Human colorectal cancer (CRC) cell lines were obtained from the Cancer and Immunogenetics Laboratory (CIL) cryogenic collection at the Weatherall Institute of Molecular Medicine (WIMM), Oxford, UK. All the cell lines used in this study are listed in Table 2.1.

Cell line name	Source and catalogue number
C10[1]	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C106	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C10A[1]	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C10S[1]	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C125PM	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C2BBE1[2]	ATCC® CRL-2102™
C32	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C70	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C75	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C80	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C84	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
C99	Cancer Immunogenetics Laboratory (W. Bodmer and D. Bicknell)
CACO2[2]	ATCC® HTB-37™
CAR1	JCRB0207
CC20	Kind gift from Dr. Anne R. Kinsella, Paterson Laboratories, Manchester M20 9BX, UK
CCK81	JCRB0208
CCO7	JT Immunotech, USA
CL14	DSMZ: ACC-504
CL40	DSMZ: ACC-535

Cell line name	Source and catalogue number
COLO201[3]	ATCC® CCL-224™
COLO205[3]	ATCC® CCL-222™
COLO206[3]	DSMZ: ACC-21
COLO320DM[4]	ATCC® CCL-220™
COLO320HSR[4]	ATCC® CCL-220.1™
COLO678	DSMZ: ACC-194
CW2	RCB0778
CX1[5]	DSMZ: ACC-129
DLD1[6]	ATCC® CCL-221™
GP2D[7]	ECACC: 95090714
GP5D[7]	ECACC: 95090715
HCA46	Kind gift from Dr. S.Kirkland, Royal Postgrad School, London
HCA7	Kind gift from Dr. S.Kirkland, Royal Postgrad School, London
HCC2998	Obtained from NCI-FREDERICK CANCER DCTD tumour/cell line repository, vial designation: 0507244
HCC56	JCRB1037
HCT8	ATCC®
HCT116	ATCC® CCL-247™
HCT15[6]	ATCC® CCL-225™
HDC111	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC114	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC135	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC142	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC143	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC54[8]	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC57[8]	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC73	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC8	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany

Cell line name	Source and catalogue number
HDC82	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HDC9	Obtained from Prof. Manfred Schwab under an MTA, DKFZ, Hedelberg, Germany
HRA19	Kind gift from Dr. S.Kirkland, Royal Postgrad School, London
HT29[5]	ATCC® HTB-38™
HT55	ECACC: 85061105
ISRECO1	Kind gift from Pr. Richard Hamelin, INSERM U434-CEPH, 75010 Paris, France
JHCOLOYI	RCB1706
KM20L2	Obtained from NCI-FREDERICK CANCER DCTD tumour/cell line repository, vial designation: 0503425
LIM1215	Kind gift from Dr. Joan K. Heath, Ludwig Institute for Cancer Research, Australia
LIM1863	R.Whitehead, Australia
LIM2405	Kind gift from Dr. Joan K. Heath, Ludwig Institute for Cancer Research, Australia
LOVO	ATCC® CCL-229™
LS1034	Kind gift from Dr. Laurent Suardet, ISCRE, Switzerland
LS123	ATCC® CCL-255™
LS174T[9]	ECACC: 87060401
LS180[9]	ECACC: 87021202
LS411	Kind gift from Dr. Laurent Suardet, ISCRE, Switzerland
LS513	ATCC® CRL-2134™
NCIH508	ATCC® CCL-253™
NCIH548[3]	ATCC® CCL-249™
NCIH716	ATCC® CCL-251™
NCIH747	ATCC® CCL-252™
OUMS23	JCRB1022
OXCO1	Kind gift from Prof. Vincenzo Cerundolo, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS
OXCO2	Kind gift from Prof. Vincenzo Cerundolo, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS
OXCO3	Kind gift from Prof. Vincenzo Cerundolo, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS
PCJW	Kind gift from Prof. Christos Paraskeva, Department of Pathology and Microbiology, University of Bristol, UK

Cell line name	Source and catalogue number
PMFKO14	RCB1426
RCM1	JCRB0256
RKO	ATCC® CRL-2577™
RW2982	Kind gift from Dr John Mariadason, Montefiore Medical Center, 111 E210th St. Department of Oncology, Bronx NY
RW7213	Kind gift from Dr John Mariadason, Montefiore Medical Center, 111 E210th St. Department of Oncology, Bronx NY
SKCO1	ATCC® HTB-39™
SNU1235	KCLB No. 01235.1
SNUC1	ATCC® CRL-5972™
SNUC2B	ATCC® CCL-250™
SW1116	ATCC® CCL-233™
SW1222	ECACC: 12022910
SW1417	ATCC® CCL-238™
SW1463	ATCC® CCL-234™
SW403	ATCC® CCL-230™
SW48	ATCC® CCL-231™
SW480[10]	ATCC® CCL-228™
SW620[10]	ATCC® CCL-227™
SW837	ATCC® CCL-235™
SW948	ATCC® CCL-237™
T84	ATCC® CCL-248™
TT1TKB	RCB1185
VACO10MS	Kind gift from J.McBain, VA Res.Ser. Madison, Wisconsin
VACO429	Obtained from Dr Elizabeth Zborowska under an MTA, Case Western reserve University, Cleveland, Ohio
VACO4A[11]	Kind gift from J.McBain, VA Res.Ser. Madison, Wisconsin
VACO4S[11]	Kind gift from J.McBain, VA Res.Ser. Madison, Wisconsin
VACO5	Kind gift from J.McBain, VA Res.Ser. Madison, Wisconsin
WIDR[5]	Kind gift from J.McBain, VA Res.Ser. Madison, Wisconsin

Table 2.1 List and source of studied CRC cell lines. The table depicts a list and source of all CRC cell lines used in this study. Cell lines originated from the same patient were indexed

with an identical number. ATCC: American Tissue Culture Collection, DSMZ: German Collection of Microorganisms and Cell Cultures GmbH, ECACC: European Collection of Authenticated Cell Cultures, DKFZ: Deutsches Krebsforschungszentrum (The German Cancer Research Centre), JCRB: Japanese Collection of Research Bioresources, KCLB: Korean Cell Line Bank, RCB: Research Cell Bank.

2.1.2 Cell culture conditions and maintenance

The cell lines were cultured in 75 cm^2 sterile tissue culture flasks (Corning® flasks) with 10 millilitre (ml) complete Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco®) and 1% penicillin/streptomycin (Lonza). Cells were also grown in 6-well, 12-well, 24-well, 48-well, and 96-well cell culture plates (Corning® tissue culture-treated plates). Flasks and plates were kept in a humidified incubator (Heracell 150 incubator, Thermo Scientific, Massachusetts, USA) at 37°C, 10% CO_2 environment. In order to avoid cross-contamination, a maximum of five cell lines were incubated at any given time.

All the cell lines in Table 2.1 except for NCIH716 are adherent cell lines. The medium was changed every 2-3 days. When the cells reached around 70-80% confluency as observed under the microscope, they were passaged. Cell passaging starts with removing cell culture medium and carefully washing the cells with 10 ml 1X Phosphate Buffered Saline (PBS). After the washing step, cells were treated with 1 ml 0.5% 10X Trypsin-EDTA (Gibco, Ref. 15400-054) for 3-5 minutes (min) to achieve detachment from the flask. Then, 5 ml of complete medium was added to neutralize the effect of the dissociation reagent. The cell suspension was harvested and centrifuged at 1000 rpm for 5 min at room temperature (RT). The supernatant was discarded, the cell pellet was gently tapped and then resuspended in 10 ml complete medium by pipetting up and down gently to break cell clumps. Depending on cell line doubling time, the dilution

factor varied from 1:5 to 1:20 and the cells were transferred into a new 75 cm^2 cell culture flask containing pre-warmed medium.

For the suspension cell line, the medium was replaced every 2-3 days. When cell clumps were apparent under the microscope, cells were collected and centrifuged at 1000 rpm for 5 min at RT. After the removal of the supernatant, the cell pellet was resuspended with complete medium and appropriately sub-cultured.

2.1.3 Cell line storage and retrieval

Cell lines were kept in liquid nitrogen for long-time storage after the first passage. When cells were 70-80% confluent, they were washed with 1X PBS and underwent trypsinisation. Then, we resuspended the cells with 5 ml complete medium and centrifuged them at 1000 rpm for 5 min at RT. The supernatant was removed, and the cell pellet was resuspended in 1 ml of freezing medium, consisting of 10% Dimethyl sulfoxide (DMSO, Fisher Bioreagents™) and 90% FBS. Five hundred microliter (µl) of cell suspension were then transferred into sterile cryogenic vials (Corning, USA), placed in a Mr Frosty Freezing Container (Thermo Fisher Scientific, UK) and stored at -80°C overnight to achieve slow-freezing. The next day, the tubes were transferred to liquid nitrogen.

In order to recover frozen stocks, cryogenic vials have to be thawed in a water bath at 37°C for 1 min prior to resuspending in 5 ml of complete medium. Cells were then centrifuged at 1000 rpm for 5 min at RT and the supernatant was carefully removed. Ten ml of complete culture medium was added to the cell pellet and cells were transferred into 75 cm^2 sterile tissue culture flasks. The medium was replaced the

following day and thawed cells were generally only used for experiments after two passages.

2.1.4 Cell counting

Cell counting was performed in the Cellometer Auto T4 device (Nexcelom Biosciences, USA) by loading 20 µl of cell suspension into a cell counting chamber (SD100, Nexcelom Biosciences, USA). The chamber is then inserted into the cell counter and the cells were imaged. The machine uses brightfield images of eight individual areas by considering cell size and morphology to calculate cell concentration per ml.

2.1.5 Mycoplasma contamination testing

Cell lines were routinely tested for mycoplasma contamination using the MycoStrip™ Mycoplasma Detection Kit (InvivoGen, UK). The test principle of the kit was based on isothermal PCR of the 16S rRNA gene which is most frequently detected in mycoplasma species.

Cells were cultured at least 48 hours before the mycoplasma test. One ml of the cell culture supernatant was transferred to a microcentrifuge tube. The medium was centrifuged at 16,000 g for 5 min. The supernatant was gently discarded without touching the pellet. Next, 500 µl of sterile 1X PBS was added to the pellet and mixed well. In order to increase sensitivity, previous steps were repeated starting with centrifugation. Afterwards, 5 µl of Reaction Buffer and 5 µl of the prepared sample were added into the tube containing 15 µl of Reaction Mix aliquot. The tube was incubated in a heat block at 65°C for 40 min. In the next step, we added 200 µl of

Migration Buffer into the tube and mixed it well before loading the mixture into the cassette. Positive and negative controls were also used according to the protocol. One band (C' band only) in the cassette represents that mycoplasma was not detected in the sample, while two bands (C' and T' bands) indicate contamination of cell culture by mycoplasma.

2.2 Drug studies

2.2.1 Notch gamma-secretase inhibitor treatment

Dibenzazepine (DBZ), a Notch gamma (γ)-secretase inhibitor was purchased from Merck Millipore and 1 mg of DBZ was reconstituted in 1 ml DMSO. Then, it was aliquoted and stock solutions were stored at -20°C. Before treatment, cell lines were passaged and counted as previously described. An appropriate number of cells (30000-50000 cells per ml) depending on the cell line's doubling time were seeded in 24-well and 96-well plates prior to the treatment. After 24 hours, the medium was removed and DBZ at a working concentration of 200 nanomolar (nM) within the appropriate amount of medium was added to the cells. The cells were cultured for an additional 24 and 48 hours to analyse the effect of the drug. DMSO was used as a solvent control.

2.2.2 5-aza-2'-deoxycytidine treatment and recovery

5-Aza-2′-deoxycytidine (5-AZA-CdR, decitabine) is an epigenetic agent that prevents DNA methyltransferase activity resulting in DNA demethylation. The drug solution was prepared by dissolving 50 milligram (mg) of 5-AZA-CdR powder (Merck Millipore) in 1 ml of DMSO. Aliquots of the stock solution were stored at -20°C.

According to the cell line growth rate, the cells (20000-50000 cells in 1 ml) were seeded before 24 hours of drug treatment. The next day, the cells were incubated with complete medium containing 5-AZA-CdR at a final concentration of 10 micromolar (μM). DMSO was used as a control and followed the same procedures as the drug. The medium containing the drug was added for different incubation times (24-, 48-, 72-, 96-, and 120 hours) before being fixed for immunostaining analysis.

During the recovery experiment, the cells were incubated in 24-well and 96-well plates and were allowed to adhere to the surface for 24 hours. Then, they were treated with 5-AZA-CdR for 72 hours as previously described. After 3 days, the drug treatment was stopped and the cells were allowed to recover for varying amounts of time (24-, 48-, 72-, 96-, 120 hours) by replacing the complete medium every 2 days. Then, some plates were fixed for immunostaining and others were prepared for different experiments.

2.2.3 Sulforhodamine B protein staining assay

The Sulforhodamine B (SRB) assay was widely used and optimised in our lab and by other scientists (Vichai & Kirtikara, 2006). It evaluates the total amount of cellular protein content which corresponds to cell number. The SRB assay was conducted in a 96-well plate following the appropriate drug treatments to study the effect of various concentrations and lengths of treatment on the cell lines.

The cells were washed with 1X PBS once and were prepared for the assay by being fixed with Paraformaldehyde (PFA, Thermo Fisher Scientific, UK) for 1 hour. After removing the PFA, the cells were washed with water four times and plates were tapped on a tissue paper to remove the remaining water. Then, 100 μl SRB solution (0.057%

Sulforhodamine B sodium salt (Sigma, S1402) in 1% acetic acid (Alfa Aesar, Thermo Fisher Scientific) was added to the cells and it was allowed to interact with proteins for 30 min. Any remaining SRB was removed from the wells and the wells were then washed four times with 200 µl 1% acetic acid. Afterwards, 200 µl of 10 millimolar (mM) Tris base was added to each well and the absorbance was measured at 520 nanometre (nm) on a µQuant™ spectrophotometer (BioTek Instrument, USA).

2.3 Transient small interfering RNA transfection

Small interfering RNA (siRNA) transfection was conducted using Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, UK). Pre-designed siRNAs against ATOH1 (Assay ID. 147381 and #M-008915-00 [Dharmacon™]), SPDEF (Assay ID. 115943 and sc-45845 [Santa Cruz Biotechnology]), CDX1 (Assay ID. 106590), CDX2 (Assay ID. 106423), OLFM4 (Assay ID. 140038), TFF3 (Assay ID. s14040), and GATA6 (Assay ID. 115635) were purchased from Thermo Fisher Scientific (UK) in addition to the ones shown in square brackets. The tube containing 5 nanomole (nmol) of siRNA powder was briefly centrifuged and 250 µl nuclease-free water was added into the tubes to prepare the stock concentration of 20 µM. The prepared siRNA solution was mixed well, aliquoted, and stored at -20°C.

In the siRNA transfection experiments, 20 nM siRNA was used as a final concentration. As a first step, the transfection mix was prepared for 96-well and 24-well plates as follows: 0.84 µl siRNA (stock 20 µM), 1.4 µl Lipofectamine RNAiMAX, and 140 µl Opti-MEM™ Reduced Serum Medium (Gibco). The mixture was gently pipetted up and down and incubated for 15 min. Subsequently, 50.000 cells (24-well plate) and 10.000 cells (96-well plate) were diluted in an antibiotics-free medium with

10% FBS and ‘reverse’ transfected with the appropriate siRNA. A total volume of 700 µl diluted cells (1 x 10^5 cells per ml) was added to the siRNA transfection mix. We plated 600 µl and 120 µl into each well of 24- and 96-well plates, respectively. The plates were incubated for 72 hours by replacing the medium with complete medium after 24 hours of seeding. When siRNA combinations were analysed, the volume of siRNAs (20 nM each) increased twice. As a control, untreated cells were used alongside scrambled RNA treatments.

2.4 RNA methods

2.4.1 RNA extraction

RNA extraction was performed according to the ReliaPrep™ RNA Cell Miniprep System protocol (Promega, UK). Cultured cells in a 24-well plate were washed with 1X PBS and 250 µl BL buffer with 1-Thioglycerol was added to break nucleoprotein complexes. Once the lysate has been pipetted up and down several times, 85 µl of 100% isopropanol was applied. The lysate was then loaded onto a Minicolumn in a collection tube and centrifuged at 13.000 g for 30 seconds (sec) which allows nucleic acids to bind to the column membrane. Next, 500 µl RNA Wash solution was added and centrifuged as in the previous step. In order to break down contaminated genomic DNA, 30 µl of RNase-free DNase I incubation mix (24 µl of Yellow Core buffer, 3 µl of $MnCl_2$, 3 µl of DNase 1 per sample) was applied straight to the membrane with 15 min incubation time. The total RNA was washed with Column Wash solution once and RNA Wash solution twice to be purified from salts, cellular components, and other contaminants. Then, the column was moved to an Eppendorf tube, and 30 µl of RNase-free water was added to the centre of the column membrane. After a 2 min incubation time, the samples were centrifuged at 13.000 g for 1 min to elute the total

RNA. The concentration and quality of RNA were checked using NanoDrop™ One^{C} Microvolume UV-Vis Spectrophotometer (Thermo Scientific) and kept at -20°C.

2.4.2 Quantitative real-time polymerase chain reaction

Quantitative real-time polymerase chain reaction (qRT-PCR) was conducted using TaqMan® RNA-to-CT™ 1-Step Kit (Applied Biosystems). TaqMan® Gene Expression Assay probes for all the genes of interest were purchased from Thermo Fisher Scientific (Table 2.2).

Target gene	Reference	Source
ATOH1	Hs00944192_s1	Applied Biosystems
CDX1	Hs00156451_m1	
CDX2	Hs01078080_m1	
DLL4	Hs00184092_m1	
GAPDH	Hs02786624_g1	
GATA6	Hs00232018_m1	
HES1	Hs00172878_m1	
SPDEF	Hs00171942_m1	
TFF3	Hs00902278_m1	

Table 2.2 List of TaqMan® Gene Expression Assay (20X) probes. The table describes an overview of TaqMan® Gene Expression Assay probes used in this study, as well as their references and sources.

In the first step, RNA was extracted, measured on the NanoDrop as previously described and diluted with nuclease-free water to 20 nanogram/µl (ng/µl). Then, the reaction master mix containing 1 µl of TaqMan® Gene Expression Assay (20X) probe, 10 µl of TaqMan® RT-PCR Mix (2X), 0.5 µl of TaqMan® RT Enzyme Mix (40X), and

4.5 µl of nuclease-free water per sample was prepared. A total of 16 µl reaction master mix was added to each well of a MicroAmp Fast Optical 96-well Reaction Plate (Applied Biosystems, Ref. 4346906) Each well was then loaded with 4 µl of diluted RNA (20 ng/µl) after adding the master mix. The plates were sealed with MicroAmp Optical Adhesive film (Applied Biosystems, Ref. 4311971) and centrifuged briefly before running the reaction. Experiments were conducted in triplicate and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a reference gene or internal control. The qRT-PCR was performed in a 7500 Fast Real-Time PCR System (Applied Biosystems, Ref. 4351107) with a cycling condition described in Table 2.3.

Step	Temp (°C)	Time	Stage
Reverse transcription	48	15 min	Holding
Enzyme activation	95	10 min	Holding
Denature	95	15 sec	Cycling (40 cycles)
Anneal/Extend	60	1 min	
Storage	4	Infinite	

Table 2.3 Thermal cycling conditions for qRT-PCR. TaqMan® RNA-to-C_T™ 1-Step qRT-PCR was performed by using the condition described in the table.

qRT-PCR results were analysed by using the delta-delta Ct method, also known as the 2^(–delta-delta CT) method. Cycle threshold (Ct) values were normalised to GAPDH, the internal control. Following were the calculations involved:

1. Calculate the delta Ct value (∆Ct) which is the difference between the Ct mean value of the gene of interest and the Ct mean value of the reference gene for a given sample.

∆Ct = Ct mean (gene of interest) – Ct mean (housekeeping gene)

2. Compute the delta-delta Ct (∆∆Ct) value which defines the difference in ∆Ct values of treated and untreated samples.

   ∆∆Ct = ∆Ct (gene of interest) – ∆Ct (control)

3. Identify the relative expression (fold change) of the gene of interest.

   Relative quantification (RQ) = 2^-(∆∆Ct)

2.4.3 Microarray gene expression analysis

We have microarray mRNA expression data of 95 CRC cell lines that were analysed by former researchers of Prof. Walter Bodmer's laboratory. The experiment was conducted using the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix, USA) and detailed information about the procedure was described by Wilding et al. (2010). In brief, the manufacturer's instructions were followed for the extraction of total RNA using the Qiagen RNeasy Mini kit (Qiagen). Next, 20 microgram (µg) of RNA from each cell line was delivered to the Paterson Institute for Cancer Research (UK) to study gene expression profiling, as instructed by the manufacturer (Affymetrix).

2.5 Molecular cloning methods

2.5.1 List of gene expression plasmids and vector backbones

The open reading frame (ORF) clones of ATOH1, SPDEF, CDX1, CDX2, HES1, GATA6, and TCF4 genes were obtained from the OriGene company. Table 2.4 shows an overview of the plasmids used in this thesis.

Plasmid name	Selection	Restriction sites used	Source and Catalog number	Notes and plasmid map
pCI-neo	AMP	variable	Promega E1841	Results section
pCMV6-ATOH1	KAN	EcoRI/NotI	OriGene RC210192	Results section
pCMV6-SPDEF	KAN	EcoRI/NotI	OriGene RC232468	Results section

Table 2.4 List of tagged/untagged ORF clones and vector backbones. The table describes information on the bacteria used to select the vectors, the restriction sites for cloning, and the source of the vector.

2.5.2 Restriction digests

The ORFs were cloned in pCI-neo empty vector (Promega, UK) to have the same backbone. A complete set of restriction enzymes and buffers was purchased from New England Biolabs (NEB). Among the ORFs, only the CDX1 construct was prepared using a single restriction enzyme digestion; the others were cut by two restriction enzymes.

The digestion reaction for ORF clones and vector backbone was conducted at 37°C for 1 hour according to the conditions described in Table 2.5. We then analysed linearised backbone DNA on a 1% ultra-pure agarose gel, followed by extraction using the QIAquick Gel Extraction Kit (Qiagen, UK).

A.

Component	Reaction
Restriction enzyme	2 µl
DNA	5 µg
Restriction enzyme buffer	5 µl
Nuclease-free water	up to 50 µl

B.

Component	Reaction
Restriction enzyme	1.5 µl each
DNA	5 µg
Restriction enzyme buffer	10 µl
Nuclease-free water	up to 50 µl

Table 2.5 Reaction condition for restriction digest. The table presents a reaction condition for single (A) and double (B) restriction enzyme digestion.

PCI-neo has been dephosphorylated for CDX1 to prevent self-ligation. Once the restriction reaction was finished, 6.1 µl of Cut Smart 10X (NEB, Ref. B7204S) and 5 µl of shrimp alkaline phosphatase (NEB, Ref. M0371S) were added to the restriction mix and incubated for an hour at 37°C.

2.5.3 Agarose gel electrophoresis

Linearised DNA fragments were analysed by agarose gel electrophoresis. A 1% agarose gel was prepared by melting 1 gram (g) of electrophoresis-grade agarose (Invitrogen, USA) in 100 ml 1X Tris-Acetate-EDTA (TAE, Fisher Scientific, Ref. BP1332-1) buffer. The solution was microwaved for two min while checking every 30 sec and swirling the flask to ensure all powder was dissolved. It was then allowed to cool before adding 10 µl of SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific, Ref. S33102). The solution was poured into a gel tray with appropriate comb sizes and set until solidified. Next, the comb was removed and solidified gel was transferred to a MINI SUB™ DNA CELL electrophoresis unit (Bio-Rad). The gel box was filled with 1X TAE buffer and 10 µl of GeneRuler DNA Ladder Mix (Thermo Fisher Scientific, Ref. SM0333) was carefully added to the gel. Each DNA sample was mixed with 6X TriTrack DNA Loading Dye (Thermo Fisher Scientific, Ref. R1161) before being loaded into the wells. Samples were run at 70-100 Volt until the dye line reached around 80% of the gel's length. In the final step, DNA fragments were visualised using a Fusion Fx gel imager (Vilber).

2.5.4 DNA extraction from agarose gel

Extraction of DNA fragments from the agarose gel was performed using the QIAquick Gel Extraction kit (Qiagen, Ref. 28704). The gel was put on the UV light Transilluminator (Alpha Innotech, LM-26E) and excised with a sterile and sharp scalpel. A gel slice was weighed, and three volumes of Buffer GC were added before being incubated for 10 min at 50°C. The gel slice was melted completely and mixed with one gel volume of isopropanol (Fisher Scientific, UK). After loading the sample on a QIAquick spin column, it was centrifuged at 17900 g for 1 minute and flow-through was discarded. In the following step, 500 µl of Buffer QG was added to the QIAquick column and centrifuged for a minute. The column was then washed with 750 µl of Buffer PE and centrifuged for 1 minute. To remove the residual buffer, the QIAquick spin column was centrifuged for another minute and transferred into a clean 1.5 ml microcentrifuge tube. For the elution of DNA, 30 µl of Buffer EB was added and the column was incubated for 1 minute before being centrifuged at 17900 g for 1 minute. Through nano-drop spectrometry, the quality and concentration of DNA were measured.

2.5.5 Ligations

The ligation reaction was performed at RT for 30 min as described in Table 2.6. The concentrations of digested vector and insert DNA needed for ligation were calculated using the NEB Cloning calculator (https://nebiocalculator.neb.com/#!/ligation).

Component	Reaction
T4 DNA Ligase Buffer (10X)	2 µl
Vector DNA	50 ng

Insert DNA	variable
T4 DNA ligase	1 µl
Nuclease-free water	up to 20 µl

Table 2.6 Reaction set-up for ligation. The amount of restriction-digested DNA insert was determined based on the insert size.

2.5.6 Bacterial transformation

For the transformation of plasmid DNA into *Escherichia Coli (E.Coli)*, 5 µl of ligation product was mixed with 50 µl of thawed NEB 5-alpha competent *E.Coli* (NEB, Ref. C2987H) cells and incubated for 5-10 min on ice. Next, the cells were heat shocked by placing the tubes in a 42°C water bath for 45-60 sec. The tubes were immediately transferred back into ice and incubated for 2-3 min. Afterwards, the cells were added to 1 ml of Lysogeny broth (LB) medium and incubated at 37°C for an hour at a speed of 220 rpm in a bacterial incubator shaker (New Brunswick™ Scientific, Innova® 44). The tube was centrifugated at 13000 g for 1 minute and most of the supernatant was removed. The remaining 100 µl sample was resuspended, plated on an LB agar plate containing the correct antibiotics for each plasmid and incubated overnight at 37°C.

2.5.7 Plasmid DNA extraction from *E.Coli*

A single colony was picked from an LB agar plate and transferred to 2 ml antibiotic-selective LB broth. After growing them overnight at 37°C in the incubator (Gallenkamp), plasmid DNA was extracted and purified using a QIAprep® Spin MiniPrep kit (Qiagen, UK). Briefly, we centrifuged the bacterial overnight culture at 6800 g for 3 min at RT. The bacterial pellet was resuspended in 250 µl P1 buffer and transferred to a new Eppendorf tube. This was followed by the addition of 250 µl P2

buffer and 4-6 times inversions of the tube. After lysing cells for approximately five min, 350 µl of N3 buffer was added to stop the reaction and immediately mixed by inverting the tube 4-6 times. Samples were centrifuged at 17000 g for 10 min at RT. A total of 800 µl of supernatant was applied to a QIAprep 2.0 spin column, which was centrifuged for 1 minute at 17000 g at RT. After washing the column with 500 µl of PB buffer, it was centrifuged again for one minute. The column was filled with 750 µl of PE buffer and centrifuged for one minute. A second centrifugation was carried out for 1 minute to remove any residual wash buffer from the column. For eluting the DNA, the column was put in a new Eppendorf tube and incubated in 30 µl EB buffer (10 mM TrisCl, pH 8.5) for 2 min before centrifugation for 1 min. The concentration and quality of the DNA were determined by Nano-Drop Spectrophotometry. The correct vector sequences for the DNA plasmids were verified by Sanger sequencing and restriction digestion as described in chapter section 2.5.2.

2.5.8 Sanger sequencing

Plasmid DNAs were sent for sequencing in the MRC WIMM Sequencing Facility to confirm the accuracy of our results. A summary of the primers used in sequencing is provided in Table 2.5. Sequence analysis and visual representation were performed using ApE software (A Plasmid Editor by M. Wayne Davis).

2.5.9 Isolation of high-quality plasmid DNA for in vitro experiments

As soon as the sequences were verified, we aimed to achieve high-quality and large quantities of plasmid DNA using the PureLink HiPure Plasmid Midiprep kit (Thermo Fisher Scientific, Ref. K210004).

Initially, a bacterial transformation was performed similarly to the procedure described in section 2.3.4. A small amount of plasmid DNA (1-2 µl) was added to 50 µl of competent E. Coli cells and incubated on ice for 5-10 min. The tubes were placed in a water bath at 42°C for 1 minute and then immediately moved to the ice for a few min. As a next step, the cells were incubated in 1 ml LB medium for one hour at 37°C in a bacterial incubator shaker. After an hour, the cells were poured into a flask containing 50 ml of LB medium along with 50 µl of specific antibiotics and allowed to grow overnight in a bacterial incubator shaker.

The next day, the midiprep procedure was carried out to obtain high quality and yield of plasmid DNA. Firstly, we applied 10 ml of equilibration buffer to the HiPure Midi Column and allowed it to drain by gravity. The overnight LB culture was centrifuged at 4000 g for 10 min and the supernatant was discarded. The cell pellet was resuspended in 4 ml of resuspension buffer with RNase until it became homogenous. Afterwards, the cells were mixed gently with 4 ml lysis buffer by inverting the tube five times and were incubated for 5 min at RT. During precipitation, 4 ml precipitation buffer was added to the cells and the mixture was homogenized by inverting the tube. The lysate was centrifuged at 13000 g for 10 min at RT and the supernatant was loaded onto the equilibrated column for binding. After washing twice with 10 ml wash buffer, the column was placed in a sterile 15 ml centrifuge tube. The purified DNA was eluted through gravity by applying 5 ml of elution buffer to the column. The eluate was mixed well with 3.5 ml isopropanol and centrifuged at 13000 g for 30 min at 4°C. Following the removal of the supernatant, 3 ml of 70% ethanol was added to the pellet and centrifuged at 13000 g for 5 min at 4°C. The supernatant was discarded, and the pellet was air-dried. The purified plasmid DNA was resuspended in 200 µl TE buffer, measured by Nano-Drop Spectrophotometry, and stored at -20°C.

2.6 Site-directed mutagenesis and screening

ATOH1 and SPDEF clones contained Myc and Flag tags before the stop codon. The tags were removed using site-directed mutagenesis as they may affect the 3D structure and function of the protein. Site-directed mutagenesis was performed using the In-Fusion® HD Cloning Kit (Takara Bio Inc., Ref. 639650).

Initially, a linearized vector was generated using a double digestion technique as shown in section 2.3.4. After digestion, the linearized vector was run in agarose gel and purified using the QIAquick Gel Extraction kit (Qiagen, Ref. 28704). Then, gene-specific primers containing 15 bp extensions (5' end of the primer) that are homologous to vector ends were designed in the In-Fusion Cloning Primer Design Tool (https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools). Genes of interest were amplified using Polymerase Chain Reaction (PCR). DNA template (10 ng) was added to a reaction master mix containing 12.5 µl of CloneAmp HiFi PCR Premix, relevant forward and reverse primers (0.3 µM each), and distilled water to a final volume of 25 µl per well and run using the following condition (Table 2.7).

Following PCR, the amplified target fragment was verified by agarose gel electrophoresis and purified. The In-Fusion Cloning procedure was conducted once a single band of the desired size had been obtained. For each sample, a total of 5 µl mix containing 100 ng of purified PCR fragment, 50 ng of linearised backbone vector, 1 µl of 5X In-Fusion HD Enzyme Premix, and distilled water was incubated for 15 minutes at 50°C. Negative and positive controls were included with each assay.

Step	Temp (°C)	Time	Number of cycles
Initial denaturation	94	5 min	
Denaturation	98	10 sec	35
Primer annealing	55	15 sec	
Extension	72	1 min	
Final extension	72	5 min	

Table 2.7 PCR cycling parameters. The table shows polymerase chain reaction conditions used in site-directed mutagenesis.

For screening, the new construct was transformed into Stellar™ competent cells (*E.Coli* HST08 strain, Takara Bio Inc., Ref. 636763). Competent cells were mixed with 2.5 µl of the In-Fusion reaction product and placed on ice for 30 min. The cells were heat-shocked for 1 min at 42°C and then kept on ice for 2 min. Following this, 1 ml of warm Super optimal broth with catabolite repression (SOC) medium was added to each tube and incubation was carried out by shaking at 220 rpm for 1 hour at 37°C. The samples were centrifuged at 13000 g for 1 min at RT and most of the supernatant was discarded. After resuspending the remaining 100 µl of the sample, it was plated on the LB agar plates containing the appropriate antibiotics for each plasmid and incubated overnight at 37°C.

On the next day, a single colony was selected from an LB agar plate and placed in 2 ml of antibiotic-selective LB broth. After growing them overnight at 37°C, plasmid DNA was extracted and purified exactly as described in section 3. To determine whether an insert had been inserted, restrictions digestion and DNA sequencing were performed.

2.7 Transient mammalian cell transfection

Genes of interest were exogenously overexpressed by adding DNA-Polyethylenimine (PEI) mix to the cells. The following mixes were prepared for a 96-well plate: 0.5 µg DNA diluted in 25 µl OPTIMEM, and 1.5 µg PEI (from 1 µg/ul stock) diluted in 25 µg OPTIMEM. After mixing diluted PEI with diluted DNA, the tube was incubated at room temperature for 30 min. The PEI-DNA mixture was then carefully added to each well of adherent cells and kept for 6 hours before replacing the medium with complete medium. Protein overexpression was analysed using immunofluorescence after 5 days.

2.8 Western blotting

2.8.1 Preparation of cell lysates and supernatants

Cells were plated in 6-well or 12-well plates and were cultured for 3 days until they reached 80% confluency. Prior to washing cells with 1X PBS once, the plates were kept on ice and the medium was collected. Cells were lysed in 50-200 µl of RIPA buffer (0.5% sodium deoxycholate, 150 mM NaCl, 1% Triton X-100, 50 mM Tris-HCL (pH 8.0), 0.1% sodium dodecyl sulphate (SDS)) depending on cell confluency for 30 min by pipetting up and down every 10 min. Cell lysates were collected and centrifuged at 10000 g (Eppendorf, Centrifuge 5417R) for 10 min at 4°C. Supernatants were carefully transferred into a new Eppendorf tube for Western blotting.

Protein secretion was examined by collecting the medium before washing the cells with PBS and centrifuging at 1000 g for 5 min. Supernatants were transferred to a new tube and prepared for Western blotting.

2.8.2 Bicinchoninic acid (BCA) assay

The concentration of protein collected from lysates and supernatant of cell lines was checked based on the BCA™ Protein Assay (Pierce, USA). Briefly, a set of protein standards (0-2000 µg/ml) were prepared from an ampule of 2 mg/ml Albumin Standard. The working reagent was made up of 50 parts of BCA Reagent A and 1 part of BCA Reagent B. Twenty-five µl of standards and protein samples in replicates were added to a 96-well plate and mixed with 200 µl of working reagent. Plates were covered, incubated at 37°C for 30 min and cooled to RT. Then, the absorbance was measured at 562 nm on a plate reader (µQuant, Bio-Tek Instrument, USA). To identify each sample's concentration, a standard curve was calculated after subtracting a blank standard measurement.

2.8.3 SDS-polyacrylamide (SDS-PAGE) gel electrophoresis

As part of sample preparation, a reducing/denaturing condition for TFF3 was applied as follows: 10 µl of Dithiothreitol (DTT), 20 µl of SDS sample buffer (0.25 M Tris-HCl (pH 6.8), 10% SDS, 50% glycerol, and a small grain of bromophenol blue), and 70 µl of lysate/supernatant. Samples were then heated at 90°C for 5 min and loaded alongside molecular markers into the gel wells. A non-reducing condition was used for MUC2, which excludes DTT and sample heating.

Depending on the size of the protein, Tris-glycine (Novex™ WedgeWell™ 14%) and Tris-Acetate (NuPAGE 3 to 8%) gels were used. Samples were run in 1X Running buffer in a Mini-PROTEAN®-II vertical electrophoresis unit (Bio-Rad, UK) at 130 V with approximately 20-25 µl of sample per lane. A running buffer was prepared according to the gel type (for 20X Tris-Acetate SDS Running buffer: 89.5 g of Tricine, 60.6 g of

Tris base, 10.0 g of SDS fill up to 500 ml with ultrapure water; for 1X Tris-Glycine SDS-PAGE Running buffer, 100 ml of 10X Running buffer (144.0 g of glycine, 30.0 g of Tris base, fill up to 1000 ml), 900 ml of ultrapure water and 5 ml of 20% SDS).

2.8.4 Immunoblotting

SDS-PAGE gel electrophoresis was followed by wet transfer (tank transfer) of the proteins to polyvinylidene difluoride (PVDF) membrane (Amersham Biosciences, USA). Initially, the membrane was activated in 100% methanol, washed briefly with distilled water, and then immersed in 1X Transfer buffer (for 3-8% Tris-Acetate gels: 50 ml of 20X Transfer buffer (52.4 g of Bis-Tris, 40.8 g of Bicine, 3.0 g of EDTA in 500 ml of water), 100 ml of methanol, 850 ml of water; for 14% Tris-Glycine gels: 50 ml of 10X Tris-Glycine SDS Running buffer, 200 ml of methanol, 750 ml of water). After that, the membrane was compressed in a transfer buffer-soaked sandwich of sponge-paper-membrane-gel-paper-sponge and the air bubbles were removed. The gel should be positioned closest to the negative electrode, and the membrane closer to the positive electrode. Proteins were transferred for 60 min at 115 V and subsequently visualized using Ponceau Red Stain (0.1% Ponceau S (Sigma-Aldrich) in 5% glacial acetic acid) for a short time.

Following a rinse with MilliQ water, the membrane was blocked in 4% semi-skimmed milk dissolved in 1X PBS for an hour at RT. The membrane was incubated with the primary antibody diluted in PBST buffer (0.1% Tween-20 in 1X PBS) containing 4% semi-skimmed milk overnight at 4°C. After three washes (10 min each) with PBST, the membrane was incubated with the HRP-conjugated secondary antibody (1:3000) diluted in PBST buffer consisting of 4% milk for an hour at room temperature. For the

final step, the membrane was washed three times (10 min each) with PBST before being incubated with ECL™ Western Blotting Detection Reagents (VWR International Ltd, RPN2209) for 3-5 min. The blot was detected in a dark room with X-ray films (CL-XPosure™ Film, Life Technologies Ltd) processed by the ECOMAX X-Ray Film Processor (Protec).

2.8.5 Antibodies used in Western blotting

A list of antibodies used in Western blotting is presented in Table 2.8.

1° Ab	Species	Source and catalogue number	Dilution	2° Ab	Source and catalogue number	Dilution
MUC2 (PR5D5)	mouse	In-house	1:250	GAM-HRP	Bio-Rad 1706516	1:3000
MUC2 (CCP58)	mouse	Cell Signaling Tech. #88686	1:2000			
Monoclonal TFF3	rabbit	Abcam ab108599	1:1000	GAR-HRP	Bio-Rad 1706515	1:3000
Polyclonal TFF3	rabbit	Raybiotech, Inc 130-10147	1:500			
Polyclonal HSP90	rabbit	Cell Signaling Tech. #4874	1:1000			

Table 2.8 Summary of antibodies used in Western blotting. GAM-HRP: Goat-anti mouse IgG (H+L)-Horseradish peroxidase (HRP) conjugate, GAR-HRP: Goat-anti rabbit IgG (H+L)-HRP conjugate.

2.9 Immunofluorescence and immunohistochemistry methods

2.9.1 Immunofluorescence staining in CRC cell lines

An appropriate number of cells was seeded on the plate and allowed to grow for three days before carrying out immunofluorescence staining. The medium was removed after 72 hours, and the cells were washed twice with 1X PBS. Afterwards, fixation was performed by using 4% Paraformaldehyde (PFA, Thermo Scientific™, Ref. 28908) for 20 min, followed by three washes with 1X PBS. The cells were permeabilized with 0.2% Triton X-100 (Sigma) for 5 min at RT and washed three times using 1X PBS. To reduce non-specific interactions, the cells were blocked with 1% bovine serum albumin (BSA, Fisher BioReagents™, Ref. BP9700-100) for an hour at RT. Following this, the primary antibody was diluted in 1% BSA in PBS and incubated at room temperature for an hour or overnight at 4°C. The cells were washed three times with 1X PBS and incubated with Alexa Fluor-conjugated secondary antibodies for 30 min. After the final washes and incubation with 4′,6-diamidino-2-phenylindole (DAPI-10 mg, Invitrogen™, Ref. D1306) at 1:10000 dilution for 10 min, the samples were imaged.

2.9.2 Immunofluorescence staining in FFPE human tissue sections

Formalin-fixed paraffin-embedded (FFPE) human tissue sections were collected from normal and cancer patients for the study titled “Investigating molecular markers in samples collected from patients who took part in QUASAR 2, VICTOR and SCOT trials”. Human tissue sections were obtained from Dr. David Church (Wellcome Centre for Human Genetics, University of Oxford), and approved by the North West – Liverpool Central Research Ethics Committee.

The immunofluorescence analysis was carried out on 4 normal and 21 colorectal cancer FFPE human tissue sections. We first heated the slides at 60°C for an hour so that the paraffin loosens from the tissue. Then, tissue sections were deparaffinized in xylene three times for 5 min and rehydrated in 100% ethanol/Industrial Methylated Spirit (IMS), 90% ethanol/IMS, and 70% ethanol/IMS, respectively, for 5 min each. The slides were rinsed with distilled water three times before an antigen retrieval step. The antigen unmasking process was carried out using a heater containing a beaker filled with 1X citrate buffer (10 mM citrate buffer, pH 6.0) and the slides were sub-boiled for 10 min inside the beaker. The slides were then allowed to cool at room temperature in the antigen retrieval buffer for 20 min. Afterwards, they were washed twice in 1X PBS and permeabilized with PBSTX buffer (0.1% Triton X-100 in 1X PBS) for 10 min at RT before being rinsed in 1X PBS. The slides were blocked with 5% BSA in 1X PBS for an hour, then incubated overnight with primary antibodies diluted in 1% BSA in 1X PBS. Following three rinses with PBST buffer (0.1% Tween-20 in 1X PBS) for 5 min, the slides were incubated in the humidifier for an hour with secondary antibodies diluted in 1% BSA in 1X PBS. After three times washing with PBST buffer, mounting medium containing DAPI (Fluoromount-G™ with DAPI, Invitrogen, Ref. 00-4959-52) was applied before placing a coverslip, and the slides were then scanned.

2.9.3 Immunohistochemistry staining in CRC cell lines

The CRC cells were plated in an 8-well chamber slide until they reached 70-80% confluency. For immunohistochemistry (IHC) staining, they were washed once with 1X PBS and fixed with 4% PFA for 30 min at RT. The cells were incubated in 3% hydrogen peroxide for 20 min before being washed with 1X PBS. For blocking, 5% BSA diluted in 1X PBS was used for an hour, and then the cells were incubated with primary

antibody diluted in 1% BSA in 1X PBS overnight at 4°C. Following 3 washes with PBST buffer, a horseradish peroxidase (HRP) conjugated secondary antibody was applied for an hour at RT. After the removal of the secondary antibody, the cells were washed two times with PBST and once with 1X PBS for 5 min each. HRP-catalysed oxidation of 3,3'-diaminobenzidine (DAB) by hydrogen peroxide was detected using a Pierce™ DAB Substrate Kit (Thermo Scientific, Ref. 34002). To prepare the 10X DAB solution, 2.5 ml of DAB was thoroughly mixed with 22.5 ml of Stable Peroxide Substrate buffer. A brow' precipitate was formed by the oxidised DAB at the HRP location and was visible under a light microscope. The signal was observed under a light microscope and the cells were then rinsed with distilled water before being mounted and scanned.

2.9.4 Immunohistochemistry staining in FFPE human tissue sections

The IHC analysis was performed on 1 normal (obtained from Dr. Shazia Irshad) and 3 colon cancer (provided by Dr. David Church) FFPE sections. The tissue sections were first heated at 60°C for an hour and deparaffinized in xylene three times for 5 min each. Afterwards, the slides were rehydrated in 100%, 90%, and 70% ethanol for 5 min, respectively, and rinsed in distilled water three times for 5 min. The slides were treated with 3% hydrogen peroxide for 20 min at room temperature to block endogenous peroxidase activity and then washed once with distilled water. The antigen retrieval process was performed by sub-boiling the slides in 1X citrate buffer (10 mM citrate buffer, pH 6.0) for 10 min and the slides were then cooled at the same buffer for 20 min at RT. As a next step, the slides were rinsed twice in 1X PBS, permeabilized with PBSTX for 10 min at RT, and washed once in 1X PBS once for 5 min. The steps for blocking, primary and secondary antibody incubation are identical

to those in IHC on CRC cell lines. After the washing steps, the DAB solution was prepared as shown in the previous section and used immediately. The signal was developed within 2-7 min before being washed with distilled water. The slides were then dehydrated in 70%, 90%, and 100% ethanol for 10-15 sec each, and air-dried for 30 min. As a final step, the slides were mounted on Fluoromount-G™ with DAPI (Invitrogen, Ref. 00-4959-52) and scanned after the coverslips had been applied.

2.9.5 Haematoxylin and eosin staining in FFPE human tissue sections

A total of 4 normal and 21 colon cancer human FFPE sections were stained using Haematoxylin and Eosin (H&E) Staining Kit (Abcam, Ref. ab245880) according to the manufacturer's instructions at the Translational Histopathology Laboratory (University of Oxford).

Briefly, the FFPE sections were first deparaffinized and hydrated with distilled water. The tissues were then counterstained with Mayer's Haematoxylin (Lillie's Modification) and incubated for 5 min. Excess stain was removed from slides by rinsing them twice in distilled water. Next, the tissue sections were thoroughly covered in Bluing Reagent for approximately 10-15 sec and washed twice with water. After dipping the slides in absolute alcohol and blotting off the excess, they were incubated with Eosin Y Solution (Modified Alcoholic) for 2-3 min. The slides were rinsed with absolute alcohol and dehydrated through three changes of absolute alcohol. As a final step, we mounted the slides and scanned them on a PhenoImager scanner from AKOYA Biosciences.

2.9.6 Antibodies used in IF and IHC staining

Table 2.9 provides an overview of antibodies used for IF and IHC staining.

1° Antibody	Source and Catalogue number	Species	Dilution	Application
MUC2 (PR5D5)	In-house	mouse	1:250	IF
TFF3 monoclonal	Abcam Ab108599	rabbit	1:1000	IHC, IF
TFF3 polyclonal	Raybiotech, Inc 130-10147	rabbit	1:500	IF
EpCAM (AUA1)	In-house	mouse	1:500	IF
ATOH1	Atlas Antibodies HPA049400	rabbit	1:500	IF
c-Myc AF488 conjugated	Santa Cruz Biotechnology sc-40 AF488	mouse	1:500	IF
2° Antibody	**Source and Catalogue number**	**Species**	**Dilution**	**Application**
Alexa Fluor 488	Thermo Fisher Scientific A11001	GAM	1:500	IF
Alexa Fluor 568	Thermo Fisher Scientific A10042	DAR	1:500	IF
Alexa Fluor 647	Thermo Fisher Scientific A31571	DAM	1:500	IF

Table 2.9 Summary of antibodies used in IF and IHC staining. GAM: Goat anti-mouse, DAR: Donkey anti-rabbit, DAM: Donkey anti-mouse.

2.10 Microscopy

Immunofluorescence imaging was performed at 20X, 40X, 100X, and 5X magnifications using a Zeiss Axio Observer Z1 inverted fluorescence, Zeiss LSM 880 confocal laser scanning, and Zeiss spinning disk confocal microscopes (WIMM, University of Oxford). The H&E-stained tissue sections were scanned on a PhenoImager scanner from AKOYA Biosciences (Old Road Campus, University of Oxford). The FFPE tissue sections and 8-well chamber slides stained with IHC were imaged using Olympus BX60 microscope (WIMM, University of Oxford).

2.11 Lumen formation assay in 3-dimensional (3D) culture

The procedure was performed in Matrigel (Corning, Ref. 354234) as a base for a 3D cell culture system according to Ashley et al. (2013). Firstly, Matrigel was mixed with 1:1 ice-cold DMEM and a Matrigel/DMEM layer (40 µl per well in a 96-well plate) was allowed to solidify for 30 min at 37°C. Five hundred single cells were obtained by filtration through 20 micrometre (µm) filters (Celltrics, Partec, Germany) and were loaded to 20 µl of 1:1 DMEM/Matrigel mixture. Matrigel-coated cells were then incubated at 37°C for an additional 30 minutes before the culture medium was added. The cells were grown at 37°C for two weeks, with the medium being changed every three days.

IF staining was initiated by removing the medium and adding 100 µl of 4% PFA in PBS for 20 minutes. The fixed cells were washed with PBS and permeabilised in 0.2% Triton X-100 for 10 minutes. After washing the cells 3 times with 50 mmol/L glycine in PBS, the colonies were stained with primary (MUC2 and TFF3) and secondary antibodies. The colonies were then rinsed 3 times with PBS and a DAPI solution was added to each well. As a final step, the plate was imaged using a Zeiss Axio Observer Z1 inverted fluorescence.

2.12 Antibody purification and conjugation

MUC2 (PR5D5) and EpCAM (AUA1) antibodies were purified before being labelling with Alexa Fluor dyes. Initially, the antibodies were measured by Nanodrop and diluted in 10X PBS at a 1:1 ratio. Next, 0.5 ml of the immobilised Protein G agarose (Pierce™ Protein G Agarose, Thermo Scientific™, Ref. 20397) was pipetted into a 15 ml tube and mixed with 10 ml of 1X PBS until a homogenous solution was formed. The mixture

was centrifuged at 500 g for 1 min and the liquid was removed with a needle. A second round of washing was also performed. A diluted primary antibody was added to Protein G and placed on a rotor in the cold room for 30 min. Following centrifugation at 500 g for 1 min, the mixture was washed twice with 1X PBS, and the supernatant was removed. Two ml of the Pierce™ IgG elution buffer (Thermo Scientific™, Ref. 21004) was added, mixed well, and centrifuged at maximum speed for 1 min. The supernatant was transferred to a clean tube and 300 µl Tris solution (1M, pH 8, Santa Cruz Biotechnology, Ref. sc-296649) was added. The resin was resuspended with the elution buffer (1 ml) and the tubes were placed on the rotor for 5 min. After a centrifugation step, the supernatant was collected and immediately mixed with 150 µl Tris. The collected antibody was measured and stored at 4°C. The antibodies (500 µl/per cassette) were loaded into Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific™, Ref. 66383) to enable buffer exchange and the removal of salts and small contaminants. The cassettes were maintained in 1X PBS by changing the buffer every few hours. After 3 rounds of buffer exchange, the purified antibodies were collected and transferred to new tubes.

For labelling the protein, APEX™ Alexa Fluor® 647 (Invitrogen™, Ref. A10475) and Alexa Fluor™ 488 (Invitrogen™, Ref. A20181) Antibody Labeling kits were used. A 1 molar (M) solution of sodium bicarbonate was first prepared by resuspending Component B (sodium bicarbonate) in 1 ml of distilled water. The antibody concentration was adjusted to 1.0 mg/ml before adding 1/10th volume of the sodium bicarbonate solution. Then, 100 µl of the antibody solution was gently mixed with the vial of Alexa Fluor® dye and incubated for 1 hour at RT. The vial was gently inverted several times throughout the labelling process. In the meantime, the purification column was prepared, and 1.5 ml of resin bed was formed using Component C

(purification resin). The column was then centrifuged at 1100 g for 3 min. The reaction mix was added to the column and centrifuged at 1100 g for 5 min to obtain the labelled antibody.

2.13 Single-cell clonal selection analysis

Serial dilutions and fluorescence-activated cell sorting (FACS) were performed to identify the time-dependent expression of MUC2 and TFF3. A combination of the BD FACS Aria™ III and Fusion was used to carry out single cell sorting using a 100 µm nozzle at 20 pounds per square inch (psi). For serial dilutions, a single cell in 100 µl medium was seeded in 96-well plates. The cells were fixed at different time points and analysed by IF.

For the time-dependent experiment, an appropriate number of the cells (day 1: 20.000 cells/well; day 3: 15.000 cells/well; day 5: 10.000 cells/well; day 7: 5.000 cells/well; day 9: 2.500 cells/well) were passed through a 20 µm filter (Celltrics, Partec, Germany) and seeded in 96-well plates. Cell mixes were prepared by using a series of sequential dilutions.

2.14 Fixed and Recovered Intact Single-Cell RNA (FRISCR) sequencing

Single-cell mRNA sequencing (scRNA-seq) was conducted to interpret patterns of expression of MUC2, TFF3, and other genes that may explain the intra-cell line variations. A 50:50 mix of MUC2-positive (LS180) and -negative (COLO678) cell lines were chosen, sorted, and prepared for sequencing according to the protocol previously published (Thomsen et al., 2016) and fixed and recovered intact single-cell RNA (FRISCR) sequencing was performed by Neil Ashley (Sequencing Facility,

WIMM). There were 50.000 barcodes (single cells) and 25.345 Unique Molecular Identifiers (UMIs) (genes) returned as output, representing the gene expression level in counts. A previous DPhil student of the Bodmer Laboratory, Dr Jeff (Ta-Chun) Liu developed a new GMMchi-based single-cell sequencing pipeline which allows a better separation of cell types of similar origin. A detailed description of the pipeline was provided in Dr Jeff''s DPhil thesis (Liu, 2021).

2.15 Computational and statistical analysis

Tables and statistical tests were analysed and visualised by the GraphPad Prism version 9.5.0 software (California, USA, www.graphpad.com). P-values were obtained using a t-test from biological replicates. $P \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, $p \leq 0.0001$, $p > 0.05$ were denoted by *, **, ***, **** and 'NS' respectively.

Microarray data from 79 CRC cell lines were analysed by Gaussian Mixture Model (GMM). Python (GMMchisquare library) and R programming languages were used to process the microarray data. Details about GMMchi are described in the literature (T. C. Liu et al., 2022), and the Python package can be downloaded from GitHub at https://github.com/jeffliu6068/GMMchi. I am grateful to Dr Jeff (Ta-Chun) Liu for making his GMM with a tail-fitting algorithm available for use in this thesis.

The molecular profiles of CRC patients and normal samples were compared using the TCGA RNA-seq data, which contained 689 CRC and normal colon samples (51 non-tumour and 638 tumour samples). Furthermore, the scRNA-seq data that was generously shared by Professor Alison Simmons was analysed in addition to our LS180-COLO678 single-cell data (Parikh et al., 2019).

Fiji/ImageJ (https://imagej.net/software/fiji/) and QuPath (https://qupath.github.io/) software were used to visualise IF and IHC images. Quantification of MUC2/TFF3-positive cells was performed using the CellProfiler 3.0.0 software (https://cellprofiler.org/). Sequencing results were visualised using the ApE (https://jorgensen.biology.utah.edu/wayned/ape/) software. DNA sequences of genes of interest were obtained from UniProt (https://www.uniprot.org/).

CHAPTER 3

CHARACTERISATION OF HUMAN CRC CELL LINES

CHAPTER 3: CHARACTERISATION OF HUMAN CRC CELL LINES

3.1 INTRODUCTION AND AIMS

Cell lines are valuable in vitro models to study biological processes and identify novel markers for various diseases. They can also be used to assess hypotheses for personalised cancer therapy and drug toxicity in preclinical studies. Cell lines provide many advantages, such as affordability, ease of use and propagation, abundant supply, and lack of ethical concerns regarding animal and human samples. Several studies have demonstrated that cancer-associated changes identified in tumour tissues can be efficiently mapped in cancer cell lines, making them efficient models for understanding cancer progression and invasion (Iorio et al., 2016; Mirabelli et al., 2019). While cancer cell lines imitate a variety of mutations, chromosomal changes, mRNA expression and DNA methylation patterns observed in cancer, there are no genetic changes in major driver genes after prolonged in vitro cultivation (Mouradov et al., 2014; Wilding & Bodmer, 2014). It is important to acknowledge that cross contamination or mislabelling of cell lines remains a serious issue.

Consensus molecular subtype (CMS) classification based on differential gene expression profiles is widely used to identify tumour subtypes in CRC. In this classification, CRC is divided into four subtypes (CMS1-4), each with different molecular and biological characteristics (Ten Hoorn et al., 2022; Valenzuela et al., 2021). CMS1 is characterised by immune activation, microsatellite instability, and high frequency of BRAF mutations; CMS2 is associated with Wnt and MYC signalling activation; CMS3 exhibits epithelial characteristics with enriched KRAS mutations and metabolic abnormalities; and CMS4 has mesenchymal features with angiogenesis and stromal invasion (Ten Hoorn et al., 2022; Valenzuela et al., 2021).

Several studies have demonstrated that CMS3 subtype colon tumours are highly enriched with goblet cell markers (Buechler et al., 2020; Miller et al., 2021; Ragulan et al., 2019). Particularly, mucinous adenocarcinomas are only associated with high expression of MUC2 compared to CMS3-classified tumours that display increased expression of all goblet cell markers, such as MUC2, TFF3, and FCGBP (Miller et al., 2021). It is possible that metabolic pathways are stimulated in CMS3 subtype tumours because of the high metabolic demands of goblet cells, or that the factors released by goblet cells might also trigger metabolism in non-goblet cells (Miller et al., 2021).

Single cancer stem cells originated from certain CRC cell lines can form crypt-like structures ("lumens") under 3D conditions in a Matrigel-based assay (Ashley et al., 2013; Richman & Bodmer, 1988; Yeung et al., 2011; Yeung et al., 2010). These structures consist of cell-free central lumens surrounded by polarised columnar cells similar to crypts in the normal colon. Lumen formation is characterised by cellular polarisation, expression of cytoskeletal proteins, brush border markers, and tight junctions (Ashley et al., 2013). A tumour's ability to form lumens can be considered to assess its differentiation pattern with respect to enterocytes.

Lumen-forming colonies maintain self-renewal ability, as well as being able to give rise to three differentiated cell types (Ashley et al., 2013). CRC cell lines that are unable to make lumens contain poorly differentiated cells. For example, HCT116, DLD1, and HT29 CRC cell lines were identified as non-lumen-forming cell lines (Ashley et al., 2013). Compared to lumen-forming colonies, cells derived from non-lumen colonies showed enhanced tumorigenic ability in mouse xenografts due to lack of differentiation (Ashley et al., 2013; Yeung et al., 2010). It was identified that hypoxia maintains the clonogenicity of cancer stem cells and inhibits cell differentiation (Yeung et al., 2011).

The hypoxic condition (1% O_2) also prevented lumen formation in CRC cell lines, such as LS180, CCK81, and SW1222, which can differentiate into lumen-forming colonies under normoxia (Yeung et al., 2011). During hypoxia, these cell lines developed small and non-lumen forming colonies similar to those found in HCT116, where nearly all cells are cancer stem cells (Yeung et al., 2011).

Despite the importance of mucins and goblet cells, there is no clear classification of colorectal tumours in terms of goblet cell differentiation. It creates a gap in understanding of cell differentiation within colorectal cancers, as well as the mechanisms involved in goblet cell differentiation. To address this issue, **Chapter 3** aims to categorise a large panel of human CRC cell lines in terms of goblet cell differentiation by using immunofluorescence and microarray expression data. Screening of CRC cell lines was done using in-house PR5D5 antibody against MUC2 and commercial Anti-TFF3 (Raybiotech, Inc). Furthermore, **Chapter 3** explores the correlation between goblet cell differentiation and lumen formation. The pattern of cell differentiation in various time points was also tracked by using single cells. This will enable us to define important time points and establish a clear hierarchy between MUC2 and TFF3 expression.

3.2 RESULTS

3.2.1 Categorisation of CRC cell lines based on MUC2 and TFF3 mRNA expression levels

Microarray data from 79 CRC cell lines were analysed using software developed in our lab called GMMchi, which converts continuous expression data into categorical regions by optimising a Gaussian Mixture Model (GMM) on the data. GMM is a probabilistic model that uses an expectation-maximization algorithm to fit either one or two normal distributions to the observed expression levels across the whole cell line panel for any given gene. It detects when there are non-normally distributed tails to an observed distribution, and eliminates observations in a tail, starting from the most extreme outlier, until the remaining distribution can be fitted by either one or two normal distributions. When the fit involves two normal distributions, the threshold for separating high and low expression is determined by the point where the two fitted normal distributions intersect. A more comprehensive description of the GMMchi pipeline can be found in Dr Jeff Liu's thesis and recently published paper (Liu, 2021; T. C. Liu et al., 2022). A Python library was developed by Dr Jeff Liu to enhance the practicality and accessibility of the GMMchi pipeline (https://github.com/jeffliu6068/GMMchi).

The initial analysis was conducted on our microarray data by assessing the gene expression levels of the primary goblet cell-associated markers, namely MUC2 and TFF3. Sixteen cell lines showed *MUC2* expression, of which the top three (CL40, RW7213, and NCIH508) exhibited the highest levels of *MUC2* gene expression (Figure 3.1). In Figure 3.1, the continuous line represents a bimodal distribution. The fraction of CRC cell lines expressing *MUC2* is about around 23% of the total as

depicted in Figure 3.1. Based on the GMMchi-based cut-off threshold ($2^{7.08}$), two groups were categorised in Bodmer microarray data: *MUC2*-positive/high and *MUC2*-negative/low cell lines (Table 3.1).

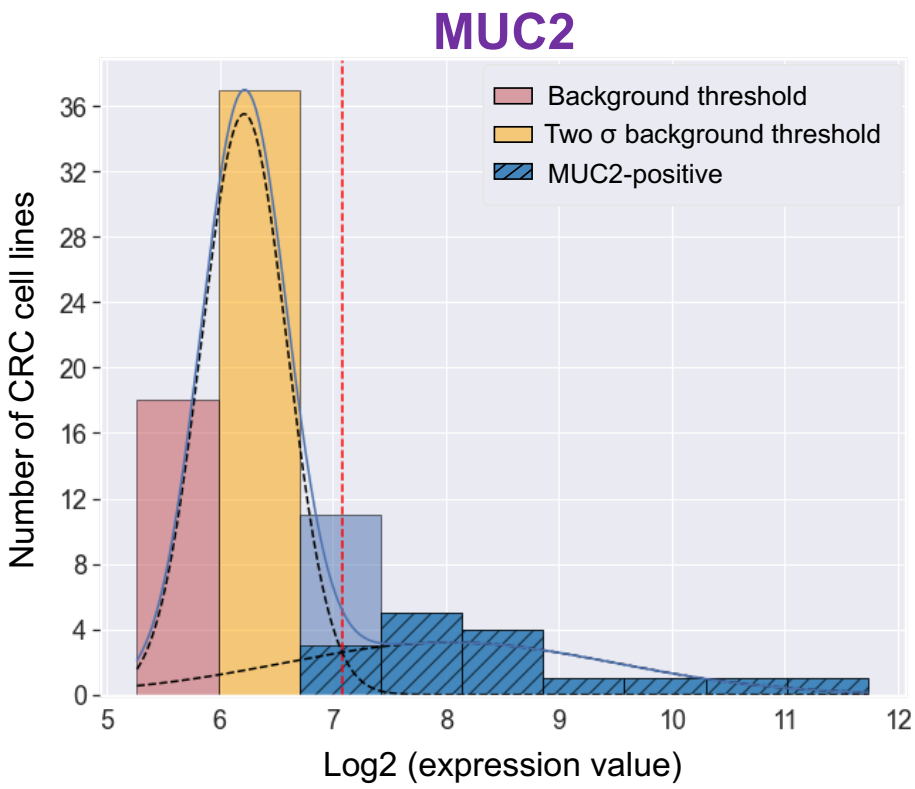


Figure 3.1 *MUC2* gene expression in CRC cell lines. Histogram shows *MUC2* expression levels in our microarray data analysis for a panel of 79 CRC cell lines. The x-axis of the histogram represents gene expression levels based on our microarray data measured on the log2 scale, whereas the y-axis displays the number of CRC cell lines corresponding to the given expression levels. The continuous curve is the bimodal mixed Gaussian distribution. The yellow bar indicates near background levels, while the pink bar refers to the background threshold detected across a panel of cell lines. The vertical dotted red line delineates the two groups: *MUC2*-positive and *MUC2*-negative (or low) cell lines at the value $2^{7.08}$.

MUC2-positive		MUC2-negative cell lines					
C125PM	LOVO	C10	CL14	HDC114	JHCOLOY1	PMFK014	SW48
C80	LS174T	C106	COLO201	HDC135	LIM1863	RCM1	SW480
C84	LS180	C32	COLO320DM	HDC142	LS1034	RKO	SW837
CL40	LS411	C70	COLO678	HDC57	LS123	RW2982	SW948
HCA46	LS513	C75	CW2	HDC8	NCIH716	SKCO1	VACO10MS
HDC111	NCIH508	C99	DLD1	HDC82	NCIH747	SNUC1	VACO429
HDC143	RW7213	CACO2	GP2D	HDC9	OUMS23	SNUC2B	VACO4A
HDC73	T84	CAR1	HCA7	HRA19	OXCO1	SW1116	VACO5
		CC20	HCC2998	HT29	OXCO2	SW1222	
		CCK81	HCC56	HT55	OXCO3	SW1417	
		CCO7	HCT116	ISCEROL	PCJW	SW403	

Table 3.1 Categorisation of CRC cell lines based on *MUC2* mRNA microarray expression. The table describes 79 CRC cell lines categorised as *MUC2*-positive and *MUC2*-negative cell lines according to GMMchi characterisation on our microarray dataset.

Following that, a *TFF3* gene expression analysis was carried out, and the results are presented in Figure 3.2. In the figure, the dotted curves demonstrate the two fitted normal distributions, while the continuous line represents the fitted mixed binormal distribution. Compared to *MUC2*-positive cell lines, there are more CRC cell lines that express *TFF3* (45 CRC cell lines versus 16 CRC cell lines).

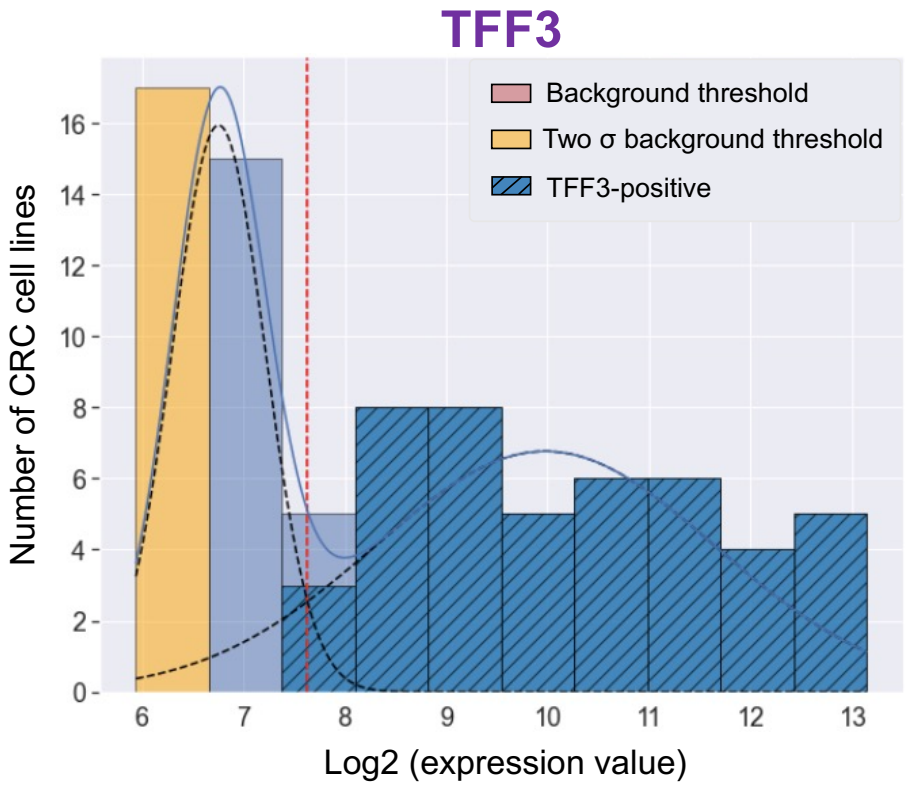
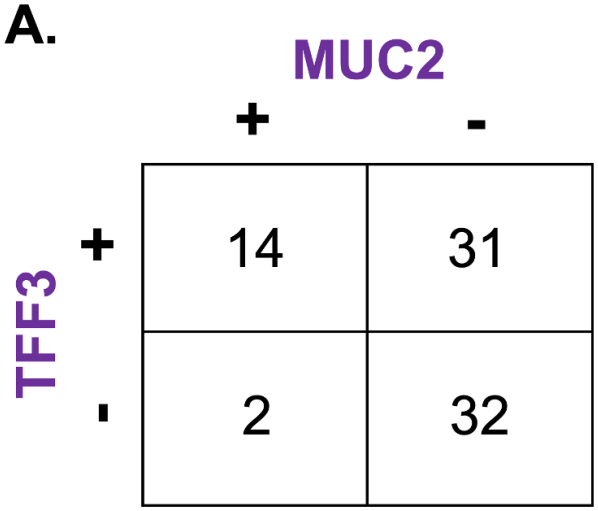


Figure 3.2 *TFF3* gene expression in CRC cell lines. Histogram illustrates *TFF3* mRNA expression levels across a panel of 79 CRC cell lines. In the figure, the x-axis shows the log2 transformation of gene expression values according to our microarray data, whereas the y-axis indicates the number of CRC cell lines. The vertical dotted red line separates the two groups of cell lines, *TFF3*-positive and *TFF3*-negative/low, based on the two distinct fitted normal distributions with a threshold value of $2^{7.63}$. The yellow bar refers to near background levels, based on a separate analysis of the overall pattern of gene expression observed in the cell lines.

We then studied the correlation between *MUC2* and *TFF3* gene expression in our microarray data. Table 3.2A describes a 2x2 table analysis that provides a pairwise comparison between *MUC2* and *TFF3*, and Table 3.2B shows the different groups according to the expression levels of both *MUC2* and *TFF3*. A total of 14 cell lines exhibited both *MUC2* and *TFF3* expression (*MUC2*+*TFF3*+). Interestingly, *TFF3* is expressed without *MUC2* in 31 CRC cell lines (*MUC2*-*TFF3*+), however, only two cell lines (*MUC2*+*TFF3*-), namely C125PM and LOVO, demonstrate *MUC2* expression

without apparent *TFF3*. *MUC2-TFF3+* cell lines are a major novel subset that needs further characterisation and to see whether this subset clearly exists, as would be expected, in primary tumour tissue and whether such CRCs can be considered incompletely or partially differentiated goblet cell-containing cell lines or cancers. It is also important to consider what the implications of this characterisation of CRCs are for clinical outcome.

A.



Chi-squared p-value = 9.38E-03

B.

MUC2+ TFF3+		MUC2+ TFF3-	MUC2- TFF3+			MUC2- TFF3-		
C80	LS174T	C125PM	C106	HT29	SKCO1	C10	HCA7	OUMS23
C84	LS180	LOVO	C70	HT55	SNUC1	C32	HCC2998	OXCO1
CL40	LS411		C99	JHCOLOY1	SNUC2B	C75	HCT116	OXCO3
HCA46	LS513		CL14	LIM1863	SW1222	CACO2	HDC114	RKO
HDC111	NCIH508		COLO201	LS1034	SW403	CAR1	HDC135	SW1116
HDC143	RW7213		CW2	NCIH716	SW948	CC20	HDC8	SW1417
HDC73	T84		Gp2D	OXCO2	VACO10MS	CCK81	HDC82	SW48
			HCC56	PCJW	VACO429	CCO7	HRA19	SW480
			HDC142	PMFK014	VACO4A	COLO320DM	ISCEROL	SW837
			HDC57	RCM1		COLO678	LS123	VACO5
			HDC9	RW2982		DLD1	NCIH747	

Table 3.2 Chi-square analysis and categorisation of CRC cell lines in terms of *MUC2* and *TFF3* mRNA expression detected by microarray analysis. (A) A 2x2 table shows the number of cell lines that fall into each category. (B) A total of 79 CRC cell lines were divided into 4 categories, namely *MUC2+TFF3+*, *MUC2+TFF3-*, *MUC2-TFF3+*, *MUC2-TFF3-*.

We also used RNA-seq data of CRC cancer cell lines in the Cancer Genome Atlas (TCGA) to identify expression levels of various goblet cell-associated genes. TCGA is a large database that includes molecular analyses of more than 20,000 primary cancer samples, together with normal samples. One major drawback of the TCGA is the inclusion criteria for clinical samples. It is essential for tumour samples to contain at least 80% tumour nuclei, which implies that up to 20% of patient samples may contain non-tumour tissue, including normal healthy tissue, immune cells, fibroblasts, endothelial cells, and blood cells (Aran et al., 2015).

Our lab developed a model to eliminate non-tumour expression from bulk expression RNA-seq data in the TCGA. Detailed information about the pipeline can be found in Dr Jeff Liu's thesis (Liu, 2021). Briefly, copy number variation in CRC tumours and the cell-type composition of the paired normal expression data were used as reliable indicators to distinguish normal tissue from cancerous samples. After GMMchi-based TCGA purification, we obtained a total of 637 CRC samples and 51 non-tumour normal samples.

Figure 3.3 illustrates that *MUC2* and *TFF3* genes are highly expressed in normal samples from the TCGA dataset. The distribution of *MUC2* gene expression exhibits a clear bimodal pattern, indicating distinct positive and negative groups. *TFF3* gene expression demonstrates a positive unimodal distribution with a tail, indicating that the majority of CRC samples exhibit *TFF3* expression. This observation correlates with the findings from our microarray data.

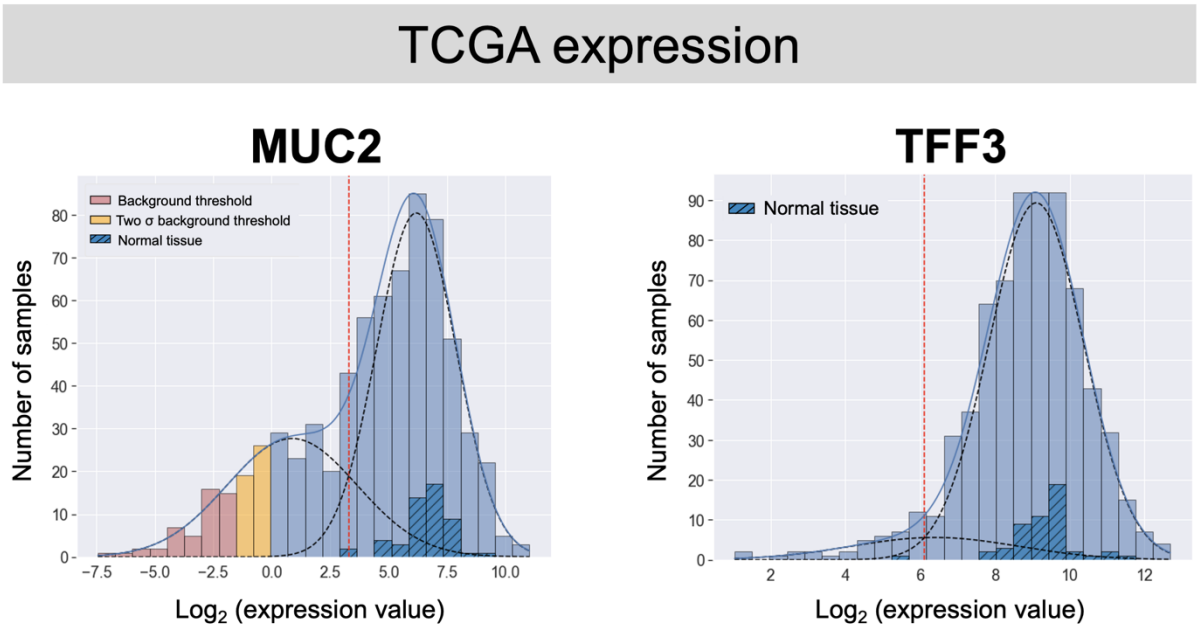


Figure 3.3 *MUC2* and *TFF3* gene expression in the TCGA dataset. The histograms depict the distribution of *MUC2* and *TFF3* gene expression in combined paired CRC tumours and normal samples from the TCGA dataset. The x-axis shows the log2-transformed values of gene expression, while the y-axis represents the number of normal and CRC samples. In the figure, the overall distribution in the lighter blue includes both normal and tumour specimens. Hatched healthy tissue highlighted in dark blue was defined based on our analyses (Liu, 2021).

Next, a 2 by 2 contingency table (2x2) was produced to define binary correlations between *MUC2* and *TFF3* expression in the CRC patients from the TCGA database (Table 3.3). Patients were grouped according to their gene expression levels, where "+" represents positive or high expression levels, and "-" represents negative or low expression levels.

Our 2x2 table analysis generates three main outputs: a Pearson correlation coefficient, an association pattern between two categorical variables and a p-value. The Pearson correlation coefficient, also known as Pearson's R or R-value, measures the direction of the correlation. It falls between -1 and 1, where a value closer to +1 indicates a strong positive correlation, a value closer to -1 indicates a strong negative correlation,

and a value closer to 0 suggests a weak or no linear correlation. The 2x2 table association pattern exhibits the counts of observations, allowing us to determine the relationship, inclusion, and exclusion criteria between the variables. In the case of a positive correlation (R>0), we can define whether the combinations of "+/-" and "-/+" are skewed towards one direction, while in the case of a negative correlation (R<0), we measure whether "+/+" and "-/-" are skewed towards a certain direction. We used the formula $\frac{(a-b)^2}{a+b}$ (where *a* and *b* are the combinations described above) and compared the value to the 99% confidence interval of a chi-squared distribution (k=1, p-value=0.01) to assess the inclusion and exclusion criteria (Bodmer & Cavalli-Sforza, 1976). The p-value is assigned from Fisher's exact test and evaluates the significance level of the correlation between the variables.

According to the Table 3.3, there is an overlap between *MUC2* and *TFF3* expression in most patients (468 out of 698) across the TCGA CRC RNA-seq dataset (Table 3.3). The presence of a very high proportion of *MUC2/TFF3*-positive tumours versus negative tumours is almost certainly due to contamination by up to 20% of normal cells. Despite that, the MUC2 and TFF3 association is striking and fits entirely with our microarray data on the relatively high proportion of *MUC2-TFF3+* tumours. Based on the analysis, 192 (27.5%) out of 698 samples express *TFF3* without the presence of *MUC2* (*MUC2-*/*TFF3+*) (Table 3.3).

MUC2	+/+	+/-	-/+	-/-	P-value	R-value	Inclusion criteria
TFF3	468	9	192	29	7.47E-09	0.23	Significant

Table 3.3 Association between *MUC2* and *TFF3* gene expression in the TCGA CRC dataset. 2x2 table shows that a positive correlation between MUC2 and TFF3 expression with inclusions (skewed -/+). The p-value is calculated from Fisher's exact test and indicates the

significance of a correlation between two variables. A Pearson correlation coefficient, or R-value, represents the direction in which correlations occur.

3.2.2 Further categorisation of a large panel of CRC cell lines based on MUC2 and TFF3 protein expression levels

In order to determine whether CRC cells differentiate into goblet cells, we analysed a large panel of CRC cell lines in immunofluorescence using antibodies against MUC2 (PR5D5, in-house) and TFF3, the main markers of goblet cells in a normal colon.

To begin with, the LS180 cell line was selected since it is a well-characterized CRC cell line, as well as the first goblet cell line transcriptome profiled in the Bodmer laboratory collection. Only some LS180 cells were co-stained with MUC2 and TFF3 antibodies (Figure 3.4). These cells were larger than the other cells and had a shape similar to what has been previously described for goblet cells. All the cells that were positive for MUC2 staining were also positive for TFF3, confirming that LS180 belongs to the MUC2+TFF3+ category.

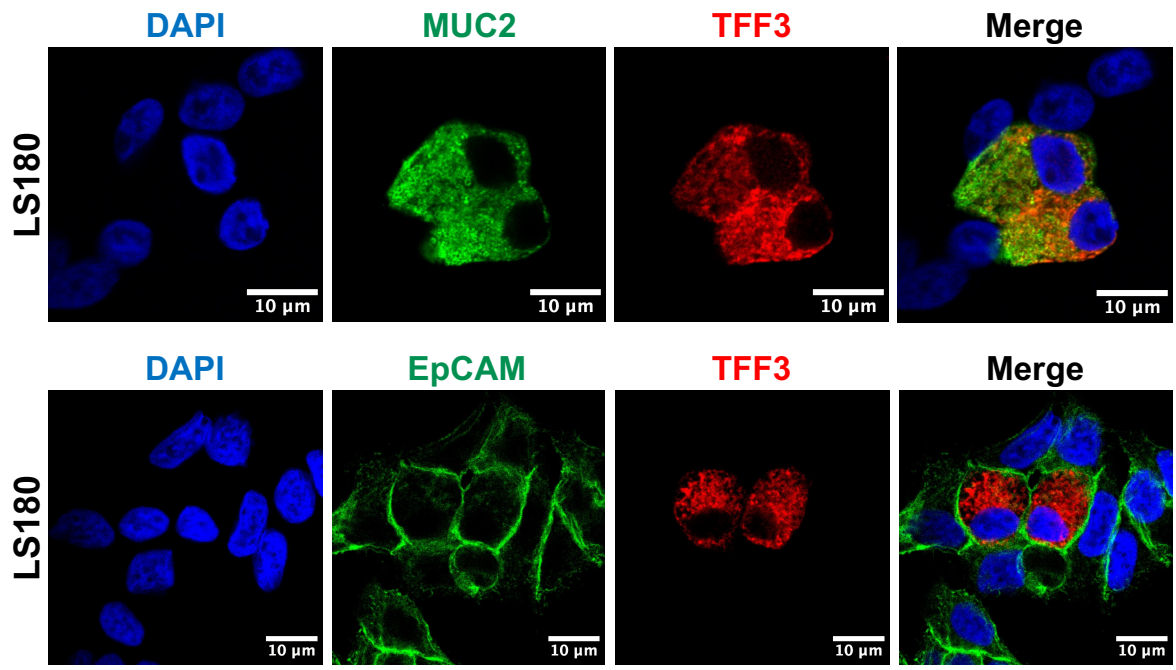
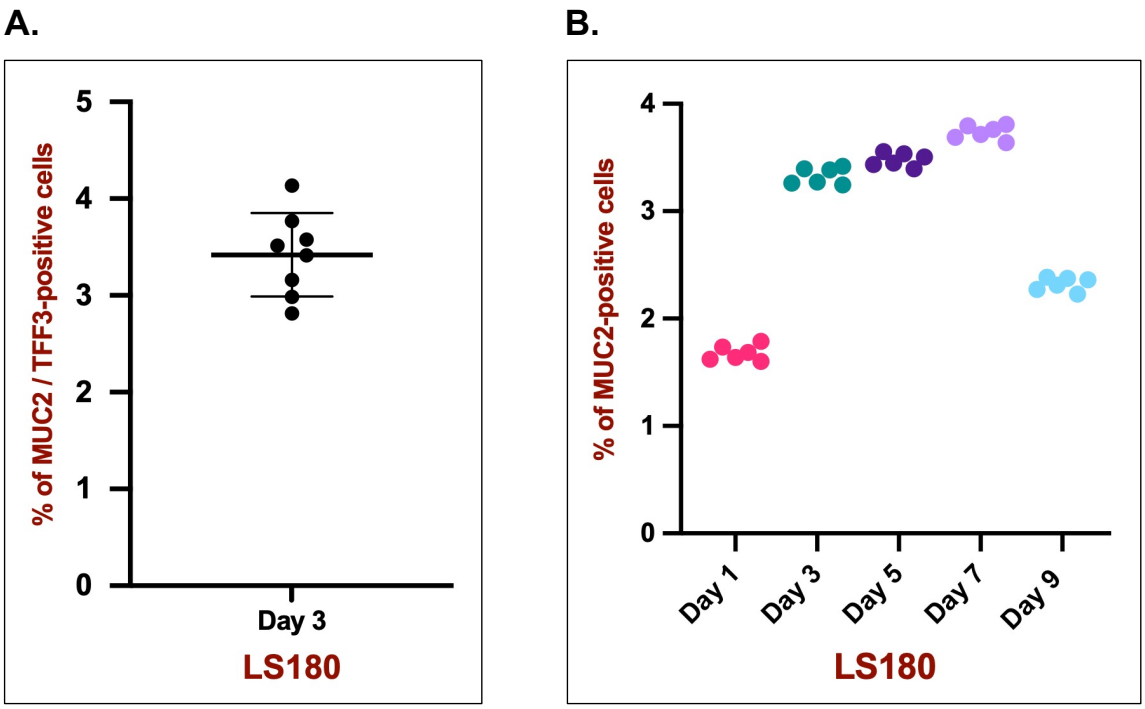


Figure 3.4 Staining of MUC2 and TFF3 and of Epithelial cell adhesion molecule (EpCAM) in LS180. The cells were grown for three days before fixing and staining. EpCAM was used

as an epithelial specific cell surface marker that is expressed in nearly all CRC cell lines. Scale bar: 10 µm.

Next, we quantified the number of goblet cells (by MUC2 and TFF3 staining) in LS180 and found that it varied between 2.81% and 4.14% at three days after seeding (Figure 3.5A). To determine whether time in culture affected the number of goblet cells, we analysed the cells over a period of nine days (Figure 3.5B). To achieve similar confluency on the day of analysis we seeded the cells at different densities for different time points. On day 0, the cells were allowed to attach after plating the cells and then fixed from starting on day 1. The time in culture influenced the percentage of goblet cells, with a lower percentage at days 1 and 9 as compared with the rest of the time points (Figure 3.5B). Figure 3.5C provides examples of representative images obtained at different time points.



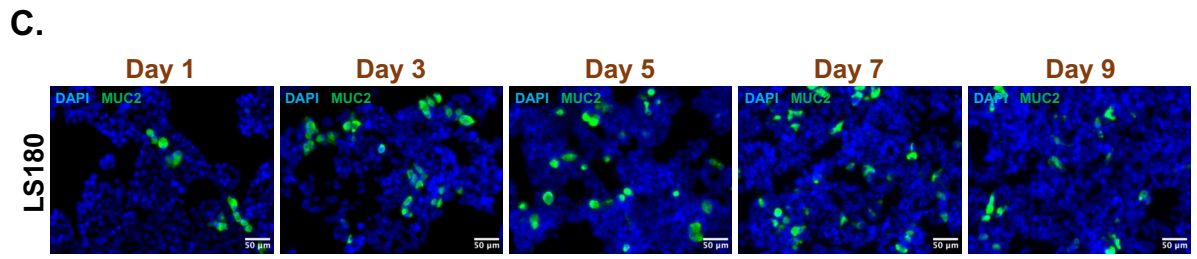


Figure 3.5 The percentage of goblet cells in LS180. Goblet cells were identified by MUC2 staining in immunofluorescence. (A) Eight independent experiments were quantified, with ten different fields imaged and analysed per experiment (n=3000-5000 cells per experiment). The nuclei, stained by DAPI, were counted using CellProfiler 3.0 and goblet cells, stained by MUC2, were counted manually. Error bars: mean with standard deviation. (B) The percentage of goblet cells was calculated as in (A) for six independent experiments during the indicated time course. (C) Examples of representative images at various time points are illustrated. Scale bar: 50 µm.

A similar analysis was also carried out on the RW7213 CRC cell line which is one of the highest MUC2 expressing cell lines according to our microarray analysis, presumably derived from a highly mucinous carcinoma. MUC2 and TFF3 showed overlapped patterns in the RW7213 cell line as was seen for LS180 (Figure 3.6).

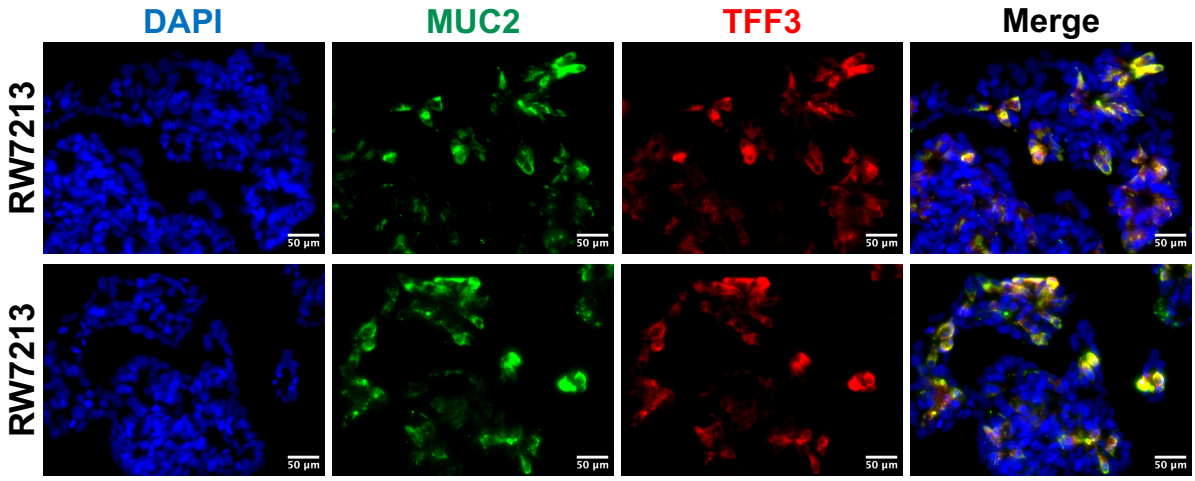


Figure 3.6 Staining of MUC2 and TFF3 in RW7213. The cells were cultured for three days before being stained with MUC2 and TFF3. Scale bar: 50 µm.

As compared to LS180, a much higher proportion of cells were co-stained with MUC2 and TFF3 antibodies. Indeed, quantification after three days showed that 22.62% to 23.13% of cells were co-stained with MUC2 and TFF3 (Figure 3.7A). The percentage of goblet cells varied at different days of culture, with a slight increase from day 3 to day 7, followed by a decrease at day 9. There was a substantial increase (nearly two-fold) between day 1 and day 3 (Figure 3.7B). Examples of representative images taken over different time points are shown in Figure 3.7C.

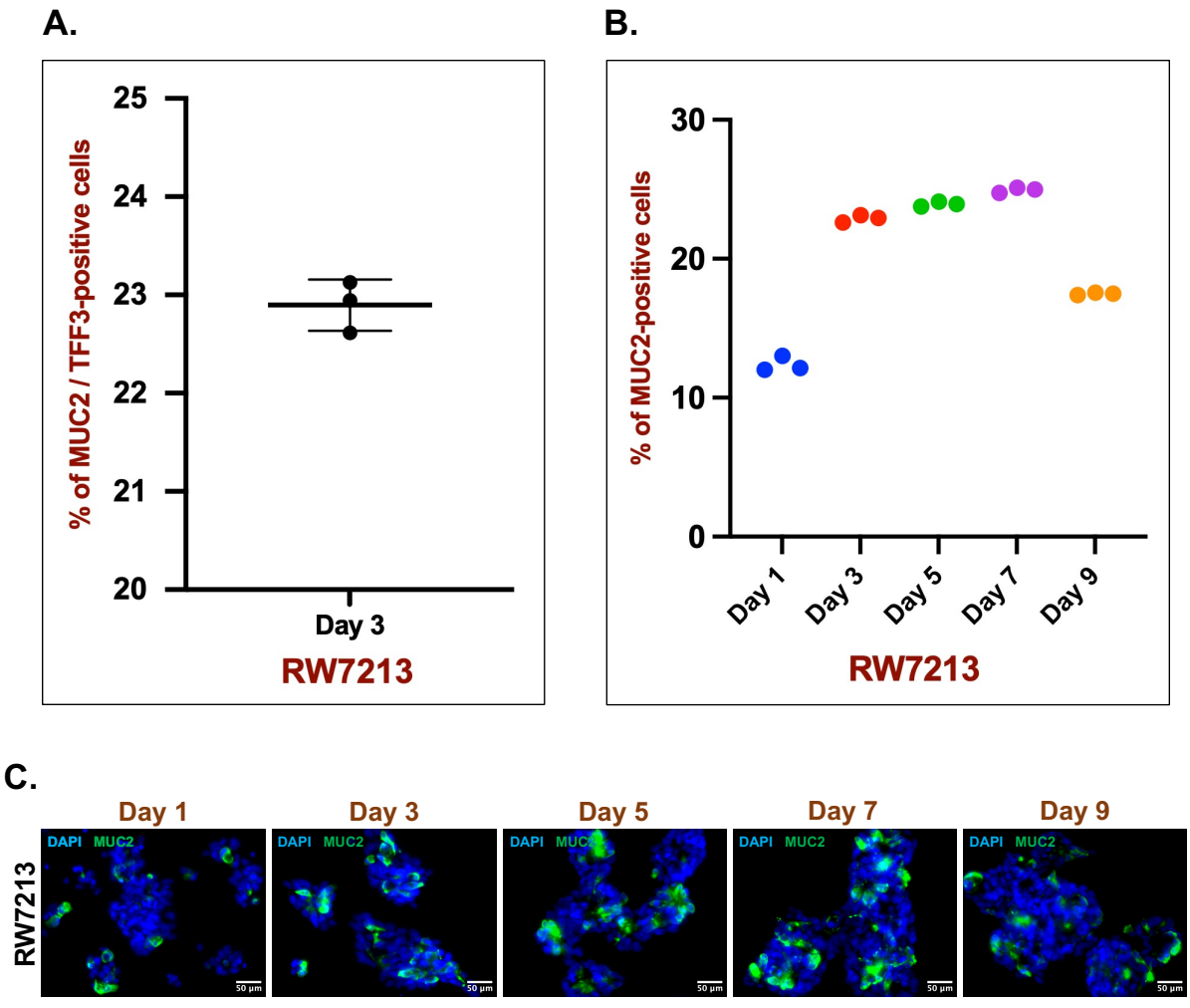


Figure 3.7 The percentage of goblet cells in RW7213. Goblet cells were quantified based on MUC2 staining in immunofluorescence. (A) Three independent experiments were performed, with ten different fields imaged and processed per experiment (n=2000-4000 cells per experiment). The nuclei were stained with DAPI and measured using CellProfiler 3.0 while

goblet cells were stained with MUC2 and counted manually. Error bars: mean with standard deviation. (B) The percentage of goblet cells was calculated as in (A) for three independent experiments over the indicated periods. (C) The figure displays a representative image from each of the time periods. Scale bar: 50 μm.

After identifying the key time point for goblet cell differentiation, we analysed 100 CRC cell lines and quantified cells expressing MUC2 and TFF3 proteins. An immunofluorescence analysis based on MUC2 and TFF3 staining in a large panel of CRC cell lines was compared with mRNA expression data of MUC2 and TFF3. During screening, at least three independent replicates were carried out for each experiment using PR5D5, an in-house antibody against MUC2 and a commercial TFF3 antibody.

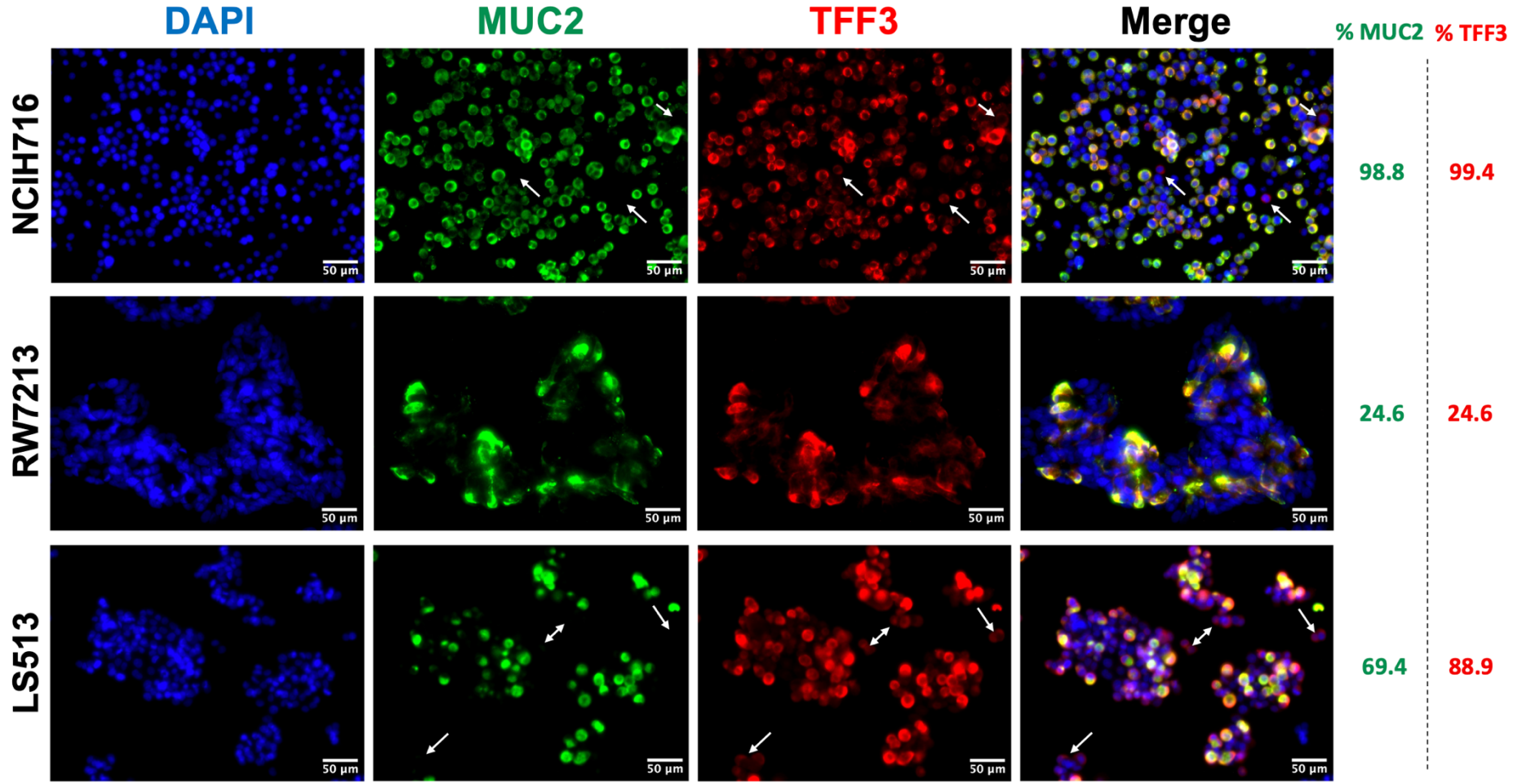
According to our analysis on a panel of 100 CRC cell lines, we categorised 5 distinct groups in terms of goblet cell differentiation based on the proportions of MUC2 and TFF3 positive cells in the lines. Representatives of the 5 groups are illustrated in Figure 3.8 and five-fold tabulation of all the lines, giving the observed percentages of MUC2 and TFF3-positive and negative cells in each line is shown in Table 3.4. A representative image of each categorised CRC cell line, as well as classification table are provided in Appendix Figure 1 and Appendix Table 1.

Group 1 includes 10 cell lines that exhibited high levels of MUC2 and TFF3 expression, suggesting them being highly mucinous carcinoma derived cell lines. In Group 1, cell lines can be subdivided into two groups, with some producing more goblet cells than others marked with an * in Table 3.4. Group 2 cell lines expressed medium levels of MUC2 and TFF3. Within the group, some cell lines (C84, HT29, and HDC73) expressed more TFF3 than MUC2, while HDC114 showed the opposite. A total of 14 out of 100 cell lines belonged to Group 3 that produced low levels of both MUC2 and

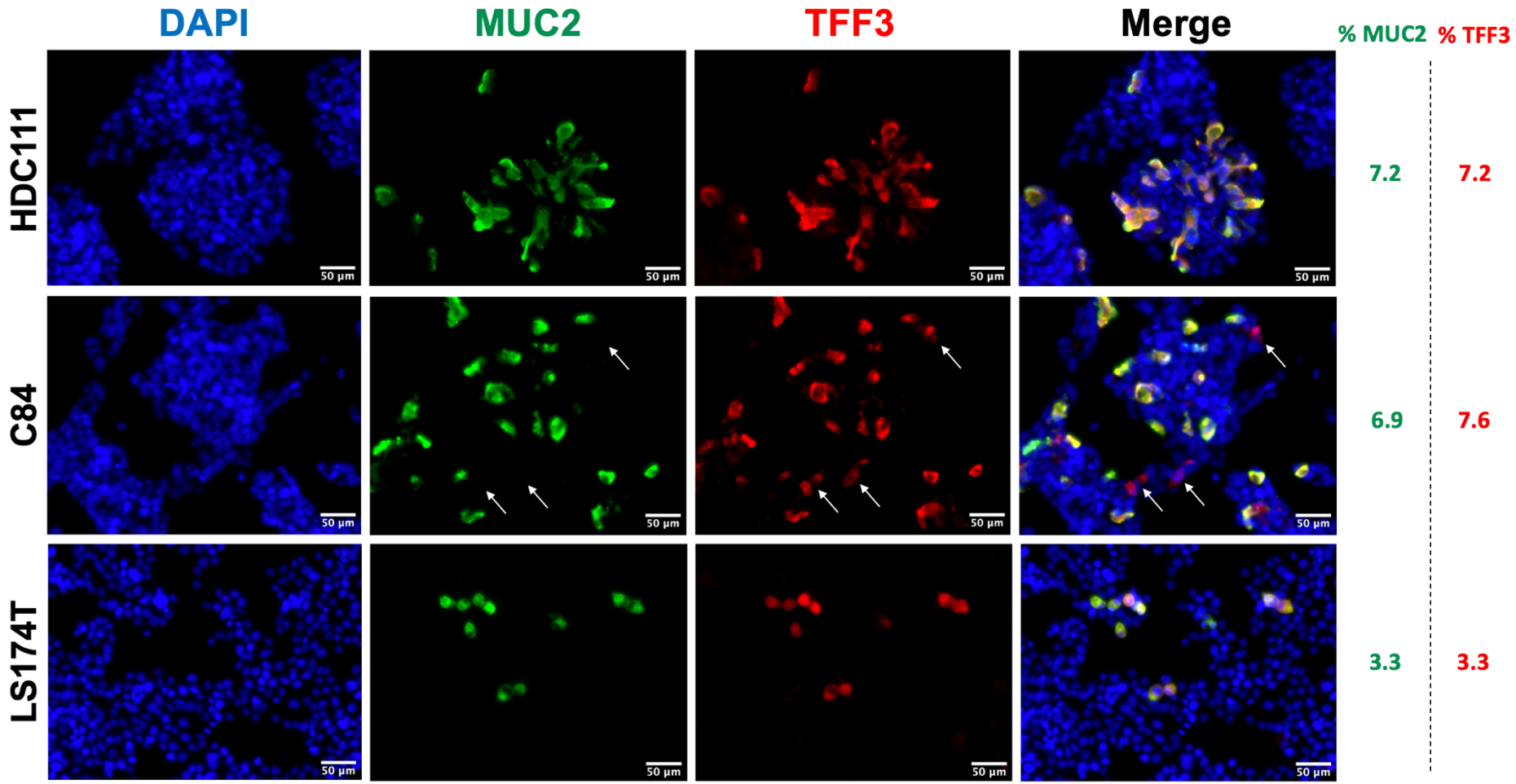
TFF3. Group 4 covered the cell lines that did not express MUC2, but almost all cells stained for TFF3. These MUC2-negative, TFF3-positive cell lines represent a unique subset of CRC cell lines, and so presumably CRCs, that have not been previously described. There was a subset of cell lines in group 4 (labelled as 4*) expressing relatively low levels of MUC2 but a large amount of TFF3. The SNUC2B cell line (indicated as 4**) was the only line that was not stained with MUC2 and had only about 17% staining with TFF3. The remaining 37 cell lines were classified as both MUC2- and TFF3-negative cell lines and were listed as group 5, the largest category. These cell lines were used as negative controls in various assays.

According to mRNA expression analysis, LOVO is one of only two *MUC2*-positive and *TFF3*-negative cell lines (Table 3.2B). In contrast to the *TFF3* mRNA data, some cells in LOVO were TFF3-positive in immunofluorescence staining. In addition, there were MUC2 positive cells that did not express TFF3 (Figure 3.8, Group 3 cell line: LOVO, cells indicated by arrows). This shows that gene expression analysis can sometimes categorise gene expressions as negative or having very low expression levels, especially when the gene is only expressed in a very small subset of cells.

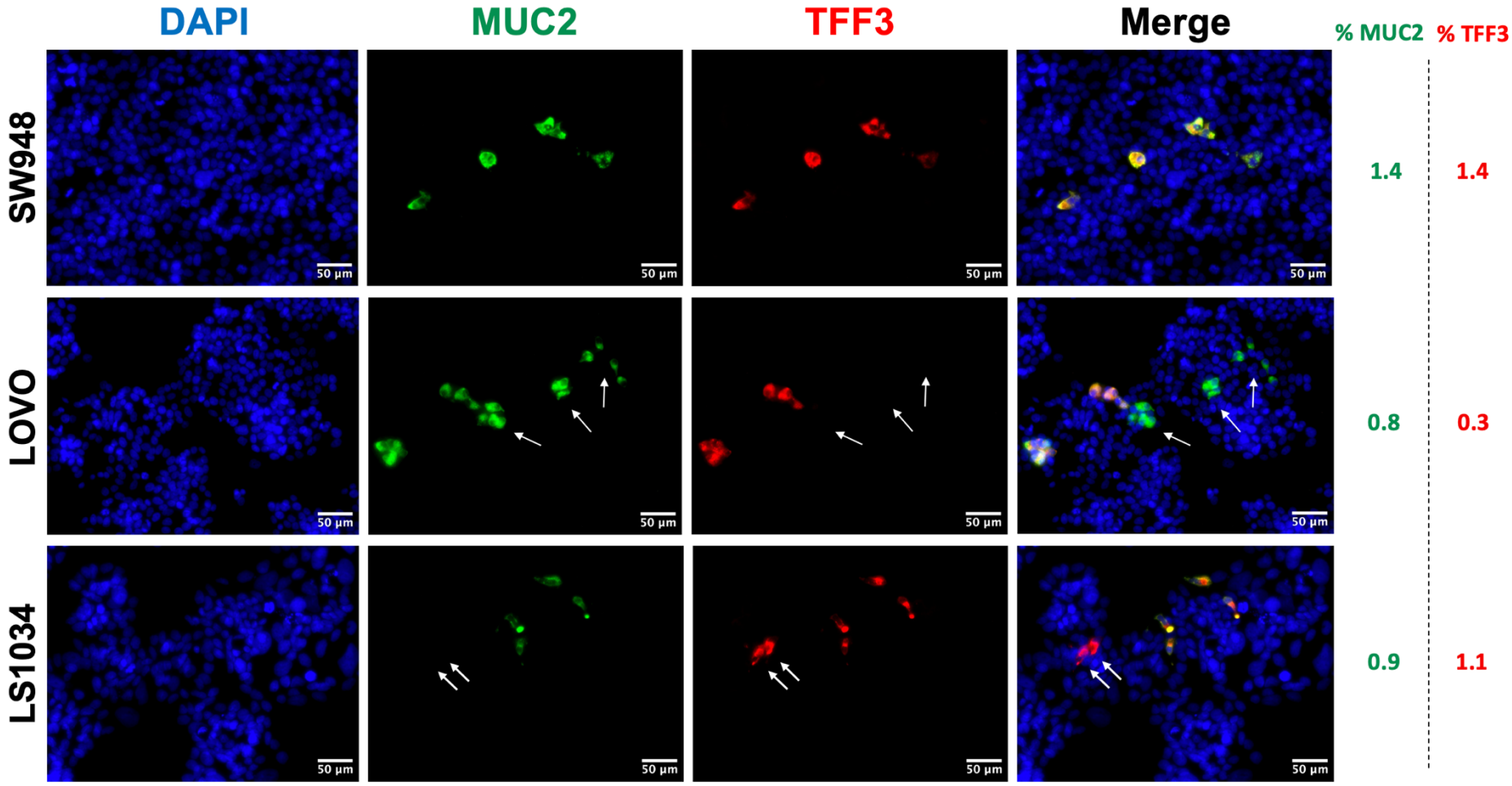
Group 1


DAPI
MUC2
TFF3
Merge
% MUC2
% TFF3
NCIH716
RW7213
LS513
50 µm
98.8
99.4
24.6
24.6
69.4
88.9

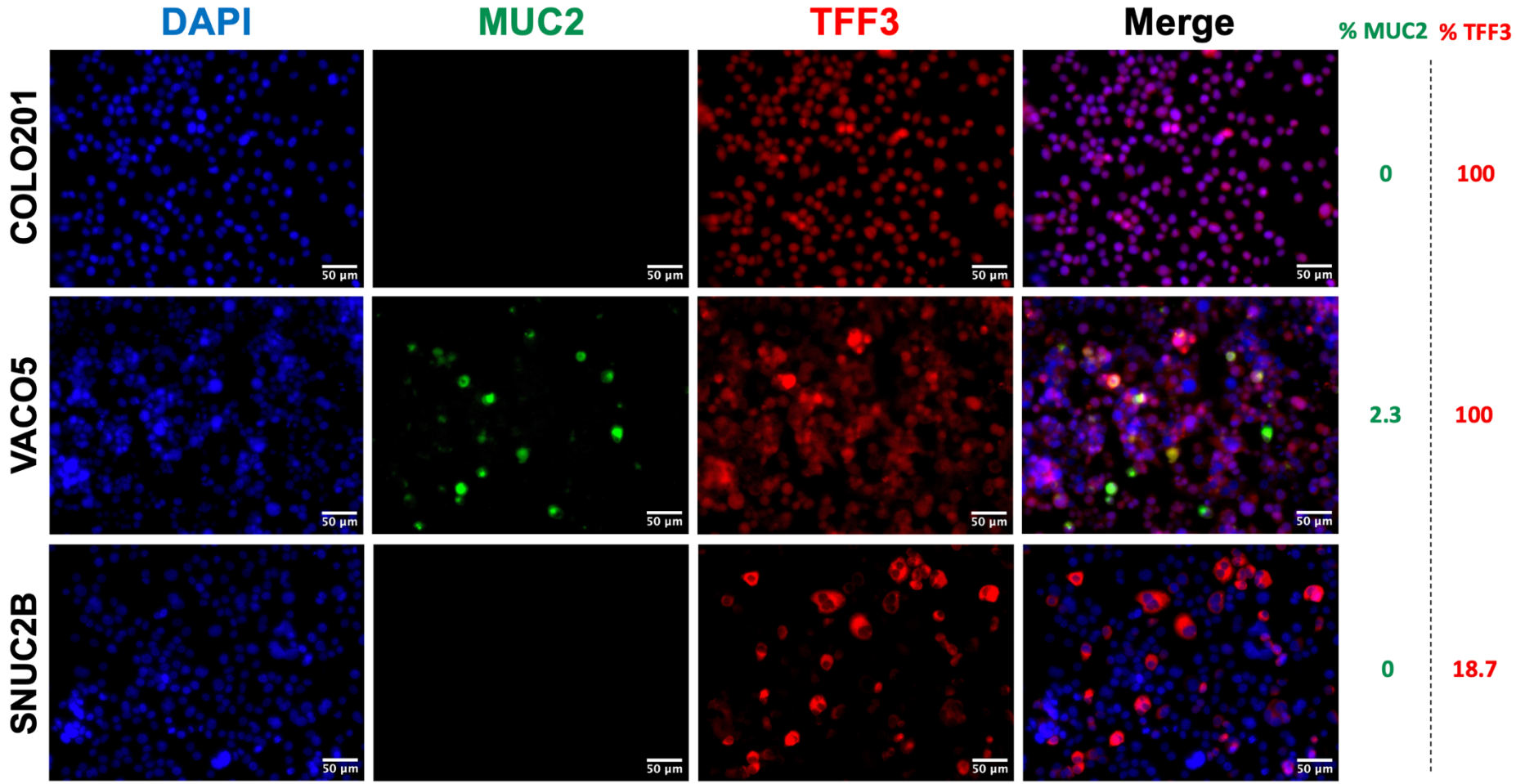
Group 2


DAPI
MUC2
TFF3
Merge
% MUC2
% TFF3
HDC111
C84
LS174T
50 µm
7.2
7.2
6.9
7.6
3.3
3.3

Group 3


DAPI
MUC2
TFF3
Merge
% MUC2
% TFF3
SW948
LOVO
LS1034
1.4
1.4
0.8
0.3
0.9
1.1
50 µm

Group 4


DAPI
MUC2
TFF3
Merge
% MUC2
% TFF3
COLO201
VACO5
SNUC2B
0
100
2.3
100
0
18.7
50 µm

Group 5

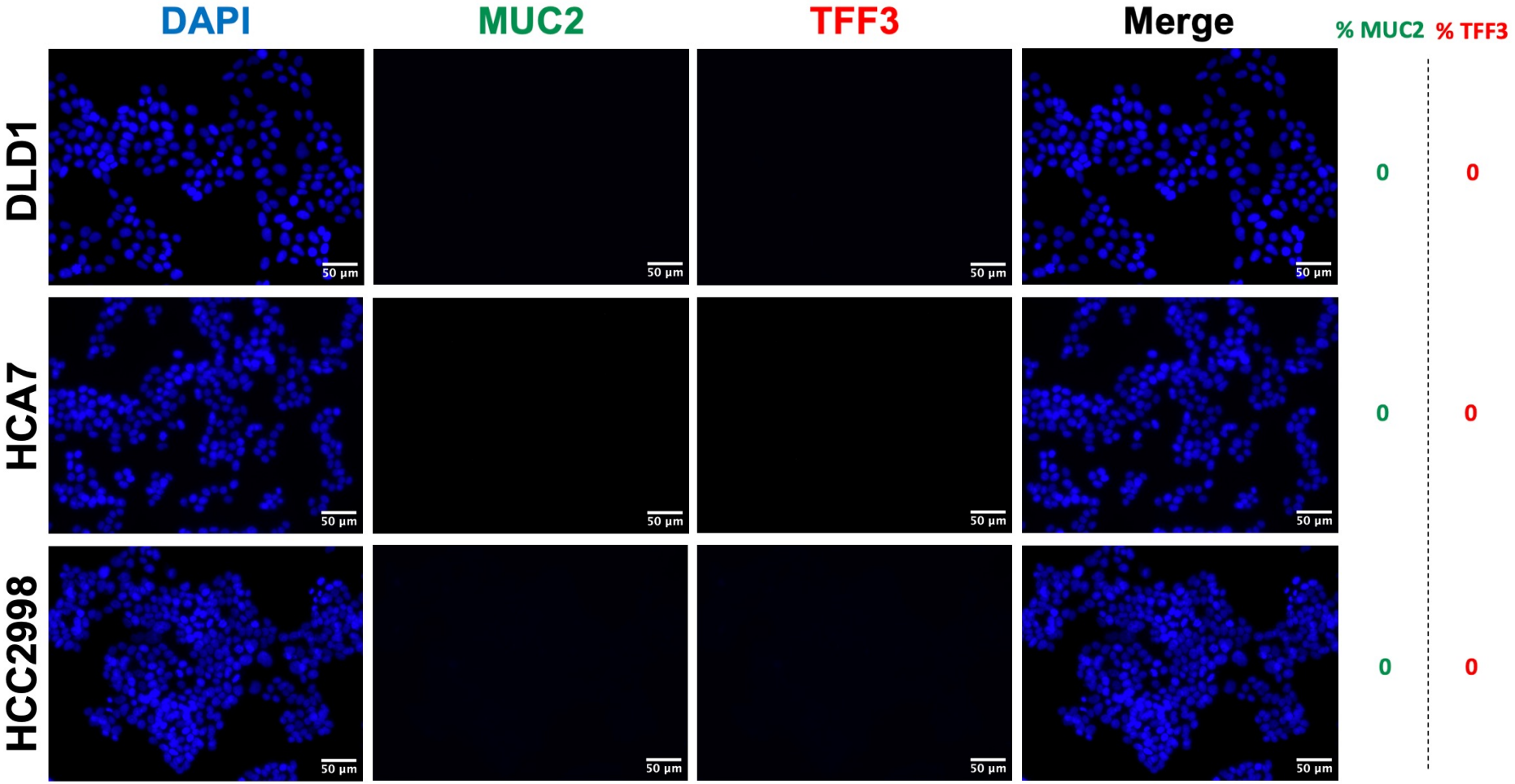

DAPI
MUC2
TFF3
Merge
% MUC2
% TFF3
DLD1
HCA7
HCC2998
50 µm
0
0
0
0
0
0

Figure 3.8 Representative images of IF staining for MUC2 and TFF3 for the 5 different categories of 100 CRC cell lines based on the proportions of cells staining positive for MUC2 or TFF3. The cells were grown for three days before being fixed and stained using MUC2 (PR5D5, 1:250) and TFF3 (1:500) antibodies. At least three independent experiments were conducted. None of these antibodies reacted with goblet cell-negative cell lines which belonged to group 5 category. Magnification: 20X, scale bar: 50 μm.

It is important to note that COLO320HSR and COLO320DM; LS174T and LS180; SW480 and SW620; GP5D and GP2D; HT29, CX1, WIDR, and KM20L2; VACO4S and VACO4A; HDC57 and HDC54; DLD1 and HCT15; C2BBE1 and CACO2; COLO201, COLO205, COLO206, and NCIH548 belong to a sets of cell lines derived from the same cancers that may mostly be expected to be of duplicates.

The protein-based classification matches perfectly with our microarray data, showing most CRC cell lines do not differentiate and MUC2-TFF3+ cell lines define a novel CRC subtype.

Table 3.4 (the following page) Five categories of 100 cell lines based on MUC2 and TFF3 staining cells in each cell line. (A) After cultivating for three days in culture, cell lines included in groups 1, 2, 3, 4, and group 5, were fixed, stained with MUC2 (PR5D5, green, 1:250) and TFF3 (red, 1:500) antibodies in immunofluorescence analysis. MUC2- and TFF3-positive cells were measured at protein level using CellProlifer 3.0 software, as well as manually. A minimum of three independent experiments were performed for each cell line and the table presents the average value of measurements based on DAPI, MUC2, and TFF3 staining. Ten different fields were imaged and analysed per experiment. (B) Table 3.4B summarizes the protein expression pattern of MUC2 and TFF3 across all five groups, with high levels at 14-100%, medium levels at 2-8%, and low levels at 0.5-1.5%.

Group 1.

Cell line	MUC2 %	TFF3 %
C99*	93.3	100.0
CL40*	94.4	100.0
HCA46*	100.0	100.0
HDC9	14.3	23.8
LS411*	38.5	99.4
LS513*	69.4	88.9
NCIH508	16.1	16.1
NCIH716*	98.8	99.4
PCJW	34.3	34.3
RW7213	24.6	24.6

Group 2.

Cell line	MUC2 %	TFF3 %
C125PM	6.0	6.0
C80	6.6	6.6
C84	6.9	7.6
HDC111	7.2	7.2
HDC114	2.5	1.8
HDC143	4.1	4.1
HDC73	3.5	6.3
HT29	1.9	2.4
LS174T	3.3	3.3
LS180	3.4	3.4
T84	4.4	4.4
VACO4A	3.8	3.8
WIDR	3.4	4.1

Group 3.

Cell line	MUC2 %	TFF3 %
CCK81	0.8	0.8
CW2	1.3	1.5
CX1	1.0	1.7
HCC56	1.0	1.0
HDC54	0.7	1.2
HDC57	1.2	1.3
JHCOLOYI	0.5	1.5
KM20L2	0.6	1.2
LOVO	0.8	0.3
LS1034	0.9	1.1
SW1222	0.7	1.0
SW1463	0.5	0.7
SW948	1.4	1.4
VACO4S	1.3	1.3

Group 4.

Cell line	MUC2 %	TFF3 %
C106	0.0	100.0
C70	0.0	100.0
CL14	0.0	100.0
COLO201	0.0	100.0
COLO205	0.0	100.0
COLO206	0.0	100.0
GP2D	0.0	100.0
GP5D	0.0	100.0
HDC142	0.0	100.0
NCIH548	0.0	100.0
PMFKO14	0.0	97.6
RCM1	0.0	100.0
RW2982	0.0	100.0
SKCO1	0.0	100.0
SNUC1	0.0	100.0

Group 4 (continued).

Cell line	MUC2 %	TFF3 %
SW403	0.0	100.0
SW620	0.0	100.0
TT1TKB	0.0	97.9
VACO429	0.0	100.0
HCT8*	0.8	100.0
HT55*	8.7	100.0
LIM1863*	0.6	100.0
OXCO2*	0.4	100.0
SNU1235*	2.9	99.0
VACO10MS*	9.2	99.0
VACO5*	2.3	100.0
SNUC2B**	0.0	18.7

Group 5.

Cell line	MUC2 %	TFF3 %
C10	0.0	0.0
C10A	0.0	0.0
C10S	0.0	0.0
C2BBE1	0.0	0.0
C32	0.0	0.0
C75	0.0	0.0
CACO2	0.0	0.0
CAR1	0.0	0.0
CC20	0.0	0.0
CCO7	0.0	0.0
COLO320DM	0.0	0.0
COLO320HSR	0.0	0.0
COLO678	0.0	0.0
DLD1	0.0	0.0
HCA7	0.0	0.0
HCC2998	0.0	0.0

Group 5 (continued).

Cell line	MUC2 %	TFF3 %
HCT116	0.0	0.0
HCT15	0.0	0.0
HDC135	0.0	0.0
HDC8	0.0	0.0
HDC82	0.0	0.0
HRA19	0.0	0.0
ISRECO1	0.0	0.0
LIM1215	0.0	0.0
LIM2405	0.0	0.0
LS123	0.0	0.0
NCIH747	0.0	0.0
OUMS23	0.0	0.0
OXCO1	0.0	0.0
OXCO3	0.0	0.0
RKO	0.0	0.0
SW1116	0.0	0.0
SW1417	0.0	0.0
SW48	0.0	0.0
SW480	0.0	0.0
SW837	0.0	0.0

B.

	MUC2	TFF3
Group 1	+++	+++
Group 2	++	++
Group 3	+	+
Group 4	-	++++
Group 5	-	-

For MUC2: +++ (14-100%) ++ (2-8%) + (0.5-1.7%)
For TFF3: +++ (14-100%) ++ (2-8%) + (0.5-1.7%) ++++ (100%)

3.2.3 Association between goblet cell differentiation and lumen formation

In this section, we aimed to understand the association between goblet cell differentiation and lumen formation in our panel of CRC cell lines. Lumen formation data had been collected by the Cancer and Immunogenetics Laboratory, WIMM, University of Oxford (Liu, 2017). In total, 60 CRC cell lines were examined, and 13 cell lines (21.7%) were able to form lumens in 3D culture (Table 3.5A/B). Lumen formation ability of 17 cell lines (C106, CACO2, CC20, CL40, COLO320DM, CX1, DLD1, HCT116, HRA19, HT29, HT55, LOVO, LS180, NCIH508, OUMS23, RW7213, SW1222) was also tested in 3D culture by Dr Johanna Michl from the Department of Physiology, Anatomy, and Genetics (DPAG, University of Oxford) and the results perfectly aligned with the previous classification (unpublished data).

Using chi-squared analysis, the association between goblet cell production and lumen formation was tested in a 2x2 table (Table 3.5). For goblet cells, we used our classification based on immunofluorescence analysis. Both MUC2 and TFF3 expression was considered goblet cell defining markers, therefore groups 1, 2, and 3 were included in the 2x2 table. Pearson's Chi-square analysis was used to determine the relationship between goblet cell production and lumen formation in a 2x2 contingency table. With a degree of freedom, the Chi-square and p-value are 0.9141 and 0.3390, respectively, less than the critical value (3.84 for 95% significance). According to Fisher's exact test, the p-value is 0.4908 which is greater than 0.05, a measure of statistical significance to support or reject the null hypothesis. There is thus no evidence of an association between goblet cell differentiation and lumen formation.

A.

Group classification	Lumen formation			
	Lumen +	**Lumen -**		
1	CL40 HCA46 RW7213	C99 NCIH508		
2	C80 LS180	C125PM C84 HDC111	HDC114 HDC73 HT29	LS174T WIDR
3	LS1034 SW1222	CX1 HDC57	JHCOLOY1 KM20L2	LOVO SW948
4 (*,)**	C106 HT55	COLO201 GP2D GP5D HDC142 LIM1863	PMFK014 RCM1 RW2982 SKCO1 SNUC2B	SW403 TT1TKB VACO10MS
5	CACO2 HRA19 OXCO1 OXCO3	C2BBE1 C32 CAR1 CC20 COLO320DM COLO678 DLD1	HCA7 HCT116 LIM1215 LIM2405 LS123 OUMS23 RKO	SW1417 SW48 SW480 SW837

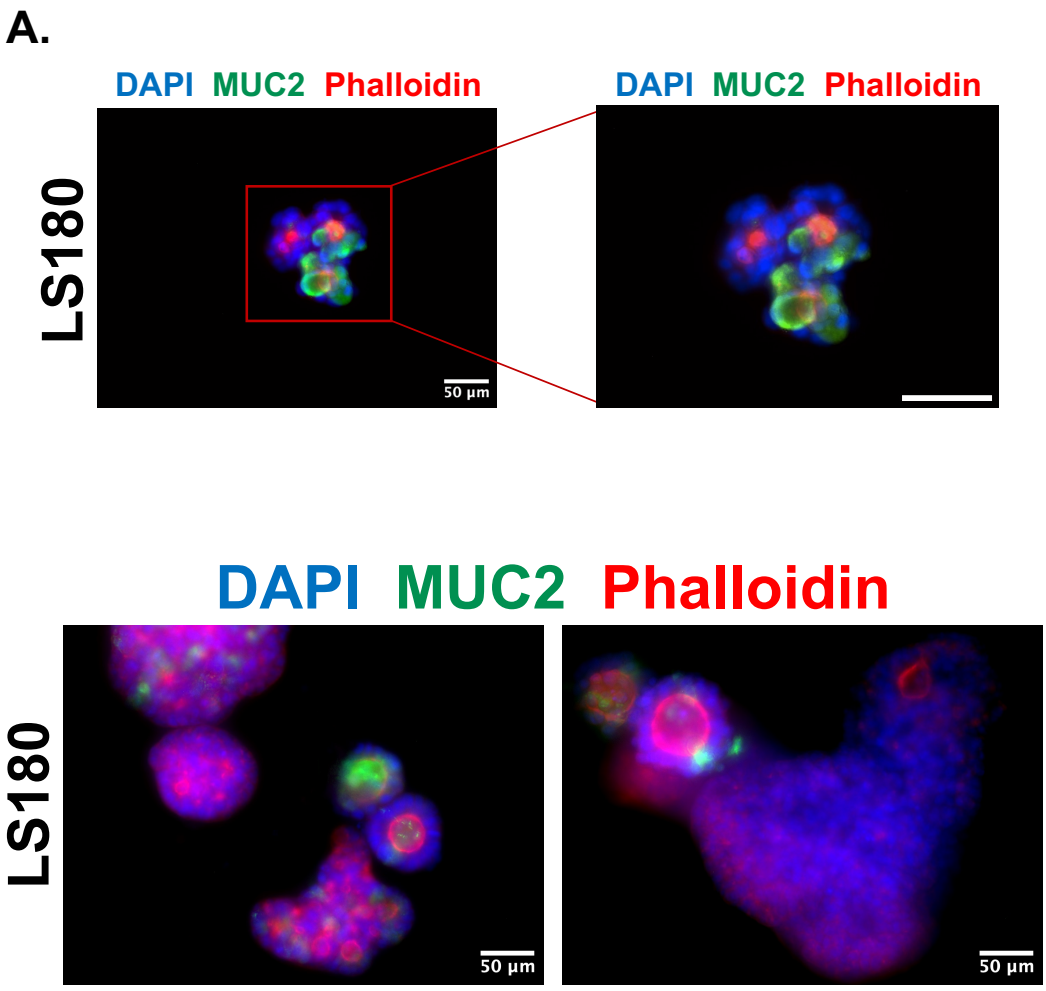
B.

	Lumen +	Lumen -	**Total**
Groups (1,2,3)	7	16	23
Group 5	4	18	22
Total	11	34	45

Table 3.5. Association between lumen formation and goblet cell differentiation. (A) Summary of the CRC cell lines in terms of lumen formation (originated from our lab data) and group classification based on our immunofluorescence analysis. (B) Number of observed phenotypes in terms of lumen formation and goblet cell differentiation were evaluated.

As shown in Table 3.5, most of the groups 2 and 3 cell lines were unable to form lumens, even though they can differentiate into goblet cells to some extent (Table 3.5). In total, 16 out of 23 (69.6%) cell lines from groups 1-3 failed to produce lumens.

We conducted a lumen formation assay on several CRC cell lines, including LS180 and RW7213. Lumens were visualised by Phalloidin staining in brush borders and MUC2 was detected within secretory granules of goblet cells or in secreted forms (Figure 3.9). The results showed that for these two goblet and lumen forming cell lines, only a small proportion of the cells within a lumen formed goblet cells, suggesting that the lumen formation was not key to goblet cell differentiation which is consistent with the lack of correlation between lumen forming and goblet cell producing cell lines.



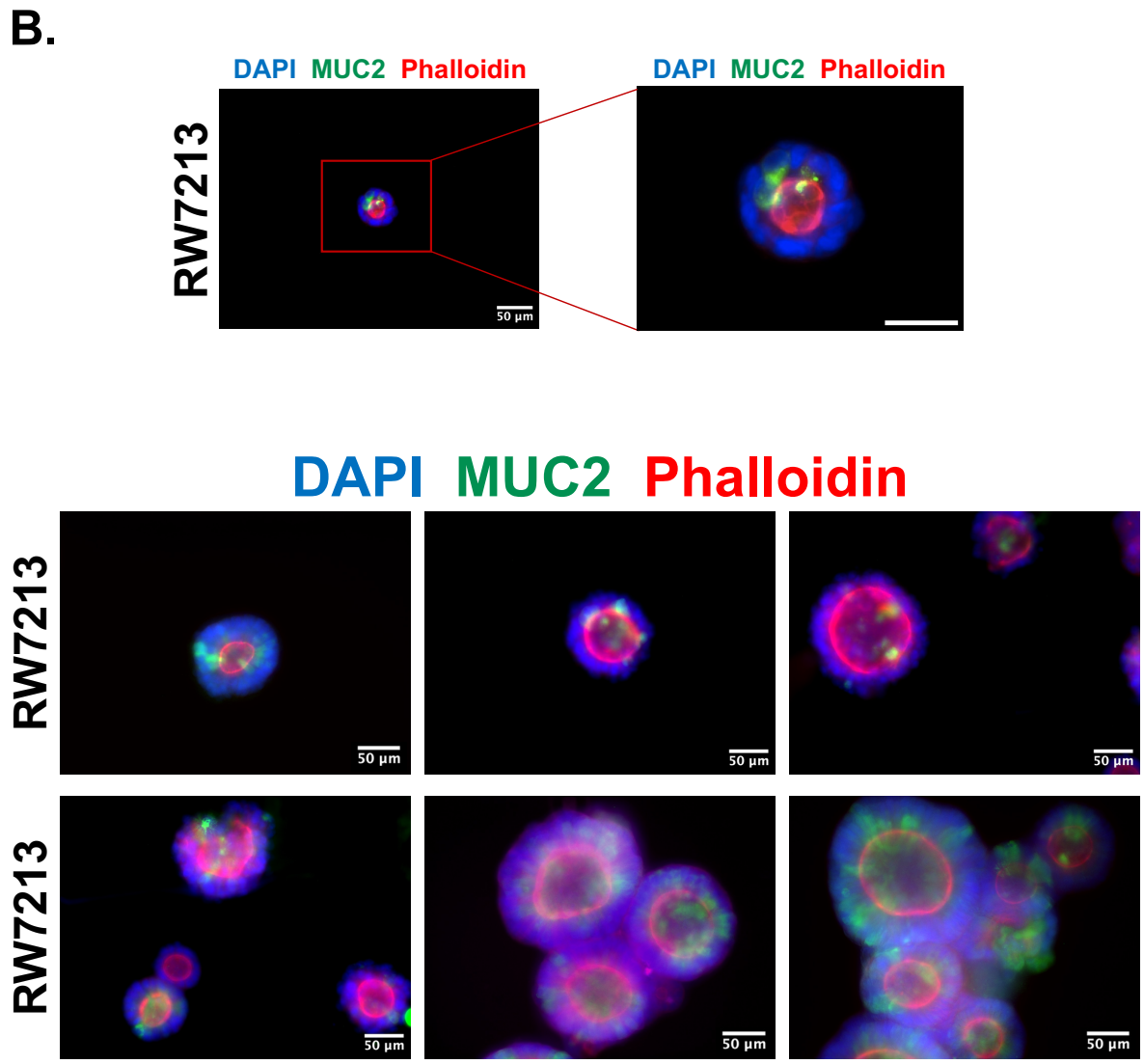


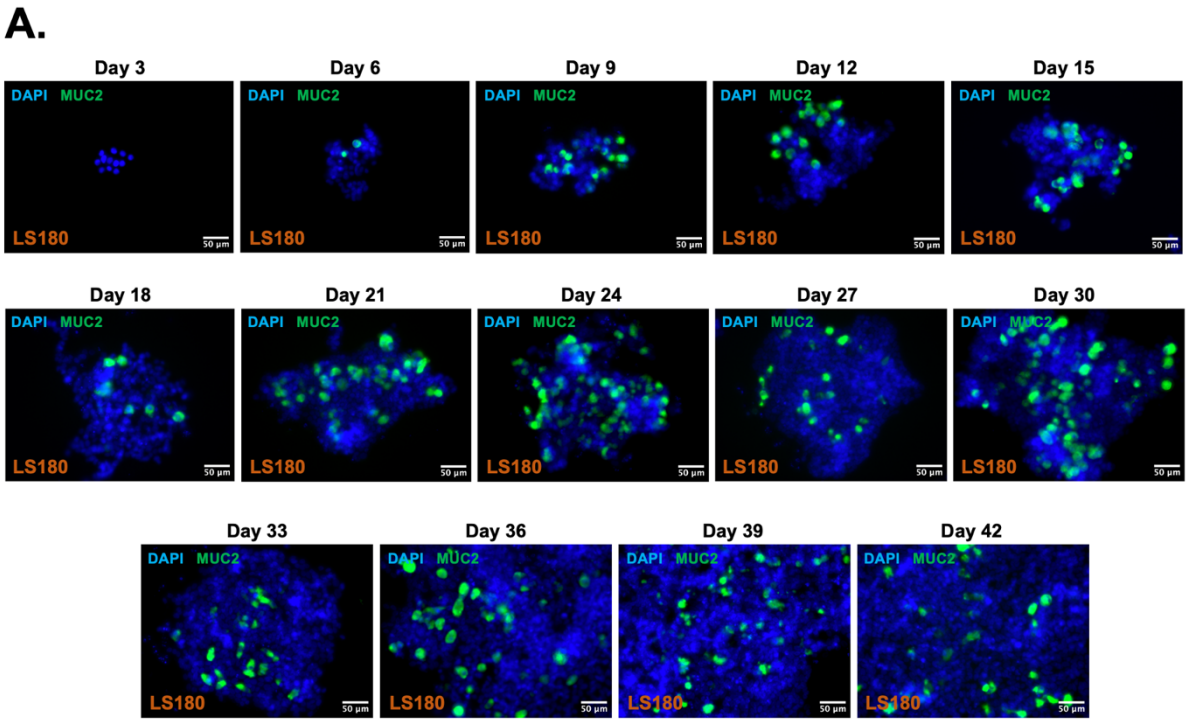
Figure 3.9 Lumen formation assay in LS180 and RW7213 CRC cell lines. Single-cell suspensions were seeded in Matrigel, incubated for 15 days, then fixed before prior to staining. The colonies were stained with DAPI (blue, 1:10000) for nuclei, Phalloidin (red, 1:1000) for F-actin, and PR5D5 (green, 1:250) for MUC2 staining. Magnification: 20X, scale bar: 50 µm.

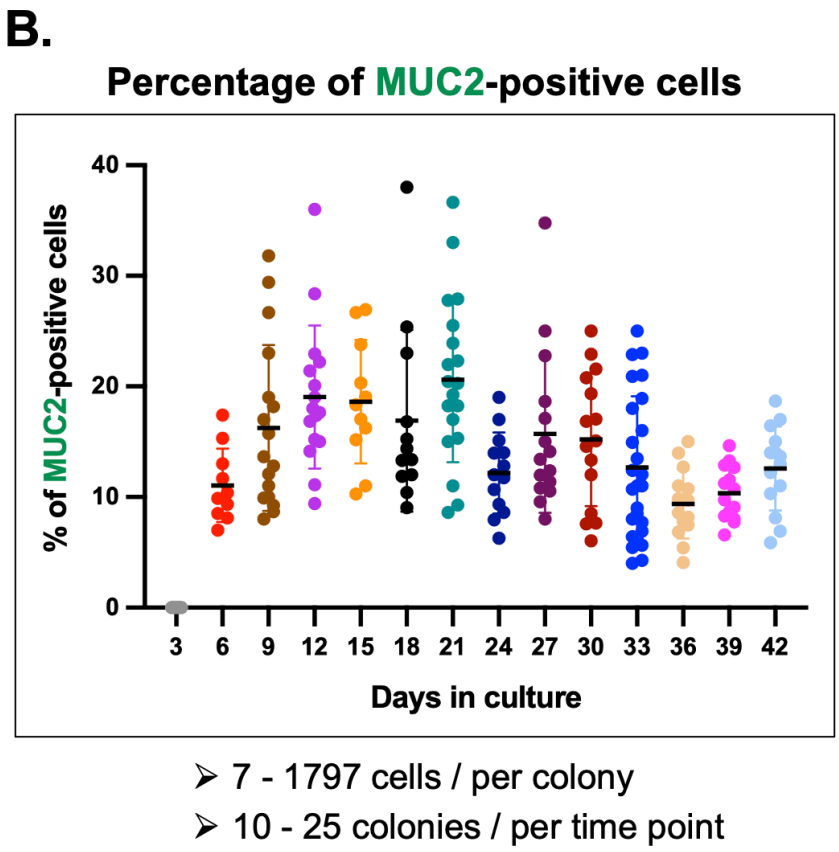
3.2.4 Single-cell clonal selection analysis in the LS180 cell line

We conducted a single-cell clonal selection analysis of LS180 CRC cell line to determine the hierarchical pattern of MUC2 and TFF3 in terms of time. Three independent serial dilutions and two FACS experiments were performed as described in the Materials and Methods section (Chapter 2.13). A combination of the BD FACS AriaTM III and Fusion was used for single cell sorting by the FACS facility (MRC WIMM, University of Oxford). After seeding individual cells in 96 well plates, the cells in the wells were fixed at various time points after initiation of growth and analysed by immunofluorescence using MUC2 (PR5D5) and TFF3 antibodies to identify MUC2 and TFF3 producing cells.

Initially, the expression of MUC2 in the LS180 CRC cell line was tested from days 3 to 42 with a 3-day interval to identify when MUC2 was first produced (Figure 3.10). Single cell suspensions obtained from serial-dilutions were seeded in a total of 60 wells per each day. Cells were fixed at the indicated time points and stained with the in-house PR5D5 antibody for MUC2 and DAPI for nuclei staining.

As illustrated in Figure 3.10A, MUC2 started showing up on day 6 and continued to appear until day 42. Approximately 50% of total colonies arising from single cell suspensions formed goblet cells in the 2D culture. We then quantified the number of MUC2-positive cells in each colony. The results showed a high percentage of MUC2 positive cells compared to the 3-4% which was what was usually seen in bulk culture of LS180. This suggested that wells were mainly assessed when they contained stem cells from the cell line and that it is only colonies derived from the LS180 stem cells that will differentiate to produce MUC2 expressing cells that are on their way to becoming goblet cells. (Figure 3.10B).



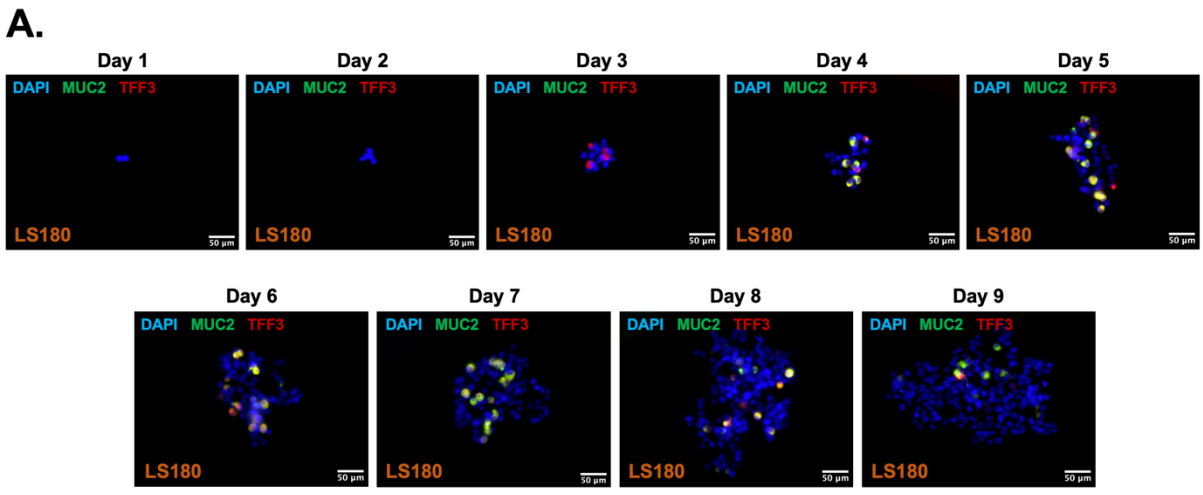


- 7 - 1797 cells / per colony
- 10 - 25 colonies / per time point

Figure 3.10 Single-cell clonal selection analysis based on MUC2 expression in LS180. (A) Cells were cultured as single cells obtained from serial dilutions. In the following step, they were fixed, and immunofluorescence was performed with antibodies against MUC2 (PR5D5,

green, 1:250) and DAPI (blue, 1:10000) for nuclei staining over the indicated time periods. Magnification: 20X, scale bar: 50 µm. (B) MUC2-positive colonies were quantified using CellProfiler 3.0.0 software and manual counting. The x axis represents the number of days cell cultures were incubated, whereas the y axis represents the percentage of MUC2-positive cells. Error bars: mean with standard deviation.

Following this, we examined when exactly MUC2 begins to be expressed and how this compares with the onset of TFF3 expression in colonies derived from single cells. LS180 cells were incubated over from 1 to 9 days, fixed, and then stained with MUC2 and TFF3 antibodies for each period. The number of MUC2/TFF3-positive colonies and the number of positive cells within each colony were counted for each of the days of culture from day 1 to 9. Figure 3.11A shows representative images of MUC2 and TFF3 from five independent single-cell experiments (three serial dilutions and two FACS experiments) using the LS180 cell line (Figure 3.11A). According to the results, TFF3 seems to be expressed earlier than MUC2 on day 3 in 2D culture while MUC2 was detected on day 4 in 2D culture. The high molecular weight and glycosylation pattern of MUC2 might be influencing this timing, though the MUC2 antibody probably recognises the MUC2 protein background before any glycosylation.



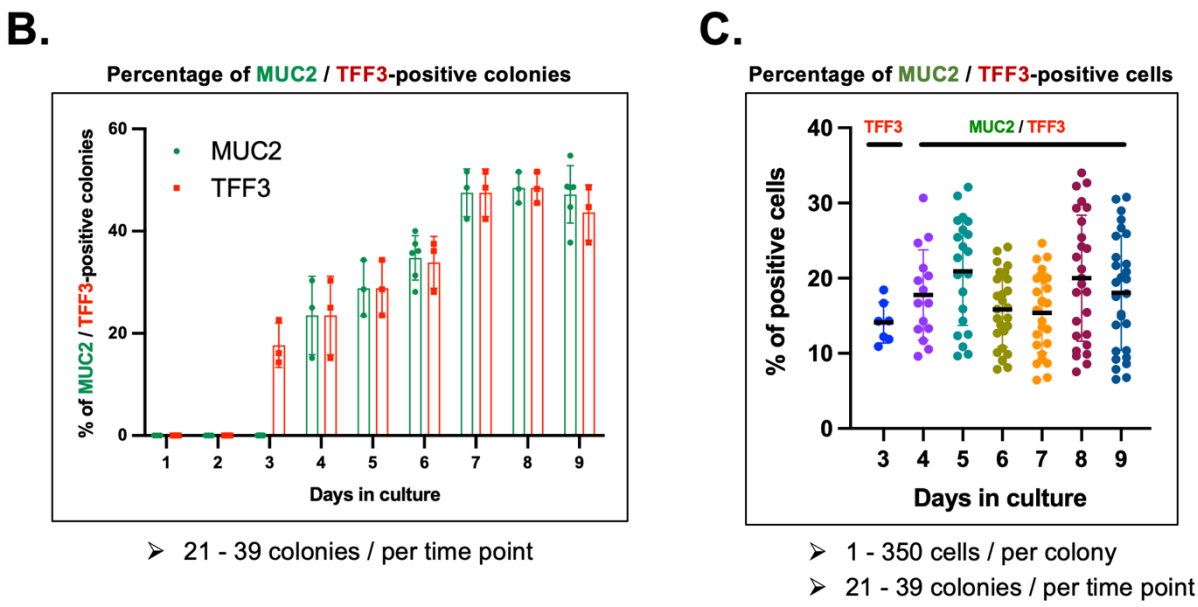


Figure 3.11 Single-cell clonal selection analysis based on MUC2 and TFF3 expression in LS180. (A) Cells were serially diluted and cultured as single cells. Then, they were fixed and stained with DAPI (blue, 1:10000), MUC2 (PR5D5, green, 1:250), and TFF3 (red, 1:500) antibodies. Magnification: 20X, scale bar: 50 µm. (B) The proportion of MUC2/TFF3-positive colonies from day 1 to day 9 was measured using CellProfiler 3.0.0 software and manual counting techniques. Error bars: mean with standard deviation for three independent experiments. (C) The percentage of MUC2/TFF3-positive cells within the colonies was also quantified. Error bars: mean with standard deviation.

Each single-cell experiment was conducted in 60 wells of 96-well plates. Approximately 50% of colonies (derived from a single cell) formed goblet cells in the 2D culture based on independent three serial dilutions and two FACS analyses (Figure 3.11B). As before, the percentage of MUC2/TFF3-expressing cells in single cell-derived colonies is much higher than in bulk culture (Figure 3.11C). This suggests that the colonies giving rise to goblet cells contain cancer stem cells and keep dividing. We also confirmed this hypothesis by quantifying colony size between goblet cell-forming and non-goblet cell colonies (Figure 3.12). Our results showed that colony sizes based on the number of cells in goblet cell-forming colonies are significantly bigger than non-goblet cell forming colonies.

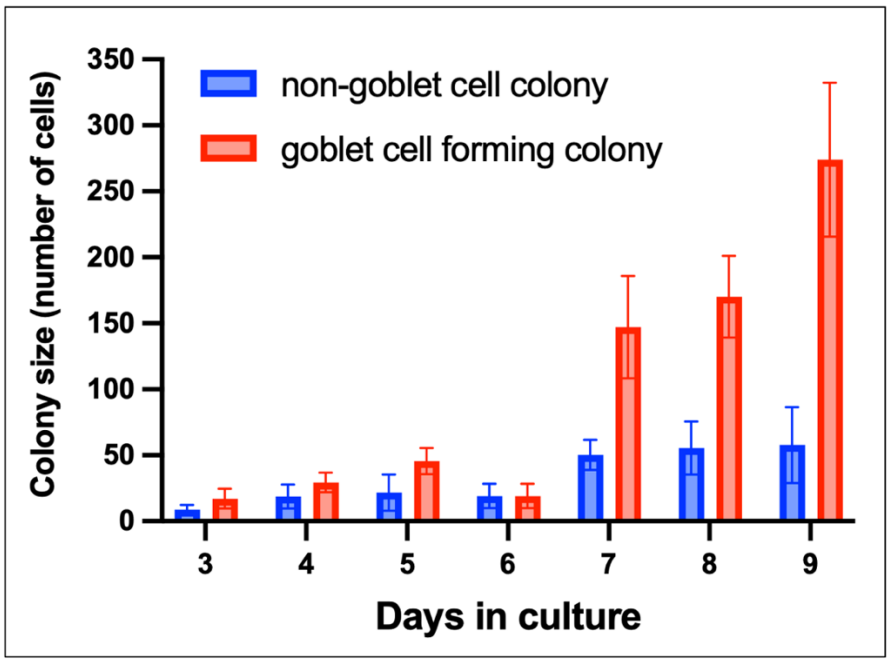


Figure 3.12 Comparison of colony size derived from single cell suspensions between goblet cell-forming and non-goblet cell-forming colonies. The x axis of the graph indicates the duration of time cells were in culture, while the y axis represents the size of colonies. Colony size was based on DAPI staining and measurement was conducted via CellProfiler 3.0.0 software. Error bars: mean of 5 independent experiments per each day with standard deviation.

ATOH1, CDX1, and CDX2 transcription factors were also studied using single-cell analysis at different time points (Figure 3.13). ATOH1, CDX1, and CDX2 were expressed at all the given times, suggesting they are involved in the early stages of differentiation. We used ATOH1 with MUC2 (Figure 3.13A) and CDX1 with TFF3 (Figure 3.13B), and CDX2 with TFF3 (Figure 3.13C) to match different species of antibodies. Analysis also indicated that MUC2 expression began at day 4, whereas TFF3 is started to be expressed on day 3.

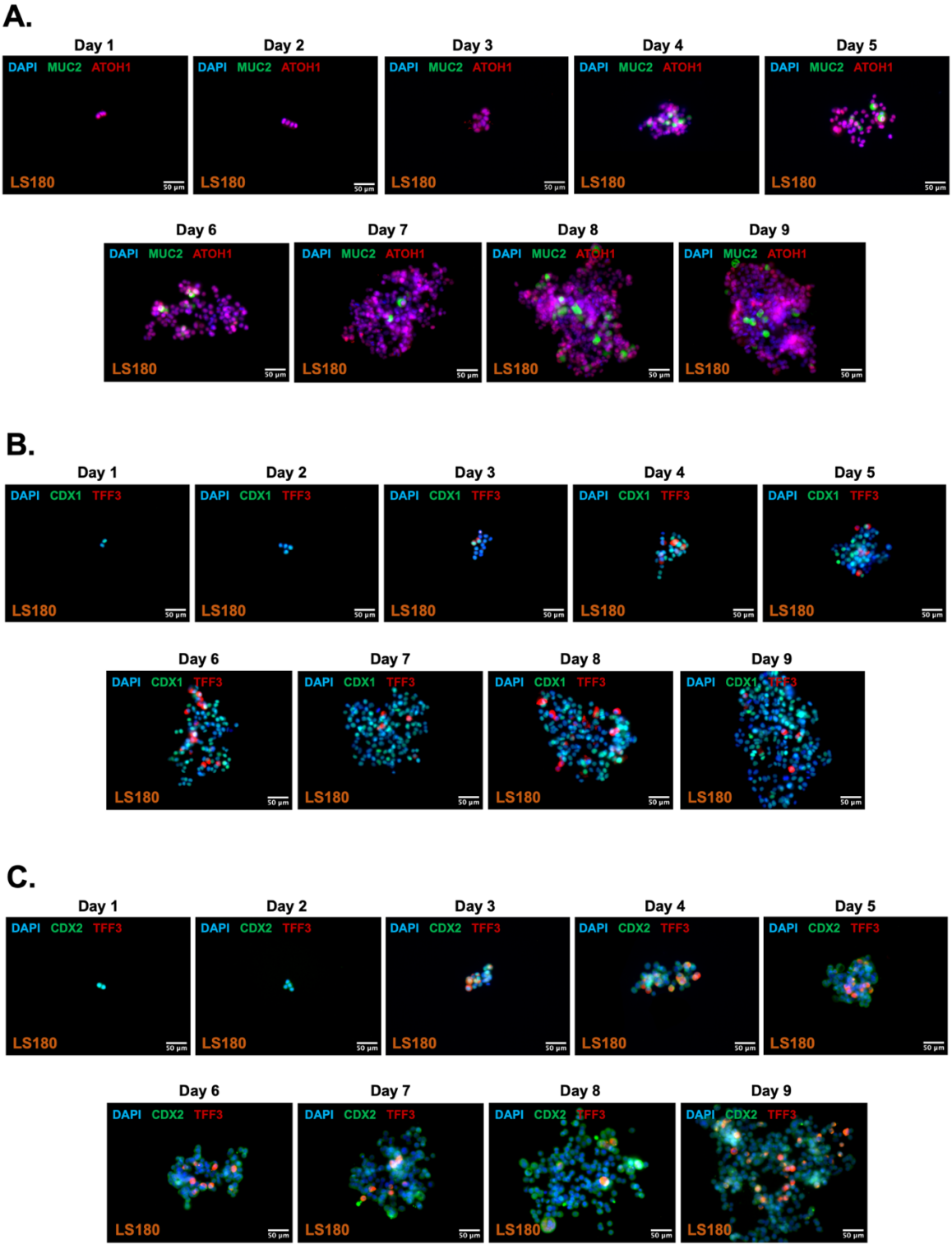


Figure 3.13 Protein expression of ATOH1, CDX1, and CDX2 transcription factors in single cell derived LS180 colonies. Cells were cultured over the indicated period and fixed prior to being stained with DAPI (blue, 1:10000), (A) ATOH1 (red, 1:500) and MUC2 (PR5D5, green, 1:250); (B) CDX1 (green, 1:500) and TFF3 (red, 1:500); (C) CDX2 (green, 1:500) and TFF3 (red, 1:500) antibodies. Magnification: 20X, scale bar: 50 µm.

3.2.5 MUC2 and TFF3 secretion assay in a panel of CRC cell lines

Our next objective was to investigate whether MUC2-positive cell lines can secrete mucin. Studying MUC2 in a secreted form was quite challenging due to its high molecular weight, complex glycosylation pattern, and difficult detection properties.

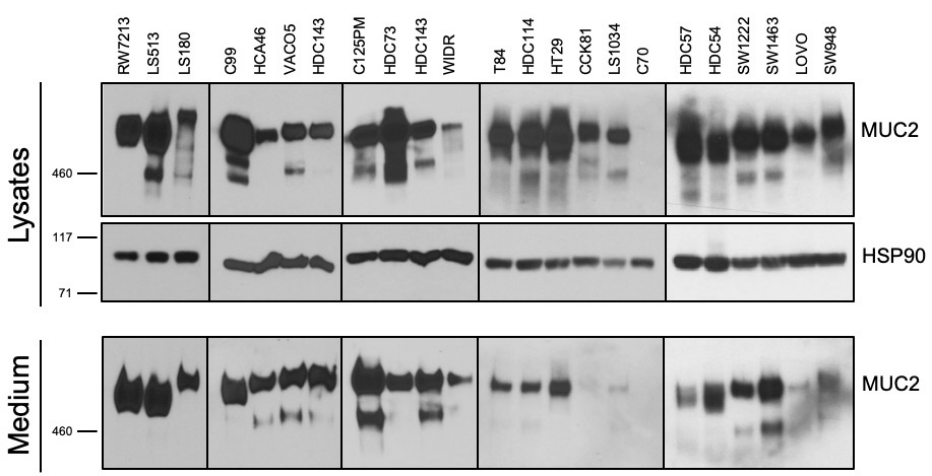
We tested multiple conditions using in-house PR5D5 (mouse, monoclonal) and commercial CCP58 (mouse, monoclonal) antibodies. The following conditions were used during Western blotting: heating samples with and without dithiothreitol (DTT), and no heating with and without DTT in the sample mix. Detailed description of the Western Blot experiment is provided in the Materials & Methods chapter (Chapter 2.8).

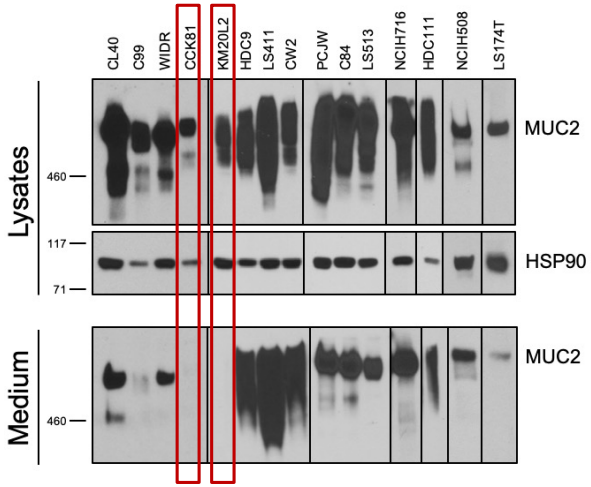
It was identified that CCP58 (mouse, 1:2000) commercial antibody worked in both heated and DTT-containing samples as opposed to PR5D5. CCP58 antibody detected over the 460 kDa molecular weight of MUC2 protein expression in MUC2-positive cell lines (Appendix Figure 2). This antibody did not show reactivity in most MUC2-negative cell lines, except in certain cell lines where non-specific or smeared bands were observed (Appendix Figure 2A). However, several issues occurred while using this antibody. First, we were unable to detect MUC2 in supernatants to reveal whether MUC2 is secreted (Appendix Figure 2B). Additionally, it failed to recognize medium/low levels of MUC2 expressed cells belonging to groups 2 and 3. A variety of reasons could account for this, such as the fact that supernatants should be concentrated, that the antibody itself has low detection sensitivity, that MUC2-positive cell lines do not release MUC2, and that CCP58 cannot react to secreted forms of MUC2 protein.

To address these issues, we concentrated the supernatants using Amicon® Ultra-0.5 centrifugal filters that perform ultrafiltration with a 100 kDa molecular weight cutoff Ultracel® membrane (Appendix Figure 2C). This approach was not successful in detecting MUC2 secretion from four MUC2-positive cell lines, including LS180, LS174T, LS513, and RW7213. The problem could be related to the specificity of the antibody or the possibility that the cell lines do not secrete mucin. Then, we conducted a deglycosylation procedure using the Protein Deglycosylation Mix II kit (P6044S, New England Biolabs) to remove glycans from MUC2 in LS180, LS174T, and LS513 CRC cell lines. Although there was a slight shift in molecular weight in lysates, no bands were detected in supernatants (Appendix Figure 2D). Due to the fact that the CCP58 antibody could not detect supernatants, we stopped using it for Western Blot analysis.

Following this, the in-house PR5D5 antibody was tested in CRC samples under no heat and no DTT conditions which showed a reactivity in both lysates and supernatants. The majority of MUC2-positive cell lines secreted mucin, with the exception of a few cell lines, including CCK81, KM20L2, HCT8, JHCOLOYI, LIM1863, and HT55 highlighted in Figure 3.14. It is important to emphasise that among group 4 cell lines, only a small subset (7 out of 27) showed the secretion of MUC2 (Figure 3.14). Absolute amount of secreted MUC2 in group 4 cell lines was mainly low.

MUC2 secretion in CRC cell lines





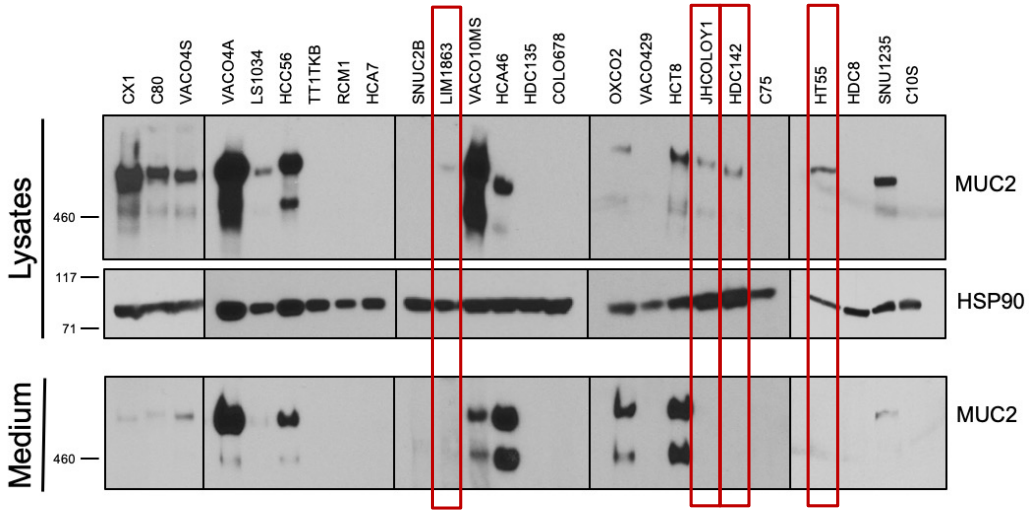


Figure 3.14 MUC2 secretion assay in a panel of CRC cell lines. In total, fifty-six CRC cell lines were analysed in terms of their mucin secretion profiles. Lysates from each cell line served as controls for intracellular MUC2, whereas secreted extracellular MUC2 was identified in the culture medium by utilising in-house MUC2 (PR5D5, 1:250) mouse monoclonal antibody. Heat Shock Protein 90 kDa (HSP90, 1:1000) was used as a loading control. Highlighted lines showed no apparent secretion.

Additionally, we quantified the amount of MUC2 in both cell lysates and supernatants after normalisation procedures (Table 3.6).

Group 1

	% of MUC2 in lysate	% of MUC2 in supernatant
C99	56.4	43.6
CL40	56.7	43.3
HCA46	17.95	82.05
HDC9	27.8	72.2
LS411	32	68
LS513	52.1	47.9
NCIH508	26.6	73.4
NCIH716	59.4	40.6
PCJW	53.7	46.3
RW7213	26.8	73.2

Group 2

	% of MUC2 in lysate	% of MUC2 in supernatant
C125PM	21.4	78.6
C80	80.3	19.7
C84	33.1	66.9
HDC111	20.6	79.4
HDC114	70.5	29.5
HDC143	41.8	58.2
HDC73	63.4	36.6
HT29	61.3	38.7
LS174T	55.7	44.3
LS180	49.5	50.5
T84	61.4	38.6
VACO4A	39.7	60.3
WIDR	27.6	72.4

Group 3

	% of MUC2 in lysate	% of MUC2 in supernatant
CW2	27.3	72.7
CX1	92.5	7.5
HCC56	45.9	54.1
HDC54	29.9	70.1
HDC57	75	25
KM20L2	100	0
SW1222	30	70
SW1463	17.4	82.6
VACO4S	50.4	49.6

Group 4

	% of MUC2 in lysate	% of MUC2 in supernatant
HCT8	9.4	90.6
HT55	100	0
LIM1863	100	0
OXCO2	5.6	94.4
SNU1235	25.5	74.5
VACO10MS	53.7	46.3
VACO5	23.6	76.4

Table 3.6 Quantification of intracellular and extracellular MUC2 across a panel of 39 CRC cell lines. The percentage of MUC2 was determined after normalisation based on the dilution factor.

We also aimed to determine whether TFF3 is released by CRC cell lines. Initially, commercial monoclonal TFF3 antibody was tested in multiple conditions to ensure that it would work in both cell lysates and cell culture medium. Three CRC cell lines, namely RW7213, PMFKO14, and SW403 were screened in the following conditions: heating samples with and without DTT, and no heating with and without DTT (Figure 3.15). We concluded that all three CRC cell lines exhibited reactivity when samples were heated and treated with DTT.

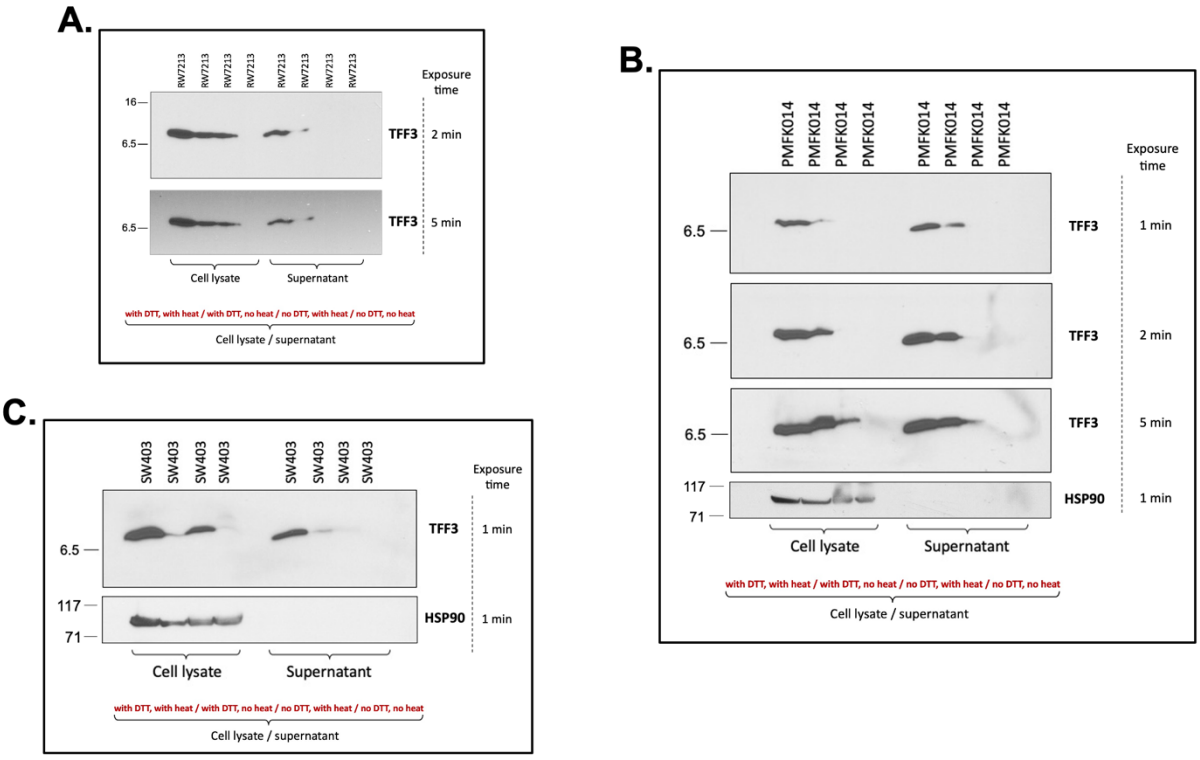


Figure 3.15 Examination of rabbit monoclonal TFF3 antibody under various conditions. Three CRC cell lines, namely (A) RW7213, (B) PMFKO14, and (C) SW403 were selected to evaluate the effectiveness of monoclonal TFF3 antibody. Assessment was conducted in both cell lysates and supernatants using following conditions, heating samples with and without DTT, and no heating with and without DTT. The antibody worked efficiently in samples that are both heated and supplemented with DTT in all three CRC cell lines. HSP90 (rabbit, diluted at 1:1000) was utilised as a loading control.

To investigate whether TFF3 is released by CRC cell lines, Western blot analysis was conducted using a monoclonal rabbit TFF3 (Abcam, catalogue number: 108599) antibody (Figure 3.16). The antibody showed specificity and revealed a band at around 10 kDa molecular weight. In the next chapter section (Chapter 3.2.6), we will discuss the reasons for choosing a different commercial TFF3 antibody in Western blot analysis. Similar to MUC2, most of analysed CRC cell lines appeared to release TFF3, except HDC114 and C125PM which are both included in the group 2. It should also be noted that KM20L2 and T84 seemed to release a small amount of TFF3.

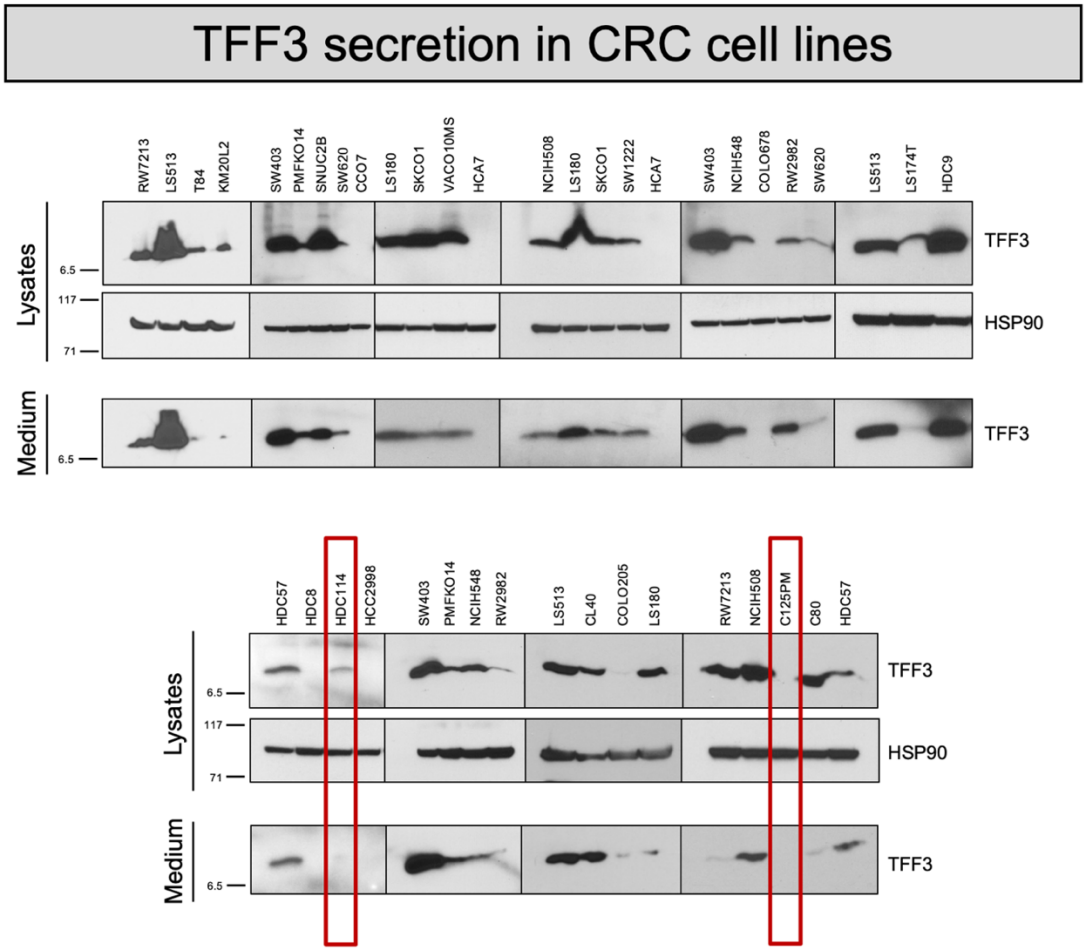


Figure 3.16 TFF3 secretion assay in a panel of CRC cell lines. In total, twenty-nine CRC cell lines were analysed based on their TFF3 secretion profiles. Intracellular and extracellular TFF3 that is released by the cells was detected in nearly all the cell lines by using a rabbit monoclonal TFF3 (1:1000) antibody. HSP90 (rabbit, 1:1000) was used as a loading control.

3.2.6 Comparison of monoclonal and polyclonal TFF3 antibodies

In this study, we used two TFF3 antibodies, including a polyclonal rabbit TFF3 (Raybiotech, Inc, catalogue number: 130-10147) and a monoclonal rabbit TFF3 (Abcam, catalogue number: 108599). As described in Chapter 3.2.5, a polyclonal rabbit TFF3 antibody was utilised to classify CRC cell lines in terms of protein expression level of TFF3 through an immunofluorescence assay. In Western blot analysis, a polyclonal TFF3 antibody showed reactivity in a band at approximately 50 kDa, despite its predicted molecular weight of 8.6 kDa according to GeneCards (https://www.genecards.org/cgi-bin/carddisp.pl?gene=TFF3) and UniProt resources (https://www.uniprot.org/uniprotkb/Q07654/entry) (Figure 3.17). This could be due to either its interaction with other proteins, such as FCGBP and MUC2, or the possibility of cross reactivity with other proteins of the polyclonal TFF3 antibody in Western blot. Polyclonal TFF3 antibody also reacted in several TFF3-negative cell lines, such as CACO2, COLO678, DLD1, and HCA7 (Figure 3.17A). Furthermore, it was unable to detect TFF3 in cell culture medium (Figure 3.17B). These statements suggest that the polyclonal TFF3 antibody did not work effectively during Western blot analysis. To validate this argument, further analysis was conducted.

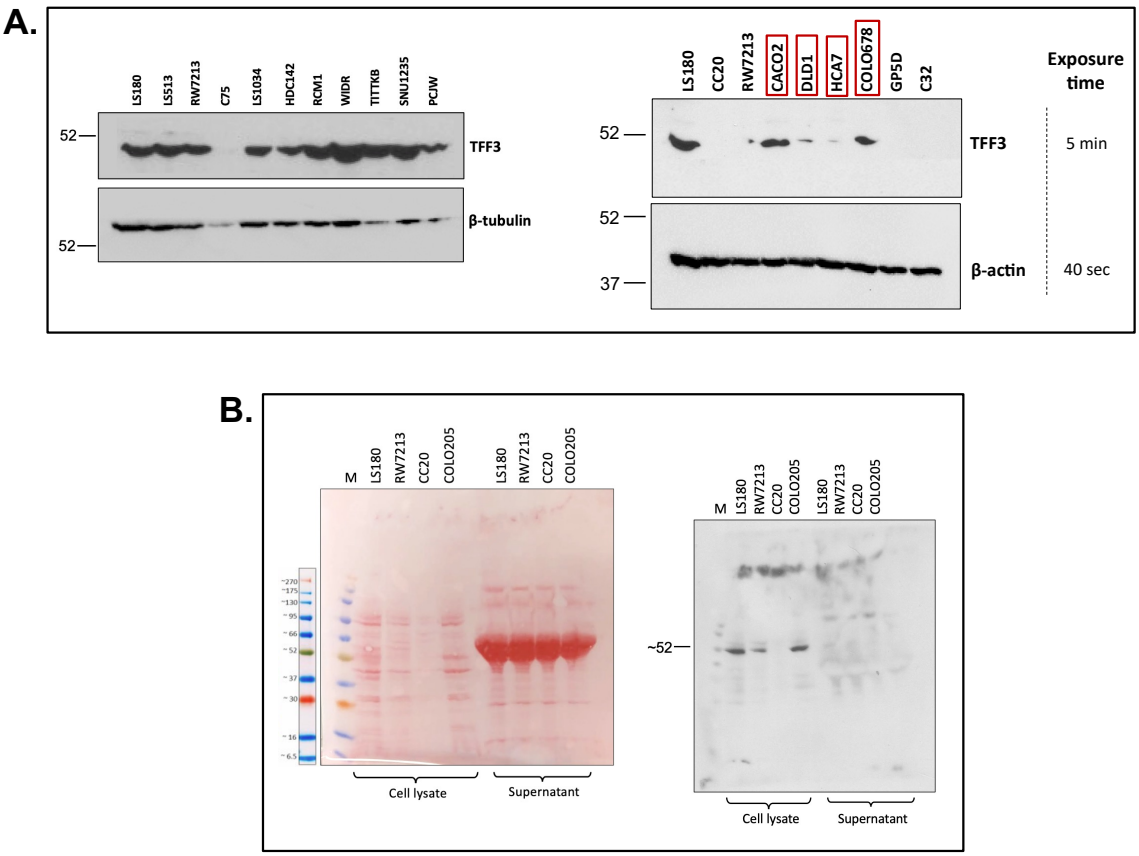


Figure 3.17 Detection of TFF3 protein expression in CRC cell lines using a polyclonal TFF3 antibody. (A) shows presumed intracellular TFF3 protein levels in various CRC cell lines. The four highlighted cell lines, classified as MUC2/TFF3-negative, showed reactivity with the polyclonal TFF3 antibody. (B) represents presumed intracellular and extracellular TFF3 expression in four cell lines, namely LS180, RW7213, CC20, and COLO205.

In order to evaluate the specificity of the polyclonal TFF3 antibody, we carried out a knockdown experiment using siRNA against *TFF3*, as well as scr-siRNA as a control. The reactivity of the antibody was assessed in both Western blotting and immunofluorescence analyses. In Western blot assay, no visible difference was observed compared to scr-siRNA, despite the knockdown being successful (Figure 3.18A). By contrast, TFF3 protein expression was significantly decreased in immunofluorescence analysis, indicating that the polyclonal antibody reacts properly during immunofluorescence assay (Figure 3.18B).

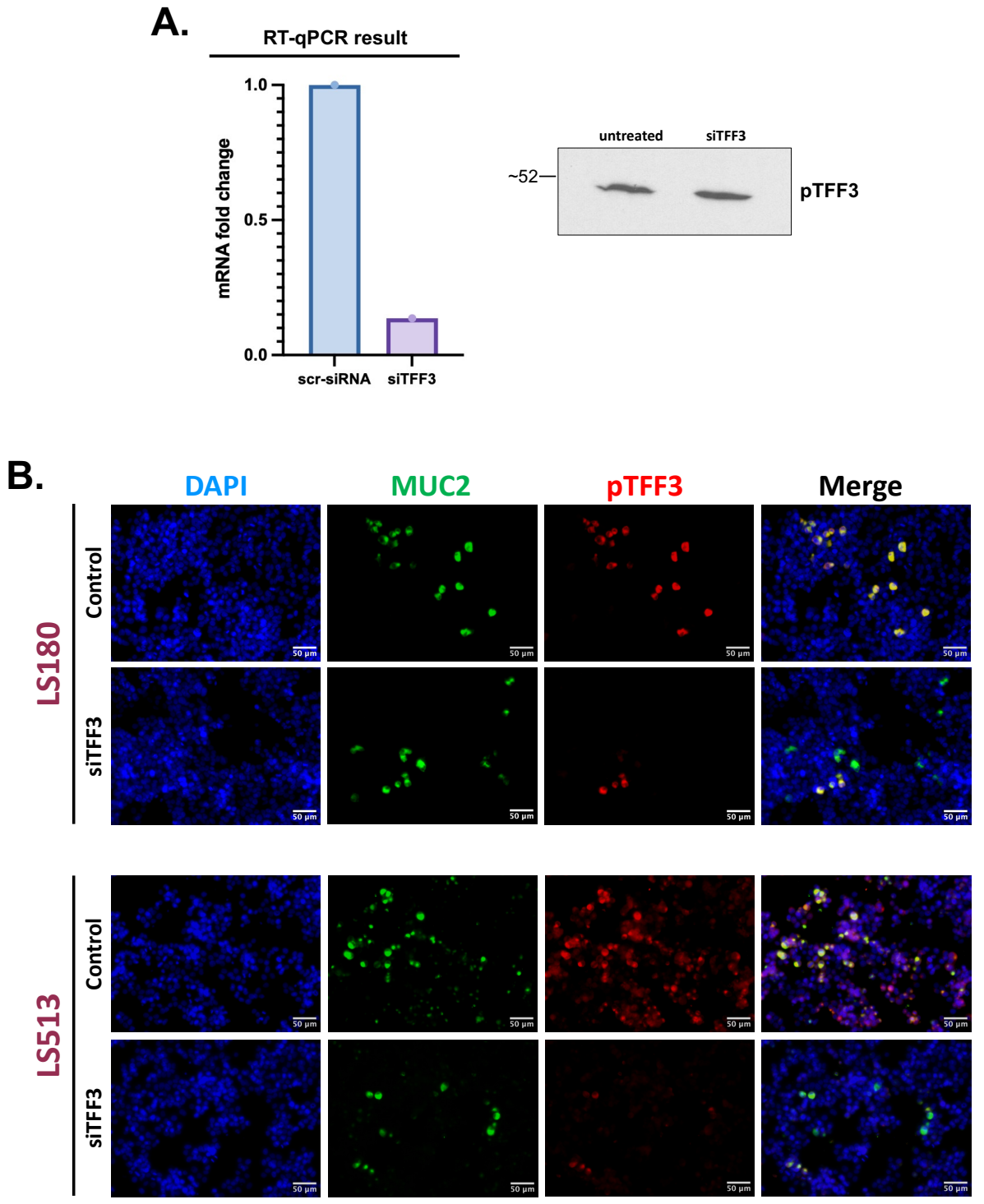
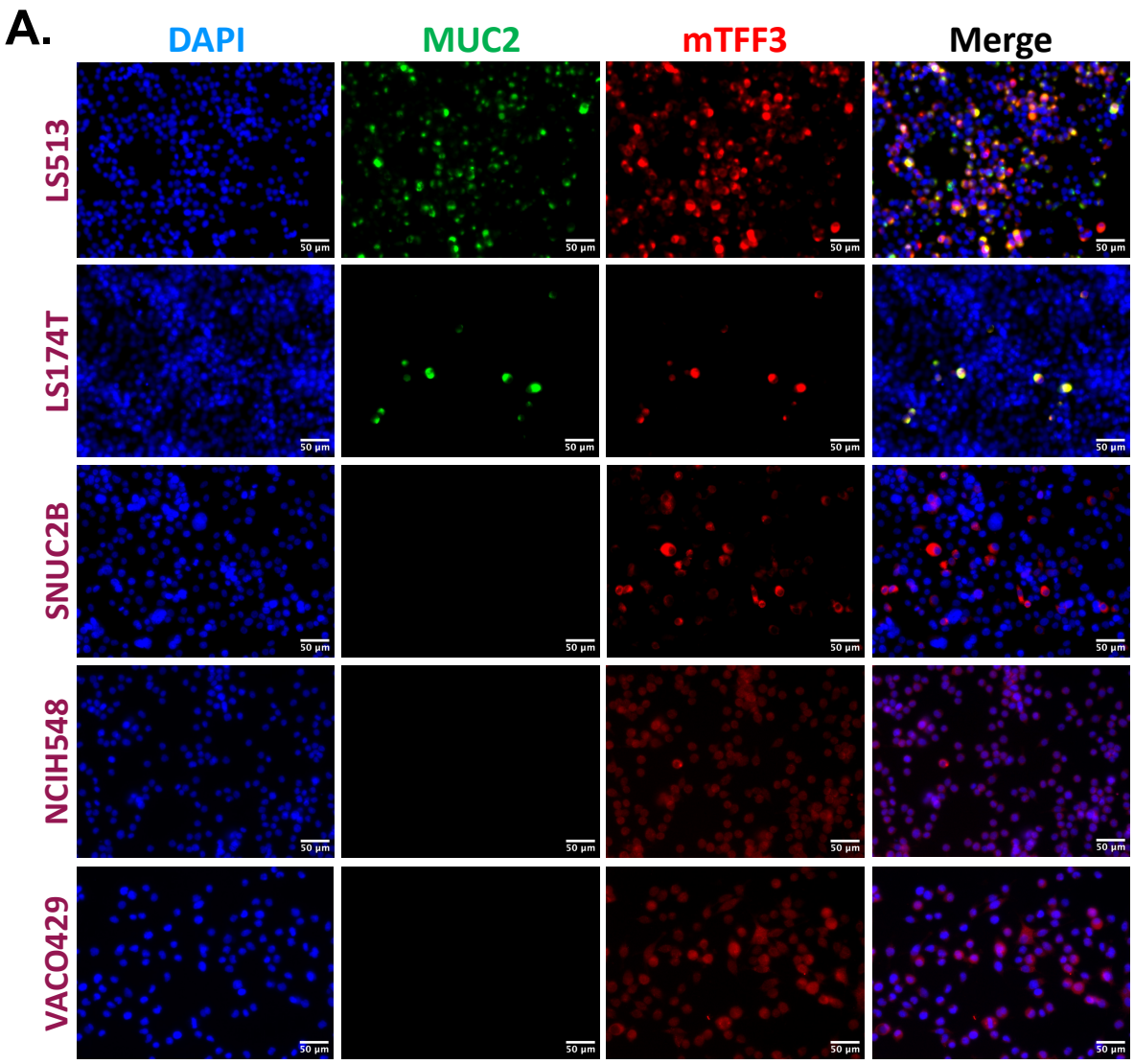


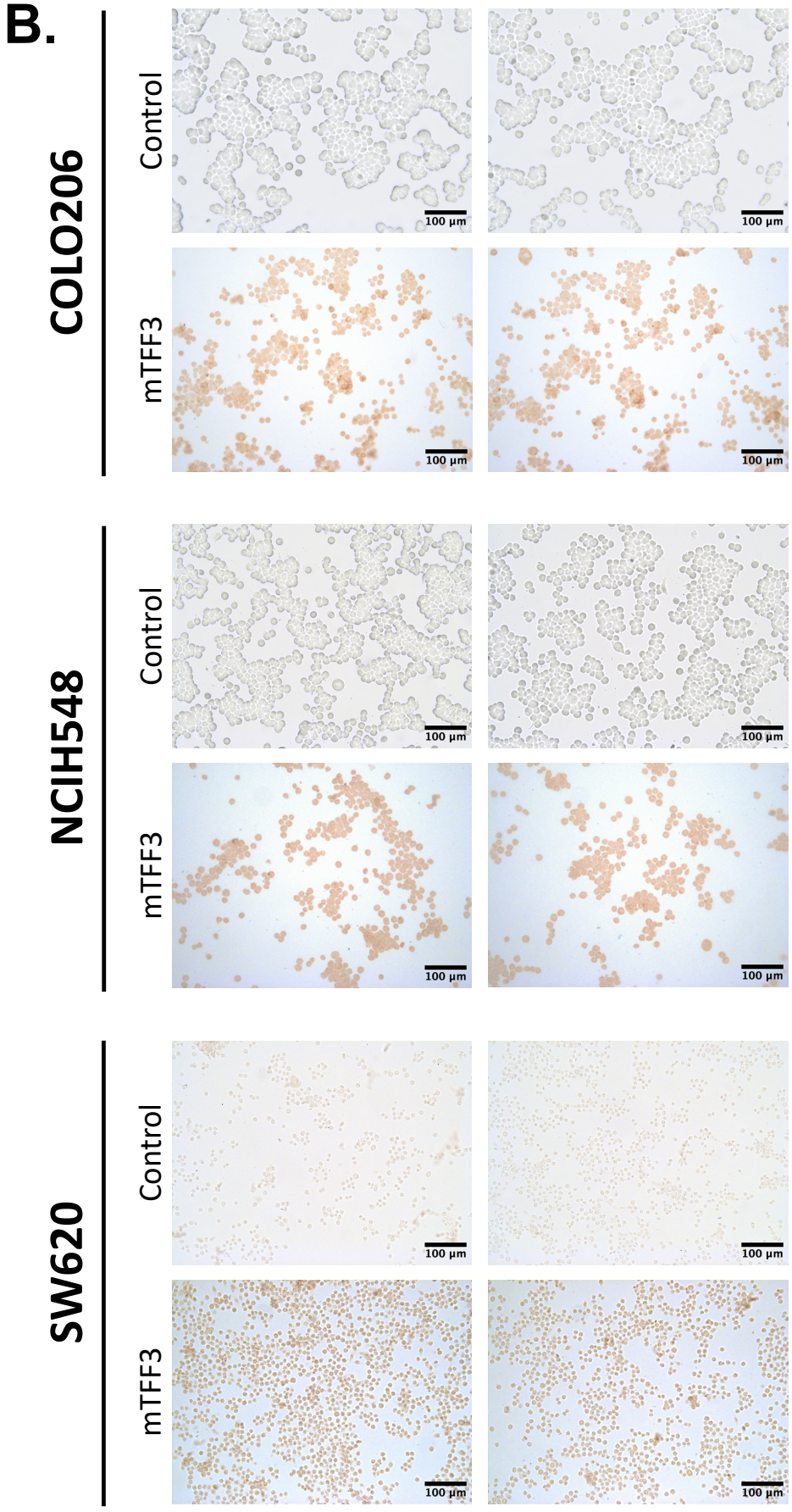
Figure 3.18 Investigation of a polyclonal antibody during the knockdown of *TFF3* mRNA in CRC cell lines. (A) depicts the results of qRT-PCR and Western blot analyses while downregulating *TFF3*. For the Western blot analysis, a polyclonal TFF3 antibody was used, and no difference was observed between the control and siRNA-treated samples. (B) illustrates immunofluorescence staining conducted utilising DAPI (1:10000), MUC2 (1:250),

and polyclonal TFF3 (pTFF3, 1:500) antibodies in both control and siRNA-treated samples of LS180 and LS513 cell lines. Scale bar: 50 μm.

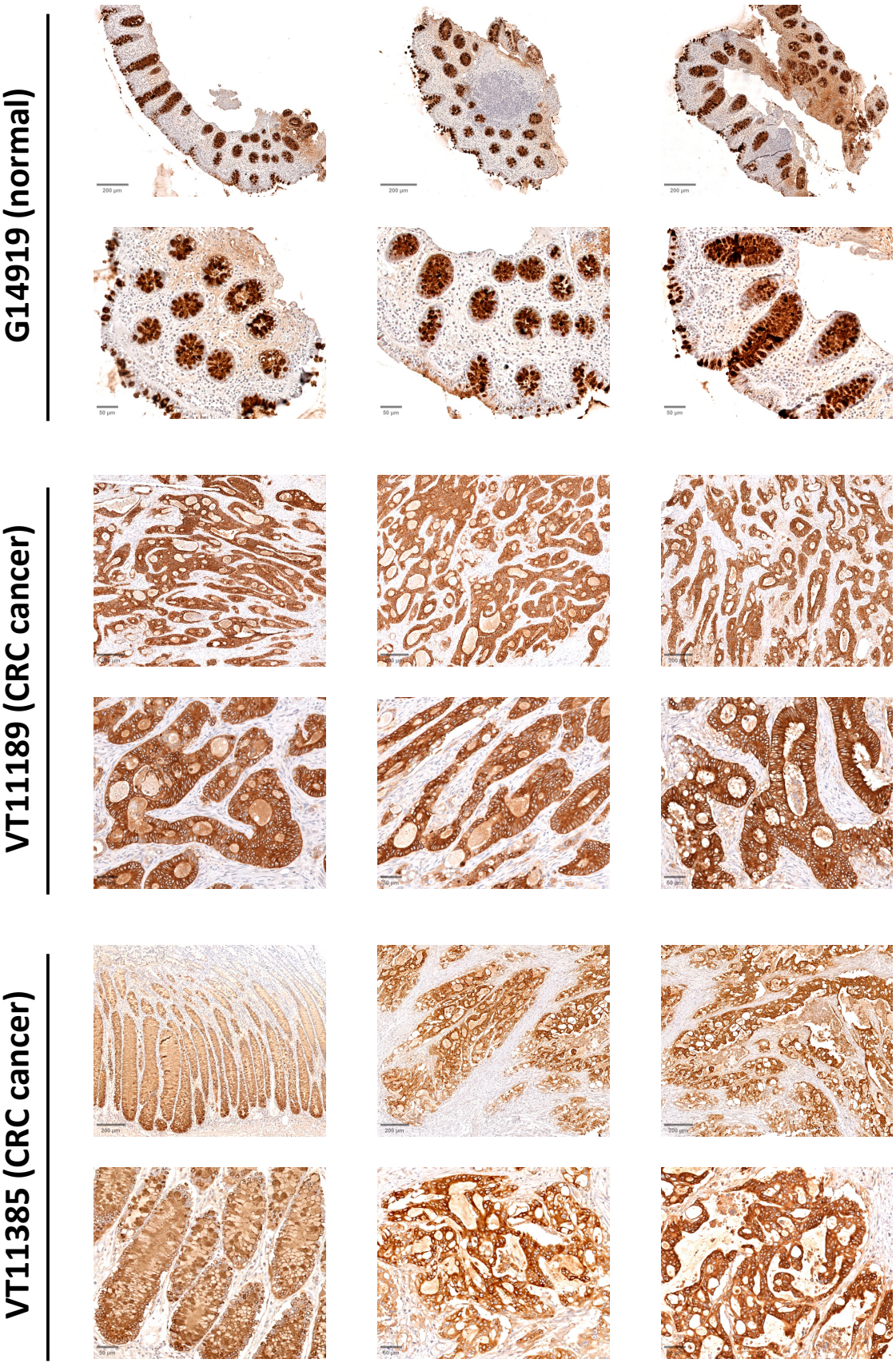
As a next step, we compared our results obtained with the polyclonal antibody relative to the monoclonal TFF3 antibody by performing immunofluorescence and immunohistochemistry analyses in multiple CRC cell lines, as well as in human FFPE tissue sections (Figure 3.19A/B/C). Additional representative images are included in Appendix Figure 3.

We confirmed the accuracy of our classification by choosing cell lines from each group and analysing them with the monoclonal TFF3 antibody. As expected, the monoclonal TFF3 antibody showed positivity with cell lines from groups 1, 2, and 4 (Figure 3.19A). Additionally, three cell lines, namely COLO206, NCIH548, and SW620 were tested by immunohistochemistry using the monoclonal TFF3 antibody and validated that nearly all cells stain with TFF3 (Figure 3.19B). Healthy and colon cancer human tissue sections were also analysed, and the results agreed with those achieved using the polyclonal antibody (Figure 3.19C).


A.
DAPI
MUC2
mTFF3
Merge
LS513
LS174T
SNUC2B
NCIH548
VACO429
50 μm


B.
COLO206
NCIH548
SW620
Control
mTFF3
100 µm

C.



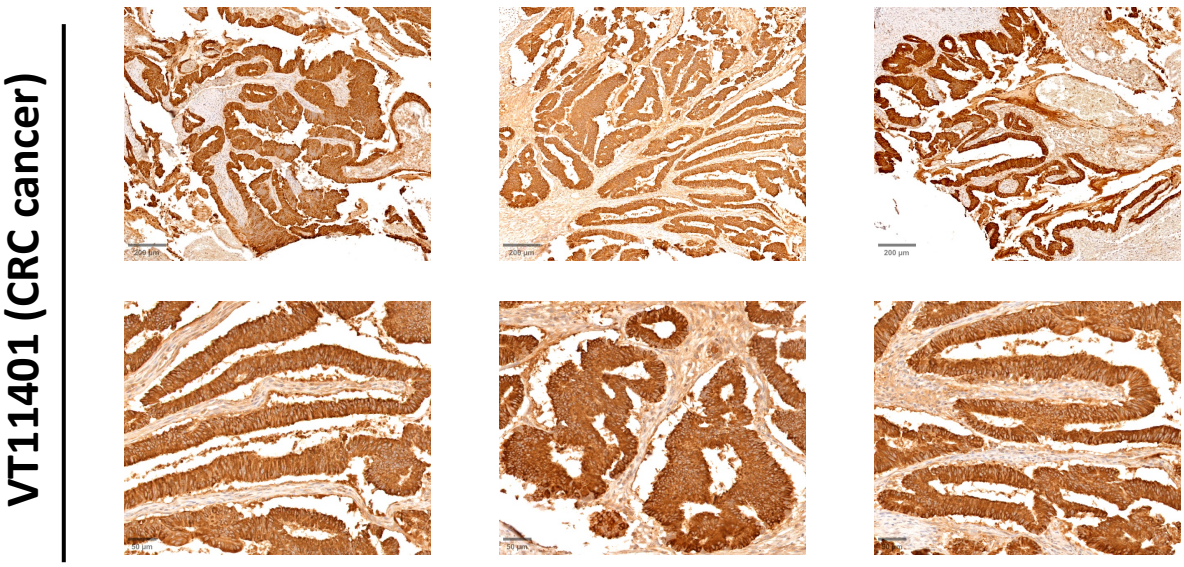


Figure 3.19 Detection of TFF3 protein expression using a monoclonal TFF3 antibody in CRC cell lines and human tissue sections. (A) illustrates detection of TFF3 protein using a monoclonal rabbit antibody (mTFF3, 1:1000) in several CRC cell lines, namely LS513, LS174T, SNUC2B, NCIH548, and VACO429. (B) and (C) describe TFF3 expression utilising a monoclonal rabbit antibody (mTFF3, 1:1000) and HRP-conjugated secondary antibody (1:3000) in CRC cell lines and human FFPE slides through immunohistochemistry.

3.3 DISCUSSION

We classified a large panel of CRC cell lines and grouped them into five distinct categories, including MUC2/TFF3-high, MUC2/TFF3-medium, MUC2/TFF3-low, MUC2-negative (or low) and TFF3-high, and MUC2/TFF3-negative. Our classification validates and improves the current mRNA-based classification of CRC cell lines in terms of *MUC2* and *TFF3* genes. We demonstrated that TFF3 can exist without MUC2 at a protein level in many CRC cell lines. Initially, this was observed based on our microarray and TCGA RNA-seq datasets and then was validated through mainly immunofluorescence, as well as immunohistochemistry assays.

Although there are previous publications indicating MUC2 and TFF3 as the primary markers of goblet cells based on their high mRNA expression levels, to our knowledge, this is the first classification that has been verified through a protein expression profile (Miller et al., 2021; Rodriguez-Salas et al., 2017; Sadanandam et al., 2013). A study by Miller et al. documented that mRNA expression of MUC2, TFF3, and FCGBP were notably elevated in CMS3 subtype TCGA colon adenocarcinoma and CRC cell lines (LS174T and HT29) (Miller et al., 2021). However, the expression levels of these goblet cell-associated markers showed heterogeneity in a descending order from the CMS1 subtype (SNUC1) to CMS2 (LOVO), and further to CMS4 subtypes (HCT116 and SW480) which was in agreement with our findings at a protein level (Miller et al., 2021).

Lumen formation is a key indicator of cancer stem cell differentiation in a 3D environment (Ashley et al., 2013; Yeung et al., 2010). We examined whether lumen formation is associated with goblet cell differentiation. According to our analysis of the

preliminary data, there was no significant correlation between the two. This result is consistent with previous findings (T. C. Liu et al., 2022; Yeung et al., 2011).

In this chapter, secretion of MUC2 and TFF3 was also investigated. Mucin is produced and secreted by intestinal goblet cells and its fast turnover protects epithelial cells against pathogens, as well as damage. It was known that MUC2 expression significantly reduced in colon cancer which is strongly correlated with increased metastasis and shorter survival rate (Bhatia et al., 2019; Li et al., 2018). The study by Ribeirinho-Soares et al. showed that stage II CRC patients with low levels of MUC2 exhibited a more favourable response to adjuvant chemotherapy compared to those with high levels of MUC2 expression (Ribeirinho-Soares et al., 2021). Patients with mucinous carcinomas, distinguished by abundant extracellular secretion of MUC2, displayed reduced sensitivity to chemotherapy (Green et al., 1993). For TFF3, several reports indicated that upregulation of TFF3 associates with the advanced stage of CRC, poor prognosis, and lymph node metastasis (Babyatsky et al., 2009; Yusufu et al., 2019; Yusup et al., 2017).

Most research on MUC2 secretion was performed following bacterial exposure and treatment with secretagogues, such as acetylcholine and carbachol (Etienne-Mesmin et al., 2019; Gustafsson et al., 2012; Halm & Halm, 2000; T. H. Jung et al., 2015). Our study shows that a few cancers may not secrete MUC2 though they produce it and likewise for TFF3. Considering that most of the CRC cell lines release MUC2, it raises several questions regarding the potential benefit of mucins for cancer cells. Numerous studies have reported mucin's role as an obstacle to drug delivery which ongoing pharmaceutical research is trying to address this issue (Boegh & Nielsen, 2015; Sigurdsson et al., 2013).

To define the hierarchical pattern of goblet cell differentiation in terms of MUC2 and TFF3 expression, single-cell experiments were conducted. The results showed that TFF3 was expressed prior to MUC2, suggesting it is most likely needed to hold mature mucin together. MUC2 expression might be delayed due to its high molecular weight and extensive glycosylation pattern of MUC2.

We also analysed the differences between monoclonal and polyclonal TFF3 antibodies to ensure that our novel classification matches both types of antibodies. It is important to carefully select antibodies for the research which is also confirmed by our study. It would be feasible to consider the use of monoclonal antibodies, whenever possible. Further discussion of what might control these differences will be described in Chapter 6.

CHAPTER 4

VALIDATION OF OUR CLASSIFICATION IN TISSUE SECTIONS FROM CRC PATIENTS

CHAPTER 4: VALIDATION OF OUR CLASSIFICATION IN TISSUE SECTIONS FROM CRC PATIENTS

4.1 INTRODUCTION AND AIMS

Results in the previous chapter were all based on evaluating CRC cell lines in terms of goblet cell differentiation. To confirm our classification, we performed extensive studies on human FFPE colon tissue sections.

Formalin-fixed paraffin-embedded (FFPE) tissue samples are essential resources for identifying biomarkers and investigating the molecular profiling of various cancers. FFPE sections have multiple advantages in comparison to fresh frozen tissue. FFPE samples can be stored at room temperature without losing their cellular and structural characteristics for several years. They are easily accessible as FFPE blocks are mostly kept in pathology laboratories under the term "pathology archive". Although during fixing and archiving procedures, FFPE sections experience certain alterations, including cross-linking and DNA fragmentation, which in turn, impact the amount and quality of the DNA that is extracted from sections (Gao et al., 2020).

It was recently reported by Gao et al. that mutation analysis of a 22-gene panel conducted on fresh frozen tissue and FFPE tissue sections can still reveal some differences, even though matched samples showed strong agreement between the two types (Gao et al., 2020). This study recommends prioritising the use of fresh frozen tissue sections for gene mutation analysis, while acknowledging FFPE sections as acceptable alternatives (Gao et al., 2020).

RNA degradation during processing of FFPE sections is another technical challenge, which subsequently affects gene expression profiling of FFPE sections. There are

currently two commonly used platforms for analysing gene expression in FFPE samples: the Nanostring nCounter® platform and the Affymetrix GeneChip® Human Transcriptome Array 2.0. A study conducted by Zhu et al. assessed the expression of 516 prognostic gene signatures (obtained from frozen CRC tissue samples) using two platforms in 42 FFPE tissue sections (Zhu et al., 2016). The findings of the study indicated that good quality RNA extracted from FFPE samples could be a reliable source for identification of prognostic genes originating from frozen tissue in CRC patients (Zhu et al., 2016). Nevertheless, it remains as a difficult task to evaluate studies on gene expression profiling involved in CRC progression.

Human tissue sections were analysed for CRC classification originated from the Vioxx® in Colorectal Therapy: Definition of Optimal Regime (VICTOR) study. VICTOR was an international trial that enrolled 7000 patients for five years (Pendlebury et al., 2003). Patients with histologically approved stage II and III CRC who received curative therapy were included in this study (Midgley et al., 2010; Pendlebury et al., 2003). The primary aim of the trial was to compare survival rates of patients following treatment with rofecoxib, a cyclooxygenase-2 inhibitor, relative to a placebo. Based on the analysis of 2434 patients, rofecoxib did not improve survival rates, which led to trial interruption (Midgley et al., 2010). Furthermore, patients showed a high frequency of cardiovascular thrombotic events after rofecoxib administration (Kerr et al., 2007).

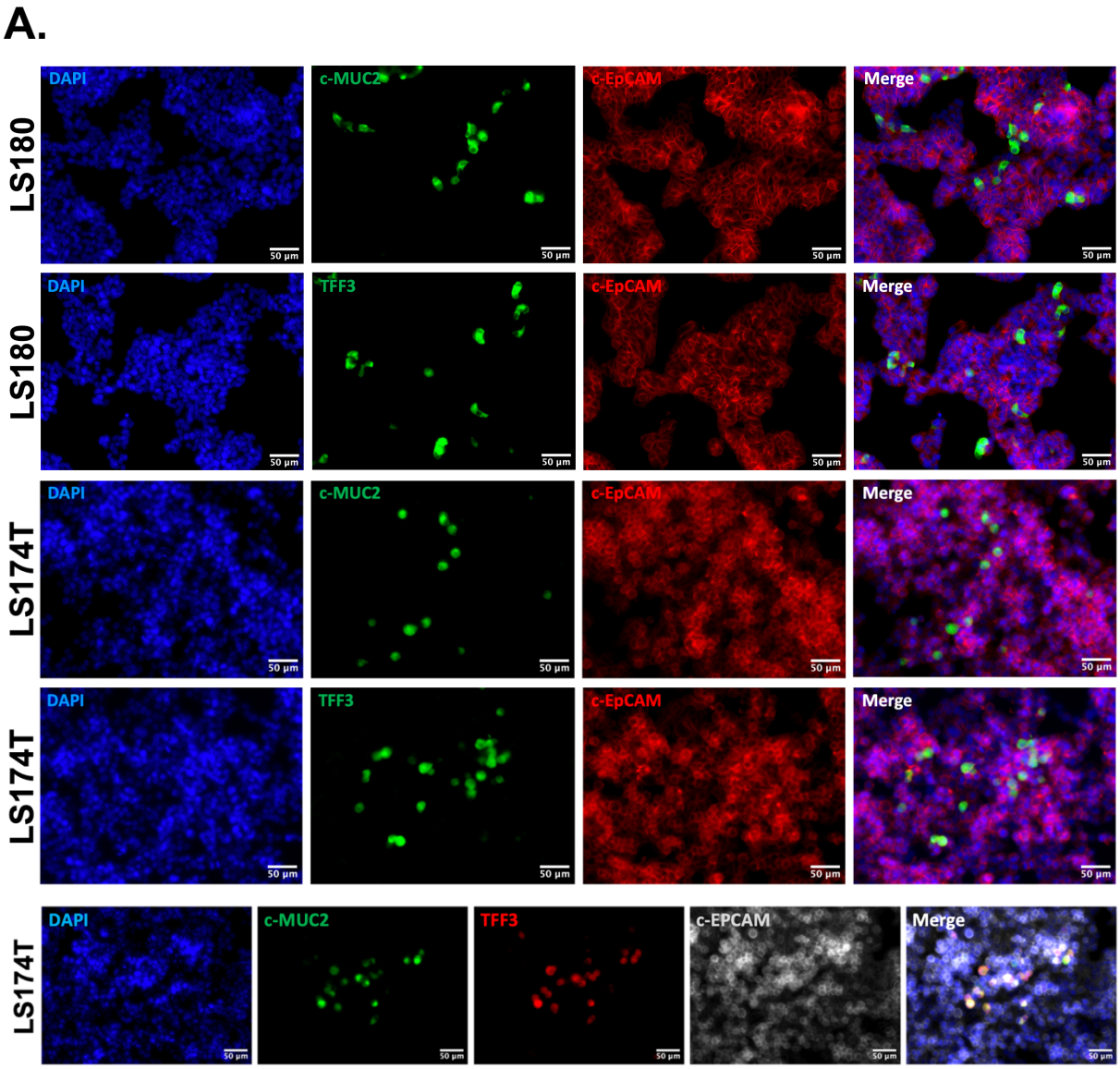
Chapter 4 aims to validate our CRC cell lines classification based on protein expression profiles of MUC2 and TFF3 in human specimens. It also examines histopathological reports and mutational data from analysed patients' tissue sections.

4.2 RESULTS

4.2.1 Analysis of tissue sections from of CRC patients to compare with the cell line characterisation - in collaboration with Professor David Kerr, Professor Marco Novelli, and Dr David Church

Prior to analysing FFPE tissue sections from the VICTOR clinical trial, multiple sections of human colonic tissue were tested in order to identify the correct conditions for MUC2 and TFF3 staining. The detailed information about immunofluorescence analysis in tissue sections was described in the Materials and Methods section (Chapter 2.9.2).

During the analysis, we used in-house PR5D5 antibody for MUC2 and a polyclonal commercial TFF3 antibody for TFF3 staining. In order to make a combination of MUC2 (mouse), TFF3 (rabbit) antibodies, and in-house EpCAM (AUA1, mouse) as a cell surface marker, we purified MUC2 and AUA1 antibodies and conjugated them with Alexa Fluor 488 and Alexa Fluor 647 dyes, respectively, following the protocol shown in Chapter 2.12. Conjugated antibodies were tested on LS180 and LS174T cell lines (Figure 4.1A), human normal (Figure 4.1B), and CRC tissue sections (Figure 4.1C). Combining conjugated EpCAM with conjugated MUC2 and unconjugated TFF3 showed some overlap between EpCAM and MUC2/TFF3 staining in a normal colon FFPE tissue sample, G15003 kindly provided by Dr Shazia Irshad from Radcliffe Department of Medicine (University of Oxford) (Figure 4.1B). There may be a technical issue involving the conjugation, or the antigen retrieval step during immunofluorescence requires different boiling temperatures for conjugated MUC2 and EpCAM. By contrast, EpCAM is nicely revealed in epithelial cell membranes when we stain tissue sections with TFF3 and conjugated EpCAM (Figure 4.1C).


A.
LS180
DAPI
c-MUC2
c-EpCAM
Merge
LS180
DAPI
TFF3
c-EpCAM
Merge
LS174T
DAPI
c-MUC2
c-EpCAM
Merge
LS174T
DAPI
TFF3
c-EpCAM
Merge
LS174T
DAPI
c-MUC2
TFF3
c-EPCAM
Merge
B.
DAPI
c-MUC2
TFF3
c-EpCAM
Merge
G15003
G15003
G15003

C.

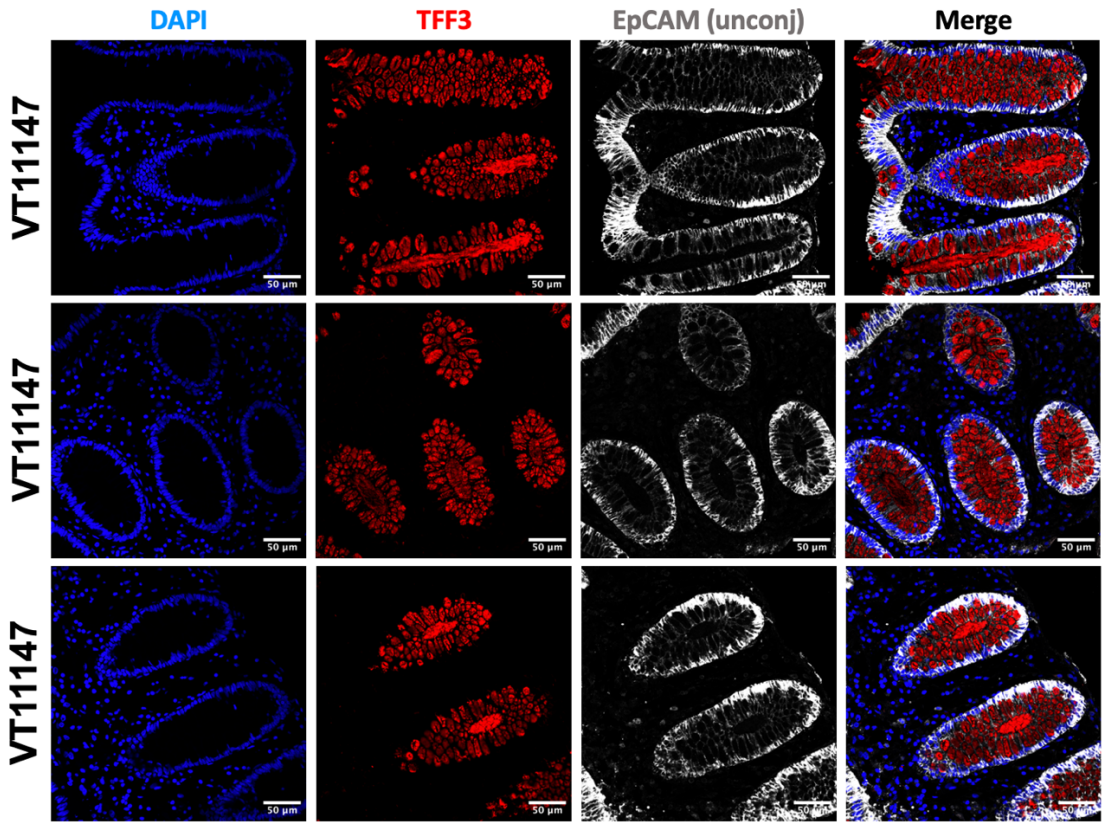


Figure 4.1 Staining of conjugated and unconjugated antibodies in fixed CRC cell lines and human FFPE tissue sections. (A) Immunofluorescence analysis of conjugated MUC2 (c-MUC2/c-PR5D5, mouse, 1:250), unconjugated TFF3 (rabbit, 1:1000), and conjugated EpCAM (mouse, 1:500) antibodies in LS180 and LS174T cell lines. (B) Same staining was also conducted in G15003 normal colon. (C) Confocal images of VT11147, human colonic mucosa, stained by TFF3 (rabbit, 1:1000) and conjugated EpCAM (mouse, 1:500) antibodies. Magnification: 20X, scale bar: 50 µm.

We then decided to use CellBriteTM Fix membrane stain (Biotium), which covalently binds to membrane proteins in epithelial and non-epithelial cells. CellBrite is easily detectable under both boiling and unboiling conditions during the antigen unmasking process. Multiple tests were conducted to determine optimal conditions for staining MUC2 and CellBrite (Figure 4.2).

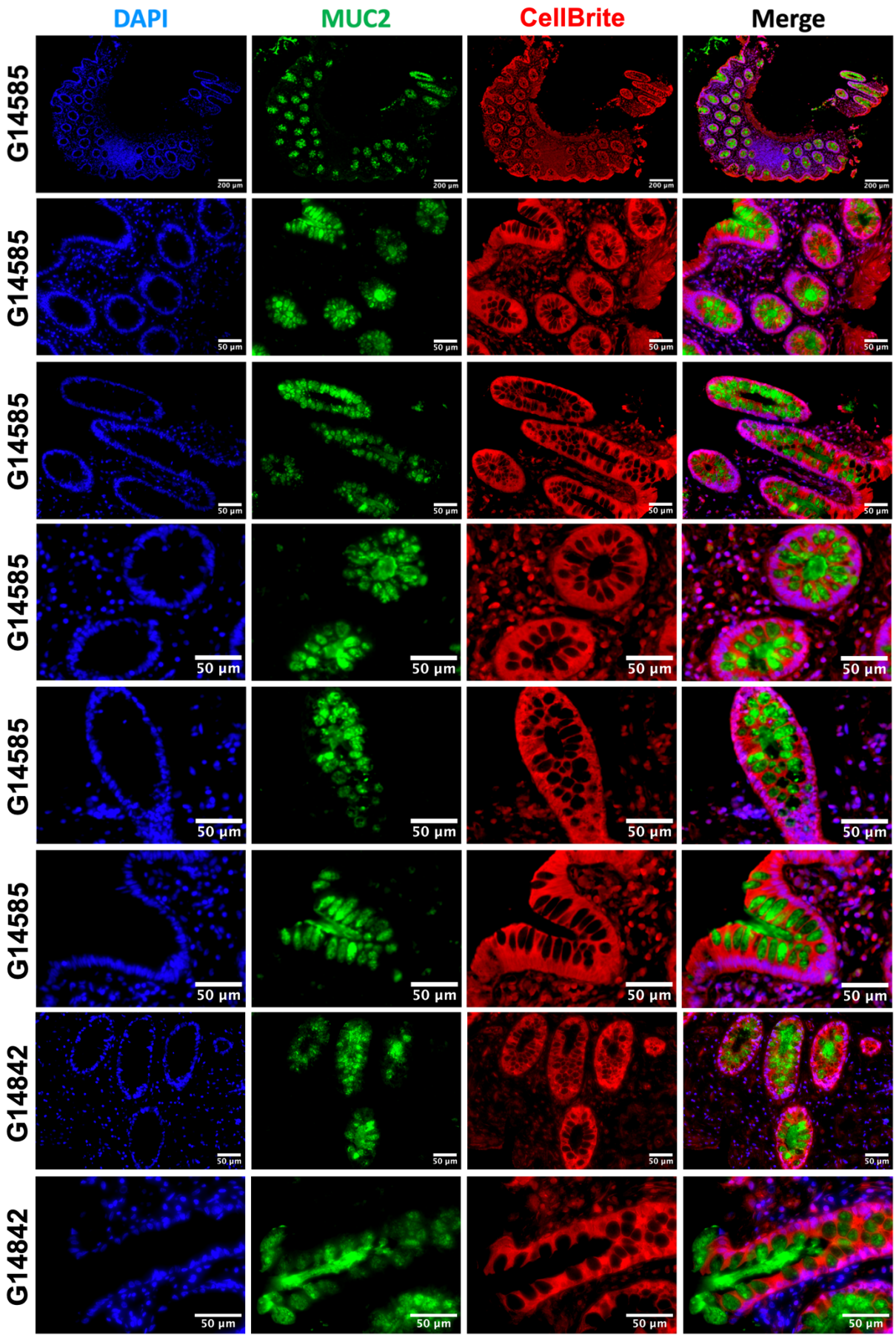

DAPI
MUC2
CellBrite
Merge
G14585
G14585
G14585
G14585
G14585
G14585
G14842
G14842
200 µm
50 µm

Figure 4.2 Staining of MUC2 and CellBrite in two normal colon tissue sections, namely G14585 and G14842. Immunofluorescence analysis was conducted using MUC2 (PR5D5, 1:250) and CellBrite (1:2000) in G14585 and G14842 normal colon FFPE tissue sections. Scale bar: 50 µm.

A combination staining of MUC2, TFF3, and CellBrite was also tested on the IBD2566 tissue section kindly provided by Dr Shazia Irshad and the results are shown in Figure 4.3. More representative images are provided in Appendix Figure 4. Next, we confirmed that MUC2 and TFF3 overlap in normal colon tissue (Figure 4.4).

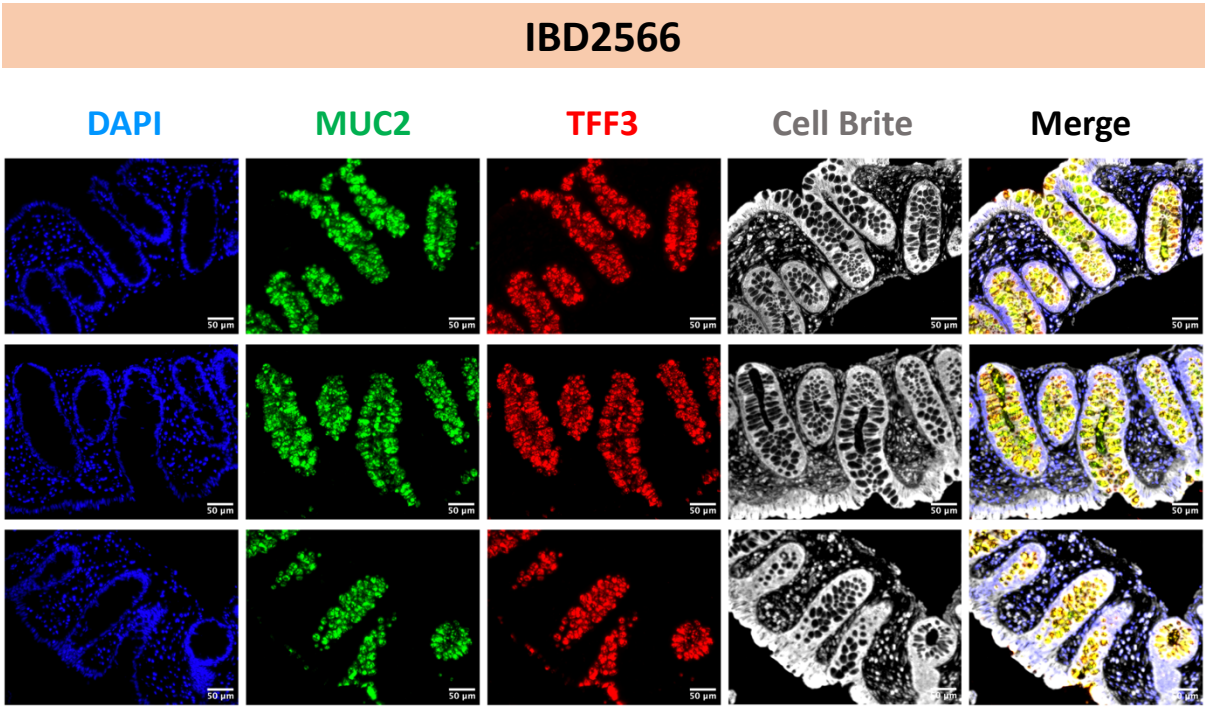


Figure 4.3 Staining of MUC2, TFF3 and CellBrite in the IBD2566 tissue section. Immunofluorescence analysis was performed using MUC2 (PR5D5, 1:250), TFF3 (1:1000), and CellBrite (1:2000) in IBD2566 FFPE tissue section. Scale bar: 50 µm.

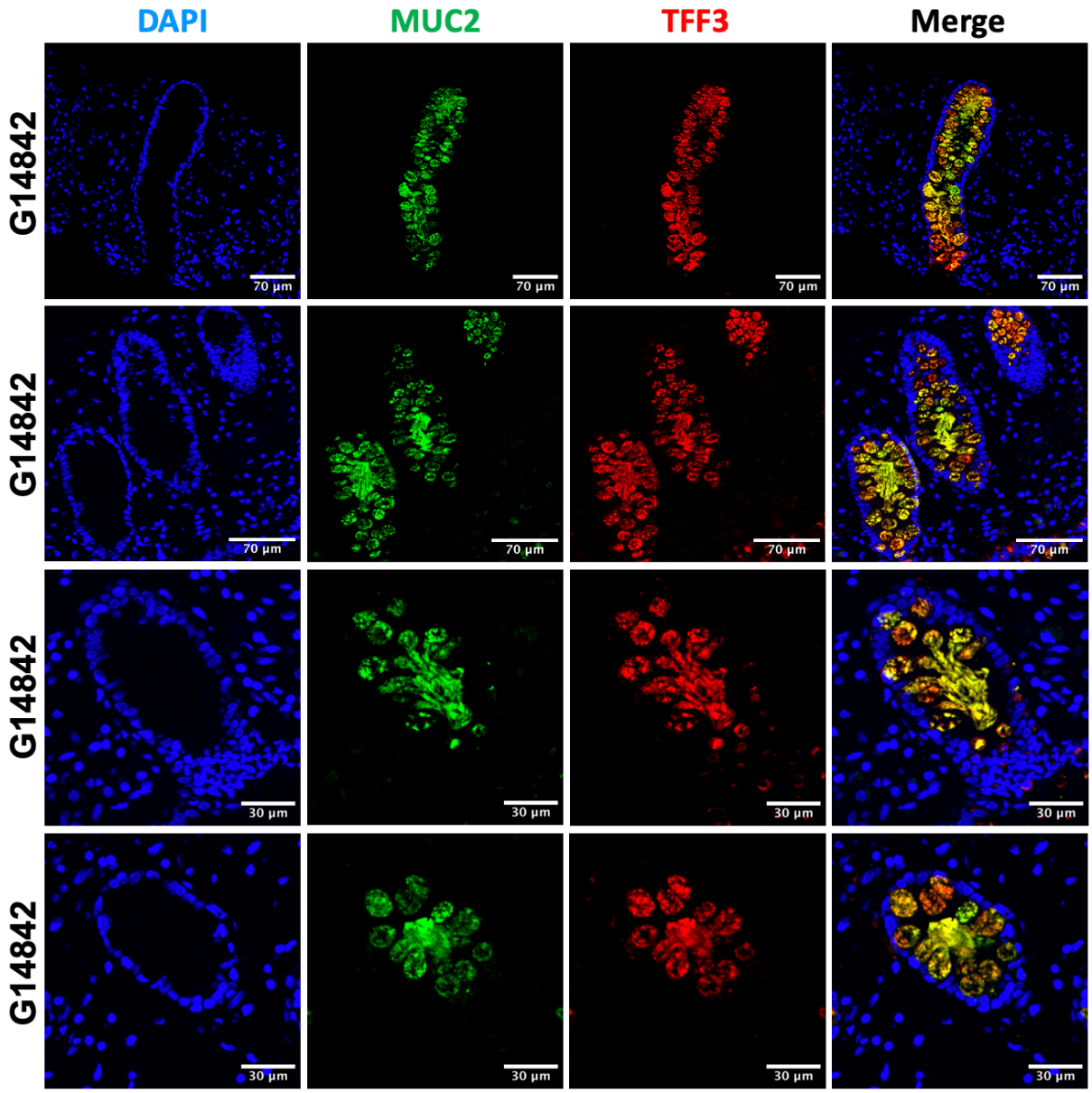


Figure 4.4 Staining of MUC2 and TFF3 in the normal colon tissue section, G14842. Immunofluorescence analysis was conducted using MUC2 (PR5D5, 1:250) and TFF3 (1:1000) in G14842 normal colon FFPE tissue section. Scale bars: 30 µm and 70 µm.

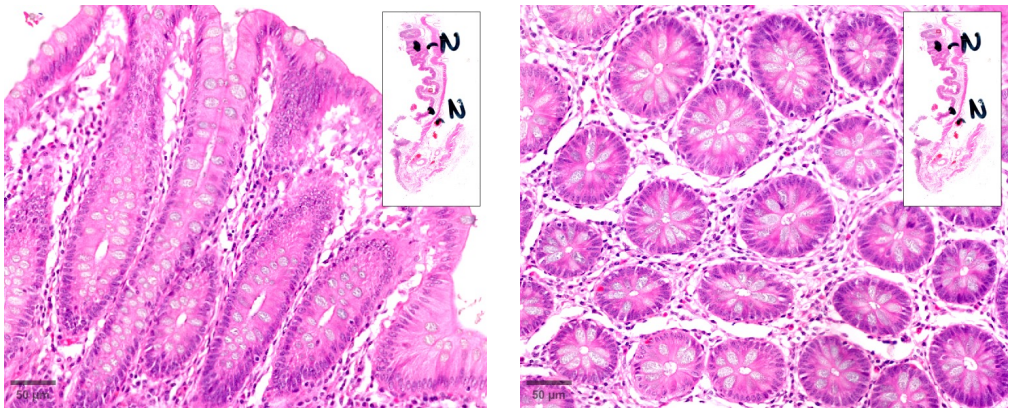
The main objective of using human CRC tissue sections was to check whether our cell line classification (based on MUC2 and TFF3 protein expression levels) matches the human tissue sections. For our analysis, we used a total of 21 human CRC FFPE tissue slides from the VICTOR study and matched normal colon tissue sections as controls. Besides immunofluorescence, we conducted H&E staining, where

haematoxylin was used to colour cell nuclei in a purplish-blue, while eosin stained the cytoplasm, extracellular matrix, and other structures in five different shades of pink. I am very grateful Leticia Campo from Translational histopathology laboratory for processing the slides and imaging them using Phenolmager from AKOYA Biosciences. A detailed description of the protocol can be found in Chapter 2.9.5.

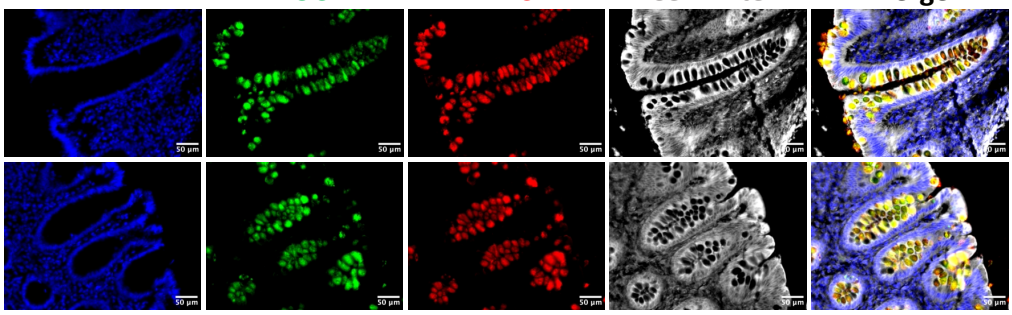
A total of 21 human colon cancer tissue sections from the VICTOR trial were analysed in our study. Immunofluorescence and H&E staining of matched healthy tissues were illustrated in Figure 4.5A/B. A total of four matched colon tissue slides, including VT11127, VT11140, VT11146, and VT11184 were stained with H&E (Figure 4.5A). Immunofluorescence analysis was conducted on two healthy colon FFPE tissue sections, namely VT11146 and VT11184 (Figure 4.5B). In each case, the H&E section is shown as (A) and the immunofluorescence staining in (B) showing good results for the combination of MUC2, TFF3, and CellBrite staining in normal colon tissue (Figure 4.5). Additional H&E and immunofluorescence images of normal colon are shown in Appendix Figure 5.

VT11146 (normal colon)

A.

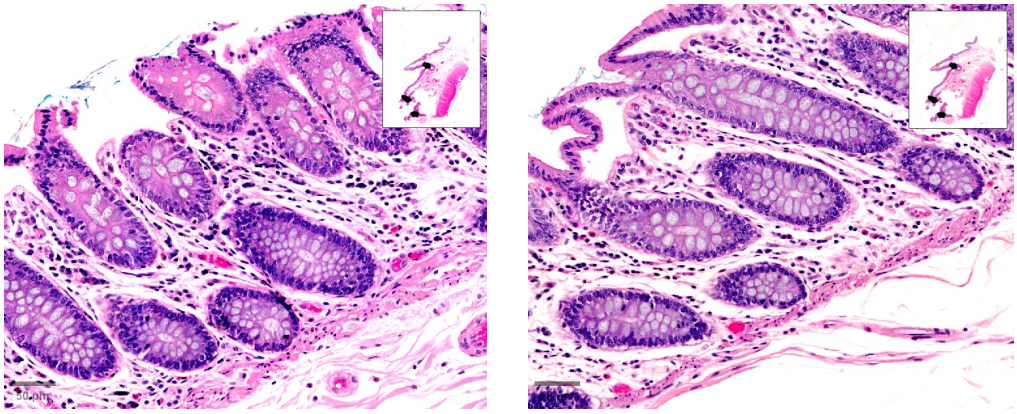


B.



VT11184 (normal colon)

A.



B.

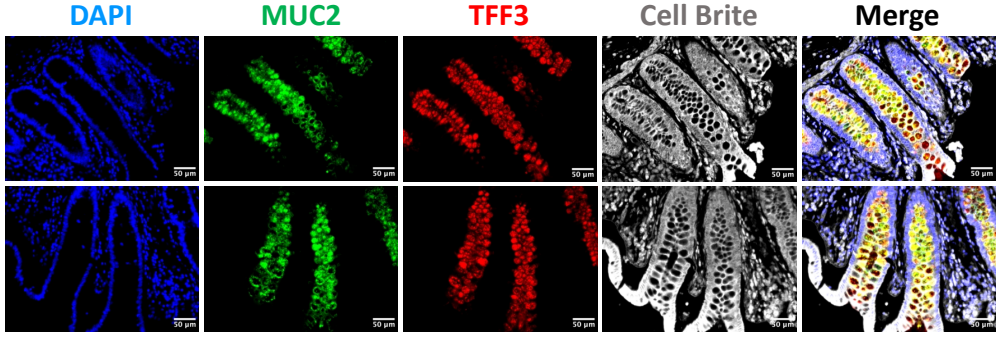


Figure 4.5 H&E and immunofluorescence staining of two matched healthy colon tissue sections, VT11146 and VT11184. (A) describes H&E staining, (B) illustrates immunofluorescence staining of MUC2 (PR5D5, 1:250), TFF3 (1:1000), DAPI (1:10000), and CellBrite (1:2000). Scale bar: 50 µm.

We first determined whether the analysed 21 CRC tissue sections included some normal colonic tissue. If there was any piece of normal tissue, the proportion of normal colon tissue within a given section based on the images taken was estimated. In addition, histological assessment in terms of cancer penetration was done by quantifying the surface area of normal colon and cancerous tissues relative to total tissue through QuPath software (Table 4.1A and Appendix Figure 6). Next, we indicated for each of the CRC tissue slides what we considered was their constituent classification with respect to the grouping of our cell line panel into 5 categories described in chapter 3 (Table 4.1B).

A.

Cancer penetration

ID number of FFPE slides	Presence of any normal colon tissue	N° of IF images with normal colon out of total	N° of IF images with colorectal carcinoma	N° of IF images in total per sample	% of IF images with normal colon	% of IF images with colorectal carcinoma	Histological assessment				
							Healthy colon tissue area (µm^2)	Cancerous colon tissue area (µm^2)	Total tissue area	Ratio of healthy versus cancerous colon tissue	% of colon tissue within total tissue
VT11179	✓	9	17	26	34.6	65.4	14459561	70950077	218845653	0.20	39.03
VT11181	X	0	44	44	0	100	0	66032963	204153129	0.00	32.34
VT11384	✓	5	16	21	23.8	76.2	33080839	72372418	378749211	0.46	27.84
VT11386	✓	3	17	20	15	85	6411765	88878156	351049346	0.07	27.14
VT11429	✓	8	15	23	34.8	65.2	72363181	50086002	291362976	1.44	42.03
VT11147*	X	0	20	20	0	100	0	180501041	261884130	0.00	68.92
VT11148*	X	0	22	22	0	100	0	123602597	250107631	0.00	49.42
VT11189	X	0	22	22	0	100	0	113936036	255655183	0.00	44.57
VT11197	X	0	30	30	0	100	0	116314355	264850538	0.00	43.92
VT11294	X	0	34	34	0	100	0	146680409	205078686	0.00	71.52
VT11385	✓	1	19	20	5	95	21981474	146476957	327225091	0.15	51.48
VT11387	X	0	20	20	0	100	0	29317451	297806318	0.00	9.84
VT11401	X	0	20	20	0	100	0	50261955	107034595	0.00	46.96
VT11150	✓	18	2	20	90	10	51748596	645303	211734526	80.19	24.75
VT11195	X	0	20	20	0	100	0	149099140	303330950	0.00	49.15
VT11276	✓	20	10	30	66.7	33.3	28072057	32604574	137133525	0.86	44.25
VT11280	X	0	25	25	0	100	0	166925892	322774404	0.00	51.72
VT11282	X	0	26	26	0	100	0	83579118	189747183	0.00	44.05
VT11290	✓	7	13	20	35	65	14775130	102642797	175932090	0.14	66.74
VT11389	X	0	23	23	0	100	0	109632195	386109645	0.00	28.39
VT11393	✓	5	15	20	25	75	11561017	38482603	194705108	0.30	25.70

B.

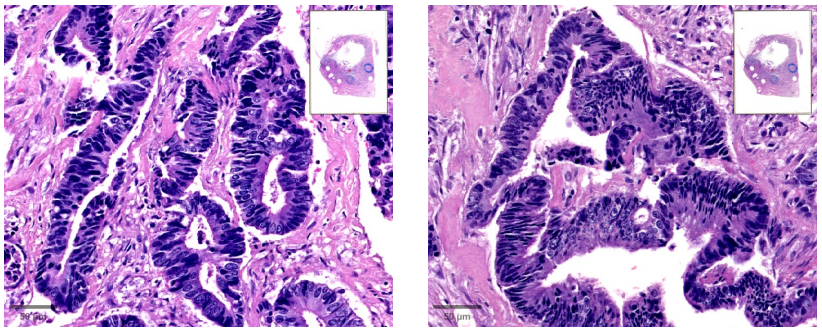
ID number of FFPE slides	MUC2 and TFF3 proteins expression profiles												Main type
	№ of images representing each category						% of each group within colorectal cancer						
	Group ID	1	2	3	4	5	Group ID	1	2	3	4	5	
	MUC2	+++	++	+	-	-	MUC2	+++	++	+	-	-	
	TFF3	+++	++	+	+++	-	TFF3	+++	++	+	+++	-	
VT11179		7	0	0	5	5		41.2	0	0	29.4	29.4	1
VT11181		43	1	0	0	0		97.7	2.3	0	0	0	1
VT11384		10	6	0	0	0		62.5	37.5	0	0	0	1
VT11386		13	0	0	0	4		76.5	0	0	0	23.5	1
VT11429		15	0	0	0	0		100	0	0	0	0	1
VT11147*		2	0	0	18	0		10	0	0	90	0	4
VT11148*		3	0	0	19	0		13.6	0	0	86.4	0	4
VT11189		0	0	0	22	0		0	0	0	100	0	4
VT11197		4	0	0	26	0		13.3	0	0	86.7	0	4
VT11294		5	2	0	27	0		14.7	5.9	0	79.4	0	4
VT11385		0	0	0	14	5		0	0	0	73.7	26.3	4
VT11387		0	0	0	16	4		0	0	0	80	20	4
VT11401		6	0	0	14	0		30	0	0	70	0	4
VT11150		0	0	1	0	1		0	0	50	0	50	5
VT11195		0	0	0	0	20		0	0	0	0	100	5
VT11276		0	0	0	0	10		0	0	0	0	100	5
VT11280		0	0	0	11	14		0	0	0	44	56	5
VT11282		0	0	9	7	10		0	0	34.6	26.9	38.5	5
VT11290		0	0	0	0	13		0	0	0	0	100	5
VT11389		0	5	6	0	12		0	21.7	26.1	0	52.2	5
VT11393		0	0	0	0	15		0	0	0	0	100	5

Table 4.1 Examination of CRC tissue sections relative to normal tissue and protein expression profiles of MUC2 and TFF3. (A) describes the proportion of cancer infiltration within a given tissue section. The first section of the table shows the presence of normal tissue defined through immunofluorescence analysis. Then, the surface area (measured in μm^2) of healthy, colon cancerous and total tissues was quantified using QuPath software. (B) illustrates the classification of FFPE tissue sections according to the protein expression patterns of MUC2 and TFF3. The last column of the table displays the predominant category based on the highest observed proportion. The slides were arranged according to their primary category. VT11147 and VT11148 highlighted in "*" are duplicate tumours.

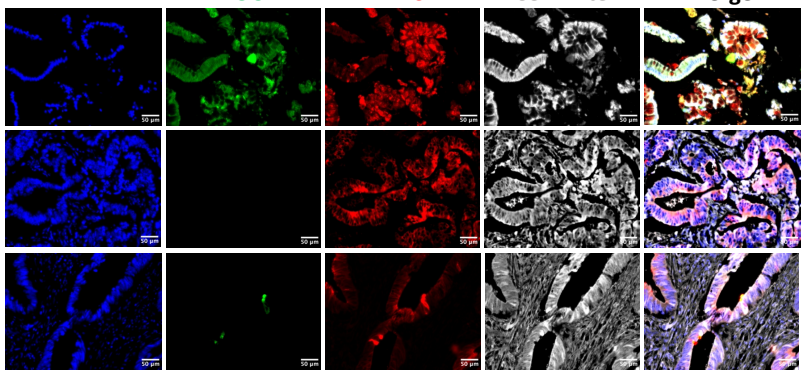
It should be noted that the slides stained with H&E and MUC2/TFF3 antibodies were originated from different sections of the same tumour. Representative images for each group within a tissue section are visualised in Figure 4.6A/B. Additional images are shown in Appendix Figure 7. According to our findings, out of 21 CRC tissue sections, 5 (23.8%) fell into group 1, 8 (38.1%) into group 4 and an additional 8 (38.1%) for group 5.

VT11147 (colorectal cancer)

A.



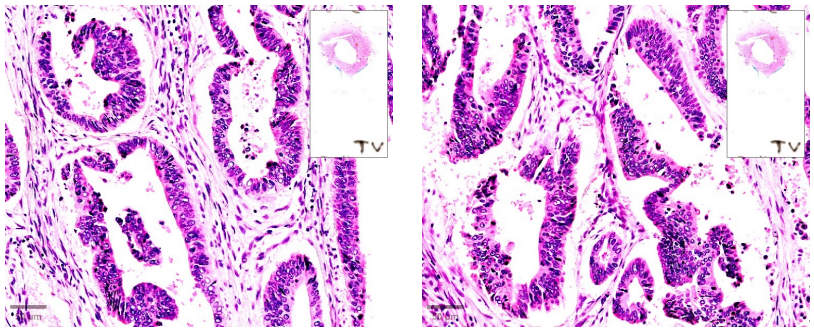
B.



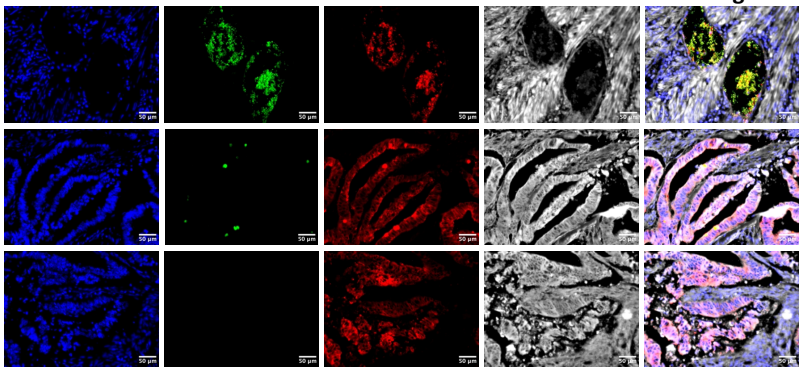
Group: 1 and 4

VT11148 (colorectal cancer)

A.



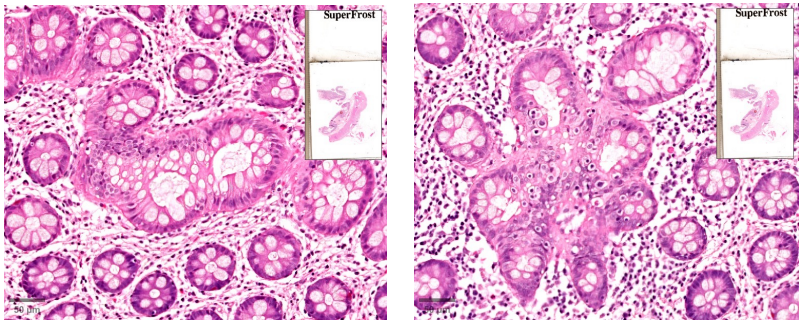
B.



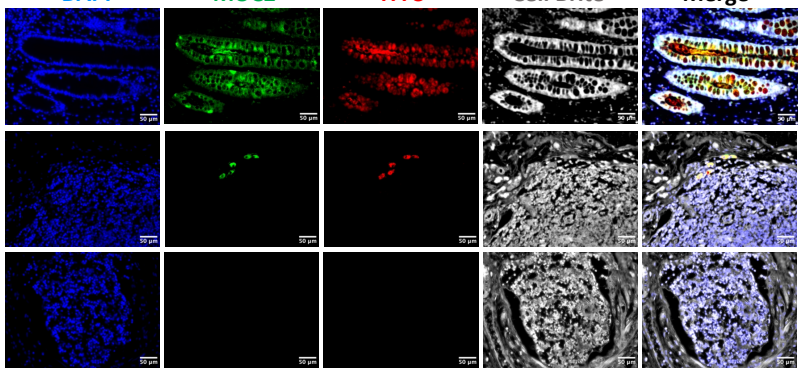
Group: 1 and 4

VT11150 (colorectal cancer)

A.



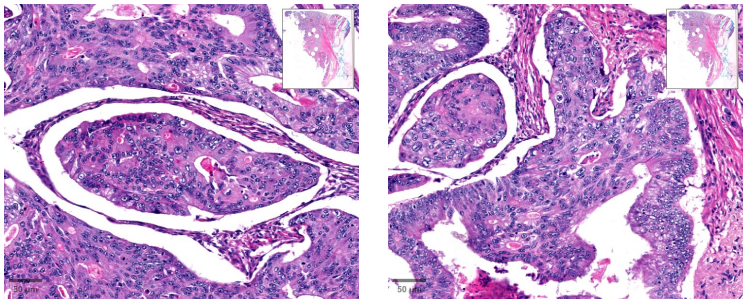
B.



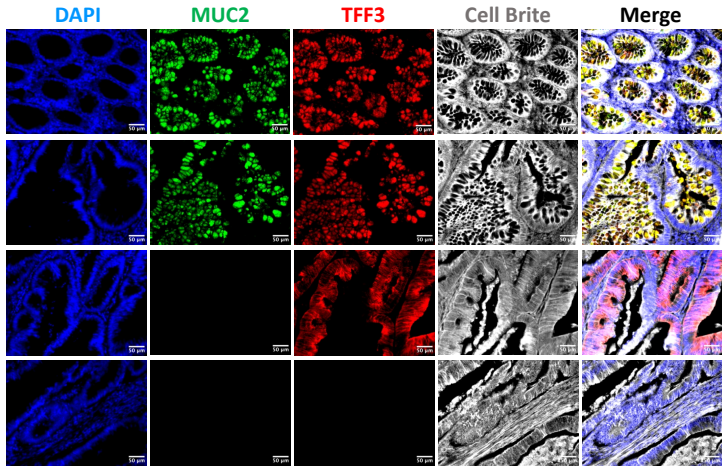
Group: 3 and 5

VT11179 (colorectal cancer)

A.



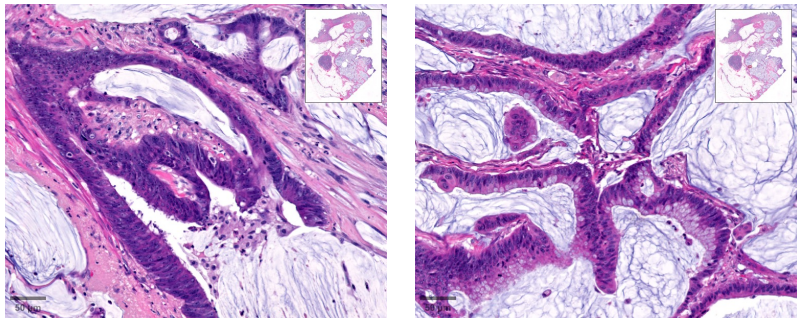
B.



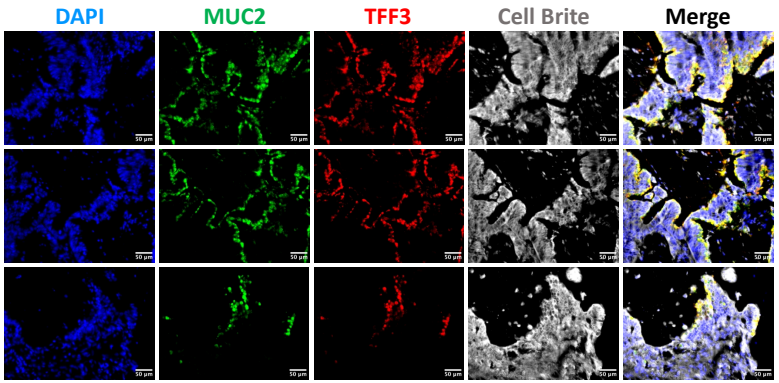
Group: 1, 4, and 5

VT11181 (colorectal cancer)

A.



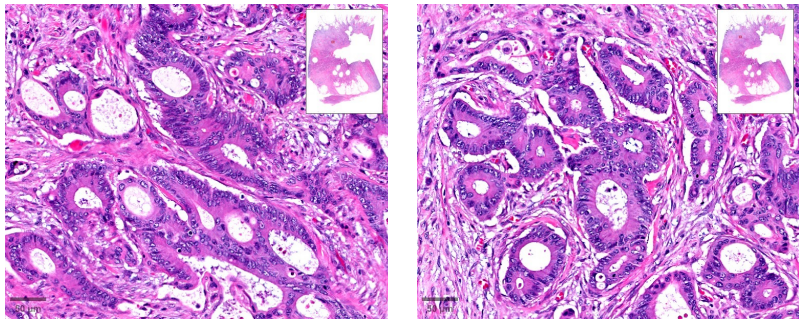
B.



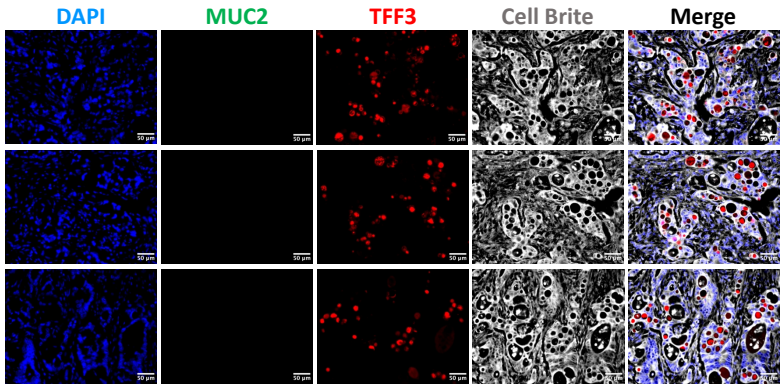
Group: 1 and 2

VT11189 (colorectal cancer)

A.



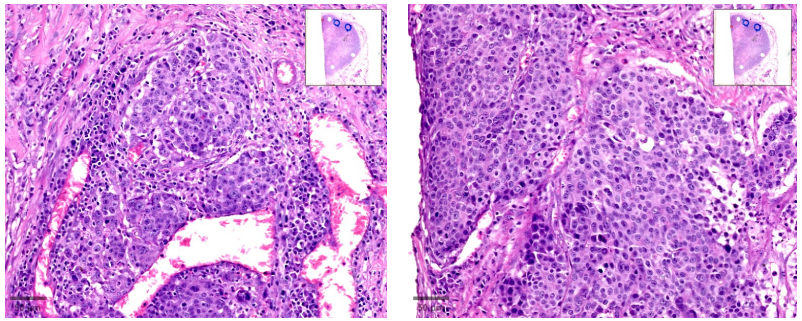
B.



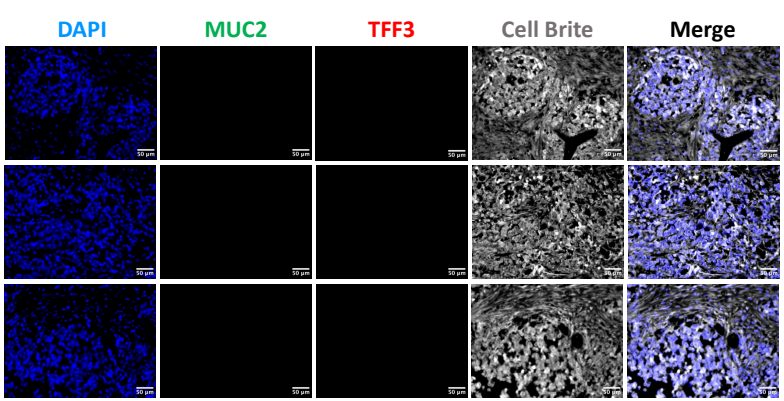
Group: 4

VT11195 (colorectal cancer)

A.



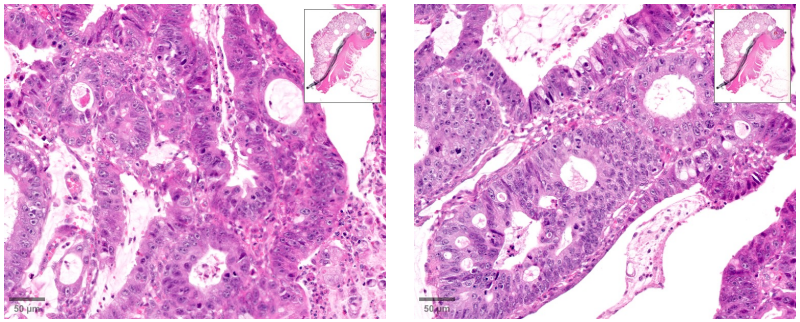
B.



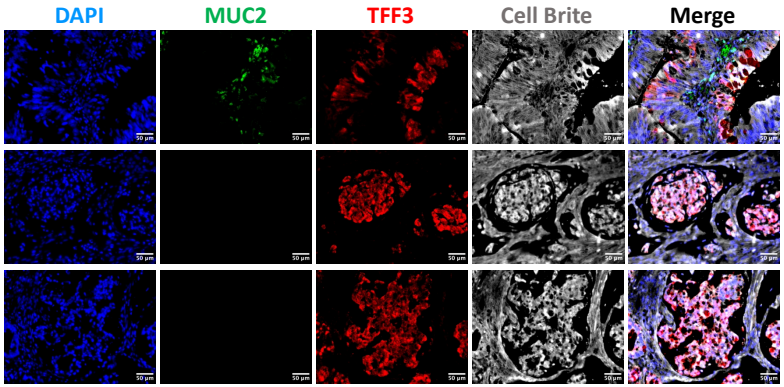
Group: 5

VT11197 (colorectal cancer)

A.



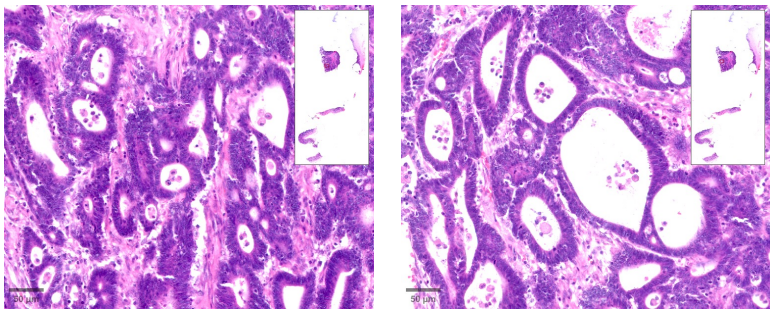
B.



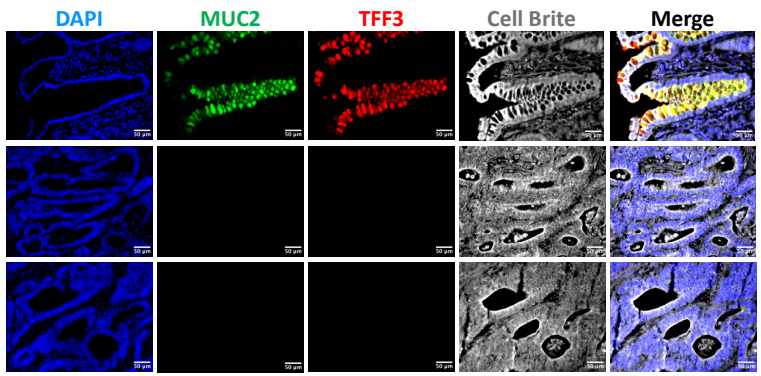
Group: 1 and 4

VT11276 (colorectal cancer)

A.



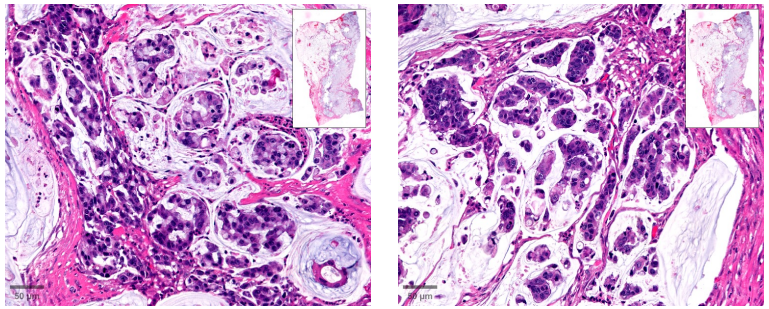
B.



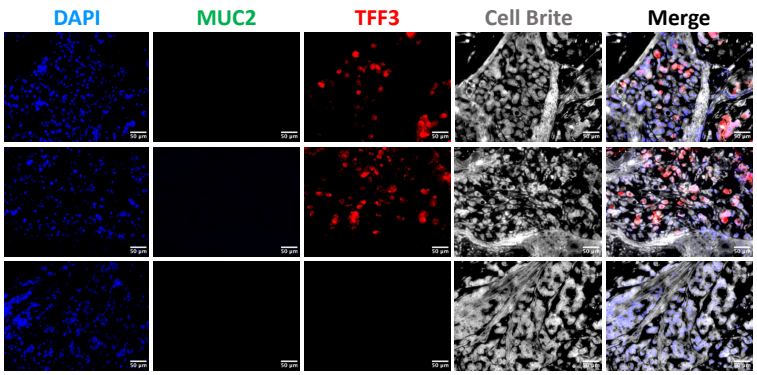
Group: 5

VT11280 (colorectal cancer)

A.



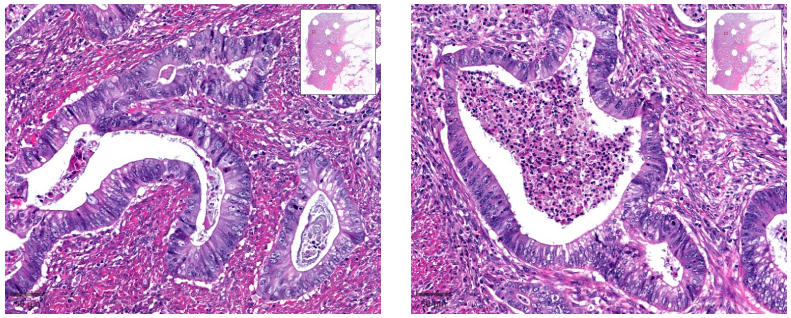
B.



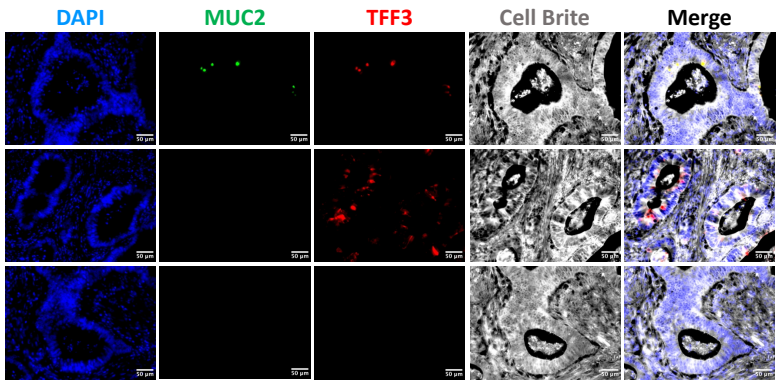
Group: 4 and 5

VT11282 (colorectal cancer)

A.



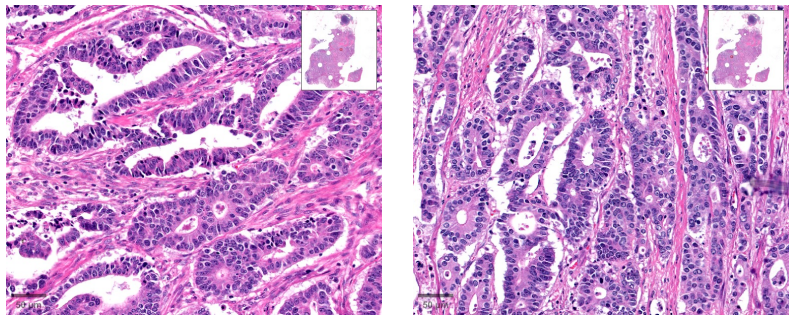
B.



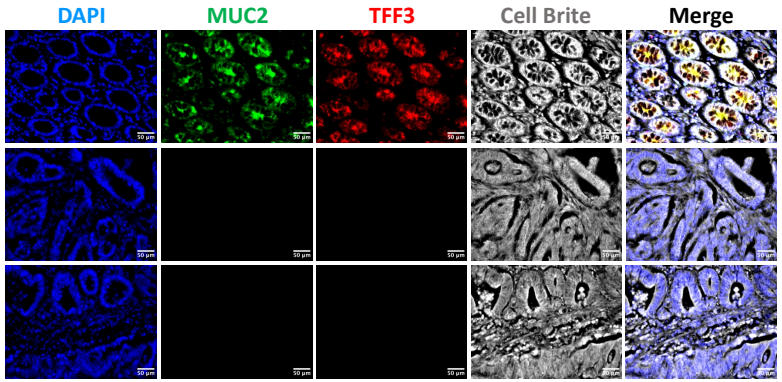
Group: 3, 4 and 5

VT11290 (colorectal cancer)

A.



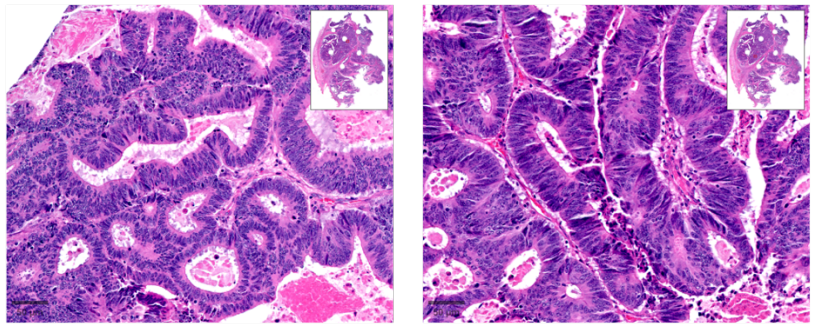
B.



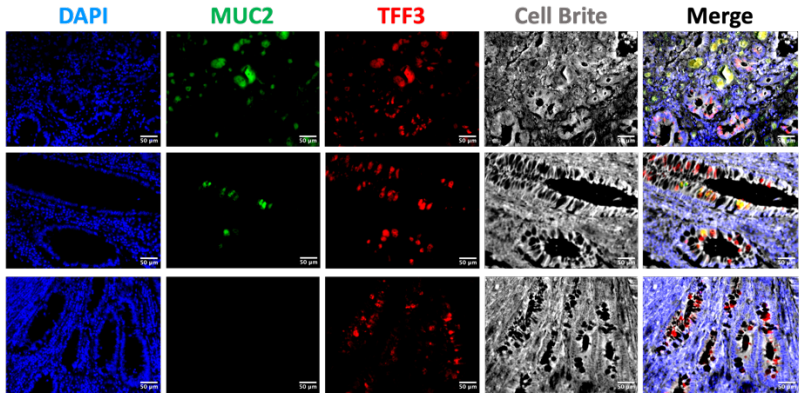
Group: 5

VT11294 (colorectal cancer)

A.



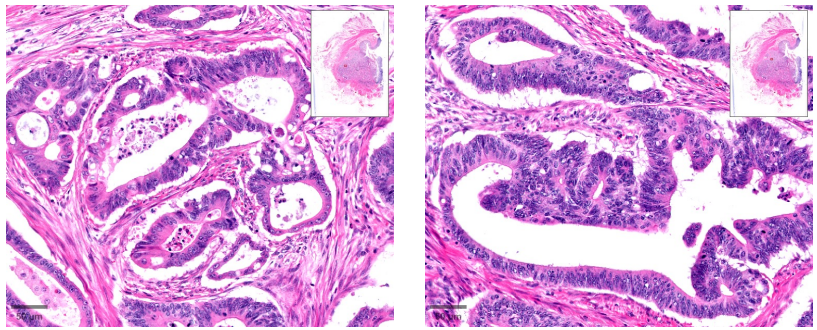
B.



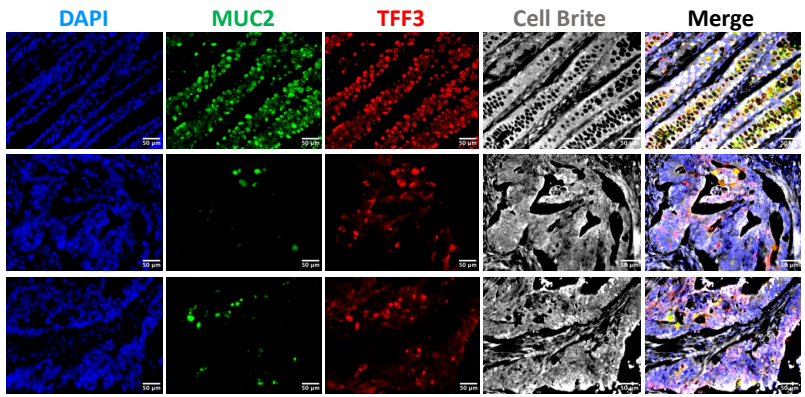
Group: 1, 2 and 4

VT11384 (colorectal cancer)

A.



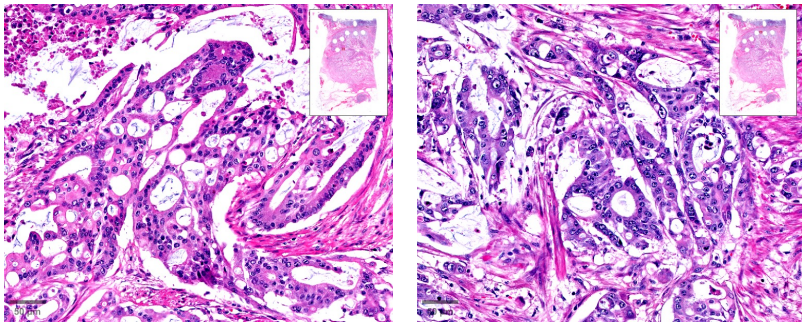
B.



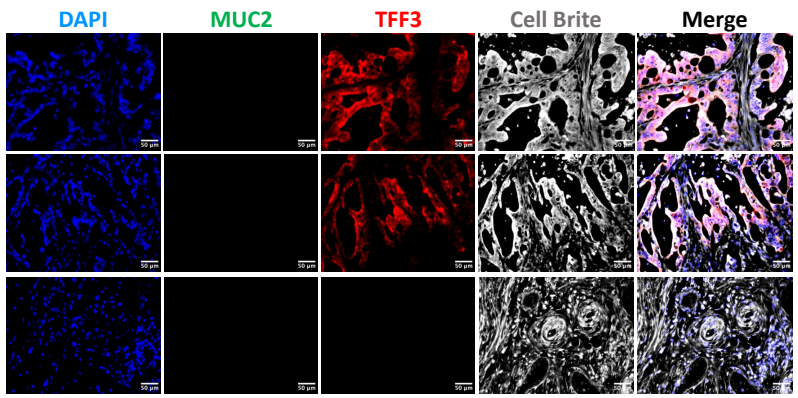
Group: 1 and 2

VT11385 (colorectal cancer)

A.



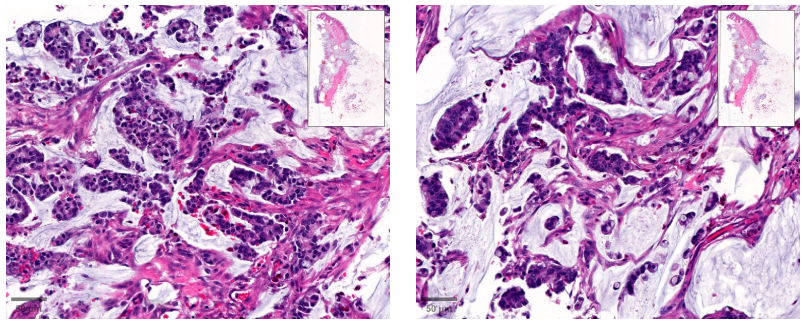
B.



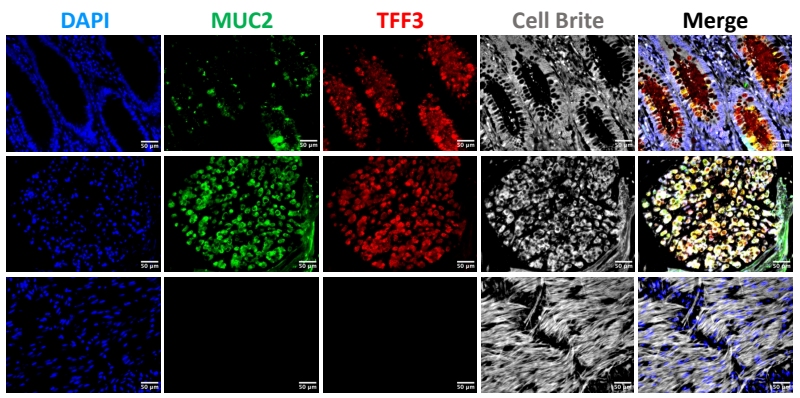
Group: 4 and 5

VT11386 (colorectal cancer)

A.



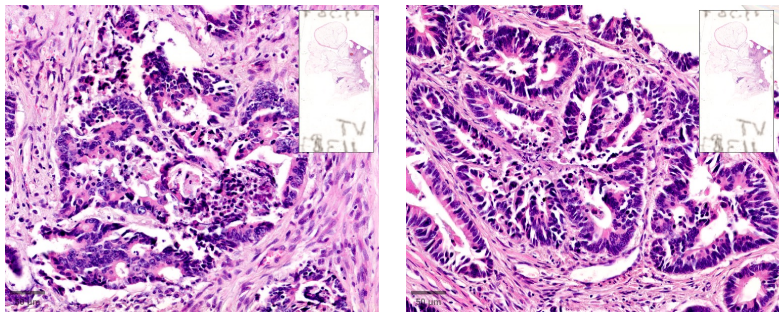
B.



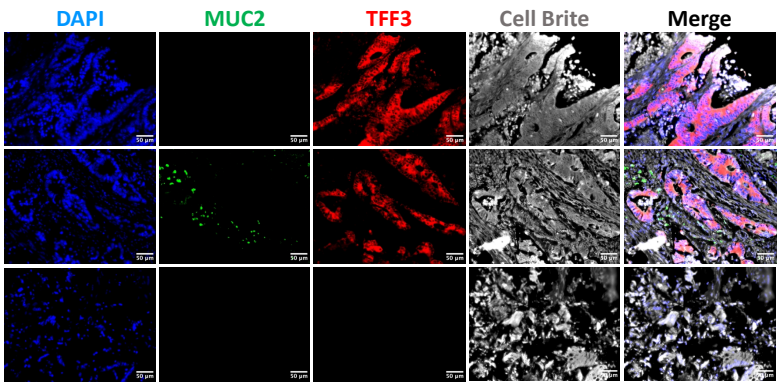
Group: 1 and 5

VT11387 (colorectal cancer)

A.



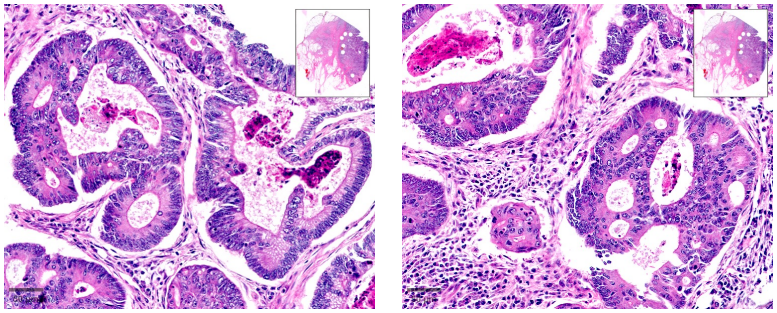
B.



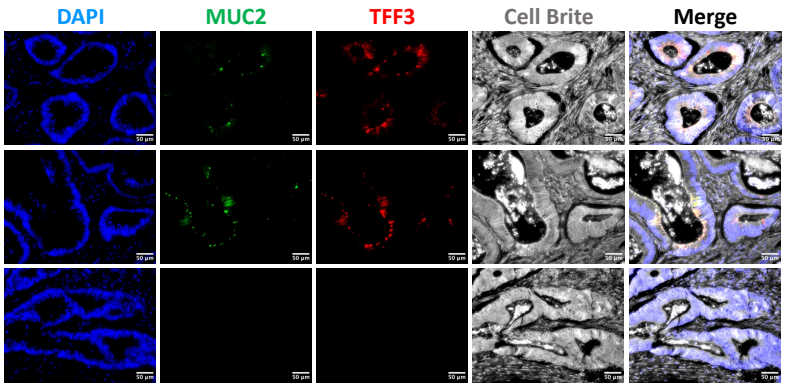
Group: 4 and 5

VT11389 (colorectal cancer)

A.



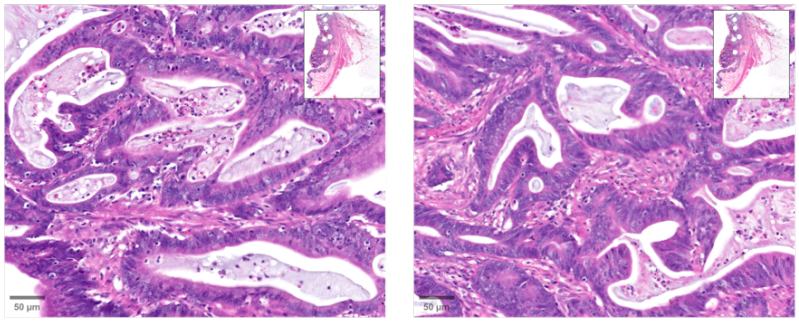
B.



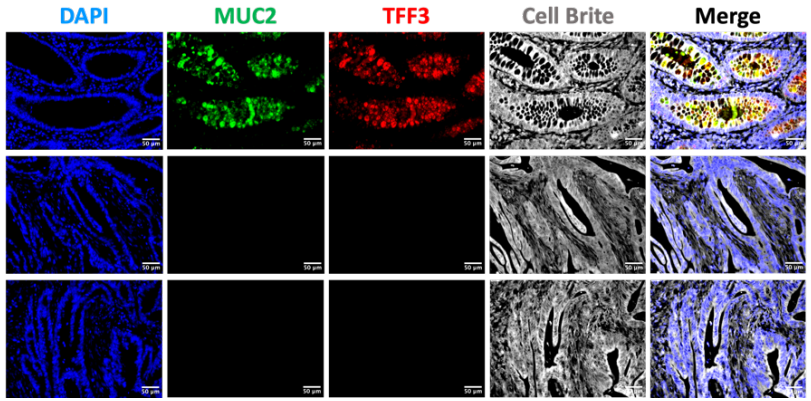
Group: 2, 3, and 5

VT11393 (colorectal cancer)

A.



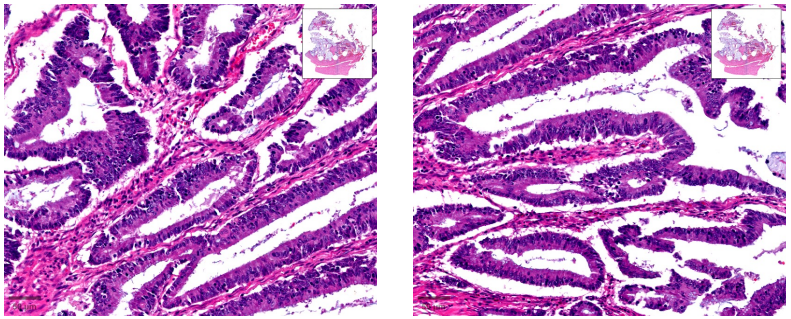
B.



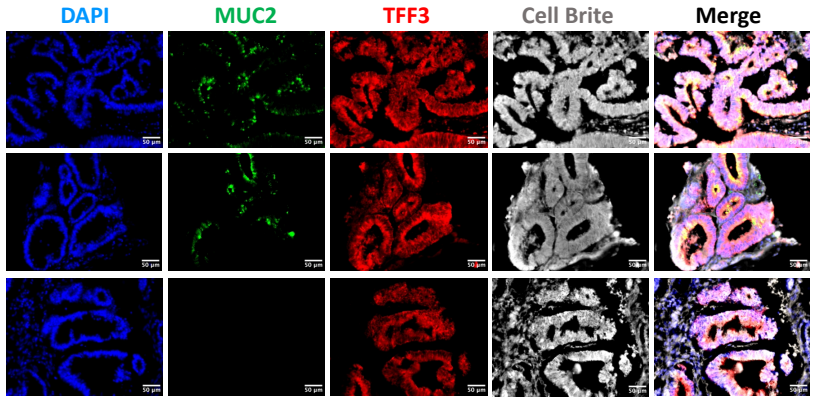
Group: 5

VT11401 (colorectal cancer)

A.



B.



Group: 1 and 4

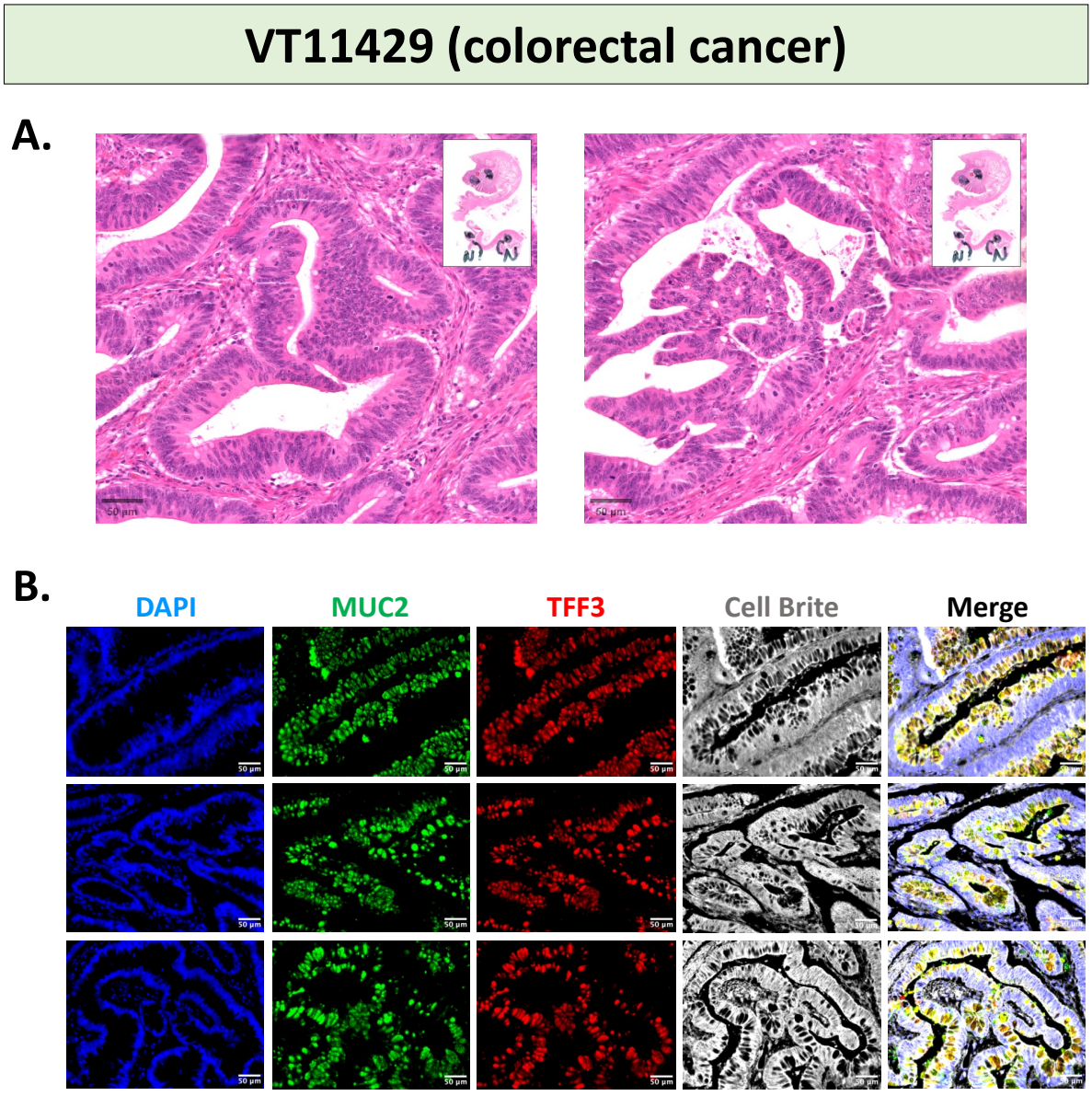


Figure 4.6 Representative H&E and immunofluorescence images of 21 colon cancer tissue sections derived from the VICTOR trial. The FFPE tissue slides were stained with H&E staining kit (Abcam) for histology assessment (A), as well as MUC2 (PR5D5, 1:250) and commercial polyclonal TFF3 (1:1000), DAPI (1:10000), and CellBrite (1:2000) for group classification (B). Figures contain examples of both normal colonic mucosa and colorectal carcinoma. Scale bar: 50 µm.

4.2.2 Analysis of the relationship of the FFPE characterisation to the pathology description

Histopathology and molecular data from the 21 CRC tissue sections were also analysed. Histopathology reports were provided by Professor Marco Novelli (University College London Hospitals) and Dr David Church (Henry Wellcome Building of Genomic Medicine, University of Oxford) and mutation screening was performed by Dr Enric Domingo. Table 4.2 summarises the available data for the analysed tissue sections. Comprehensive details regarding to the molecular profiling and the association of molecular/clinical variables of VICTOR trial samples are in the papers by Domingo et al. (2013) and Mouradov et al. (2013).

It is worth mentioning that even though VT11150 is described as normal based on H&E screening (Professor Novelli's histopathology report), in the immunofluorescence assay we can see a low proportion of tumour presence. This suggests that some tumours may be hidden deeper in the blocks, as the slides stained with H&E and antibodies were derived from distinct sections of the same tumour block. Additionally, despite VT11429 being classified as normal according to Dr Church's classification, we defined a small proportion of tumour belong to group 1 category. The molecular data for this block in Table 4.2 shows molecular alterations found in a tumour block, not the normal block. The blocks have been used for several research projects and sectioned quite a lot in some cases and less in others. It is perfectly possible that what was first thought to be a normal block now could show some bits of tumour. Or the other way round, some tumour blocks may only show now normal tissue. Overall, this emphasises the importance of using specific biomarkers during screening.

ID number of FFPE slides	Group ID	Normal tissue presence	Histopathology report_1 (Prof Novelli's classification)	Histopathology report_2 (Dr Church's classification)	KRAS	BRAF600	NRAS	PIK3CA	TP53	FBXW7	CIN	MSI
VT11179	1	✓	Moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	mut	wt	wt	wt	mut	wt	-	MSS
VT11181	1	X	Mucinous well-differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	mut	wt	wt	mut	-	-	CIN	MSS
VT11384	1	✓	Moderately differentiated adenocarcinoma (scan2)	Moderate (Grade 1)	mut	wt	wt	wt	mut	wt	CIN	MSS
VT11386	1	✓	Mucinous moderately to poorly differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	wt	wt	wt	wt	wt	wt	CIN	MSS
VT11429	1	✓	Moderately differentiated adenocarcinoma (scan2)	Normal	mut	wt	wt	wt	mut	wt	CIN	MSS
VT11147*	4	X	Moderately differentiated adenocarcinoma (scan4)	Moderate (Grade 1)	wt	wt	wt	wt	mut	wt	CIN	MSS
VT11148*	4	X	Moderately differentiated adenocarcinoma (scan2)	Moderate (Grade 1)	wt	wt	wt	wt	mut	wt	CIN	MSS
VT11189	4	X	Moderately differentiated adenocarcinoma (scan1)	Poor (Grade 2)	mut	wt	wt	wt	wt	wt	CIN	MSS

ID number of FFPE slides	Group ID	Normal tissue presence	Histopathology report_1 (Prof Novelli's classification)	Histopathology report_2 (Dr Church's classification)	KRAS	BRAF600	NRAS	PIK3CA	TP53	FBXW7	CIN	MSI
VT11197	4	X	Mucinous moderately differentiated adenocarcinoma (scan4)	Moderate (Grade 1)	wt	wt	wt	wt	mut	wt	CIN	MSS
VT11294	4	X	Moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	wt	wt	wt	wt	mut	wt	negative	MSS
VT11385	4	✓	Mucinous moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	mut	wt	wt	wt	wt	wt	negative	MSS
VT11387	4	X	Moderately differentiated adenocarcinoma (scan1/5)	Moderate (Grade 1)	wt	wt	wt	wt	mut	wt	CIN	MSS
VT11401	4	X	Mucinous moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	mut	wt	wt	mut	wt	wt	negative	MSS
VT11150	5	✓	Normal colon (scan5)	Poor (Grade 2)	wt	wt	wt	wt	-	-	CIN	-
VT11195	5	X	Lymph node with poorly differentiated adenocarcinoma (probably MSH) (scan1)	Poor (Grade 2)	wt	wt	wt	wt	mut	wt	CIN	MSS

ID number of FFPE slides	Group ID	Normal tissue presence	Histopathology report_1 (Prof Novelli's classification)	Histopathology report_2 (Dr Church's classification)	KRAS	BRAF600	NRAS	PIK3CA	TP53	FBXW7	CIN	MSI
VT11276	5	✓	Moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	wt	wt	wt	mut	mut	wt	negative	MSI
VT11280	5	X	Mucinous moderately to poorly differentiated adenocarcinoma (mostly necrotic) (scan1)	Moderate (Grade 1)	wt	mut	wt	wt	wt	wt	CIN	MSS
VT11282	5	X	Moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	mut	wt	wt	mut	wt	wt	CIN	MSS
VT11290	5	✓	Moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	wt	wt	wt	wt	wt	wt	CIN	MSS
VT11389	5	X	Moderately differentiated adenocarcinoma (scan1)	Moderate (Grade 1)	wt	wt	wt	mut	wt	wt	CIN	MSS
VT11393	5	✓	Moderately differentiated adenocarcinoma (mucinous diff) (scan3)	Moderate (Grade 1)	wt	mut	wt	wt	wt	mut	negative	MSI

Table 4.2 Histopathology reports and molecular profiling of analysed CRC tissue sections. Two histopathology reports were made in addition to the information about group classification and the presence of normal tissue. The mutation data cover mutations in CRC driver genes, including *KRAS*, *BRAF600*, *NRAS*, *PIK3CA*, *TP53*, and *FBXW7*. In the table, "wt" denotes wild type, whereas "mut" indicates the presence of a mutant form. VT11147 and VT11148 tumour blocks (highlighted in "*") belong to the same patient/tumour, hence all their associated data are duplicated. The chromosomal instability (CIN) and microsatellite instability (MSI) features were also identified for each tumour. VT11147 and VT11148 tumour blocks (highlighted in "*") belong to the same patient/tumour, hence all their associated data is duplicated.

4.3 DISCUSSION

It is notable that many of the CRCs seem to contain segments showing the presence of more than just one category, as might be expected from somatic evolutionary progression. Category 4 (high TFF3, no/low MUC2) stands out very clearly as we predicted quite early in these studies. Categories 2 and 3 are sometimes hard to distinguish and are associated very clearly with single cells that may express only MUC2 or only TFF3, as we have seen in the cell lines. We also see secreted MUC2 outside cells in a way we do not see in the cell lines because it gets washed away as soon as we do the staining. This suggests that category 4 may occasionally secrete low levels of MUC2.

We suggest that category 1 is what is called mostly mucinous while category 4 has not been previously identified. Perhaps artificial intelligence approaches could identify Category 4 just from H&E when trained on our IF-based categorisation.

Despite the relatively small number of human CRC tissue samples, we were able to validate our proposed classification based on protein expression profiles of MUC2 and TFF3 in CRC cell lines. Validation of our classification in human tissue sections proves that cancer cell lines are representative of real tissues and can be used to study biologic processes in primary tumours which is consistent with other reports (Goodspeed et al., 2016; Mirabelli et al., 2019; Wilding & Bodmer, 2014). Our results further support the suggestion that the integration of machine learning and H&E-stained tumour sections may provide a novel approach to categorising tumours when H&E sections are trained on their cellular content as obtained, for example, by immunofluorescence. This possibility has been an area of recent interest (Kleppe et al., 2022; Kleppe et al., 2021).

CHAPTER 5

SINGLE-CELL SEQUENCING ANALYSIS

CHAPTER 5: SINGLE-CELL SEQUENCING ANALYSIS

5.1 INTRODUCTION AND AIMS

Single-cell RNA sequencing (scRNA-seq) is an advanced technology that involves sequencing individual cells rather than bulk samples. It is widely applied in cancer studies to define tumour heterogeneity, rare subpopulations, and specific markers for the development of new diagnostic and therapeutic approaches (Tang et al., 2019). The cell clusters produced by single-cell sequencing allow scientists to define subset cell groups within selected cell lines and interpret patterns of gene expression that might explain the intra-cell line variations.

Cancer cells accumulate genetic and epigenetic changes from time to time, as part of their somatic evolutionary progress resulting in genetic heterogeneity. Conventional mRNA sequencing methods have some limitations to detect individual cell populations. Therefore, mRNA heterogeneity present in cancerous cells is most likely be masked or overlooked using bulk sequencing. By contrast, a time-dependent analysis of scRNA-seq data can reveal genetic and epigenetic alterations that occur in cancer over time. Becket et al. examined single-cell profiles of 27 normal tissues, 48 polyps, and 6 CRCs, and revealed that a significant portion of polyps and CRC cells display features similar to stem cell phenotype (Becker et al., 2022). Another study described biclonal origin of CRC cells detected through scRNA-seq analysis (Yu et al., 2014). The predominant CRC subtype had *APC* and *TP53* mutations, while another subtype lacked these mutations, but exhibited cell division cycle (*CDC27*) and Poly(A) binding protein cytoplasmic 1 (*PABPC1*) gene mutations.

There are several droplet-based platforms, such as Chromium Controller (10X Genomics), InDrop (1CellBio), and ddSEQ (Illumina/Bio-Rad Laboratories) that are utilised to perform scRNA-seq analysis (Haque et al., 2017). The workflow of scRNA-seq contains multiple consecutive steps to obtain reliable data (Figure 5.1). These steps include the isolation of individual cells from the sample of interest, cell lysis, preparation of complementary DNA (cDNA) through reverse transcription of poly(T)-primed mRNA, cDNA amplification, library preparation, single-cell sequencing, and data analysis using bioinformatic and computational tools (Haque et al., 2017; Klein et al., 2015).

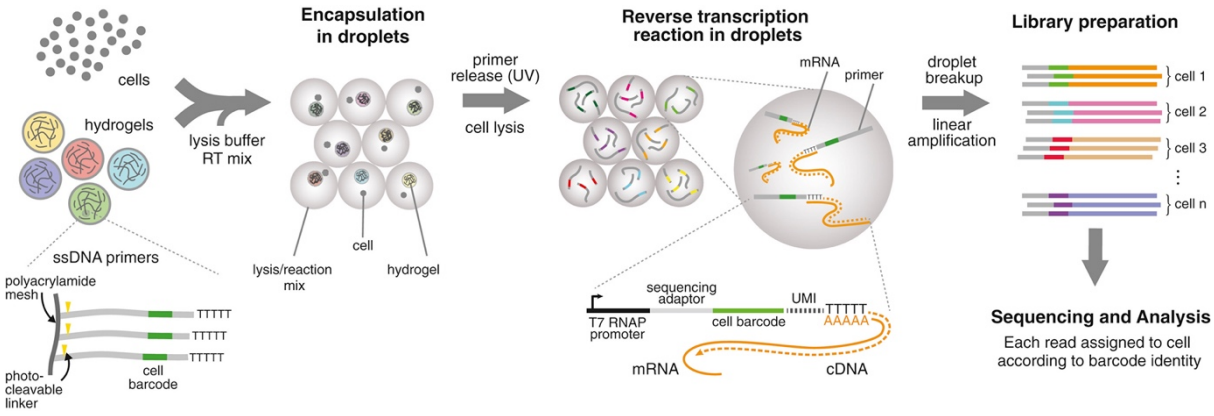


Figure 5.1 Workflow of scRNA-seq preparation. Cells are enclosed in droplets together with cell lysis buffer, reverse-transcriptase reagents, and barcoded primers. Following the encapsulation of cells in droplets and the release of primers, a reverse transcription reaction occurs. Consequently, cDNA within each droplet is labelled with a unique molecular identifier (UMI). Droplets are then ruptured which allows the contents of all cells to be linearly amplified before undergoing sequencing. The figure was adapted from Klein et al. (2015) with permission from Elsevier under the licence number: 5557120487465.

The 10X Genomics Chromium platform is a commonly used droplet-based platform, capturing gene expression profiles of single cells at high throughput. This system isolates individual cells into nanoliter-scaled droplets using a Gel-bead in Emulsion method. Each droplet is composed of a single cell, a unique barcode-embedded bead,

a reverse transcription reaction mix, a lysis buffer, and sequencing adaptors/primers. A droplet is separated from its neighbours via a carrier oil and afterwards passes through a chromium system for sequencing (Klein et al., 2015).

Although scRNA-seq is a powerful and comprehensive tool for assessing a wide range of biological processes, it requires substantial optimisation to minimise the technical noise produced during multiple steps. As such, post-processing scRNA-seq data presents some challenges such as empty droplets, multiple cells containing droplets (doublets), a “soup” of cell-free RNAs, and dealing with a high dropout rate (Young & Behjati, 2020). Current research using scRNA-seq mainly concentrates on gaining insights into diverse populations of cell types in tissue samples and biopsies. Therefore, there are several pipelines available for scRNA-seq data analysis, including using t-distributed stochastic neighbour embedding (TSNE) or uniform manifold approximation and projection (UMAP). However, there is a lack of research regarding the comparison of mixed cell lines that originate from the same tissue.

The aim of this chapter is to investigate scRNA data from LS180-COLO678 cell lines and Professor Alison Simmons data (GEO accession number: GSE116222). In this chapter, we studied the following objectives:

- Compare a goblet cell cluster with other cell types and explore variations within the goblet cell cluster in a single-cell profile of human normal colon tissue (Professor Simmons data)
- Identify intra-cell line variations of MUC2 and TFF3 genes expression and analyse their correlation with other genes in LS180 and COLO678 single-cell data

5.2 RESULTS

5.2.1 Single-cell mRNA sequencing analysis in normal colon

To study molecular profiling of the normal colon, Professor Alison Simmons' group (MRC Weatherall Institute of Molecular Medicine, University of Oxford) kindly provided us with single-cell data from healthy volunteers. First, colonic biopsy samples were taken from healthy individuals. Following the dispersion of normal colonic crypts into single cells and selection of epithelial cells using an anti-EpCAM antibody, droplet-sequencing technology was used to sequence the normal colon derived epithelial cells. Detailed information about collecting data and sequencing technology is described in Parikh et al. (2019). Sequencing data files can be accessed using the GEO accession number: GSE116222.

K-means clustering was performed on the post-processed scRNA-seq data and visualised using UMAP plots. In the normal colon, cell types generally can be classified as stem cells, transit-amplifying cells or progenitors, enterocytes or colonocytes, goblet cells, and enteroendocrine cells. We also added immune cells clusters after checking markers for lymphocytes which had not been completely removed by the anti-EpCAM selection. Overall, K-means clustering was used with k=6 which clearly separated 6 major cell type clusters that reflect different cell types within the normal colon (Figure 5.2). Every cluster was identified based on the gene expression of various markers specific to that cell type. Alongside discrete values representing all cells with any message, the five different levels of gene expression across whole cell population were described through UMAP plots. The code for the analysis carried out according to a pipeline developed by Dr Jeff Liu is shown in Code Appendix 1 (Liu, 2021).

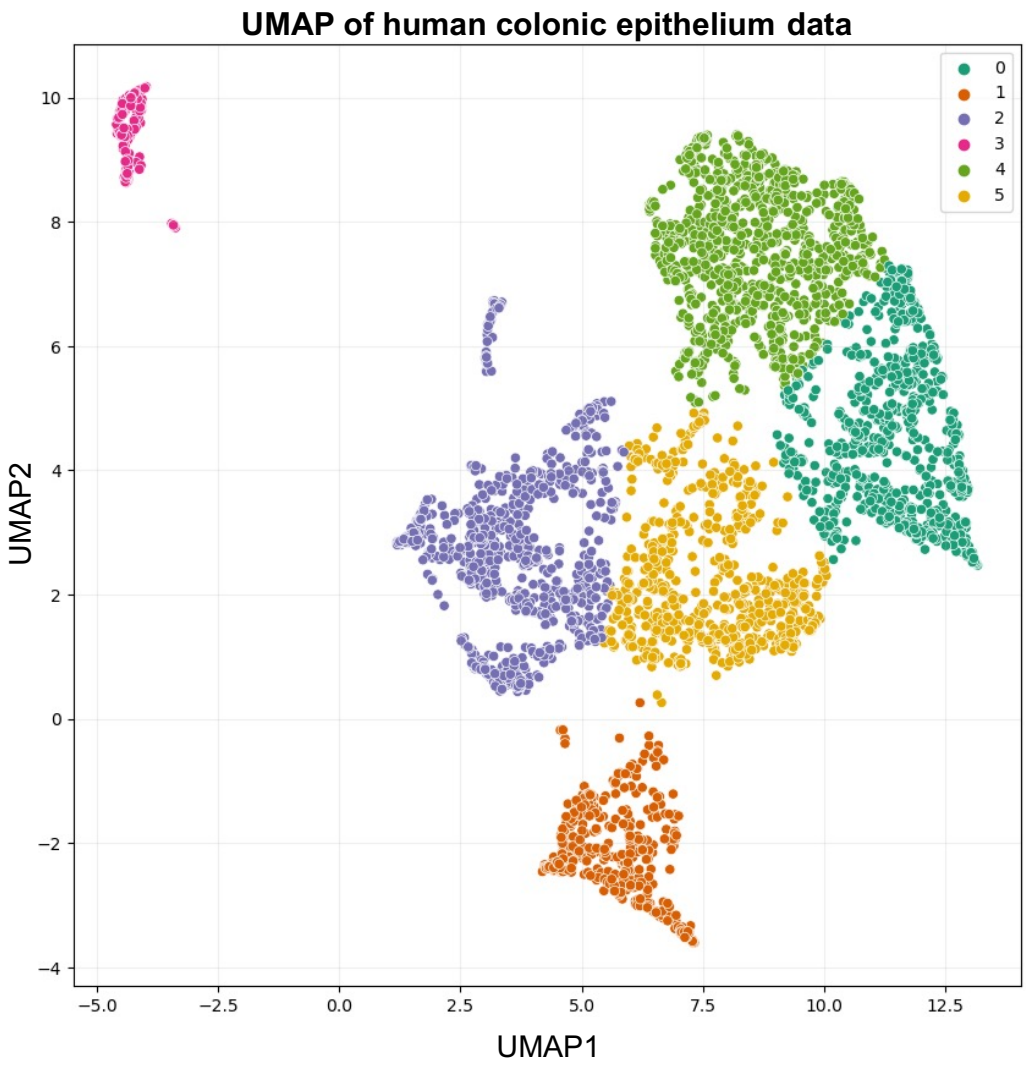


Figure 5.2 Two-dimensional K-means (k=6) display of scRNA-seq data obtained from epithelial cells from normal colon tissue of healthy individuals. The dataset was from Parikh et al. (2019) and the pipeline for the analysis was generated by Dr Jeff Liu (Liu, 2021).

Stem cells are the continually dividing cell types that are responsible for homeostasis and regeneration of the intestinal epithelium. Progenitor or transit amplifying cells are produced by the stem cells and are the cells that eventually differentiate into the three major cell types in the colonic crypts as described in Figure 1.1. We used different markers to define intestinal cell types and immune cells, as illustrated in Table 5.1.

Different biomarkers were selected and analysed to define intestinal cell types and immune cells, as illustrated in Table 5.1 and Figure 5.3. Overall, gene expression analysis suggests that clusters 0 and 4 are associated with enterocyte differentiation, cluster 1 represents goblet cells, cluster 2 probably includes the stem cell, 2 and 5 contain the transit-amplifying cells, and cluster 3 indicates immune cells, almost all lymphocytes.

	Gene description	References
Stem/progenitor cell markers	***ASCL2*** - Achaete-scute family BHLH transcription factor 2 ***AXIN2*** - Axin 2 ***CCNA2*** - Cyclin A2 ***CDCA7*** - Cell division cycle associated 7 ***EPHB2*** - EPH receptor B2 ***FOXM1*** - Forkhead box M1 **LGR5** - Leucine rich repeat containing G protein-coupled receptor 5 ***LRIG1*** - Leucine rich repeats and immunoglobulin like domains 1 ***MCM5*** - Minichromosome maintenance complex component 5 ***MEX3A*** - Mex-3 RNA binding family member A ***MKI67*** - Marker of proliferation Ki-67 ***MSI1*** - Musashi RNA binding protein 1 ***MYBL2*** - MYB proto-oncogene like 2 ***NUSAP1*** - Nucleolar and spindle associated protein 1 ***OLFM4*** - Olfactomedin 4 **PTK7** - Protein tyrosine kinase 7 ***RGMB*** - Repulsive guidance molecule BMP co-receptor B ***SMOC2*** - SPARC related modular calcium binding 2 ***TOP2A*** - DNA topoisomerase II alpha ***WDHD1*** - WD repeat and HMG-box DNA binding protein 1	Ashley et al. (2013); Barriga et al. (2017); Habowski et al. (2020); P. Jung et al. (2015); Wang et al. (2020)

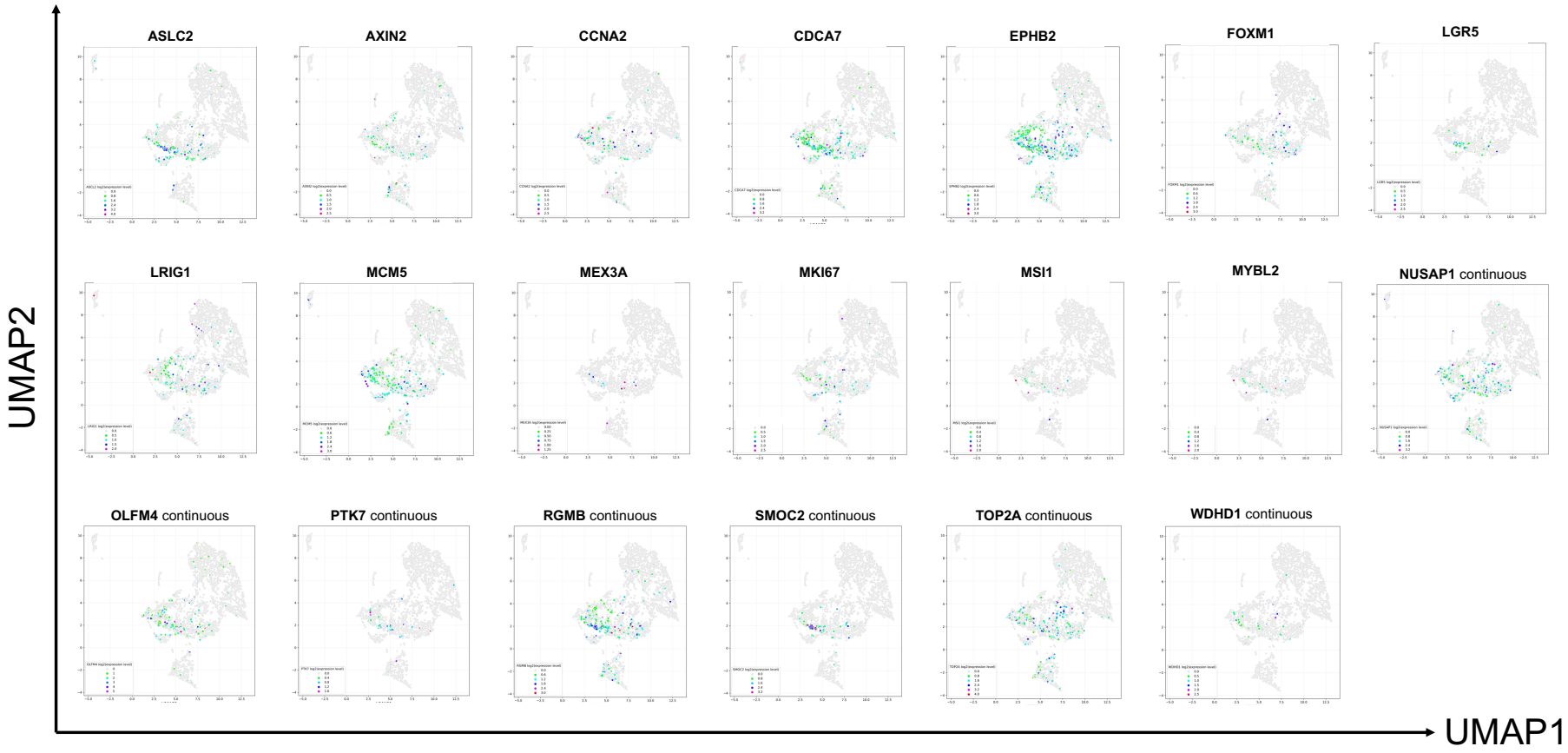
Enterocyte cell markers	***ALPI*** - Intestinal alkaline phosphatase ***AQP8*** - Aquaporin 8 ***BEST4*** - Bestrophin 4 ***CEACAM1*** - Carcinoembryonic antigen-related cell adhesion molecule 1 ***CEACAM6*** - Carcinoembryonic antigen-related cell adhesion molecule 6 ***CEACAM7*** - Carcinoembryonic antigen-related cell adhesion molecule 7 ***GUCA2A*** - Guanylate cyclase activator 2A ***GUCA2B*** - Guanylate cyclase activator 2B ***MS4A12*** - Membrane spanning 4-domains A12 ***OTOP2*** - Otopetrin 2 ***SLC26A3*** - Solute carrier family 26 member 3 ***TMEM37*** - Transmembrane protein 37	Burclaff et al. (2022); Dalerba et al. (2011); Kanke et al. (2022); Parikh et al. (2019); Wang et al. (2020)
Goblet cell markers	***ATOH1*** - Atonal BHLH transcription factor 1 ***BEST2*** - Bestrophin 2 ***CLCA1*** - Calcium-activated chloride channel regulator 1 ***FCGBP*** - Fc Gamma Binding Protein ***FFAR4*** - Free fatty acid receptor 4 ***ITLN1*** - Intelectin 1 ***MUC2*** - Mucin 2 ***REG4*** - Regenerating family member 4 ***SPDEF*** - SAM Pointed Domain Containing ETS Transcription Factor ***SPINK4*** - Serine protease inhibitor Kazal-type 4 ***TFF3*** - Trefoil factor 3 ***ZG16*** - Zymogen granule protein 16	Hu et al. (2023); Ito et al. (2013); Wang et al. (2020)

Enteroendocrine cell markers	***CHGA*** - Chromogranin A ***FEV*** - FEV transcription factor ETS family member ***INSM1*** - INSM transcriptional repressor 1 ***NEUROD1*** - Neuronal differentiation 1	Wang et al. (2020)
Immune cells markers	***CD52*** - CD52 molecule ***HOPX*** - HOP homeobox ***LSP1*** - Lymphocyte specific protein 1 ***PTPRC*** - Protein Tyrosine Phosphatase Receptor Type C	Bourque et al. (2021); Donovan and Koretzky (1993); Hermiston et al. (2003); Moutel et al. (2020); Wang et al. (2020)

Table 5.1 Genes selected to define intestinal cell types and immune cells in human normal colon scRNA-seq data. The references for the markers were provided in the last column of the table. Overall, we chose 20 markers for stem and progenitor cells, 12 markers for both enterocytes and goblet cells, and 4 markers for both enteroendocrine and immune cells.

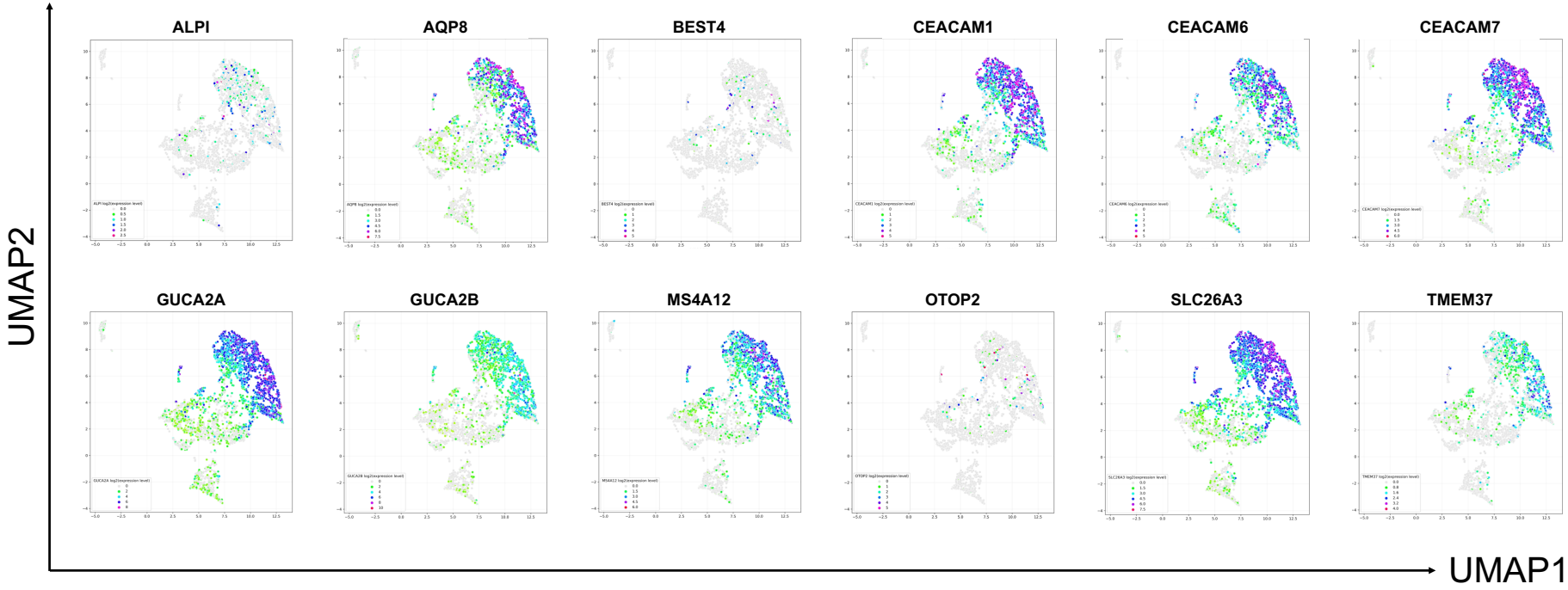
A.

Single-cell data analysis of normal colon (Stem / progenitor cell markers)



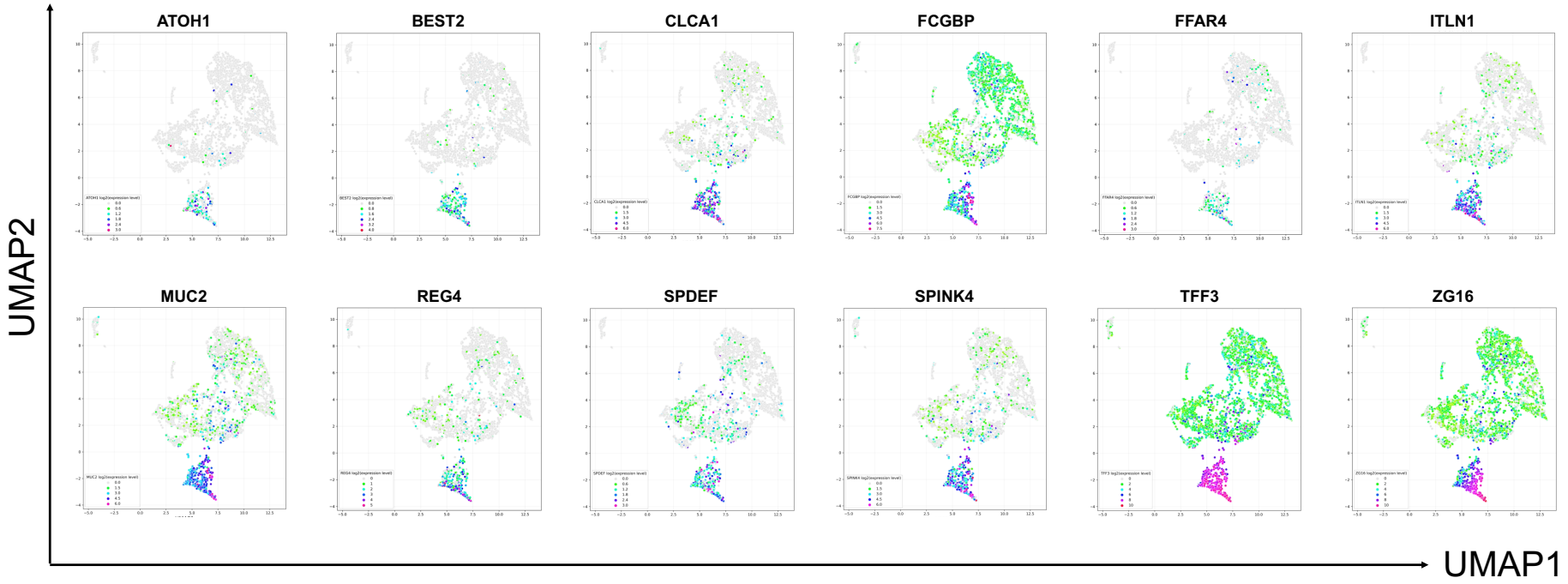
B.

Single-cell data analysis of normal colon (Enterocyte markers)



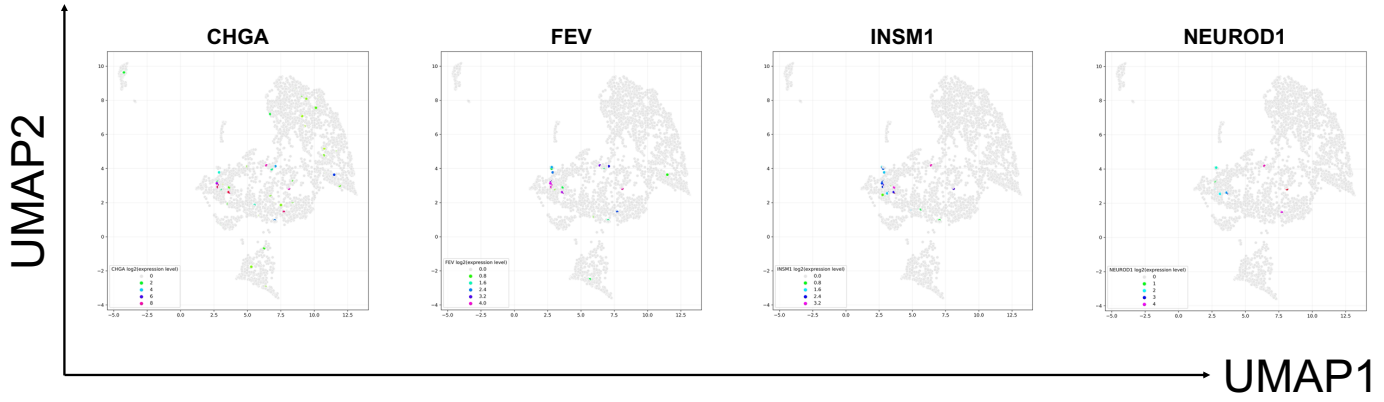
C.

Single-cell data analysis of normal colon (Goblet cell markers)



D.

Single-cell data analysis of normal colon (Enteroendocrine cell markers)



E.

Single-cell data analysis of normal colon (Immune cells markers)

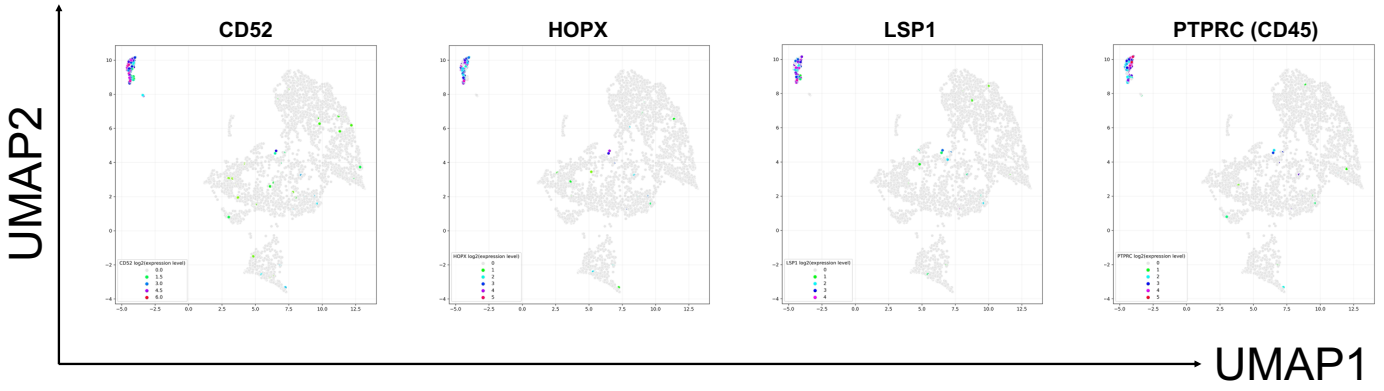


Figure 5.3 UMAP of scRNA-seq data of markers for intestinal cell types, along with immune cells. The graphs indicate the expression levels of stem/progenitor (A), enterocyte (B), goblet (C), enteroendocrine (D), and immune cell markers (E). Expression patterns were described using continuous values for each gene.

Once we defined the clusters, a volcano plot was generated to determine highly expressed genes within the goblet cell cluster as compared to the combination of all the other clusters, excluding the immune cells (Figure 5.4). The fold change and p-values of 56 selected genes from the volcano plot are listed in Appendix Table 2.

ATOH1 and SPDEF are known to be essential regulators of goblet cell differentiation. Their roles have already been discussed in previous sections, along with the importance of REG4 and FCGBP. Of the remaining highly goblet expressing genes it is not easy to determine which are the key controlling genes, which are important for goblet cell function, and which are simply passengers associated with the key controlling genes. It is important to emphasise that differences that matter may not just be presence or absence, but clear difference in levels of expression.

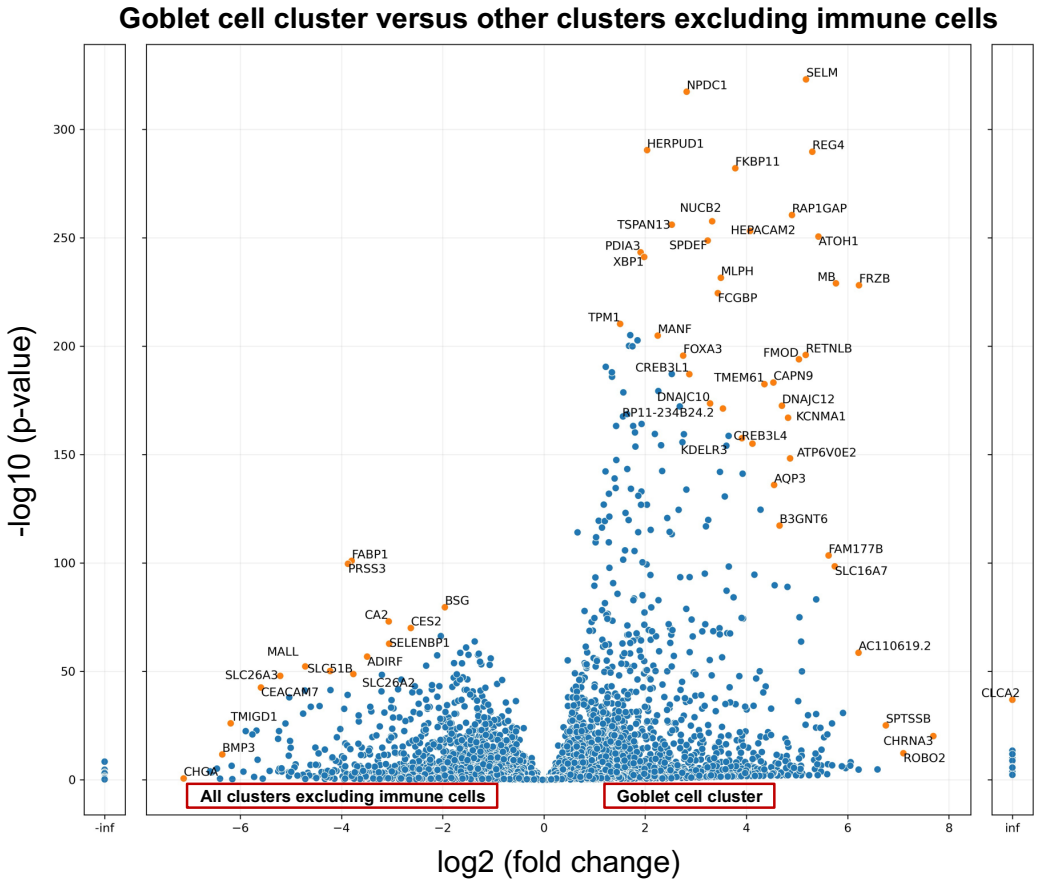
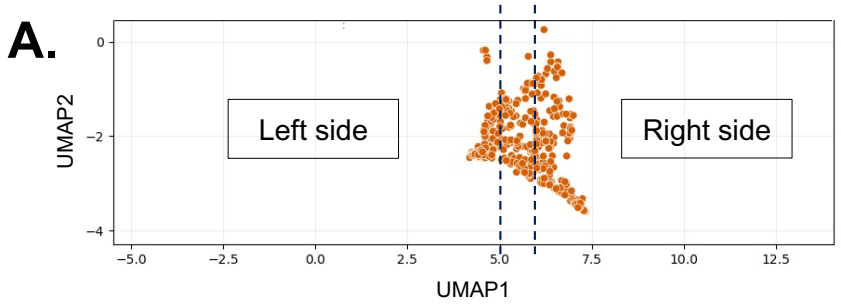


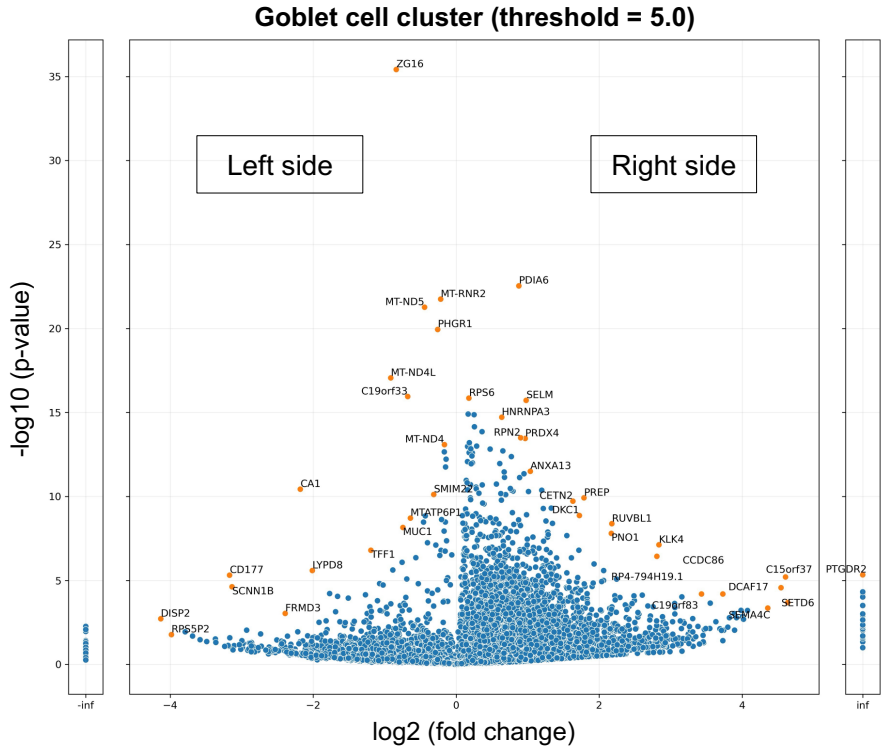
Figure 5.4 Volcano plot of gene expression analysis in goblet cell cluster versus other clusters excluding immune cells in scRNA-seq data of normal colon. The volcano plot exhibits the log2 of the fold change along the x-axis and the negative log10 of the p-value across the y-axis. Highly expressed genes were selected and visualised.

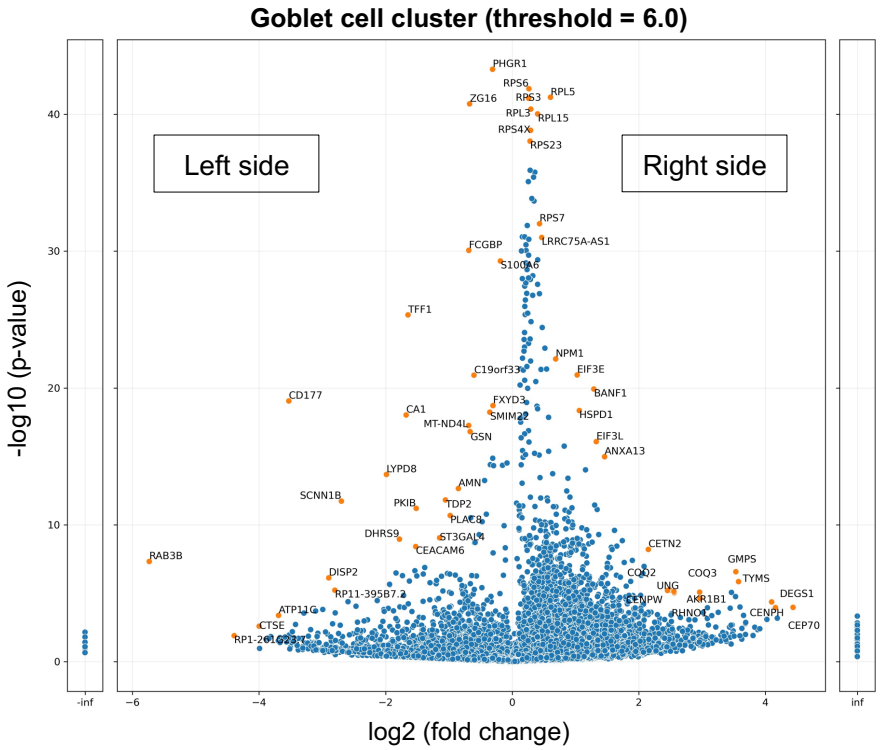
We also detected that there are spatial variations, notably between the left and right side, in expression of several genes within the goblet cell cluster. To identify genes involved in this expression difference, we used various thresholds, ranging from 5 to 6, and defined the most common genes from the right and left sides of the goblet cell cluster (Figure 5.5A/B/C). As illustrated in Appendix Figure 8, there was not much difference between the plots if we look only at the extremes, therefore, we displayed only the threshold values of 5 and 6 (Figure 5.5B). Appendix Table 3 depicts the fold change and p-values for selected genes. Middle parts of the volcano plots were excluded in our analyses.

A.


Left side
Right side
UMAP1
UMAP2

B.


Goblet cell cluster (threshold = 5.0)
Left side
Right side
-log10 (p-value)
log2 (fold change)


Goblet cell cluster (threshold = 6.0)
Left side
Right side
-log10 (p-value)
log2 (fold change)

C.

Left side of the goblet cell cluster									
Gene	**Count**		**Gene**	**Count**		**Gene**	**Count**		**Gene**	**Count**
LYPD8	11		PKIB	10		LGALS9C	3		PLAC8	1
ZG16	11		MT-RNR2	9		APOBR	2		SPTSSB	1
CD177	11		S100A6	8		MTATP6P1	2		FXYD3	1
SCNN1B	11		FRMD3	7		RP11-395B7.2	2		B3GALT5	1
CA1	11		RAB3B	7		C2orf88	2		MUC1	1
TFF1	11		MT-ND5	6		DHRS9	2		CLDN8	1
MT-ND4L	11		SMIM22	5		RPS6	1		CKB	1
PHGR1	11		RPS5P2	5		AMN	1		RP1-261G23.7	1
C19orf33	11		ATP11C	4		CTSE	1			
DISP2	10		GSN	4		ST3GAL4	1			
FCGBP	10		CEACAM6	4		TDP2	1			

Right side of the goblet cell cluster

Gene	Count	Gene	Count	Gene	Count	Gene	Count
RPS6	10	HNRNPA3	3	GNB1L	2	C19orf83	1
TYMS	9	HES6	3	CTPS1	2	RMI2	1
KLK4	9	TDP1	3	MANF	1	GINS1	1
PDIA6	8	RPL3	3	KTI12	1	RBM34	1
RPL5	8	SELM	3	SHF	1	FAM86DP	1
ANXA13	7	C15orf37	3	LYRM4	1	ZBTB14	1
HSPD1	7	COQ3	3	CCND2	1	ADO	1
AKR1B1	7	RPS7	3	DCAF17	1	MDC1	1
RPS3	7	RPL15	3	SEMA4C	1	TMEM97	1
SMARCC1	7	TCF19	3	CENPH	1	CENPK	1
EIF3E	7	SETD6	2	RP11-579D7.4	1	PTCD2	1
DKC1	7	CEP70	2	DEGS1	1	CD44	1
PRDX4	5	EIF3L	2	GMPS	1	SCPEP1	1
CETN2	4	NPM1	2	RHNO1	1	KLK12	1
ALDH4A1	4	LRRC75A-AS1	2	CENPW	1	C11orf73	1
RUVBL1	4	AP001610.9	2	COQ2	1	RPN2	1
GNPDA1	4	RIOK2	2	UNG	1	PREP	1
FAM111B	4	RP4-794H19.1	2	BANF1	1	L1TD1	1
C1QBP	3	CCDC86	2	RPS4X	1	RPS23	1
TGFBI	3	PNO1	2	SRD5A1	1		

Figure 5.5 Analysis of the goblet cell cluster using a range of threshold values. (A) Visual representation of separation interval within the goblet cell cluster. (B) Volcano plots compare the right and left sides of the goblet cell cluster with threshold values of 5.0 and 6.0. (C) The gene count was defined by the frequency of presence of genes within the threshold range of 5 to 6 during the comparative analysis of the right and left sides of the goblet cell cluster.

On the left side of the cluster, the most commonly expressed genes are LY6/PLAUR Domain Containing 8 (*LYPD8*), ZG16, CD177, Sodium channel epithelial 1 subunit beta (*SCNN1B*), Carbonic Anhydrase 1 (*CA1*), TFF1, Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4L (*MT-ND4L*), Proline, histidine and glycine rich 1 (*PHGR1*), Chromosome 19 open reading frame 33 (*C19orf33*). In addition, FCGBP, Galectin 9C (*LGALS9C*), ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (*ST3GAL4*), and MUC1 were also mainly expressed on the left side of the cluster. Considering the mature goblet cell markers described in the literature, which will be covered in detail in the Discussion section 5.3, we can hypothesise that the right side of the cluster are predominantly composed of mature goblet cells and the left side as cells still in the maturation process.

5.2.2 Single-cell mRNA sequencing analysis in LS180-COLO678 CRC cell lines

The scRNA-seq experiment in LS180 and COLO678 CRC cell lines was conducted as described in the Materials and Methods section (Chapter 2.14). This study intended to define and compare the expression of goblet cell-associated markers within and between different CRC cell lines. An initial analysis of a mixture (50:50) of goblet cell-positive (LS180) and -negative (COLO678) cell lines showed the potential of this approach and suggested novel ways to process single-cell mRNA expression data. As an input of scRNA-seq in two CRC cell lines, 50.000 barcodes and 25.345 UMIs were obtained and analysed.

Our lab successfully developed a new GMMchi (GMM with a Chi-square protocol)-based single-cell sequencing pipeline in Python programming language capable of effectively preprocessing single-cell data (Liu, 2021). This pipeline covers the following stages for better separation: a) filtering out empty barcodes and doublets, b) eliminating barcodes with low library complexity, c) extracting poor-quality barcodes through analysing the distribution of housekeeping genes, d) library normalisation, e) removing genes with little or no expression, f) excluding barcodes with high mitochondrial and ribosomal proteins encoding genes. An overall description of each step and post-processed outcomes can be found in Dr Jeff Liu's thesis (Liu, 2021).

Prior to mixing two cell lines, individual cells were not labelled with barcodes. This made it impossible to determine directly from which cell line a particular cell came from. Our microarray data and immunofluorescence analysis of LS180 and COLO678 cell lines indicate that LS180 expresses MUC2, but not placental alkaline phosphatase (ALPP), while COLO678 shows the opposite gene expression pattern. It allowed us to consider MUC2 and ALPP as reliable markers for separating two cell lines and

evaluating our post-processing pipeline efficiency. K-means clustering was done on the post-processed scRNA-seq expression data which identified two clusters that we could consider the LS180 and COLO678 cell lines by the presence of MUC2-positive and ALPP-positive cells, respectively (Figures 5.6 and 5.7A). Figure 5.7B specifies the count of cells that are positive and negative for ALPP and MUC2, along with total cells.

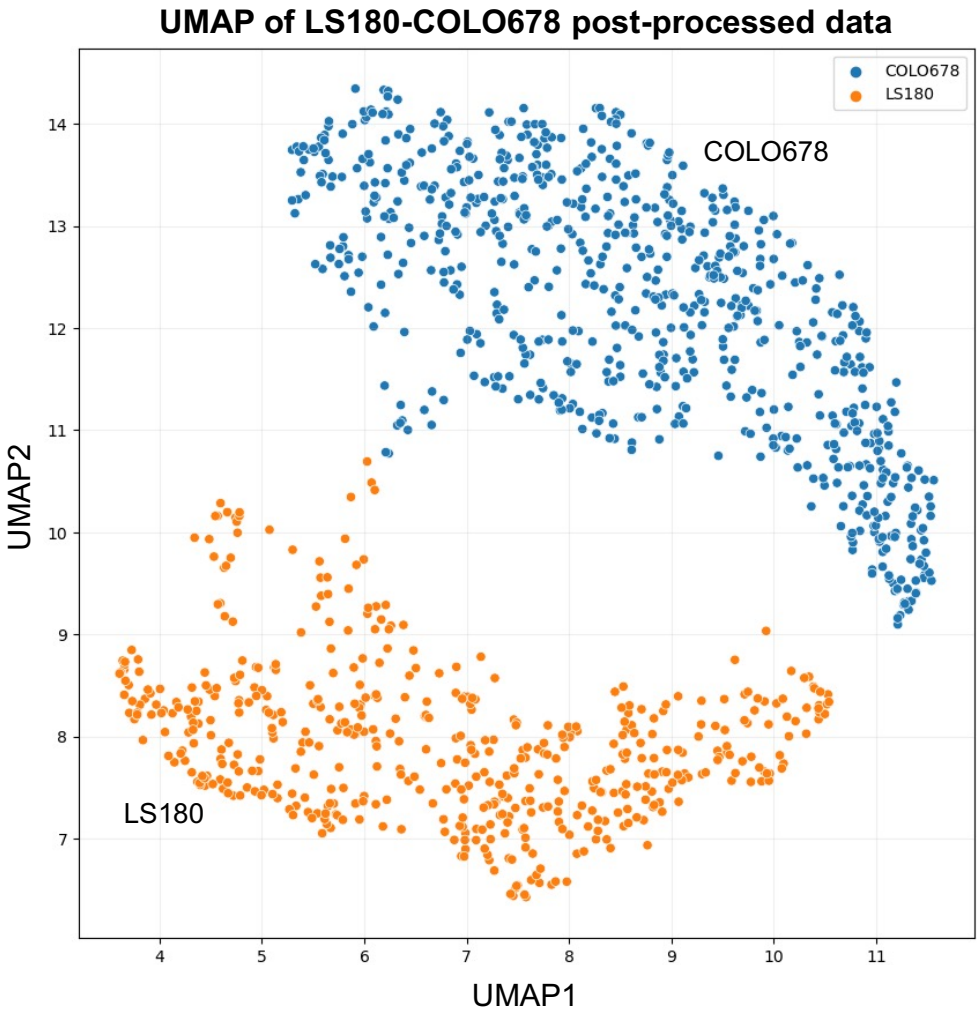
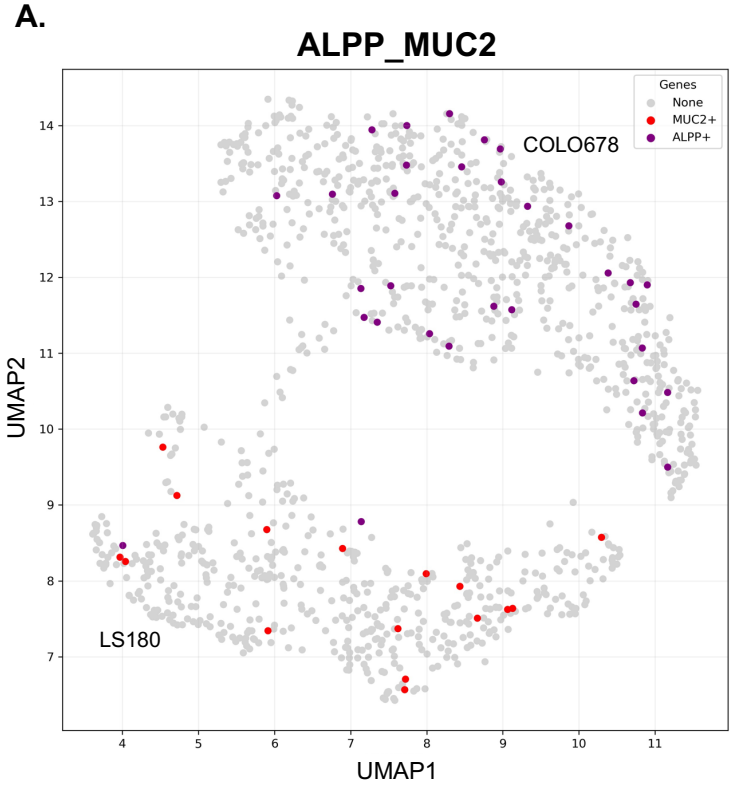


Figure 5.6 K-means (k=2) clustering on UMAP of LS180-COLO678 scRNA-seq data. K-means clustering (n=2) of our scRNA-seq data indicates two major clusters in the data, which are associated with LS180 (orange) and COLO678 (blue) cell lines.

A.



B.

Genes	Count
None	1178
ALPP+	32
MUC2+	16
Total	**1226**

Figure 5.7 UMAP of scRNA-seq data post-processed with GMMchi-based single-cell sequencing pipeline. (A) UMAP matrix of post-processed scRNA-seq data clearly distinguishes LS180 and COLO678 cell lines. MUC2-positive cells are labelled in red, while ALPP-positive cells are represented in purple. (B) The number of ALPP-positive and MUC2-positive cells is 32 and 16, respectively. The total number of cells within the two clusters is 1226.

As described in the previous section (Chapter 3.2.2), we quantified a range of MUC2 expressed cells (2.81%-4.14%) in the LS180 cell line based on IF analysis. When focusing specifically on the LS180 cluster in single-cell data, MUC2 expression is detected in a relatively small proportion, consisting of 16 cells out of a total population of 514. This complies with a subset of 3.1% of cells within the LS180 cluster that exhibit MUC2 expression, indicating a reliable estimate of MUC2-expressing cells (Figure 5.8A).

According to our IF analysis, TFF3, another goblet cell marker, is also expressed in LS180 at low levels. Within the LS180 cluster, a subset of 53 cells out of a total population of 514 express TFF3, which corresponds to nearly 10% of the cells (Figure 5.8B).

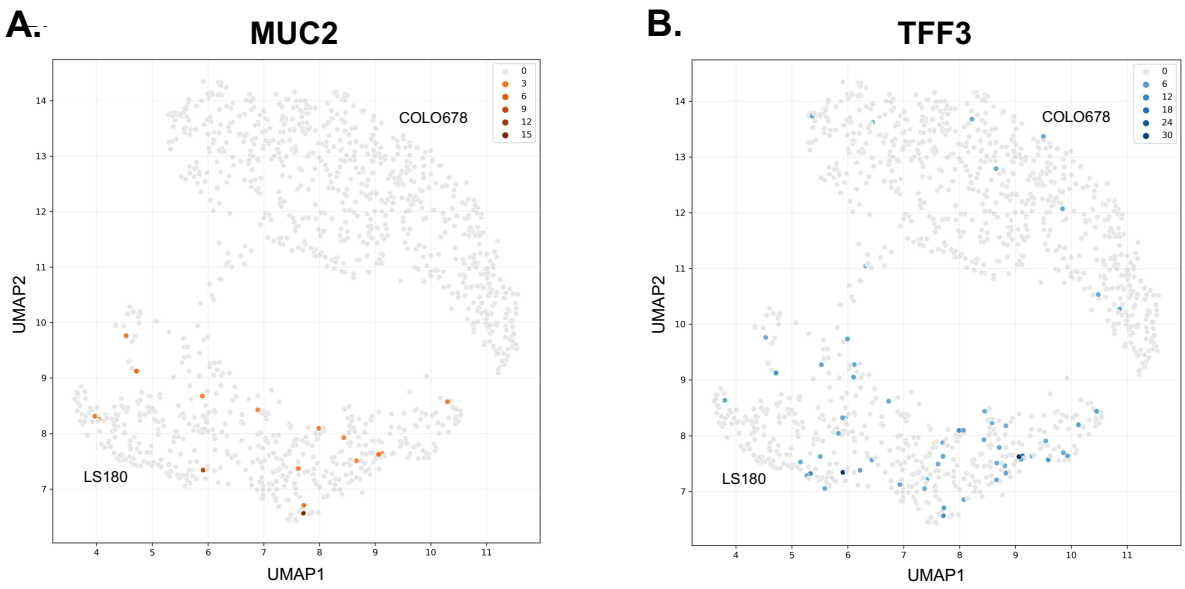
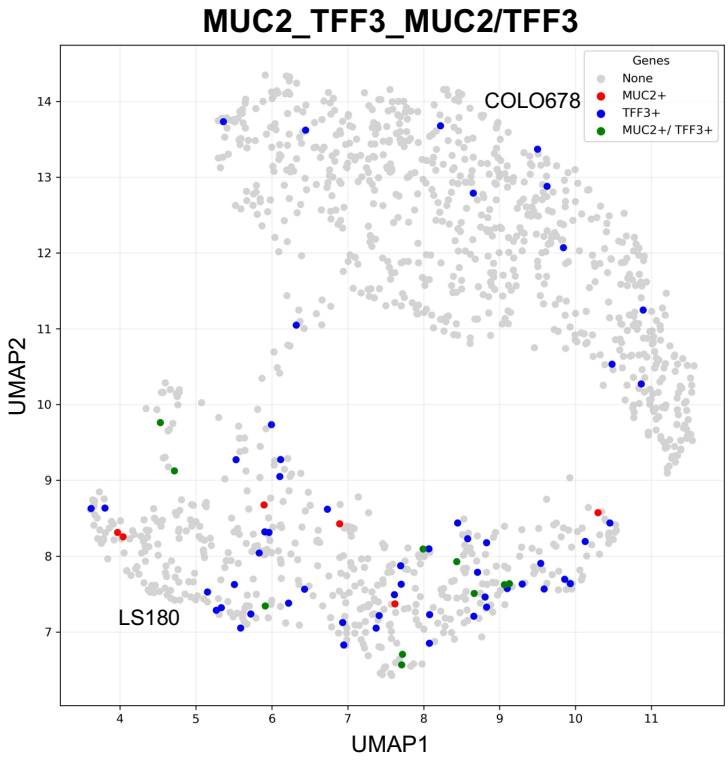


Figure 5.8 *MUC2* and *TFF3* genes expression in LS180-COLO678 single-cell data. The graph describes that only a small subset of cells expresses MUC2 (A) and TFF3 (B) within the LS180 cluster. Gene expression levels of MUC2 and TFF3 are illustrated in continuous values.

Based on our time-dependent single-cell experiment, TFF3 appears to be expressed earlier than MUC2 during differentiation (Chapter 3.2.4). There could be various reasons why TFF3 appears to represent a higher proportion compared to MUC2: a) The antibody used in IF may detect TFF3 better when it is more stable, which typically occurs in the presence of MUC2; b) The MUC2 antibody (PR5D5) used in this study specifically recognizes mature MUC2 and cannot detect immature forms of MUC2. This limitation in detection might contribute to the observed smaller proportion of MUC2-expressing cells compared to TFF3.

In the previous chapter, we identified that MUC2 and TFF3 are usually overlapped in the LS180 cell line (Chapter 3.2.2). Accordingly, when we label cells that represent both MUC2 and TFF3 expression, we can see that 10 out of 16 MUC2-positive cells also produce TFF3 (Figure 5.9A/B) which, as might be expected, is a highly significant association.

A.



B.

Genes	Count
None	1156
MUC2+	6
TFF3+	54
MUC2+/TFF3+	10
Total	**1226**

p value < 0.0001

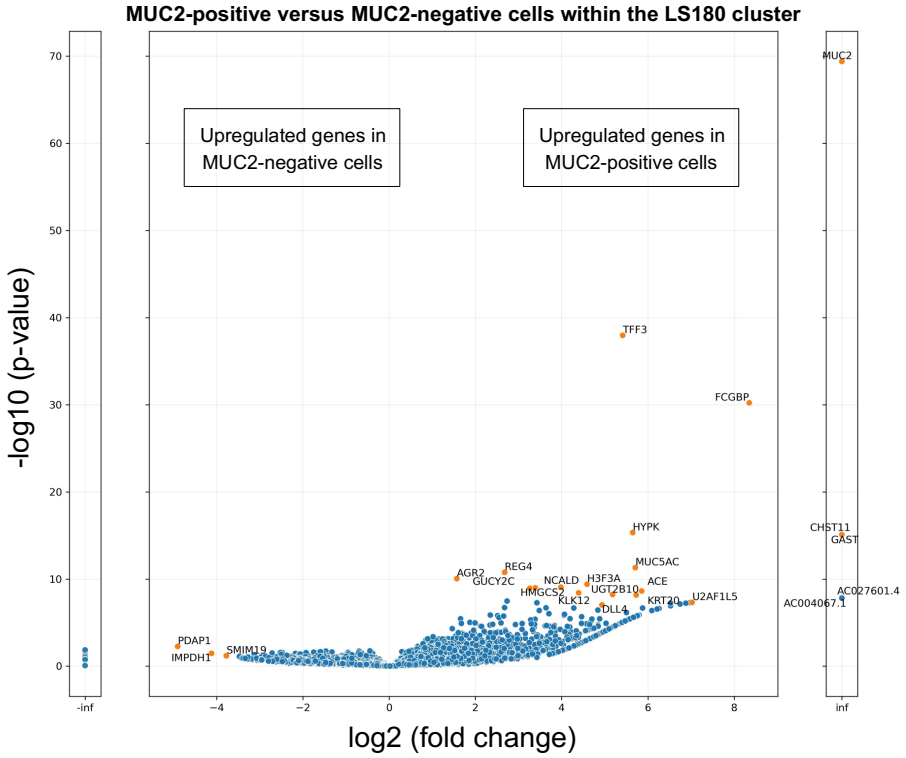
Figure 5.9 UMAP matrix of scRNA-seq data indicating count of MUC2-, TFF3-, and MUC2/TFF3-positive cells, along with a total number of cells. (A) In both clusters, cells expressing MUC2, TFF3, and both MUC2/TFF3 are labelled in red, blue, and green, respectively. (B) The number of MUC2-positive, TFF3-positive, and MUC2/TFF3-positive is indicated and a chi-square test was performed.

A volcano plot was generated to identify genes that are significantly up- and down-regulated in MUC2-positive cells compared to MUC2-negative cells within the LS180 cluster (Figure 5.10A). The code for this analysis is described in Code Appendix 2. The fold change values were determined by dividing the mean of individual gene

expression in MUC2-positive cells by that of the MUC2-negative cells in LS180. It should be noted that the p-values were not adjusted with Bonferroni correction and only genes with a p-value of less than 0.01 were considered statistically significant and included in the subsequent analysis. Appendix Table 4A shows a list of 20 genes selected from the volcano plot (Figure 5.10A) along with their corresponding fold change and p-values.

A similar analysis was conducted for TFF3, where the volcano plot was used to identify differentially expressed genes between TFF3-positive and TFF3-negative cells in the LS180 cell line (Figure 5.10B). Appendix Table 4B contains a list of selected genes from the volcano plot (Figure 5.10B) together with their fold change and p-values.

A.


MUC2-positive versus MUC2-negative cells within the LS180 cluster
Upregulated genes in MUC2-negative cells
Upregulated genes in MUC2-positive cells
-log10 (p-value)
log2 (fold change)
MUC2
TFF3
FCGBP
HYPK
CHST11
GAST
MUC5AC
AGR2
REG4
GUCY2C
NCALD
H3F3A
HMGCS2
UGT2B10
ACE
KLK12
KRT20
U2AF1L5
DLL4
AC027601.4
AC004067.1
PDAP1
IMPDH1
SMIM19

B.

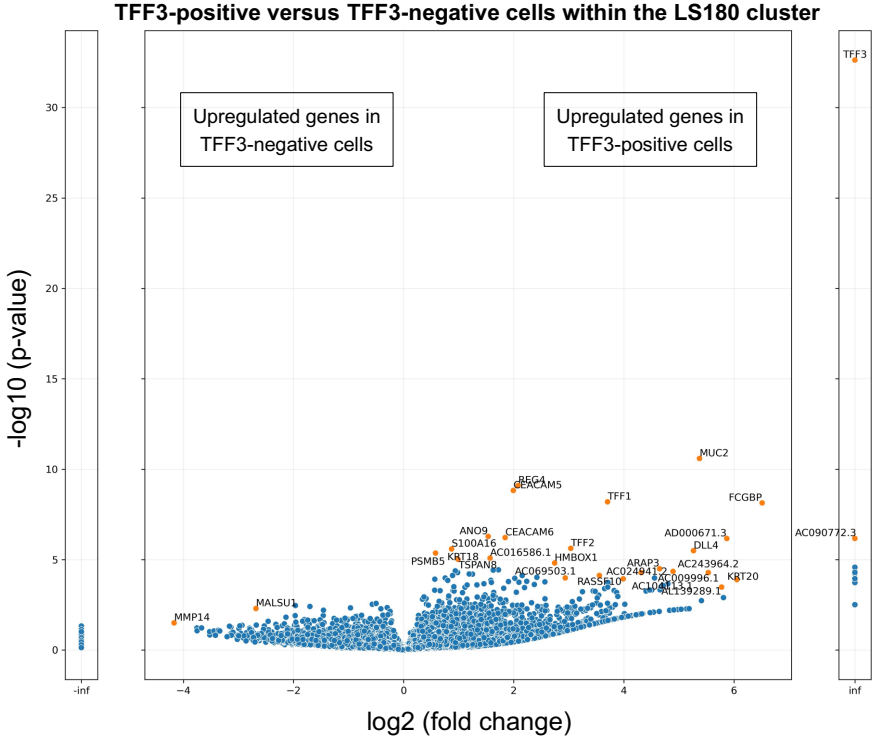

TFF3-positive versus TFF3-negative cells within the LS180 cluster
Upregulated genes in TFF3-negative cells
Upregulated genes in TFF3-positive cells
-log10 (p-value)
log2 (fold change)
TFF3
MUC2
REG4
CEACAM5
TFF1
FCGBP
ANO9
CEACAM6
S100A16
AC016586.1
TFF2
HMBOX1
AD000671.3
DLL4
AC090772.3
PSMB5
KRT18
TSPAN8
AC069503.1
ARAP3
AC243964.2
AC024941.2
AC009996.1
KRT20
RASSF10
AL139289.1
MALSU1
MMP14

Figure 5.10 Differential gene expression analysis in MUC2/TFF3-positive cells versus MUC2/TFF3-negative cells in scRNA-seq data of LS180. The log2 of the fold change is plotted against the negative logarithm (base 10) of the p-value in the volcano plot. (A) Genes that are differentially expressed between the MUC2-positive and MUC2-negative cells are labelled. Infinite log2 of the fold change values were plotted separately. (B) Up- and down-regulated genes between the TFF3-positive and TFF3-negative cells within LS180 cluster are highlighted. Infinite log2 of the fold change values were visualised in separate axes.

TFF3, FCGBP, regenerating islet-derived family member 4 (*REG4*), and anterior gradient 2 (*AGR2*) were among the upregulated genes in MUC2-positive cells, which are involved in MUC2 maturation and folding. MUC5AC is typically not expressed in the normal colon, however, its de novo expression can appear in adenomas and colorectal cancer (Bu et al., 2010). According to our data on the MSI status of cell lines, LS180 is classified as an MSI-high cell line. This finding correlates with the fact that colorectal cancer with high MSI often displays a higher rate of MUC2 and MUC5AC gene expression (Bu et al., 2010). Keratin 20 (*KRT20*) is primarily expressed in intestinal epithelial cells and has been widely investigated as a marker of cellular differentiation. According to the findings in our laboratory, KRT20 is controlled by CDX1 in 38 primary CRC cell lines (Chan et al., 2009).

MUC2, FCGBP, and REG4 are strongly associated with one another and provide a stable structure for a mucin layer. TFF1, a trefoil factor family member gene, appears to have a positive correlation with TFF3. Several studies showed TFF1 and TFF3 were significantly increased in colorectal cancer patients compared to healthy individuals (Vocka et al., 2015; Yusufu et al., 2019). Both KRT18 and KRT20 form intermediate filaments and are well-known intestinal markers. Carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) and CEACAM6 are associated with differentiation.

A 2x2 table was generated to identify the correlation between MUC2/TFF3 and other genes within LS180 cell line (Table 5.2). Cells were classified based on their gene expression levels, with "+" indicating positive/high and "-" indicating negative/low gene expression levels. In the association study, genes were selected based on the Benjamini-Hochberg adjusted p-value with a cut-off of 0.05. According to 2x2 analyses, MUC2 exhibits a strong correlation with TFF3 and FCGBP (Table 5.1A/B), as might be expected. The inclusion criteria for TFF3 represents that 10/16 (62.5%) of MUC2-positive cells express TFF3, whereas 43/53 (81.1%) of TFF3-positive cells do not express MUC2 (Table 5.1B). This strongly suggests that TFF3 is required for MUC2 maturation, but not vice versa and that many of the TFF3 only expressing cells may not yet be mature goblet cells.

Associations with MUC2 and TFF3 expression

A) Positive association of MUC2 expression with inclusions (skewed to +/-)

MUC2	+/+	+/-	-/+	-/-	P-value	R-value	Inclusion criteria
FCGBP	5	11	1	497	8.78E-08	0.50	Significant

B) Positive association of MUC2 expression with inclusions (skewed to -/+)

MUC2	+/+	+/-	-/+	-/-	P-value	R-value	Inclusion criteria
TFF3	10	6	43	455	2.96E-07	0.31	Significant

Table 5.2 Genes associated with MUC2 and TFF3 expression in scRNA-seq data of LS180. (A) and (B) describe genes positively associated with MUC2 expression that are inclusions skewed towards +/- and -/+, respectively. Genes were chosen based on the Benjamini-Hochberg adjusted p-value with a cut-off of 0.05.

5.3 DISCUSSION

Our findings in this chapter validate that single cell analysis provides a deeper understanding compared to the analysis of bulk tissue. As described in normal colon scRNA-seq data provided by Professor Alison Simmons, we can clearly separate goblet cells from stem, transit-amplifying cells, and mature enterocytes mostly based on known expression profiles.

Analysis of the goblet cell cluster difference from other normal cell types showed expression of many genes that have been clearly associated with goblet cell function - notably ATOH1 and the co-presence of MUC2 and TFF3. More detailed analysis suggested that the cluster contained two categories of goblet cells which from the pattern of differential expression indicated that one group might be mature goblet cells while the other were cells in which the maturation of MUC2 expression was happening but not yet complete.

Maturation of goblet cells is a complex process in which they undergo both morphological and chemical changes. The SPDEF transcription factor plays a key role in terminal differentiation of goblet cells. Despite the limited research on identifying markers that distinguish mature and immature colonic goblet cells, it is feasible to make some hypotheses from the existing literature. The study by Kanke et al. performed single cell sequencing analysis to examine molecular variances in colonic epithelial cell types from patients with Crohn's disease and without inflammatory bowel disease (Kanke et al., 2022). It was identified that Crohn's disease patients exhibited a high predominance of mature goblet cells based on MUC2, MUC4, and TFF1 expression among secretory cell lineages. Considering the common expression of TFF1 on the

left side of the cluster, we can hypothesise that there might be a difference between mature and not yet mature goblet cells.

ZG16 and LYPD8 were significantly upregulated genes on the left side of the goblet cell cluster. As described in the introduction section (Chapter 1.3.4), ZG16 protects the intestinal epithelium by interacting with a membrane of Gram-positive bacteria. Similarly, Ly6/plasminogen activator urokinase receptor domain containing 8 (LYPD8) provides a protective lining against Gram-negative bacteria by inhibiting flagellated bacterial mobility (Okumura et al., 2016). Mice with LYPD8 deficiency showed higher sensitivity to colitis induced by DSS (Allaire et al., 2018; Okumura et al., 2016). Furthermore, in mice lacking ZG16, the thickness of the distal colonic mucus layer remained constant, but commensal bacteria were able to localise to the colon epithelium (Allaire et al., 2018; Bergstrom et al., 2016). This suggests that ZG16 and LYPD8 contribute to intestinal protection through separating commensal bacteria from epithelial cells.

Another supportive factor for this hypothesis could be the existence of glycosylation enzymes and molecules on the left side of the goblet cell cluster, which are likely related to the late stage of mucin maturation. These include Galectin 9C (LGALS9C), ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (ST3GAL4), and Beta-1,3-galactosyltransferase 5 (B3GALT5).

LGALS9C, a soluble galactoside-binding lectin, possesses multiple functions, such as induction of an adaptive immune system, glucose homeostasis, and regulation of tumour immunity (Yang et al., 2008). It was reported to interact with either N/O glycans or both, as well as protein disulfide isomerase which involves the formation of disulfide bonds in the endoplasmic reticulum (Schaefer et al., 2017). Furthermore, mice with a

Galectin 9 (LGALS9) deficiency experienced higher susceptibility and defected cellular proliferation in response to intestinal epithelial damage (Robinson et al., 2020). By contrast, the expression of LGALS9 elevated in regenerating colonic crypts of patients during wound healing (Robinson et al., 2020). Beta-1,3-galactosyltransferase 5 (B3GALT5) is involved in translocating Gal to GlcNAc-based molecules and ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (ST3GAL4) plays an important role in the formation of N-glycans ending with Neu5Ac(α2,3)Gal through transferring sialic acid to the last Gal residue (Arike et al., 2017; Qi et al., 2020). Taking into account all of the above-mentioned statements, we can propose a hypothesis that goblet cell maturation occurs from the left to right.

As a next step, we analysed scRNA-seq data from LS180-COLO678 cell lines conducted in our laboratory. The two cell lines analysis shows how easy it is to distinguish two cell lines even though both are CRCs. Single cell analysis allows analysing variation within the cell line. It was identified that TFF3 expression was more than MUC2 but highly associated at the single cell level.

In summary, we compared the goblet cell cluster and identified certain genes associated with goblet cell formation that require further investigation. It appeared that there were differences between the right and left sides of the goblet cell cluster. In addition, the investigation of CRC cell lines at a single cell level is a powerful tool for defining variations among cells.

CHAPTER 6

STUDY OF FACTORS INVOLVED IN GOBLET CELL DIFFERENTIATION

CHAPTER 6: STUDY OF FACTORS INVOLVED IN GOBLET CELL DIFFERENTIATION

6.1 INTRODUCTION AND AIMS

Goblet cell differentiation is a complex process requiring genetic and epigenetic factors. The Notch pathway controls cellular proliferation and differentiation in the intestine (Demitrack & Samuelson, 2016; Zecchini et al., 2005). It defines intestinal cell fate, leading cells to either the absorptive or secretory lineage. Suppression of the Notch pathway caused a notable reduction in HES1 expression and accelerated secretory lineage differentiation (Ueo et al., 2012; van der Flier & Clevers, 2009).

Several transcription factors have been reported to be involved in goblet cell differentiation. ATOH1 is a key transcription factor for the secretory lineage, detected in both secretory progenitors and mature cell types (Demitrack & Samuelson, 2016). HES1 was shown to inhibit the expression of ATOH1 via direct interaction with the promoter region of ATOH1 (Demitrack & Samuelson, 2016; Zheng et al., 2011). SPDEF plays a role in the terminal differentiation of goblet cells (Gregorieff et al., 2009; Noah et al., 2010). Other transcription factors, CDX1 and CDX2 regulate the expression of MUC2 in CRC cell lines through the putative CDX binding sites in the MUC2 promoter region (Mesquita et al., 2003). DNA methylation patterns in the intestine show dynamic changes during cell proliferation and differentiation (Huang et al., 2015; Sheaffer et al., 2014). The study by Sheaffer et al. demonstrated that the number of differentiated cell types within the small intestine remained mainly unchanged during loss of methylation (Sheaffer et al., 2014). In **Chapter 6**, we investigated the transcription factors involved in goblet cell differentiation, as well as the effects of Notch pathway and methylation on goblet cell production.

6.2 RESULTS

6.2.1 Identification of differentially expressed genes based on the groups determined by microarray and immunofluorescence analysis

Based on our microarray data, MUC2- or TFF3-positive cell lines and MUC2- or TFF3-negative cell lines were selected for identifying candidate genes involved in the control of goblet cell differentiation. To achieve this, the gene mRNA expression of these cell lines was compared using volcano plots (Figure 6.1A/B). Fold change was obtained by dividing the mean of individual gene expression of the MUC2/TFF3-positive cell lines by that of MUC2/TFF3-negative cell lines. Appendix Table 5A/B provides all selected genes with their corresponding fold change and p-values, represented by the volcano plot on the Figure 6.1A/B.

Goblet cells are defined by the high levels of MUC2 and TFF3 which was also confirmed by the volcano plots. The analysis indicated that goblet-cell positive cell lines expressed 7.9-fold and 17.5-fold higher levels of MUC2 (p-value = 3.20E-06) and TFF3 (p-value = 2.30E-06), respectively, relative to non-goblet cell lines. The p-values for MUC2 and TFF3 were not corrected for multiple comparisons.

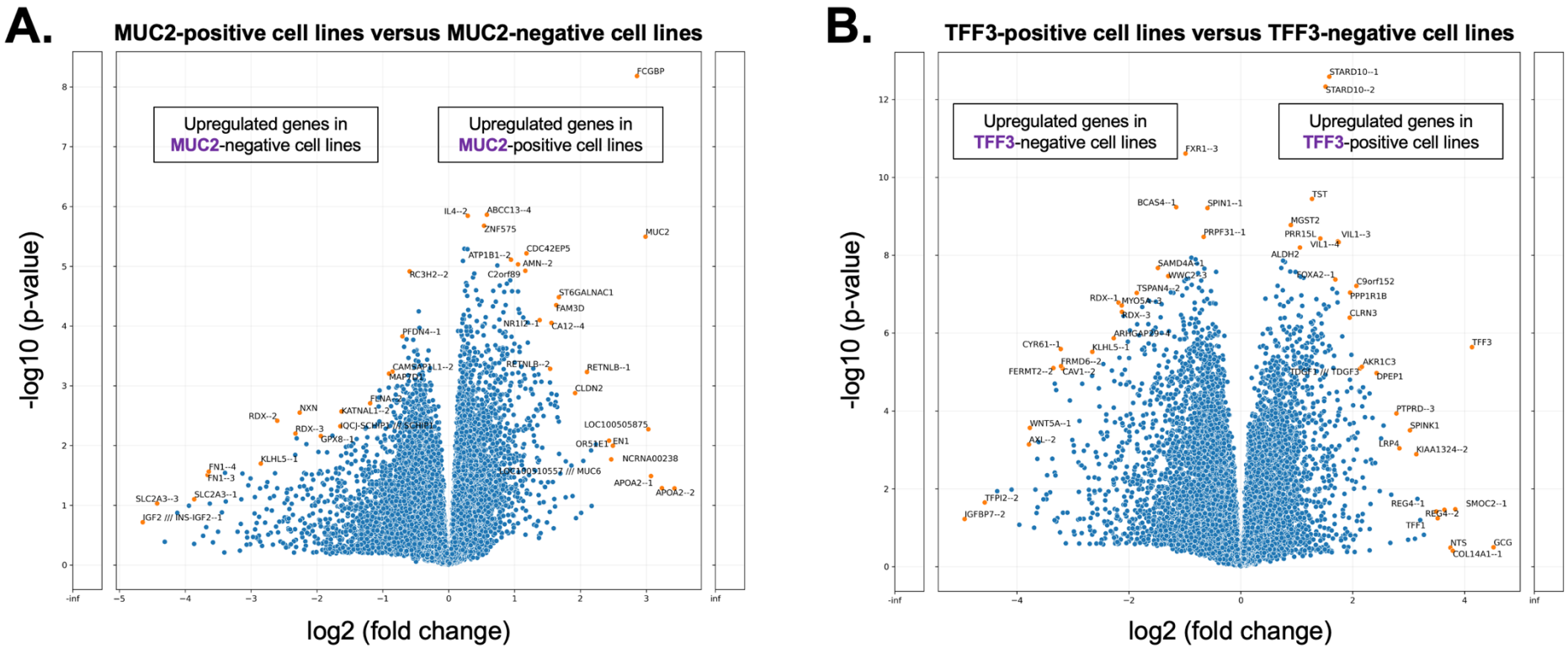
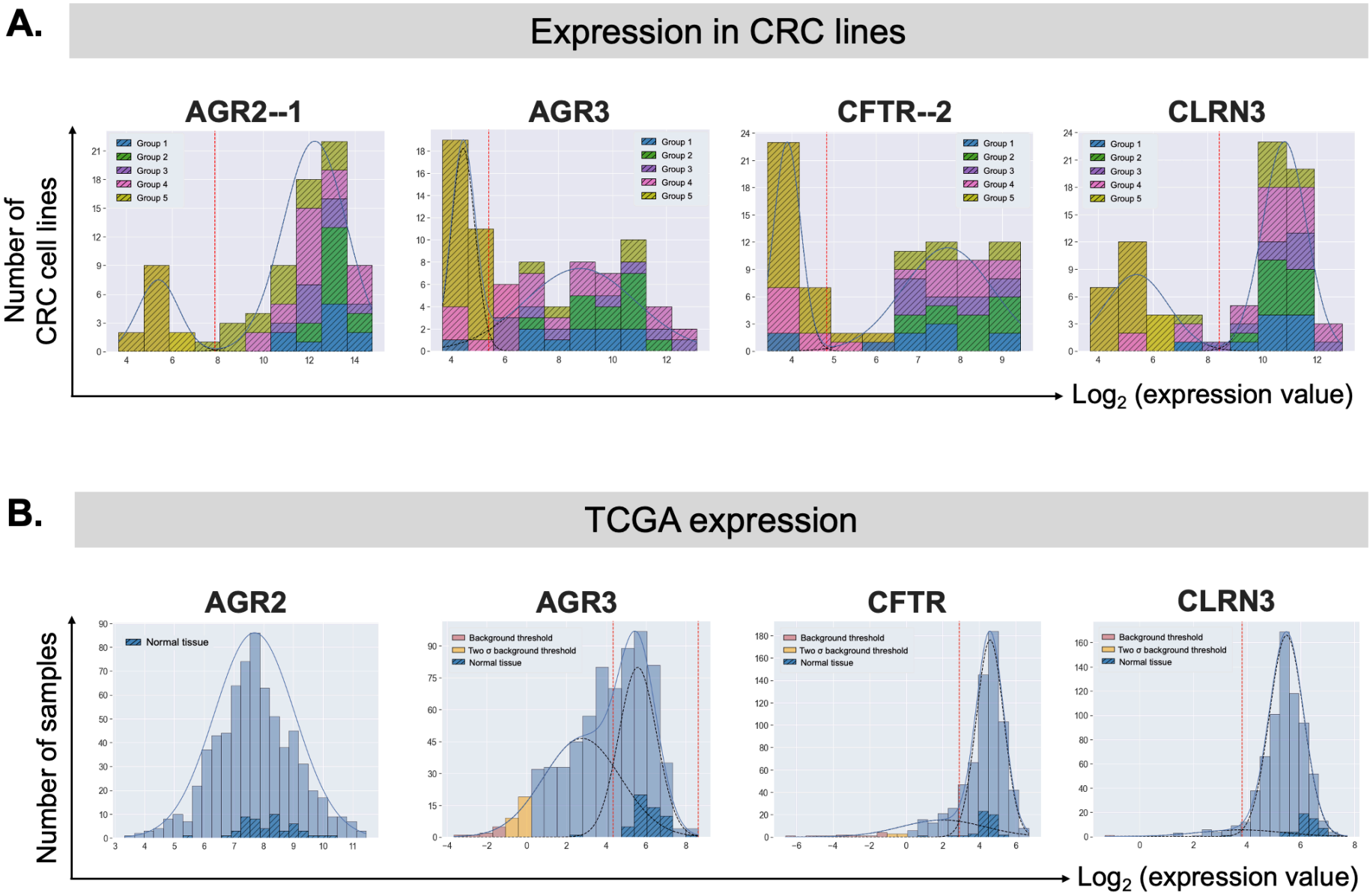


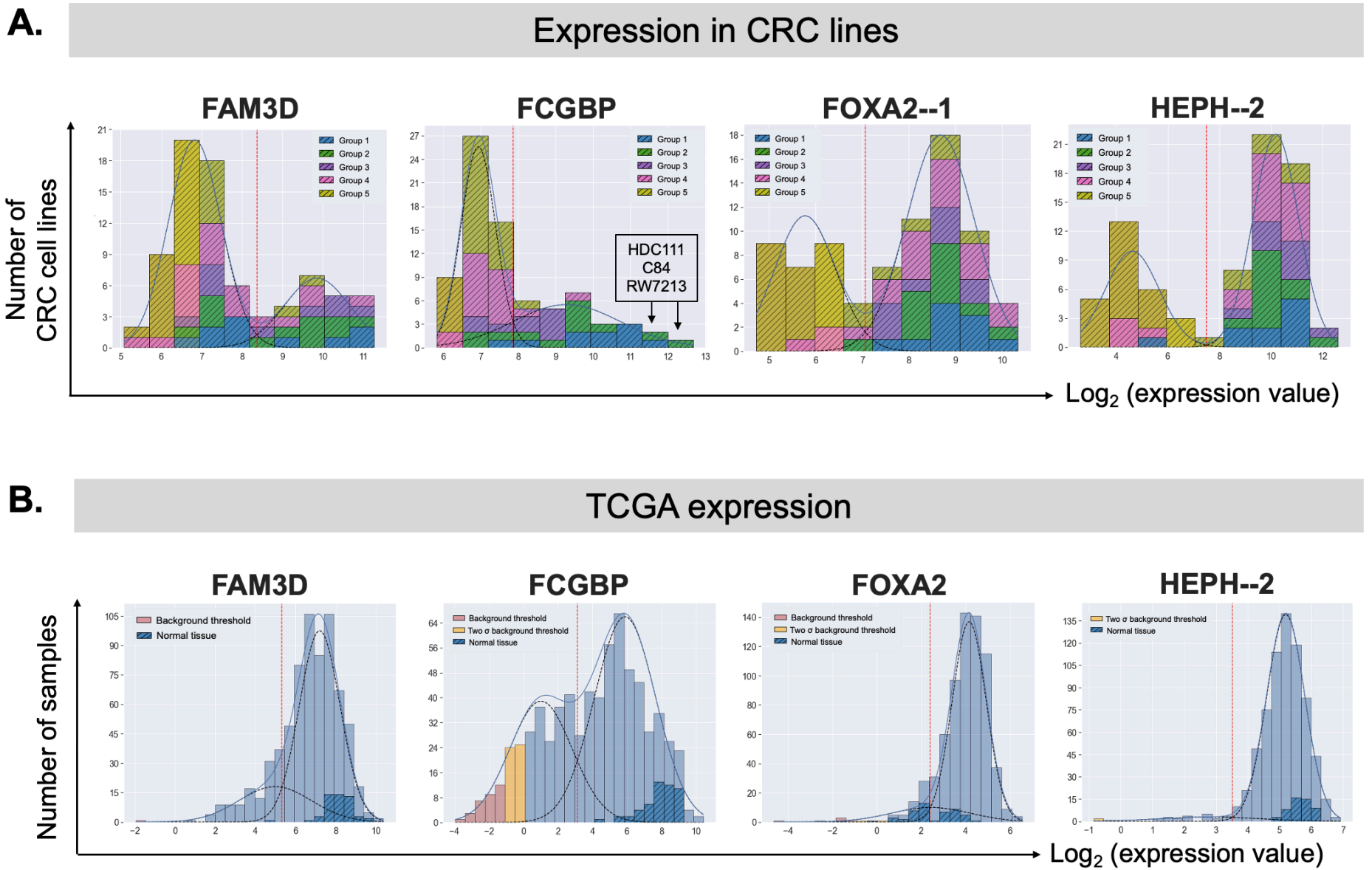
Figure 6.1 Gene expression analysis in MUC2- or TFF3-positive cell lines versus MUC2- or TFF3-negative cell lines in our microarray data. In the volcano plot, the x-axis shows the log2 of the fold change, whilst the y-axis illustrates the negative logarithm (base 10) of the p-value. Key differentially expressed genes between two groups of cell lines are labelled.

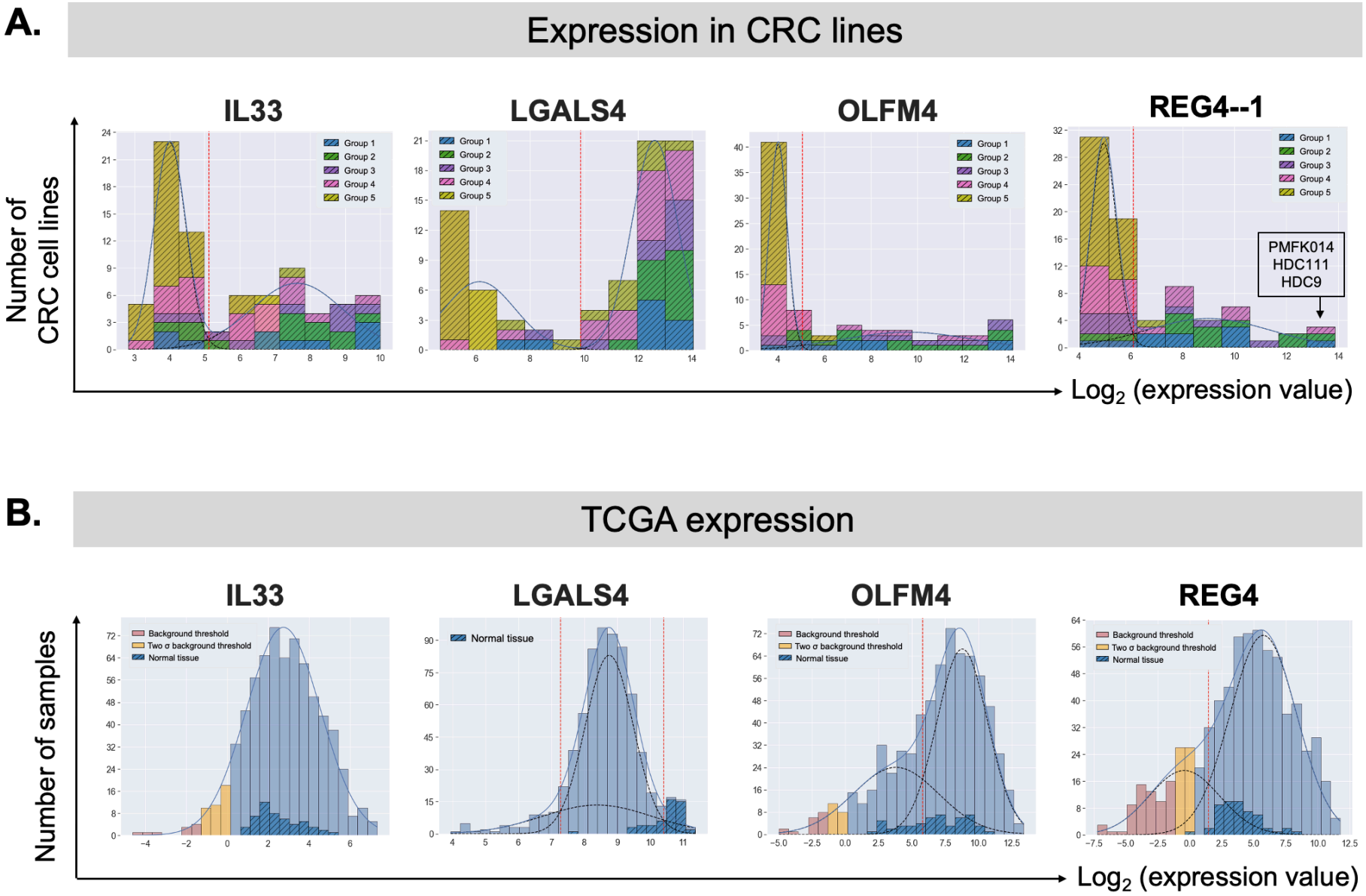
Other goblet cell-associated genes, namely FCGBP (fold change = 7.26) and REG4 (fold change = 11.28), were also found to be highly expressed in MUC2/TFF3-positive cell lines. ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (*ST6GALNAC1*) is involved in mucin glycosylation, making it significantly upregulated genes in goblet cells (Zhang et al., 2022). Anterior gradient protein 2, protein disulphide isomerase family member (*AGR2*), AGR3, Clarin 3 (*CLRN3*), FAM3 metabolism regulating signaling molecule D (*FAM3D*), Hephaestin (*HEPH*), Cystic fibrosis transmembrane conductance regulator (*CFTR*), Resistin-like beta (*RETNLB*), Interleukin 33 (*IL33*), Galectin 4 (*LGALS4*), TOX high mobility group box family member 3 (*TOX3*), and OLFM4 were among differentially expressed genes associated with goblet cells.

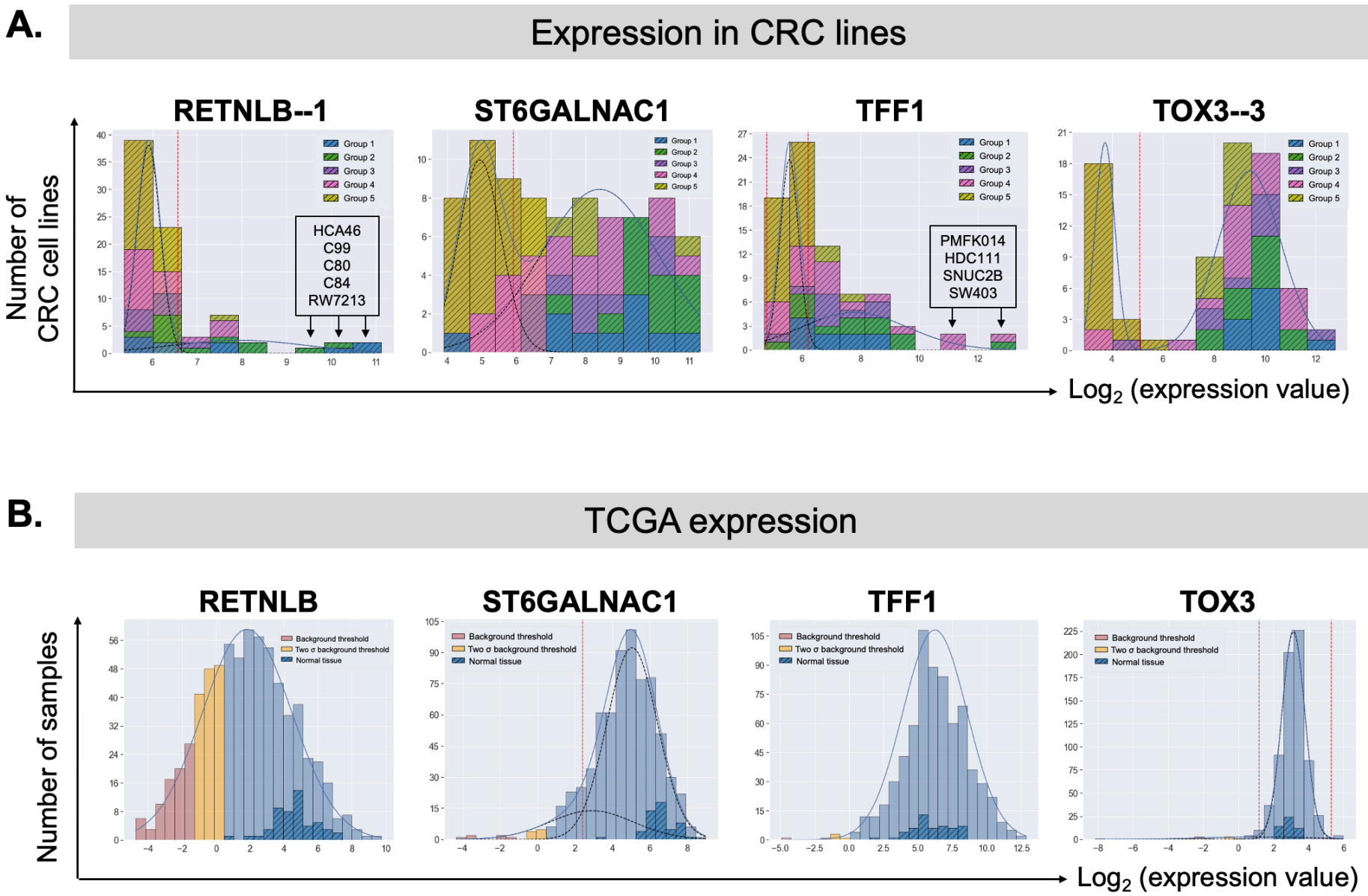
The mRNA expression levels of the indicated genes were investigated using our microarray data as shown in Figure 6.2A. Additionally, TCGA RNA-seq data was analysed to compare expression levels of the indicated genes related to normal colonic samples (Figure 6.2B).

Figure 6.2 (the following page) Expression analysis of significantly upregulated genes in MUC2- and TFF3-positive cell lines. (A) The GMMchi analysis of 16 genes that are positively correlated with goblet cells. CRC cell lines were classified based on the protein expression levels of MUC2 and TFF3, and labelled blue for group 1, green for group 2, purple for group 3, pink for group 4, and yellow for group 5 as discussed in Chapter 3. The genes identified with multiple probe sets are described with "--" followed by a number defining the probe set. When there are differences between probe sets this is due to splice variants. (B) The figures describe the corresponding TCGA RNA-seq data analysis of the same 16 upregulated genes in goblet cell-positive cell lines. These data are based on mRNA sequencing which does not distinguish splice variants.


A.
Expression in CRC lines
AGR2--1
AGR3
CFTR--2
CLRN3
Group 1
Group 2
Group 3
Group 4
Group 5
Number of CRC cell lines
Log2 (expression value)
B.
TCGA expression
AGR2
AGR3
CFTR
CLRN3
Normal tissue
Background threshold
Two σ background threshold
Normal tissue
Number of samples
Log2 (expression value)


A.
Expression in CRC lines
FAM3D
FCGBP
FOXA2--1
HEPH--2
Group 1
Group 2
Group 3
Group 4
Group 5
HDC111
C84
RW7213
Number of CRC cell lines
Log2 (expression value)
B.
TCGA expression
FAM3D
FCGBP
FOXA2
HEPH--2
Background threshold
Two σ background threshold
Normal tissue
Number of samples
Log2 (expression value)


A.
Expression in CRC lines
IL33
LGALS4
OLFM4
REG4--1
Group 1
Group 2
Group 3
Group 4
Group 5
PMFK014
HDC111
HDC9
Number of CRC cell lines
Log2 (expression value)
B.
TCGA expression
IL33
LGALS4
OLFM4
REG4
Background threshold
Two σ background threshold
Normal tissue
Number of samples
Log2 (expression value)


A.
Expression in CRC lines
RETNLB--1
ST6GALNAC1
TFF1
TOX3--3
Number of CRC cell lines
Group 1
Group 2
Group 3
Group 4
Group 5
HCA46
C99
C80
C84
RW7213
PMFK014
HDC111
SNUC2B
SW403
Log2 (expression value)
B.
TCGA expression
RETNLB
ST6GALNAC1
TFF1
TOX3
Number of samples
Background threshold
Two σ background threshold
Normal tissue
Log2 (expression value)

Among 16 upregulated genes, AGR2, AGR3, CFTR2, CLRN3, FAM3D, FCGBP, FOXA2, HEPH, LGALS4, ST6GALNAC1, REG4, IL33, and TOX3 genes displayed clear bimodal patterns in Bodmer microarray data, with higher expression in goblet cell positive cell lines. According to the TCGA dataset, all except TFF1, TOX3, and possibly HEPH-2 also showed differences between normal and cancer, but these differences do not show as clearly because of the normal cell contamination in the TCGA tumours.

Next, we performed a 2x2 association analysis between MUC2 and TFF3 expression levels versus all other genes as shown in Tables 6.1A and B. A list of the 10 most upregulated genes associated with MUC2 and TFF3 was selected for presentation (Table 6.1A). For MUC2, only REG4 and FCGBP appeared at a p-value cutoff of 0.05 following Bonferroni multiplicity correction. The 20 most upregulated genes for TFF3 (Table 6.1B), all had a p-value ≤ 0.05 after Bonferroni multiplicity correction. We excluded genes starting with "nan--" from both MUC2 and TFF3 analyses.

Associations with MUC2 expression

A) Top 10 genes positively correlated with *MUC2* expression

MUC2	+/+	+/-	-/+	-/-	Corrected p-value	R-value	Inclusion criteria
MESP2--1	3	0	1	12	7.14E-03	0.83	No
LRRC16B	6	1	1	19	1.59E-04	0.81	No
LOC645895 /// MYBL1	4	1	2	17	6.49E-03	0.65	No
C15orf38	5	1	3	20	2.53E-03	0.64	No
REG4_weighted	14	2	12	51	6.36E-07	0.59	Significant
FCGBP	14	2	13	50	1.28E-06	0.57	Significant
RGS12--4	3	2	2	30	1.17E-02	0.54	No
LOC648987	4	5	1	33	4.58E-03	0.53	No
ST6GALNAC1	16	0	24	39	4.66E-06	0.50	Significant

SPINK4	12	4	12	51	4.59E-05	0.49	Significant

Associations with TFF3 expression

B) Top 20 genes that are positively associated with *TFF3* expression (Bonferroni corrected p-value ≤ 0.05)

TFF3	+/+	+/-	-/+	-/-	Corrected p-value	R-value	Inclusion criteria
STARD10--1	39	6	2	32	9.11E-14	0.80	No
SPPL2B--4	24	4	2	21	2.31E-08	0.77	No
nan--9448	24	5	2	24	1.37E-08	0.75	No
BRP44--2	33	2	6	20	6.29E-09	0.73	No
STARD10--2	43	2	12	22	4.89E-09	0.65	Significant
LGALS4	42	3	11	23	1.28E-08	0.64	Significant
CYP39A1--1	22	12	1	30	7.92E-08	0.64	Significant
TDGF1 /// TDGF3	39	6	8	26	1.22E-08	0.64	No
nan--3808	33	4	7	19	5.47E-07	0.64	No
ALDH2	44	1	14	20	8.53E-09	0.63	Significant
AGR2_weighted	44	1	14	20	8.53E-09	0.63	Significant
LOC100134229	35	2	9	16	7.58E-07	0.63	Significant
PPP1R1B	42	3	12	22	4.50E-08	0.62	Significant
LOC113230	37	2	10	16	8.15E-07	0.62	Significant
CKMT1A /// CKMT1B	37	8	7	27	5.15E-08	0.61	No
CYP3A5_weighted	37	8	7	27	5.15E-08	0.61	No
CLRN3	41	4	11	23	4.54E-08	0.61	No
HEPH--2	41	4	11	23	4.54E-08	0.61	No
CDHR5--1	20	10	2	26	2.34E-06	0.61	Significant
PRR15L	44	1	15	19	3.20E-08	0.61	Significant

Table 6.1 Top upregulated genes associated with MUC2 and TFF3 expression in Bodmer microarray data. (A) describes the top 10 genes positively linked to MUC2 expression. (B) illustrates the top 20 upregulated genes for TFF3 with the p-value corrected by Bonferroni method. Genes were sorted based on the R value from highest to lowest.

Among the 16 genes analysed in Figure 6.2, FCGBP, REG4--2, and ST6GALNAC1 showed a strong correlation with MUC2 expression (Table 6.1A). Our experimental

data in LS180 cell line showed that only a small number of cells stain with REG4 similar to MUC2/TFF3 (Figure 6.3).

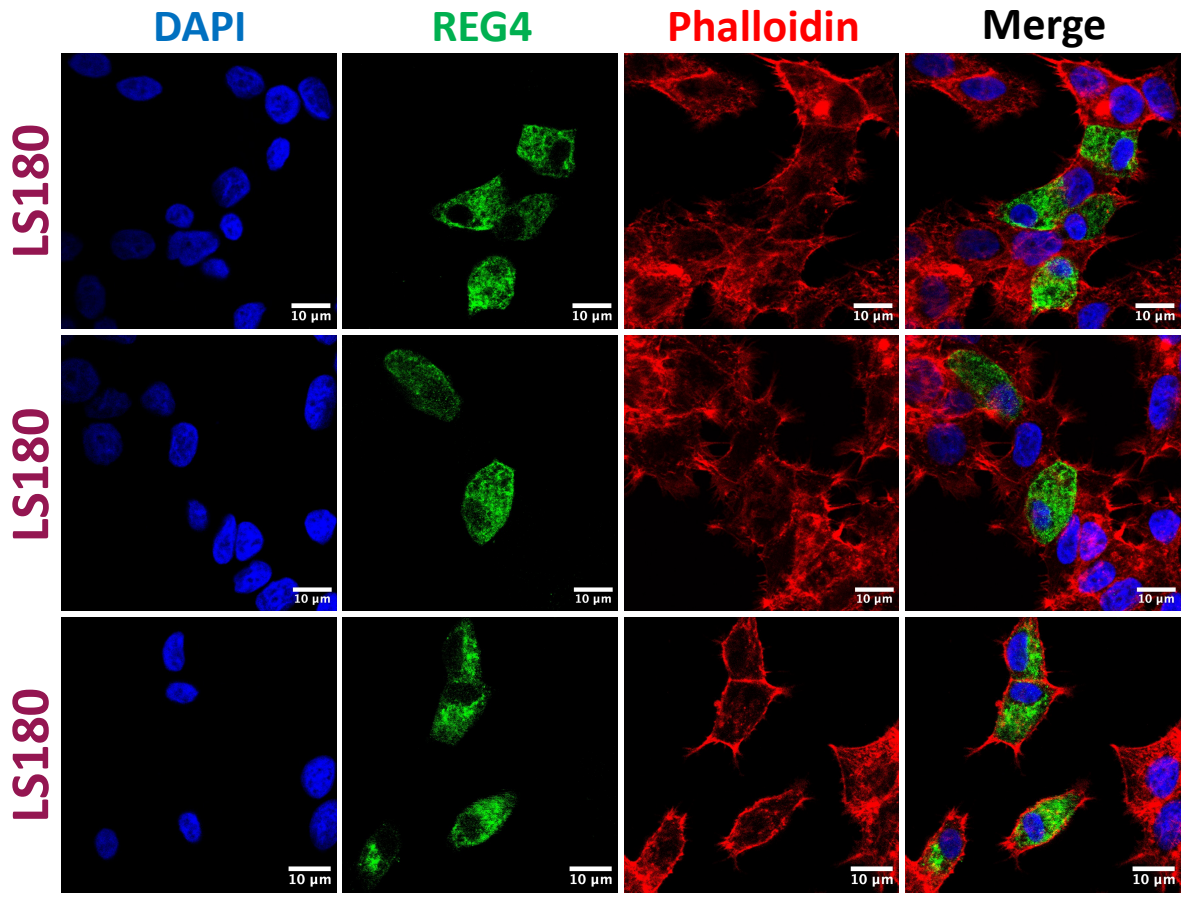


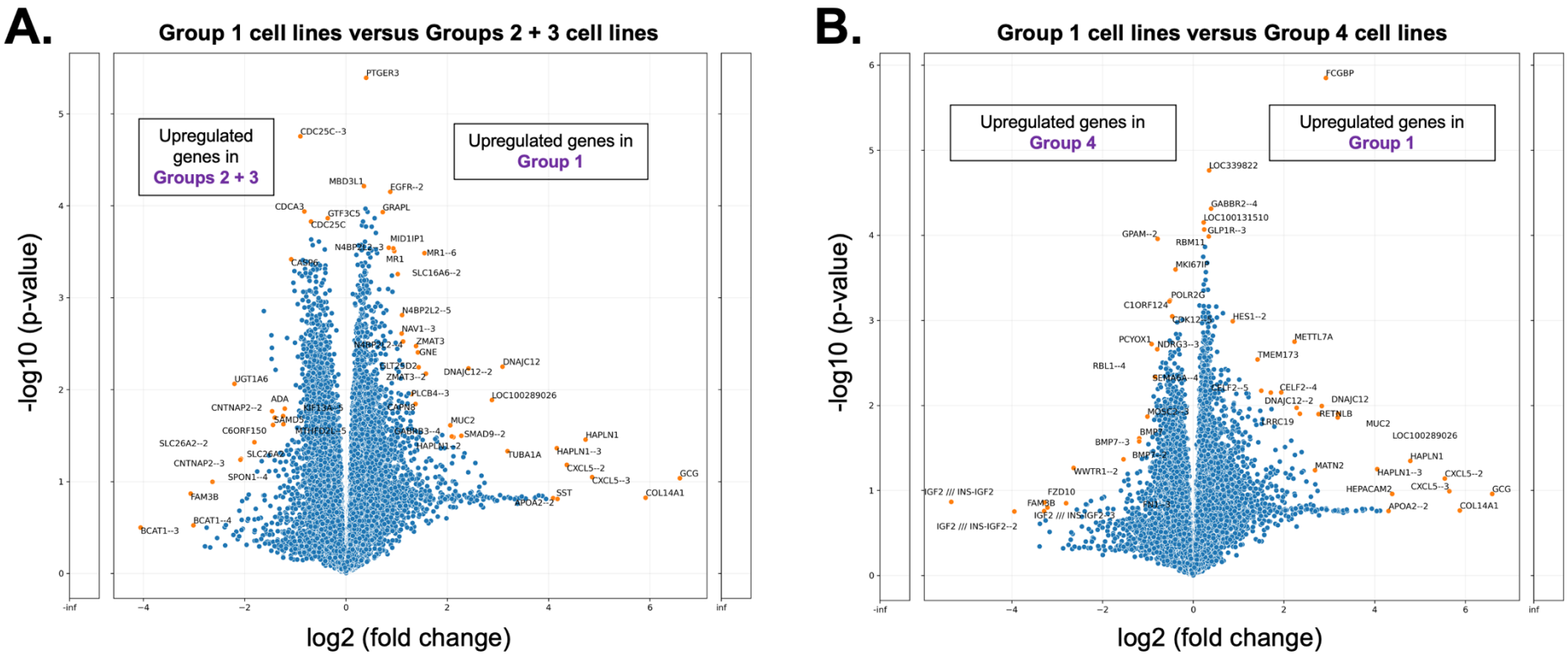
Figure 6.3 Representative images of REG4 staining in LS180 cell line. Cells were incubated over three days and fixed before staining with REG4 (1:250), Phalloidin (1:1000), and DAPI (blue, 1:10000). Scale bar: 50 µm.

The 2x2 analysis suggested that 14/16 (87.5%) of MUC2-producing cell lines express FCGBP, however 13/27 (48.1%) of FCGBP-positive cell lines do not express MUC2. There is a possibility that MUC2 forms a subpopulation under FCGBP expression. We also had a similar finding in TFF3, where almost all MUC2-positive cells express TFF3, but not vice versa. These results may indicate that FCGBP and TFF3 are necessary before MUC2 is matured and secreted into the lumen, thus requiring further

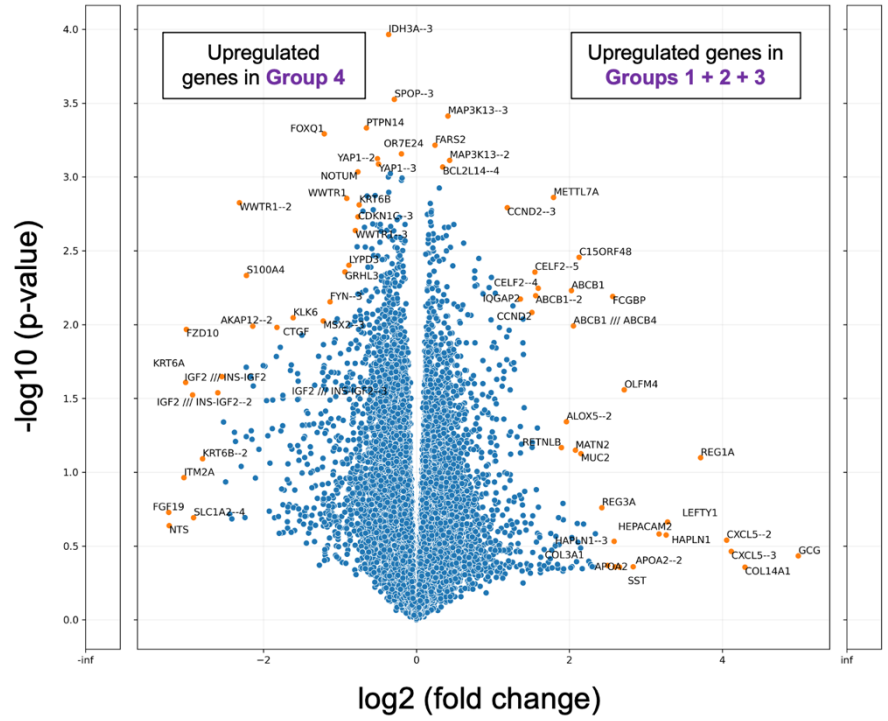
investigation. When we perform the Bonferroni correction with a p-value < 0.05, only REG4 and FCGBP appeared as only most upregulated gene for MUC2 association.

As a next step, we investigated the pattern of up- and down-regulated genes with the protein levels using our 5-group classification using volcano plots. The aim was to identify genes that contribute to differences between various groups. We compared group 1 versus groups 2 and 3; group 1 versus group 4; MUC2-positive groups (groups 1-3) versus group 4; MUC2/TFF3-positive groups (groups 1-4) versus group 5 (Figure 6.4A-D). Highlighted genes are represented in Appendix Table 5, along with their fold changes and p values (Appendix Table 6). In Appendix Table 6, genes were sorted according to their R values. The upregulated genes in groups 1+2+3+4 versus 5 were largely those we have already associated with goblet cell differentiation. For 1+2+3 versus 4, OLFM4 and perhaps HEPACAM2 are worth further investigating as also suggested by group 1 versus 4.

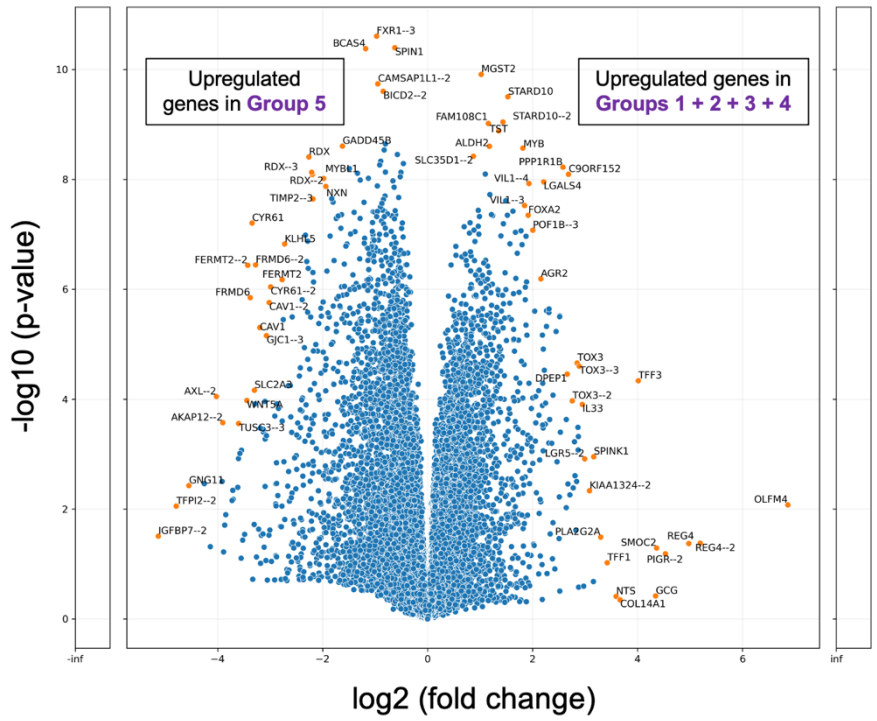
Figure 6.4 (the following page) Differentially expressed genes in various groups originated based on the 5 protein expression levels of MUC2 and TFF3. The x-axis of the volcano plot represents the log2 of the fold change, while the y-axis displays the negative logarithm (base 10) of the p-value. (A) compares high MUC2-expressing cell lines versus medium/low MUC2-expressing cell lines. (B) describes the comparison between groups 1 and 4. (C) depicts up- and down-regulated genes in MUC2-positive groups (groups 1-3) versus group 4 cell lines. (D) represents differentially expressed genes in both MUC2- and TFF3-positive cell lines versus group 5.


A.
Group 1 cell lines versus Groups 2 + 3 cell lines
Upregulated genes in Groups 2 + 3
Upregulated genes in Group 1
-log10 (p-value)
log2 (fold change)
B.
Group 1 cell lines versus Group 4 cell lines
Upregulated genes in Group 4
Upregulated genes in Group 1
-log10 (p-value)
log2 (fold change)

C.

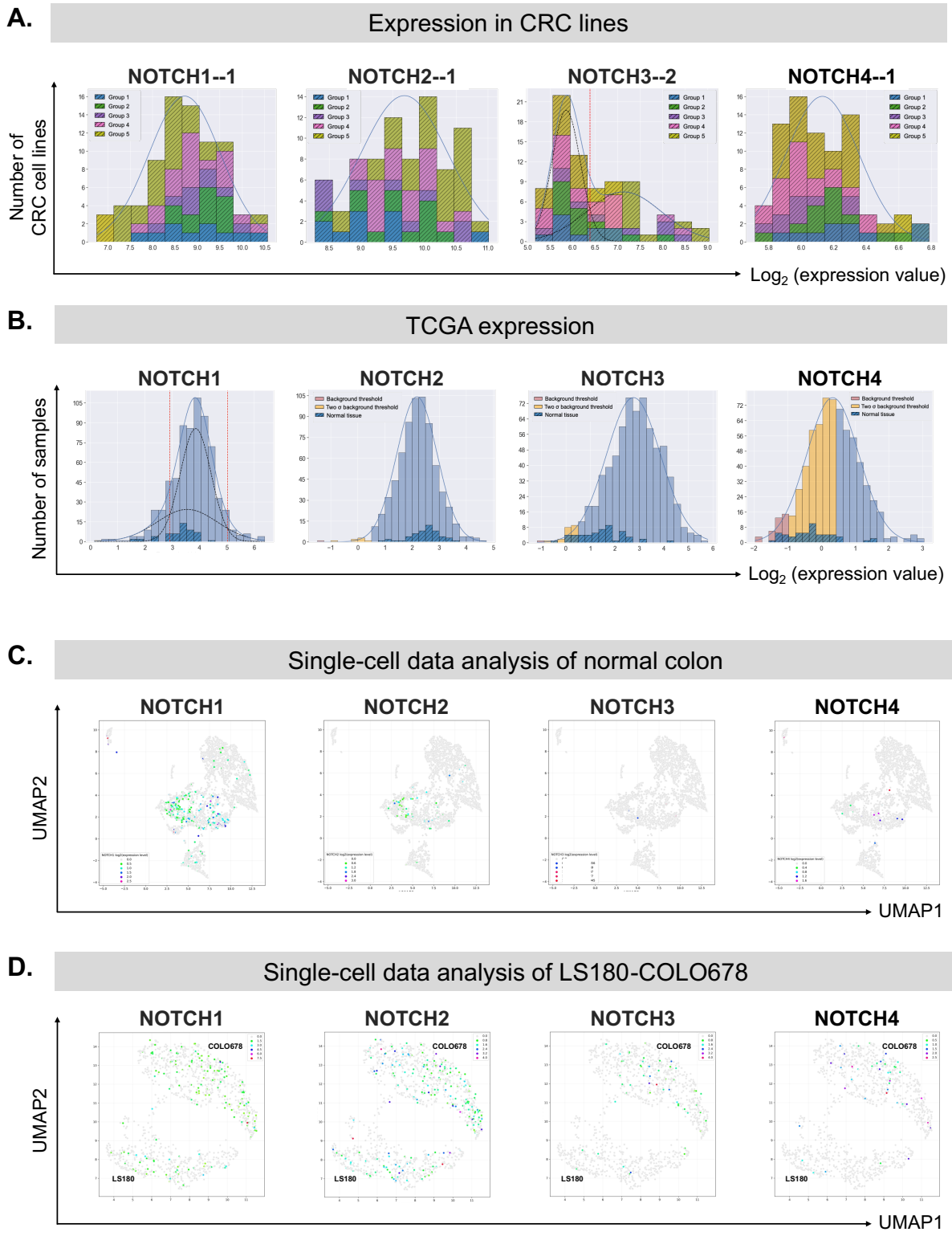

Groups 1 + 2 + 3 cell lines versus Group 4 cell lines
Upregulated genes in Group 4
Upregulated genes in Groups 1 + 2 + 3
-log10 (p-value)
log2 (fold change)

D.


Group 1 + 2 + 3 + 4 cell lines versus Group 5 cell lines
Upregulated genes in Group 5
Upregulated genes in Groups 1 + 2 + 3 + 4
-log10 (p-value)
log2 (fold change)

6.2.2 Effect of Notch pathway on goblet cell differentiation

We analysed four Notch receptors (*Notch 1*, *2*, *3*, and *4*) and five Notch ligands (*JAG* 1, 2 and *DLL* 1, 3, 4) as described in Figure 6.5.



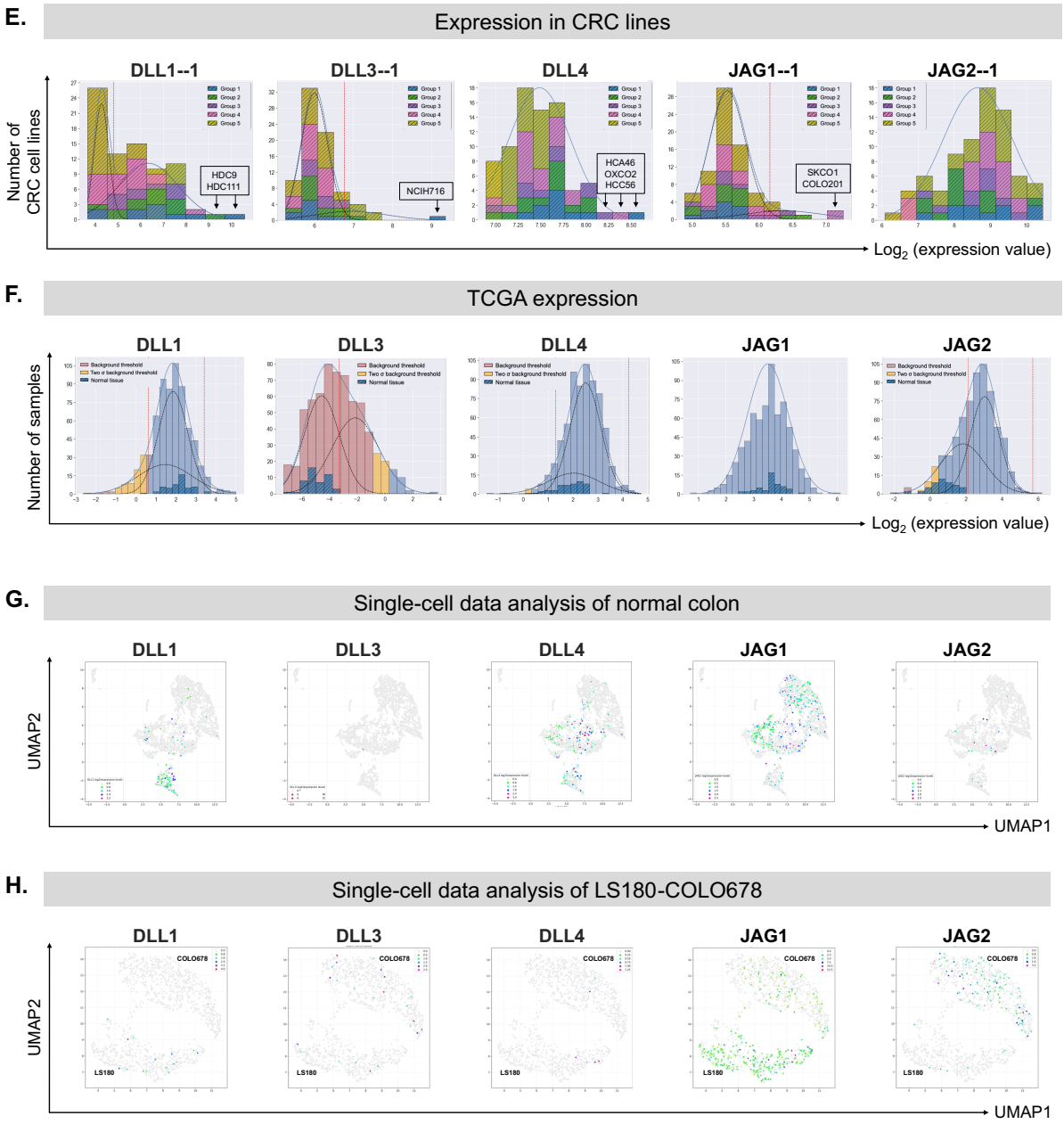


Figure 6.5 Data analysis of Notch receptors and ligands. Figures A-D show expression analysis of Notch receptors based on (A) our microarray data, (B) TCGA dataset, (C) single-cell RNA-seq data of normal colon, (D) single-cell data of LS180-COLO678 cell lines. Figures (E-H) represent gene expression analysis of Notch ligands with a similar order as Figures A-D.

Among Notch receptors, only Notch3 exhibits bimodal distribution in our microarray data of CRC cell lines. Notch1 and Notch2 shows positively unimodal pattern, whereas Notch4 demonstrates negatively unimodal distribution. According to scRNA-seq data

from normal colon tissue, these receptors are primarily expressed in the cluster of stem cells and progenitors. Goblet-cell negative cell line, COLO678 exhibits high expression of these receptors, suggesting its stem-like phenotype. Regarding the Notch ligands, DLL1 and DLL4 stand out more interesting due to their bimodal pattern based on GMMchi analysis, as well as mainly expression in the goblet cell cluster.

We investigated the role of the Notch pathway in goblet cell differentiation by using dibenzazepine (DBZ), a γ-secretase inhibitor. The results revealed a significant increase in the number of MUC2 and TFF3-positive cells following 48 hours of DBZ administration in LS180 cell line (Figure 6.6A). qRT-PCR analysis revealed that ATOH1, a key transcription factor for goblet cell differentiation, substantially increased, whereas HES1, which is known to function as a negative regulator of ATOH1, showed a significant reduction in gene expression (Figure 6.6B). DLL4, a Notch ligand, was also upregulated following DBZ treatment (Figure 6.6B).

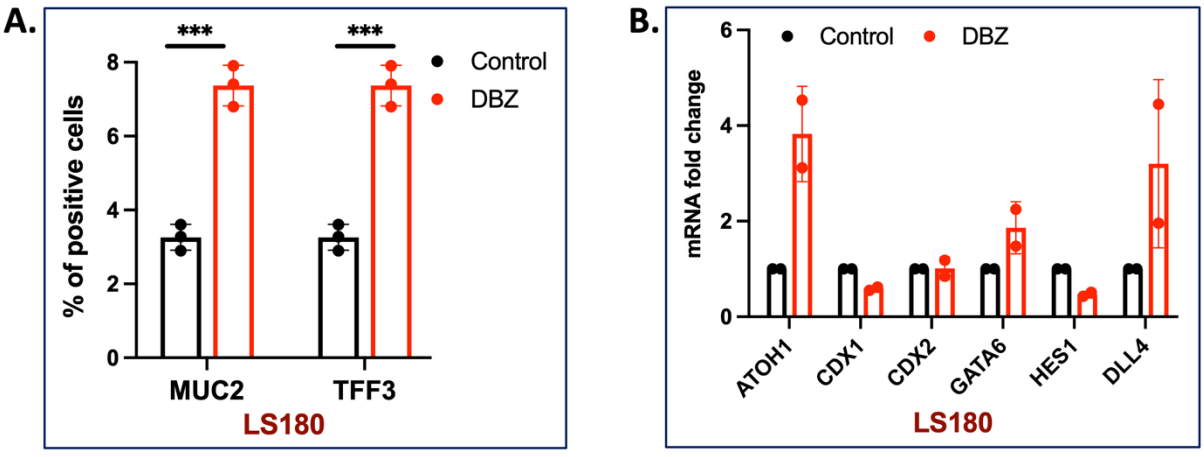
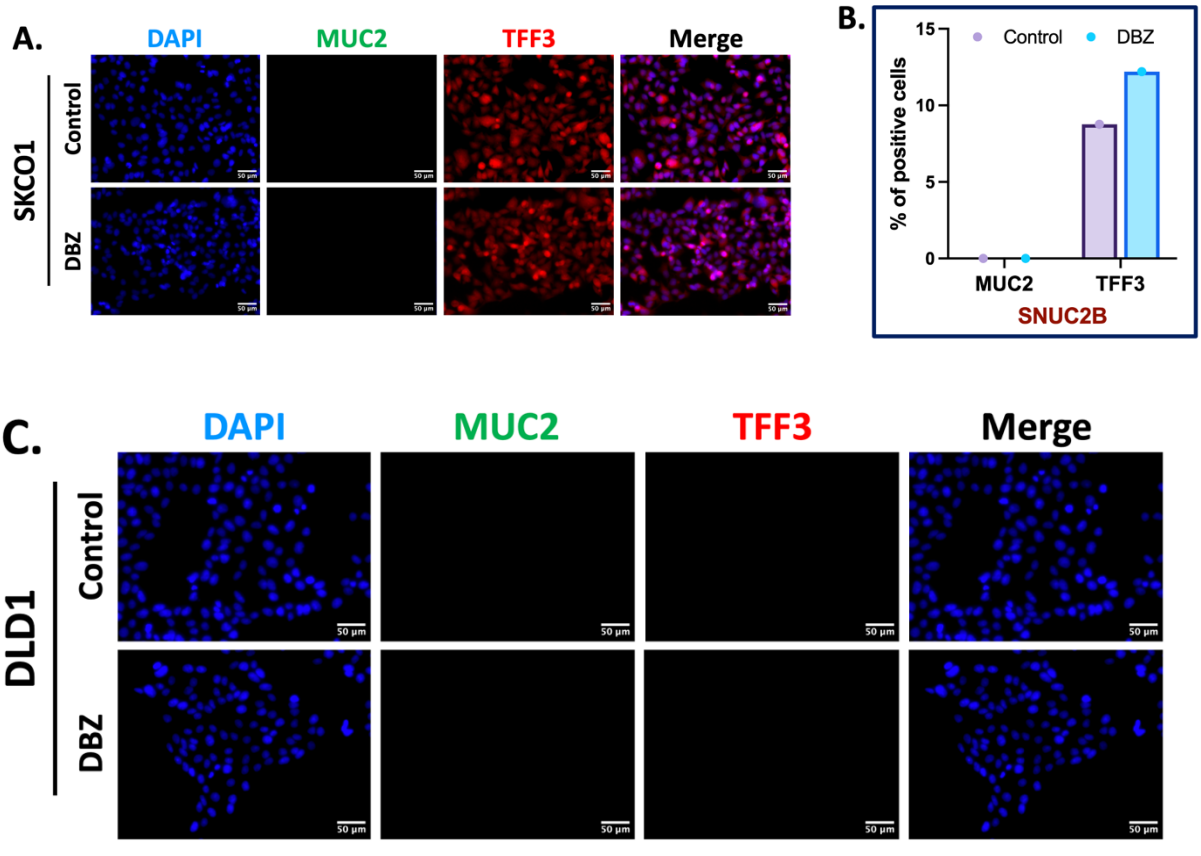


Figure 6.6 Assessment of the effect of the Notch pathway on the LS180 cell line. (A) Cells were treated with DBZ (200 nM), a γ-secretase inhibitor, for 48 hours, then fixed and stained with MUC2, TFF3, and DAPI. The number of MUC2- and TFF3-positive cells were quantified using CellProfiler for nuclei, and manually for goblet cells. (B) represents qRT-PCR results of ATOH1, CDX1, CDX2, GATA6, HES1, and DLL4. Statistical analysis was done

using a t-test to compare DBZ relative to the control group in three independent experiments. ***p ≤ 0.001.

We also assessed the effect of Notch pathway in other cell lines belong to Group 4 and Group 5. Inhibition of Notch pathway didn't induce MUC2 expression in SKCO1 cell line belonging to group 4 (Figure 6.7A). Another group 4 cell line, SNUC2B exhibited a slight increase in protein expression levels of TFF3 (Figure 6.7B). It should be noted that this experiment was conducted once, and further replications are necessary to validate the outcome. There was no notable difference in terms of MUC2 and TFF3 expression levels in group 5 cell lines (Figure 6.7C).



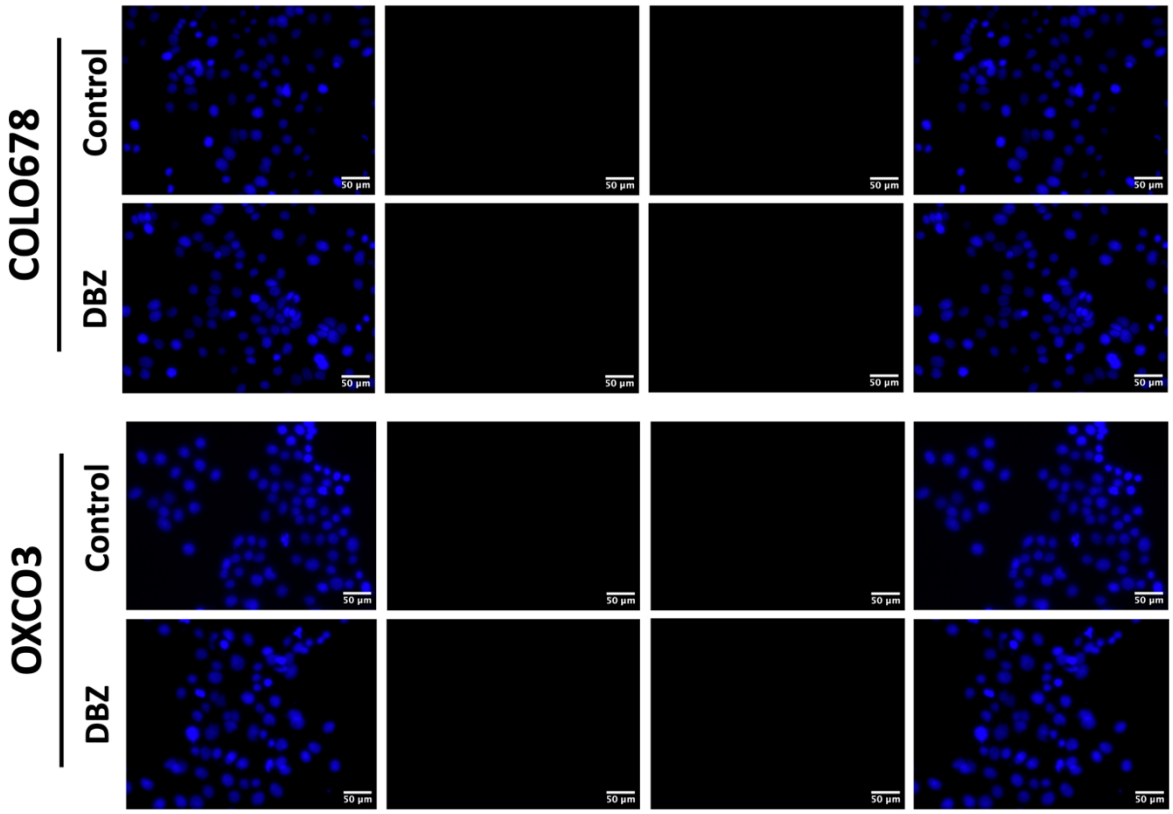


Figure 6.7 Effect of Notch pathway on the cell lines belonging to groups 4 and 5. (A) represents the effect of Notch pathway in SKCO1 cell line. SKCO1 cells were treated with DBZ (200 nM) prior to being fixed and stained with MUC2 (1:250), TFF3 (1:500), and DAPI. (B) shows the percentage of MUC2- and TFF3-positive cells in SNUC2B cell line. CellProfiler was used to quantify nuclei, and the number of TFF3-positive cells were defined manually. (C) displays the impact of Notch pathway on goblet cell differentiation in group 5 cell lines. Staining was done as shown in (A). Scale bar: 50 µm.

We also studied whether cell density influences goblet cell differentiation in bulk cultures. To test that, we plated various numbers of LS180 cells in a 96-well plate and fixed them on day 4 after seeding. According to our results, cell density didn't affect goblet cell differentiation in the LS180 cell line (Figure 6.8). It is most likely due to plating the cells in bulk culture where signalling pathways are already established.

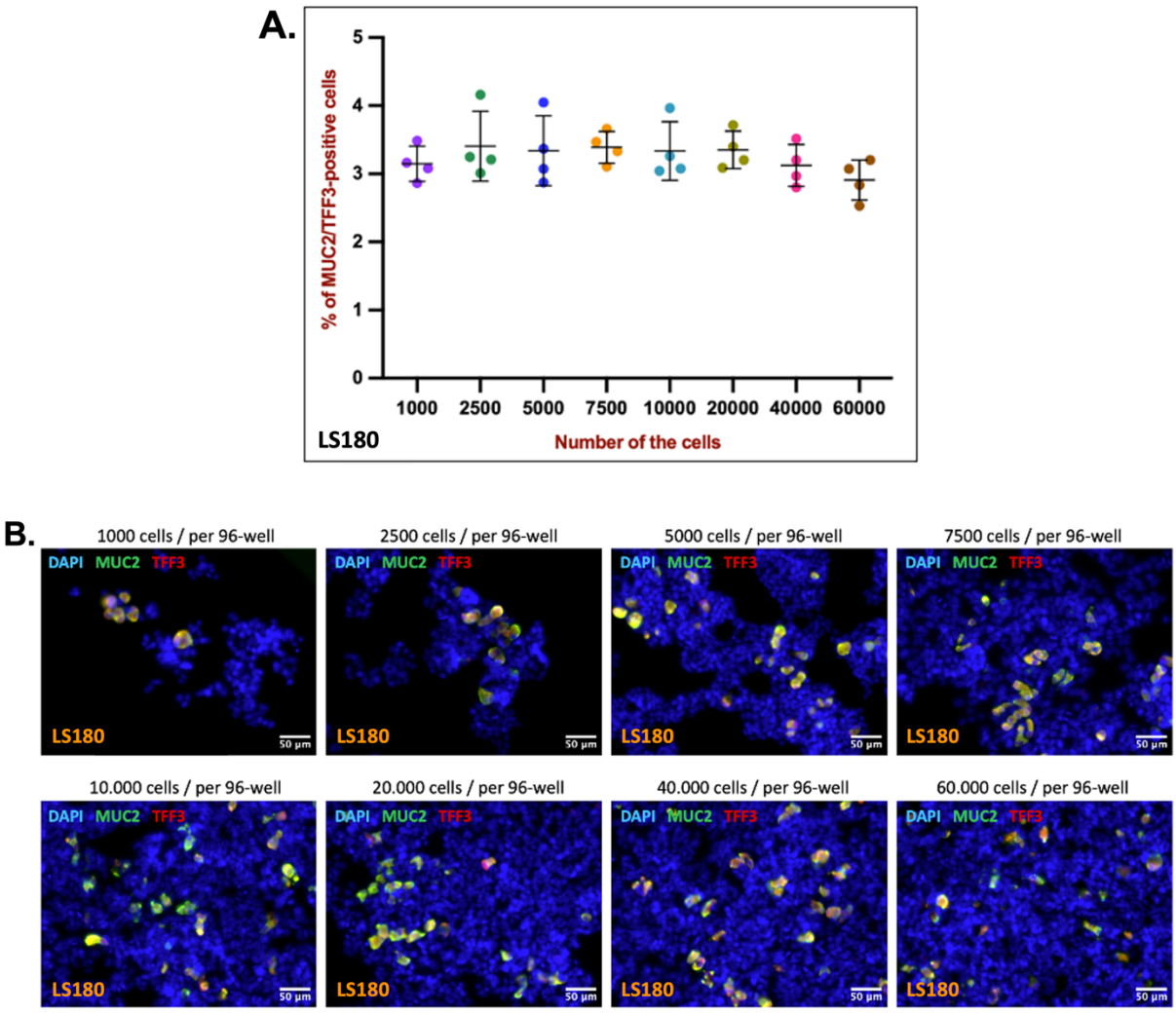


Figure 6.8 Effect of cell confluency on goblet cell differentiation in LS180 cell line. (A) displays the percentage of MUC2- and TFF3-positive cells according to indicated number of cultured cells. In the figure, the x-axis displays the indicated number of cells seeded on the first day of culturing, whereas the y-axis represents the percentage of MUC2- and TFF3-positive cells. LS180 cells were fixed after 4 days of culturing and then stained with MUC2 (1:250), TFF3 (1:500), and DAPI. (B) demonstrates representative images from each condition. Scale bar: 50 µm.

6.2.3 Classification of CRC cell lines based on expression of various transcription factors

Next, we explored how the expression of transcription factors correlates with our cell lines classification. CRC cell lines belonging to groups 1-5 were classified according to mRNA expression levels of ATOH1, SPDEF, CDX1, CDX2, GATA6, HES1, and ETS2 transcription factors, which appear to be associated with goblet cell differentiation (Table 6.2). The most striking result in Table 6.2 is the overall high frequency of lack of expression of the transcription factors ATOH1, SPDEF and CDX1 in the group 5 MUC2-TFF3- cell lines. In contrast CDX1, CDX2, GATA6 and ETS2 are nearly all positive in groups 1,2 and 3, which are the groups that express MUC2 and TFF3, though at varying levels. HES1 seems to be expressed in nearly all the cell lines. Group 4 has a lower number of cell lines expressing ATOH1, SPDEF1 and ETS2 suggesting that this combination may have something to do with the group 4's distinctive expression of TFF3 without MUC2. These results overall provide strong further evidence for the key role of these factors in goblet cell differentiation and the control of expression of TFF3 and MUC2. They furthermore suggest that it is combinations of these factors that may be the determinants of the differentiation, and variation of the combinations that explains the 5 groups found in the CRC cell lines.

Group 1.

Cell line	% MUC2	% TFF3	ATOH1	SPDEF--1	CDX1	CDX2	GATA6--1	HES1--2	ETS2--1
C99	93.3	100.0	negative	negative	positive	positive	positive	positive	positive
CL40	94.4	100.0	negative	negative	positive	positive	positive	positive	positive
HCA46	100.0	100.0	positive	negative	positive	positive	positive	positive	positive
HDC9	14.3	23.8	positive	positive	positive	positive	positive	positive	negative
LS411	23.5	99.4	positive	positive	negative	positive	positive	positive	positive
LS513	22.6	29.8	negative	negative	positive	positive	positive	positive	positive
NCIH508	16.1	16.1	positive	positive	positive	positive	positive	positive	positive
NCIH716	98.8	99.4	positive	positive	negative	positive	negative	positive	negative
PCJW	21.2	21.2	negative	negative	positive	positive	positive	positive	positive
RW7213	30.4	30.4	negative	negative	positive	positive	positive	positive	positive

Group 2.

Cell line	% MUC2	% TFF3	ATOH1	SPDEF--1	CDX1	CDX2	GATA6--1	HES1--2	ETS2--1
C125PM	6.0	6.0	negative	negative	positive	positive	positive	positive	positive
C80	6.6	6.6	positive	positive	positive	positive	positive	positive	positive
C84	6.9	7.6	positive	positive	positive	positive	positive	positive	positive
HDC111	7.2	7.2	positive	positive	negative	negative	positive	positive	positive
HDC114	2.5	1.8	negative	negative	positive	positive	positive	positive	positive
HDC143	4.1	4.1	positive	positive	positive	positive	positive	positive	positive
HDC73	3.5	6.3	positive	positive	positive	positive	positive	positive	positive

HT29	1.9	2.4	positive	positive	negative	negative	positive	positive	positive
LS174T	3.3	3.3	negative	negative	positive	positive	positive	positive	positive
LS180	3.4	3.4	positive	negative	positive	positive	positive	positive	positive
T84	4.4	4.4	positive	negative	positive	positive	positive	positive	positive
VACO4A	3.8	3.8	negative	negative	positive	positive	positive	positive	positive
WIDR	3.4	4.1	negative	positive	negative	negative	positive	positive	positive

Group 3.

Cell line	% MUC2	% TFF3	ATOH1	SPDEF--1	CDX1	CDX2	GATA6--1	HES1--2	ETS2--1
CCK81	0.8	0.8	positive	negative	positive	positive	positive	positive	positive
CW2	1.3	1.5	positive	negative	positive	positive	positive	positive	positive
CX1	1.0	1.7	positive	negative	negative	negative	positive	positive	positive
HCC56	1.0	1.0	negative	negative	positive	positive	positive	positive	positive
HDC54	0.7	1.2	negative	negative	positive	positive	positive	positive	positive
HDC57	1.2	1.3	negative	negative	positive	positive	positive	positive	positive
JHCOLOY1	0.5	1.5	negative	negative	positive	positive	positive	positive	positive
KM20L2	0.6	1.2	negative	positive	negative	positive	positive	positive	positive
LOVO	0.8	0.3	negative	negative	positive	positive	positive	positive	positive
LS1034	0.9	1.1	negative	negative	positive	positive	positive	positive	positive
SW1222	0.7	1.0	negative	positive	positive	positive	positive	positive	positive
SW1463	0.5	0.7	NA	NA	NA	NA	NA	NA	NA
SW948	1.4	1.4	negative	negative	positive	positive	positive	positive	positive
VACO4S	1.3	1.3	positive	negative	positive	positive	positive	positive	positive

Group 4.

Cell line	% MUC2	% TFF3	ATOH1	SPDEF--1	CDX1	CDX2	GATA6--1	HES1--2	ETS2--1
C106	0.0	100.0	negative	negative	positive	positive	positive	positive	positive
C70	0.0	100.0	negative	negative	positive	positive	positive	positive	negative
CL14	0.0	100.0	negative	negative	positive	positive	positive	positive	positive
COLO201	0.0	100.0	negative	negative	negative	negative	negative	positive	negative
COLO205	0.0	100.0	negative	negative	negative	negative	positive	positive	negative
COLO206	0.0	100.0	negative	negative	negative	negative	positive	positive	negative
GP2D	0.0	100.0	negative	negative	positive	positive	positive	positive	positive
GP5D	0.0	100.0	negative	negative	positive	positive	positive	positive	positive
HDC142	0.0	100.0	negative	positive	positive	positive	positive	positive	negative
HT55	0.0	100.0	negative	negative	positive	positive	positive	positive	positive
NCIH548	0.0	100.0	negative	negative	negative	negative	positive	positive	negative
PMFK014	0.0	97.6	negative	negative	negative	negative	positive	positive	positive
RCM1	0.0	100.0	negative	negative	positive	positive	positive	positive	negative
RW2982	0.0	100.0	negative	negative	positive	positive	positive	positive	positive
SKCO1	0.0	100.0	negative	positive	positive	positive	positive	positive	positive
SNUC1	0.0	100.0	positive	positive	positive	positive	positive	positive	positive
SW403	0.0	100.0	positive	negative	positive	positive	positive	positive	positive
SW620	0.0	100.0	negative	negative	negative	negative	positive	positive	negative
TT1TKB	0.0	97.9	NA	NA	NA	NA	NA	NA	NA
VACO429	0.0	100.0	negative	negative	negative	negative	positive	positive	positive

HCT8	0.8	100.0	negative	negative	positive	positive	positive	positive	negative
LIM1863	0.6	100.0	positive	negative	positive	positive	positive	positive	positive
OXCO2	0.4	100.0	negative	positive	positive	positive	positive	positive	positive
SNU1235	2.9	99.0	NA	NA	NA	NA	NA	NA	NA
VACO10MS	2.9	99.0	positive	negative	positive	positive	positive	positive	positive
VACO5	2.3	100.0	negative	negative	negative	negative	positive	positive	positive
SNUC2B	0.0	18.7	negative	positive	negative	negative	positive	positive	negative

Group 5.

Cell line	% MUC2	% TFF3	ATOH1	SPDEF--1	CDX1	CDX2	GATA6--1	HES1--2	ETS2--1
C10	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
C10A	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
C10S	0.0	0.0	negative	negative	negative	negative	negative	positive	positive
C2BBE1	0.0	0.0	negative	negative	negative	positive	positive	positive	negative
C32	0.0	0.0	negative	negative	positive	positive	positive	positive	positive
C75	0.0	0.0	negative	negative	positive	positive	positive	positive	positive
CACO2	0.0	0.0	negative	positive	positive	positive	positive	positive	negative
CAR1	0.0	0.0	negative	negative	negative	negative	positive	positive	negative
CC20	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
CCO7	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
COLO320DM	0.0	0.0	negative	negative	negative	positive	negative	negative	negative
COLO320HSR	0.0	0.0	positive	negative	negative	positive	negative	negative	negative
COLO678	0.0	0.0	positive	negative	negative	negative	negative	positive	negative
DLD1	0.0	0.0	negative	negative	negative	positive	positive	positive	positive

HCA7	0.0	0.0	negative	negative	negative	negative	positive	positive	negative
HCC2998	0.0	0.0	negative	negative	positive	positive	positive	positive	positive
HCT116	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
HCT15	0.0	0.0	negative	negative	negative	positive	positive	positive	positive
HDC135	0.0	0.0	positive	negative	negative	negative	negative	positive	negative
HDC8	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
HDC82	0.0	0.0	negative	negative	positive	positive	positive	positive	positive
HRA19	0.0	0.0	negative	negative	positive	positive	positive	positive	positive
ISCEROL	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
LIM1215	0.0	0.0	NA	NA	NA	NA	NA	NA	NA
LIM2405	0.0	0.0	NA	NA	NA	NA	NA	NA	NA
LS123	0.0	0.0	negative	negative	negative	negative	positive	positive	negative
NCIH747	0.0	0.0	negative	negative	negative	negative	positive	positive	negative
OUMS23	0.0	0.0	negative	negative	negative	positive	positive	positive	negative
OXCO1	0.0	0.0	negative	negative	negative	negative	positive	positive	negative
OXCO3	0.0	0.0	negative	negative	negative	positive	negative	positive	negative
RKO	0.0	0.0	negative	negative	negative	negative	negative	negative	negative
SW1116	0.0	0.0	negative	negative	positive	positive	positive	positive	positive
SW1417	0.0	0.0	negative	negative	negative	positive	negative	positive	negative
SW48	0.0	0.0	negative	negative	negative	positive	negative	positive	negative
SW480	0.0	0.0	negative	negative	negative	negative	negative	positive	negative
SW837	0.0	0.0	negative	negative	negative	negative	negative	positive	negative

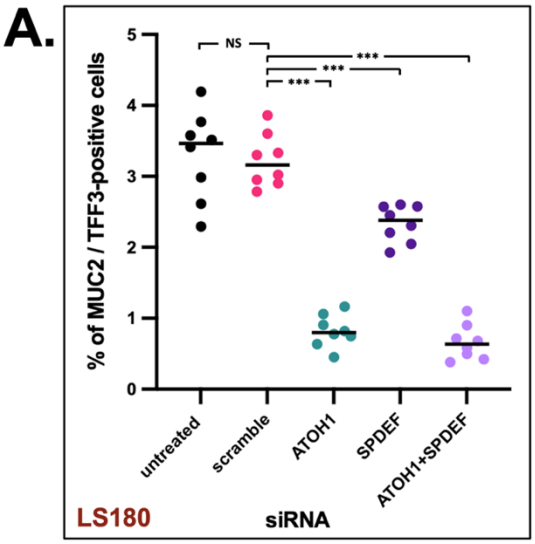
Table 6.2 Gene expression analysis of ATOH1, SPDEF, CDX1, CDX2, GATA6, HES1, and ETS2 transcription factors in a panel of CRC cell lines. The percentage of MUC2 and TFF3 was identified based on the quantification analysis of protein expression levels. Cell lines were classified into the 5 groups as before and characterised by mRNA expression levels of various transcription factors, including ATOH1, SPDEF, CDX1, CDX2, GATA6, HES1, and ETS2. The threshold values for separating positive/high and negative/low cell lines in terms of ATOH1, SPDEF, CDX1, CDX2, GATA6, HES1, and ETS2 expression are $2^{5.97}$, $2^{8.44}$, $2^{8.80}$, $2^{7.15}$, $2^{7.86}$, $2^{6.10}$, and $2^{10.06}$, respectively. “NA” indicates that mRNA expression data are not available for the cell lines, namely SW1463, TT1TKB, SNU1235, LIM1215, and LIM2405.

6.2.4 Effect of different factors on goblet cell differentiation

6.2.4.1 Effects of ATOH1 and SPDEF knockdown and overexpression

To investigate the effects of ATOH1 and SPDEF in the control of goblet cell differentiation in our cell lines, we started with LS180. Cells were transfected with siRNAs against *ATOH1* and *SPDEF* and cultured for 3 days. Next, the number of MUC2/TFF3-positive cells was quantified.

Knockdown of *ATOH1* resulted in a dramatic decrease in goblet cell differentiation, whereas downregulation of *SPDEF* had a more modest effect, suggesting a primary role for ATOH1 in the LS180 cell line (Figure 6.9A). Knockdown of *ATOH1* and *SPDEF* together did have an apparently modest, though not statistically significant, impact on goblet cell differentiation when compared to *ATOH1* knockdown alone (Figure 6.9A). The efficiency of the knockdown was checked by qRT-PCR (Figure 6.9B). The knockdown of *ATOH1* and *SPDEF* appeared to influence each other's levels, suggesting mutual interactions. The effect of *ATOH1* and *SPDEF* knockdown was also checked in LOVO (group 3), RW7213 (group 1), and SW1222 (group 3) CRC cell lines, all of which showed a significant impact on goblet cell differentiation (Figure 6.9C).



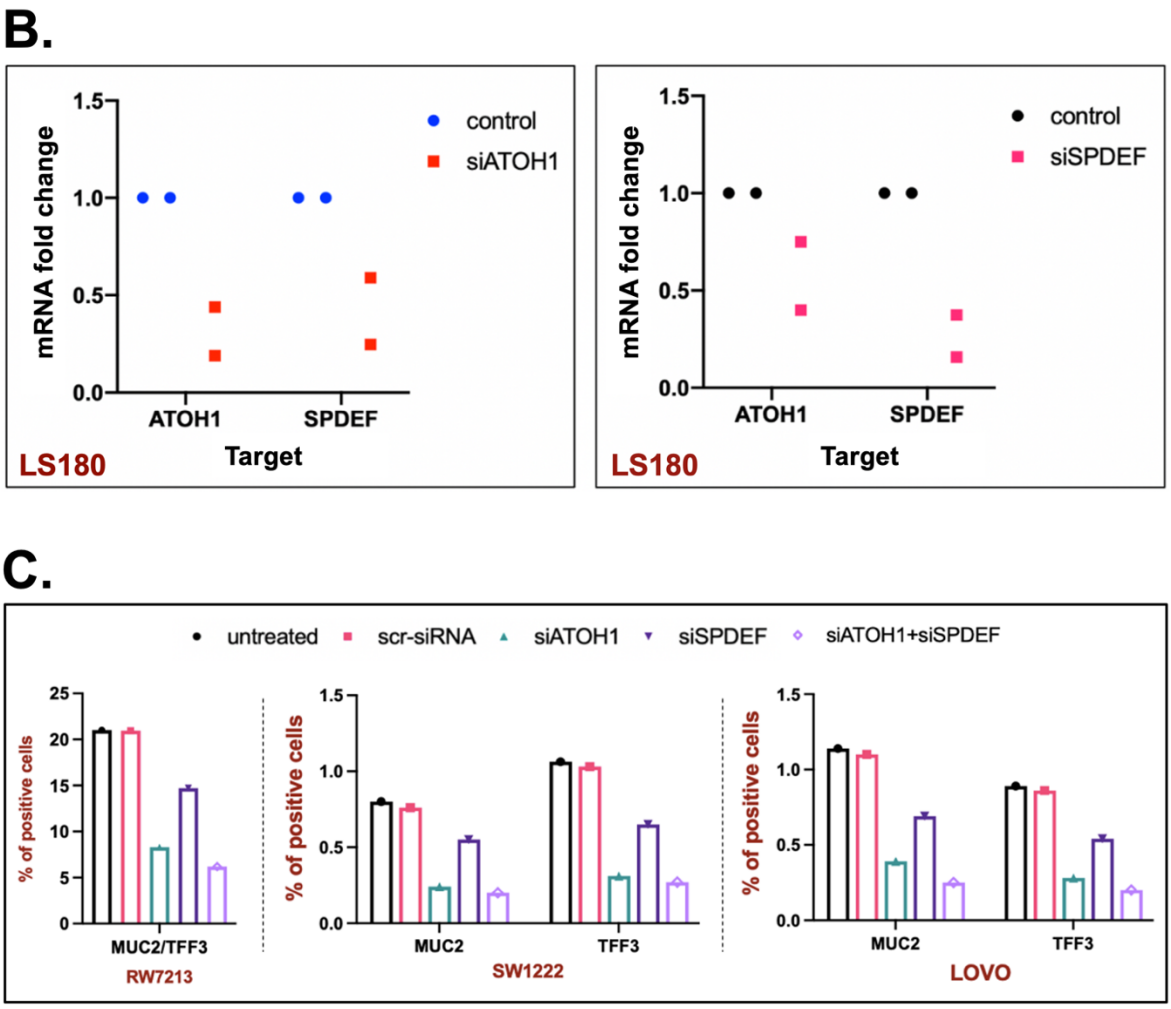


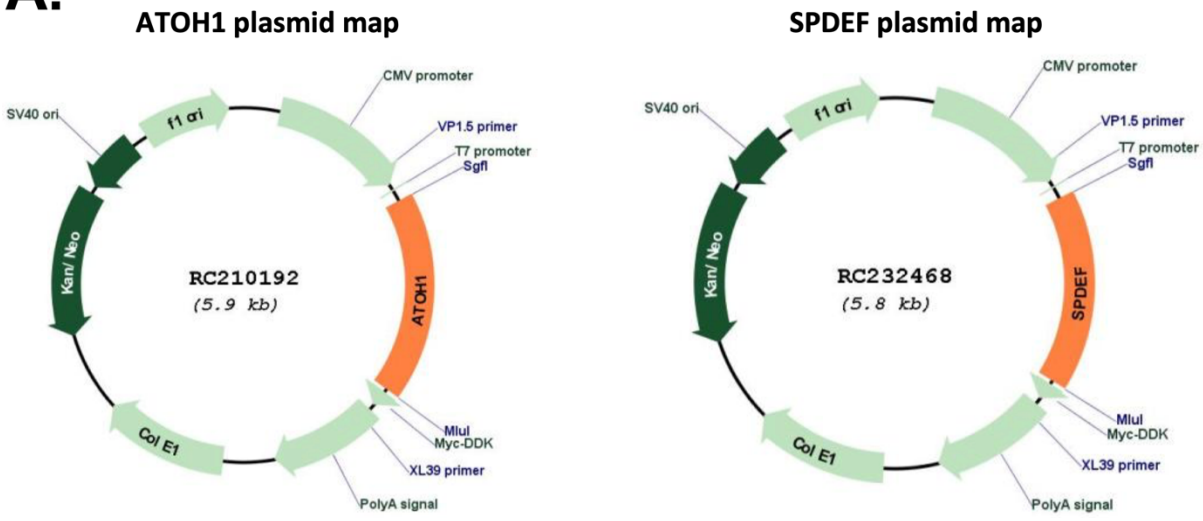
Figure 6.9. Knockdown of *ATOH1* and/or *SPDEF* in LS180 and other cell lines led to downregulation of MUC2 and TFF3 expression. (A) LS180 cells were transiently transfected with siRNAs against *ATOH1* and *SPDEF* and analysed three days post-transfection. The percentage of MUC2 and TFF3 expression was quantified as in Figure 3.5A. Statistical analysis was done using a t-test for the different functional transfections compared to scrambled siRNA (scr-siRNA) in eight independent experiments. ***p ≤ 0.001, 'NS' (p > 0.05) indicates that the comparison was not significant. (B) The efficiency of knockdown was analysed by RT-qPCR. GAPDH was used as an internal control. (C) The effects of knocking down of *ATOH1* and *SPDEF* were analysed in RW7213, SW1222, and LOVO cell lines.

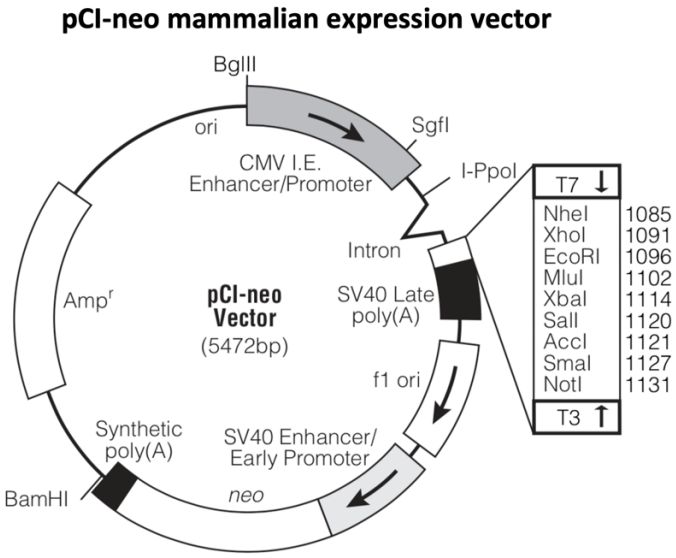
According to scRNA-seq data analysis of the normal colon, ATOH1 and SPDEF are predominantly expressed in the goblet cell cluster (Figure 5.3C). ATOH1 is a transcription factor that seems to be expressed at low levels overall in many cell lines, but probably at detectable levels in only a subset of cells. Thus, the categorisation into

“negative” cell lines needs to be interpreted with caution. This emphasises the question, namely is this a problem of *ATOH1* detection or are there additional factors, other than ATOH1, that control the expression of MUC2 in those cell lines?

It was next explored whether ectopic expression of ATOH1 and SPDEF could induce goblet cell differentiation in MUC2- and TFF3-negative cell lines. ATOH1 and SPDEF plasmids were purchased from OriGene and then reconstructed onto the pCI-neo mammalian expression vector, which served as the backbone for both plasmids. ATOH1 and SPDEF plasmids consisted of Myc and Flag tags as shown in Figure 6.10A. Therefore, we performed site-directed mutagenesis using the In-Fusion HD cloning kit (Takara Bio) to remove the tags, as they might affect the 3D structure of the protein. The general description of the procedure was described in Figure 6.10B. EcoRI and NotI were used as restriction enzymes for both ATOH1 and SPDEF. After eliminating the tags, we generated a construct by ligating the plasmid vector to the pCI-neo expression vector, and this construct was subsequently transfected into the cells.

A.





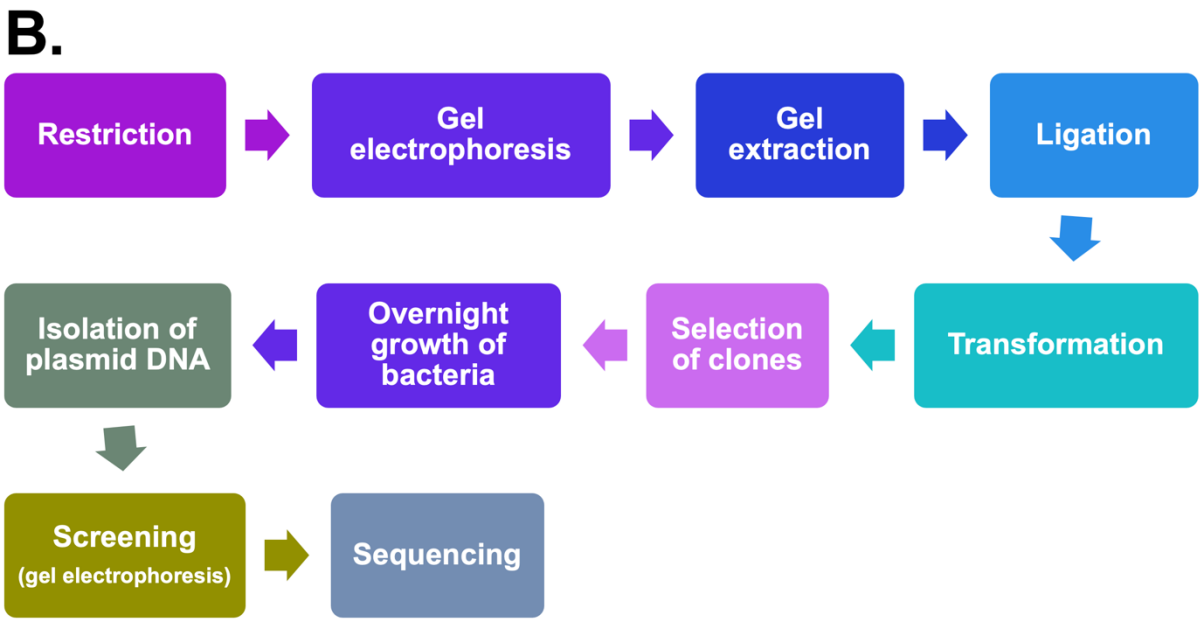


Figure 6.10 Plasmid maps of ATOH1, SPDEF, and pCI-neo and overall description of molecular cloning procedure. (A) displays schematic structure of ATOH1 and SPDEF. (B) exhibits a visual schematic representation of cloning process.

After eliminating the tags, we generated a construct by ligating the plasmid vector to the pCI-neo expression vector, and this construct was subsequently transfected into cells (Figure 6.11). Overexpression of ATOH1 and/or SPDEF led to an increase in MUC2- and TFF3-positive cells in LS180 (Figure 6.11A). By contrast, there was no

effect on goblet-cell negative cell lines, CC20, OXCO3, DLD1, and COLO678 (Figure 6.11B).

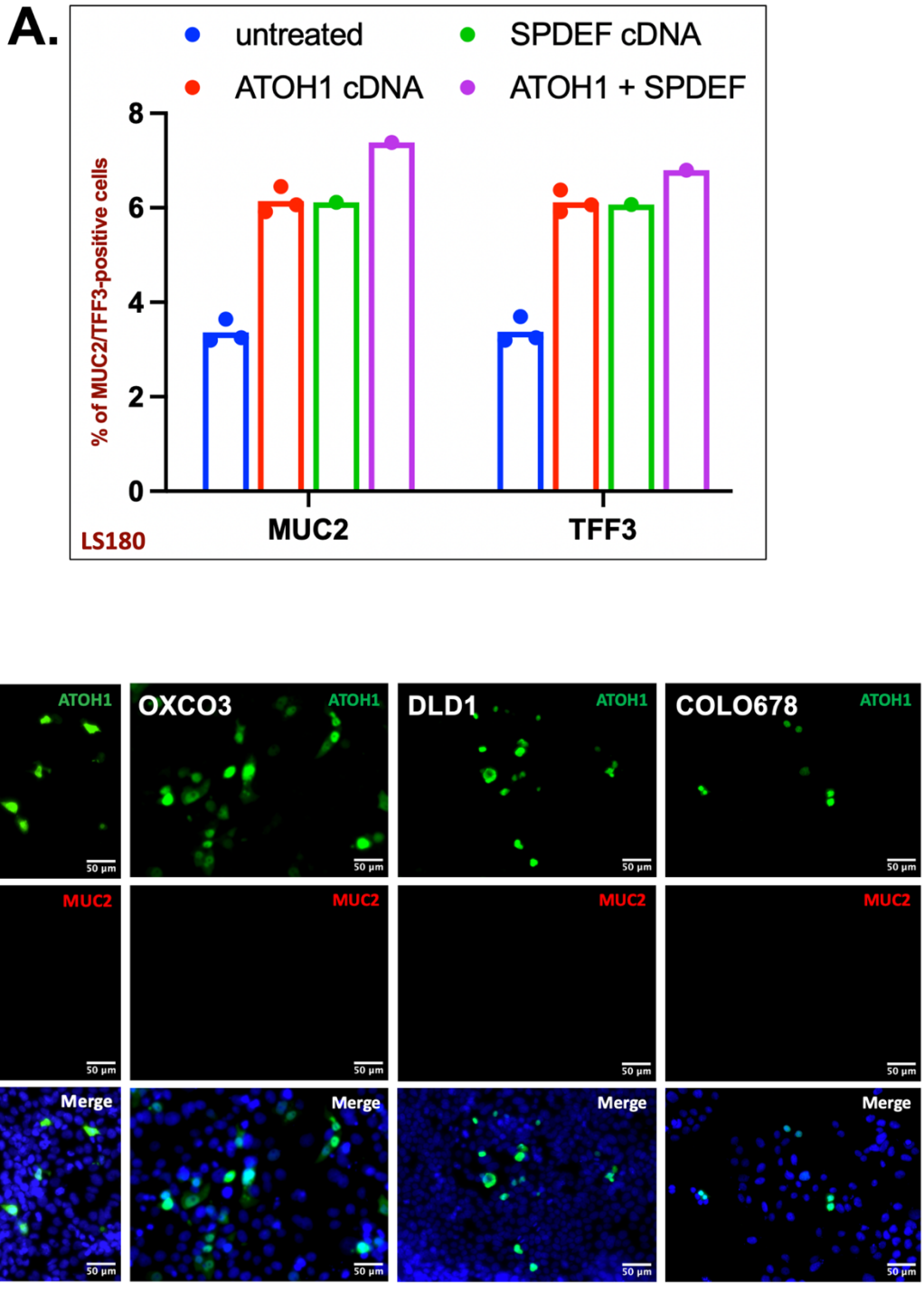


Figure 6.11 Overexpression of ATOH1 and SPDEF in CRC cell lines. A) describes the effect of ectopic expression of ATOH1 and SPDEF on their own and together on the percent of MUC2/TFF3-positive cells in the MUC2/TFF3 LS180 cell line. B) shows the lack of effect of

ectopic expression of MUC2 in four group 5 MUC2-/TFF3- cell lines. MUC2 (PR5D5, 1:250), ATOH1 (rabbit, 1:500), and DAPI were used in immunofluorescence analysis.

6.2.4.2 Effects of CDX1 and CDX2 knockdown

The mRNA expression analysis of CDX1 and CDX2 is shown in Figure 6.12. In each case, there is a strong association between high level expression of these genes and expression of MUC2 and TFF3. There is a clear suggestion that these transcription factors may be associated together in a transcription factor complex.

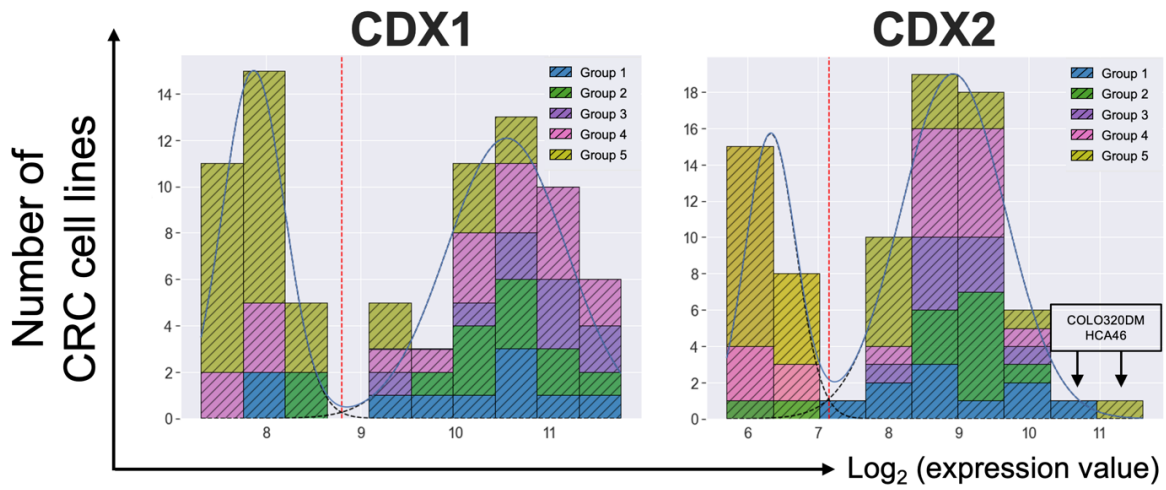


Figure 6.12 GMMChi gene expression analysis of CDX1 and CDX2. The cut-off thresholds for the difference between low and high expression are $2^{8.8}$ and $2^{7.2}$, respectively for CDX1 and CDX2.

To validate a role for CDX1 and CDX2 in the control of goblet cell differentiation we downregulated their expression levels by using siRNAs against *CDX1* and *CDX2*, respectively, in LS180, with *ATOH1* as a positive control. After three days of treatment, the percentage of MUC2- and TFF3- was quantified by immunofluorescence as before (Figure 3.5A). Single or double knockdown of *CDX1* and *CDX2*, led to a significant decrease of goblet cells in LS180, though the effect was not as strong as that of *ATOH1* knockdown (Figure 6.13A). Knockdown of *CDX2* has a stronger effect compared to knockdown of *CDX1*. The efficiency of the knockdown was checked by

qRT-PCR (Figure 6.13B). The effects of *CDX1* and *CDX2* downregulation was also investigated in LS513 (group 1), RW7213 (group 1), and SNUC2B (group 4) cell lines (Figure 6.13C). Our results revealed that knockdown of *ATOH1* did not significantly affect the levels of *CDX1*, and *CDX2*. Knockdown of *CDX1* led to an increase in *CDX2* mRNA levels, suggesting a compensatory mechanism. Knockdown of *CDX2* resulted in a dramatic reduction of *ATOH1* mRNA level.

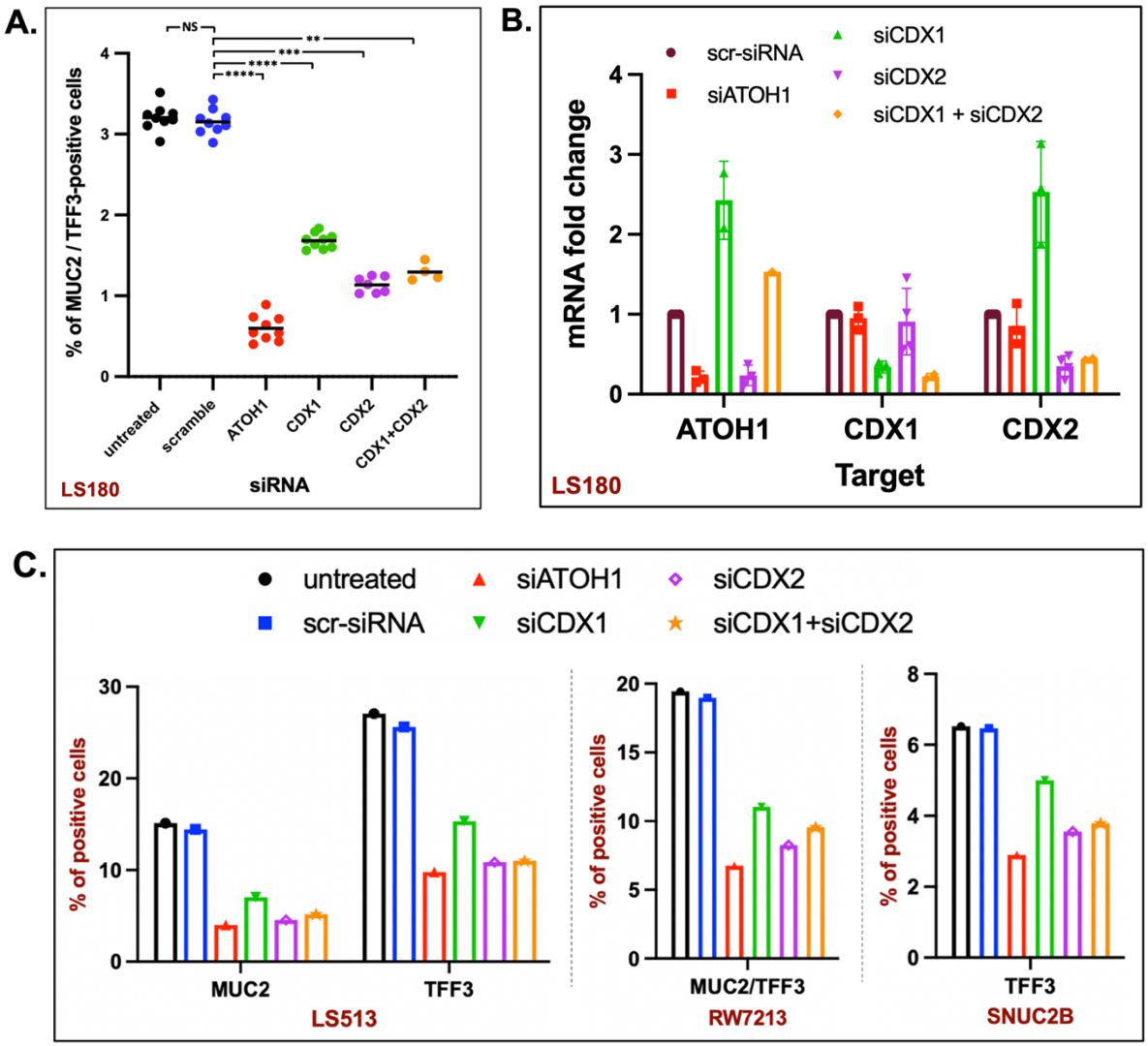


Figure 6.13 Knockdown of *CDX1* and/or *CDX2* in LS180 caused a significant reduction in MUC2 and TFF3 protein expression. (A) LS180 cells were transiently transfected with siRNAs against *CDX1* and *CDX2* and fixed three days after transfection. The proportion of MUC2- and TFF3-positive cells was defined as shown in Figure 3.5A. Statistical analysis was performed using the t-test. **p < 0.01, ***p < 0.001, ****p < 0.0001, 'NS' (p > 0.05). (B) The

efficiency of the knockdown was evaluated using RT-qPCR. GAPDH was used as an internal control. (C) The impacts of CDX1 and CDX2 knockdown were analysed in 3 more cell lines.

6.2.5 Effect of methylation on goblet cell differentiation

To study the effect of methylation on goblet cell differentiation, we used 5-AZA-CdR (5-aza-2′-deoxycytidine), a DNA-hypomethylating drug. First, different drug concentrations were tested and 10 μM was chosen as the working concentration. LS180 cell line was treated with 5-AZA-CdR for 48 hours and the percentage of MUC2- and TFF3-positive cells was quantified (Figure 6.14A). Our results revealed that the proportion of MUC2- and TFF3-positive cells significantly decreased following 5-AZA-CdR treatment, compared to the control samples. DMSO was used as the control in these experiments. Next, we tested mRNA expression analysis of ATOH1, CDX1/2, GATA6, HES1, and DLL4 (Figure 6.14B). Gene expression of ATOH1 was markedly reduced, indicating its regulation through methylation. By contrast, the gene expression of HES1 exhibited an increase.

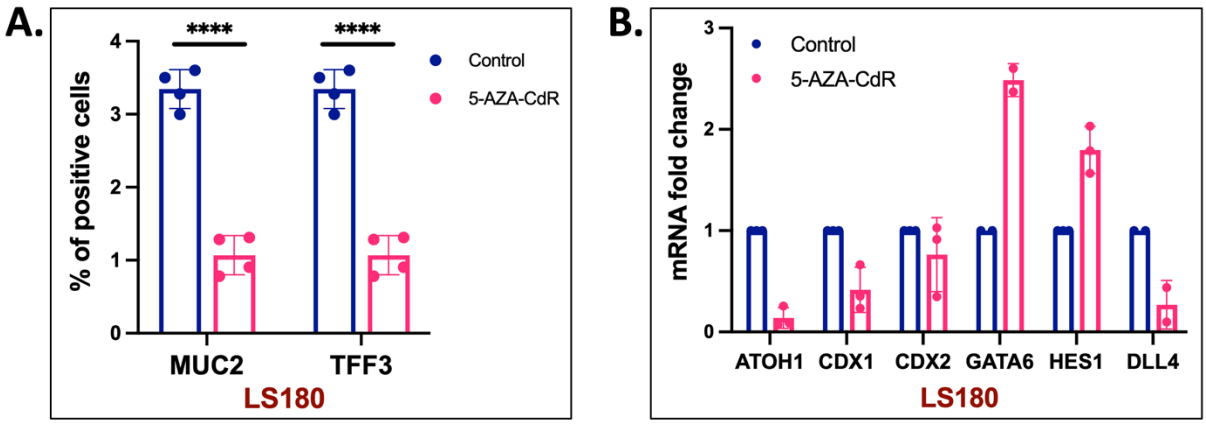


Figure 6.14 DNA methylation affects goblet cell differentiation in the LS180 cell line. The cells were treated with 10 μM of 5-AZA-CdR for 48 hours. (A) represents the quantification of MUC2 and TFF3 immunofluorescence staining in LS180. Statistical analysis was performed using a t-test of 5-AZA-CdR-treated samples compared to controls. ****p < 0.0001. (B)

displays mRNA expression analysis of ATOH1, CDX1/2, GATA6, HES1, and DLL4. GAPDH was used as an internal control.

Next, we investigated the duration of drug efficacy by conducting recovery experiments. LS180 cells were cultured in multiple 96-well plates and fixed for various time periods. Treatment was applied from 24 hours to 120 hours, with 5-AZA-CdR being added every 24 hours. The cells were fixed at indicated time points and the proportion of MUC2- and TFF3- was quantified (Figure 6.15A). For recovery experiments, we treated the cells with 5-AZA-CdR for 72 hours and then maintained them under a drug-free condition from 24 hours to 120 hours, by replacing the media every 2 days. The number of MUC2- and TFF3-positive cells reduced during the drug treatment and then began to rise following discontinuing the drug. SRB assay was carried out as described in Chapter section 2.2.3 to assess cell density (Figure 6.15B). According to our results, the samples treated with 5-AZA-CdR exhibited lower cell density than the controls, suggesting drug toxicity.

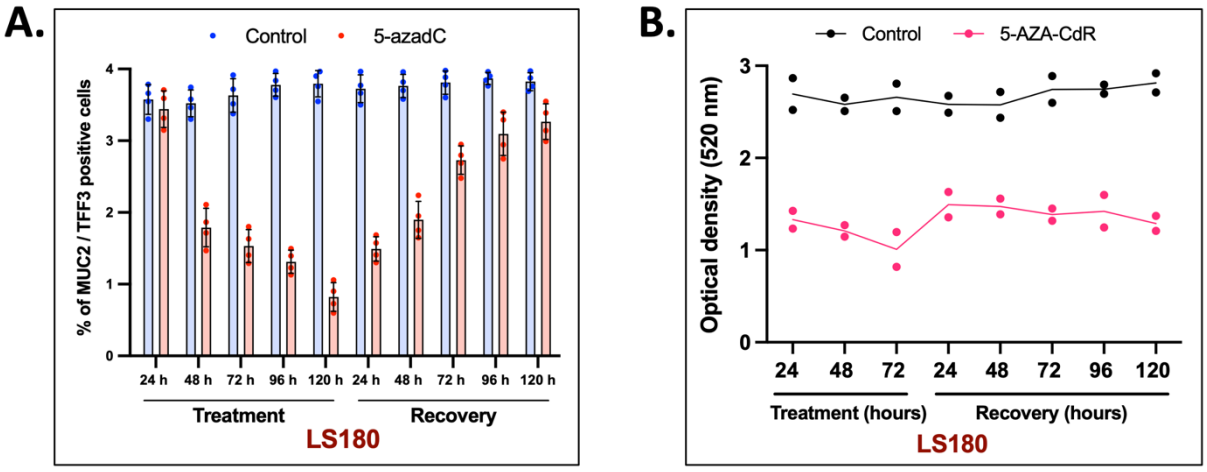


Figure 6.15 Assessment of 5-AZA-CdR efficacy on goblet cell differentiation in LS180. (A) For treatment, cells were subjected to the same procedure as shown in Figure 6.10A. During the recovery experiments, LS180 cells were administered with 10 μM of 5-AZA-CdR for 72 hours and then kept in a drug-free environment for the indicated time points. After each

time point, cells were fixed and quantification of MUC2- and TFF3- positive cells were performed as shown in Figure 3.5A. (B) SRB assay of untreated and 5-AZA-CdR-treated samples in LS180. The absorbance was measured at 520 nm. The connecting line represents the mean value of each replicate.

We also investigated the impact of demethylation on the group 4 cell line, COLO205. Cells were treated with 10 μM of 5-AZA-CdR, for time periods ranging from 24 hours to 120 hours. During the recovery experiments, we first treated the cells with the drug for 72 hours and then maintained them under a drug-free condition at indicated time intervals before fixing and staining. Representative images from three replicated experiments were shown in Figure 6.16A. Based on our results, COLO205 started to express MUC2 after 72 hours of treatment (Figure 6.16A) as opposite to LS180. The expression of MUC2 remained during the 96- and 120-hour treatment periods, but then declined gradually during the recovery phase. We can hypothesise that in group 4 cell line, COLO205, MUC2 is most likely methylated. Methylation sequencing analysis is necessary to confirm this statement and validate it in other group 4 cell lines. SRB assay was conducted, and the results exhibited a similar trend to that obtained in LS180 (Figure 6.16B). To sum up, goblet cell differentiation is affected by methylation in different ways via MUC2 or transcription factor promoters.

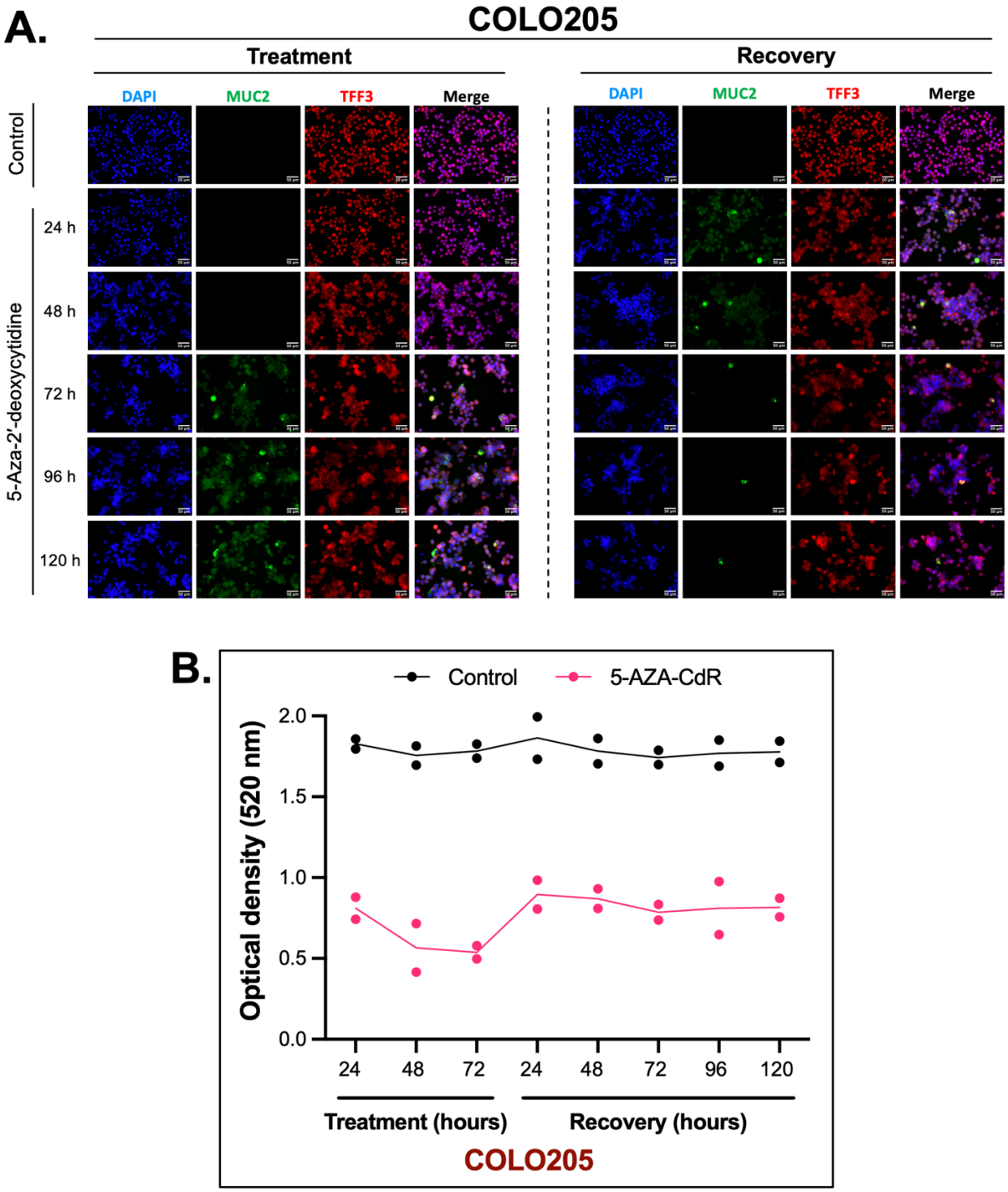


Figure 6.16 DNA methylation induces MUC2 expression in COLO205 cell line. The cells were administered with 10 μM of 5-AZA-CdR for indicated time periods. (A) describes immunofluorescence staining of MUC2 and TFF3 during untreated and treated cases in COLO205. (B) SRB assay indicates a comparison of cell density between untreated and treated samples. The connecting line indicates the mean value of each replicate.

6.3 DISCUSSION

Our results showed that the suppression of Notch signalling by γ-secretase inhibitor, DBZ, increases intestinal goblet cell differentiation. This corresponds to previous results in our lab, which showed DBZ led to enhanced goblet cell differentiation in the SW1222 cell line during normoxia and hypoxia conditions (Yeung et al., 2011). Other reports have also documented similar results following DBZ administration (van Es et al., 2005; Zecchini et al., 2005).

We confirmed that ATOH1 is an essential transcription factor for goblet cell differentiation. Our findings align with the results reported by other research groups (Kazanjian et al., 2010; Leow et al., 2004; Yang et al., 2001; Zhu et al., 2012). Kazanjian et al. demonstrated that γ-secretase inhibitors exhibit their effects in an ATOH1-dependent manner in both normal and cancerous intestinal epithelium (Kazanjian et al., 2010). According to their study, γ-secretase inhibitors did not promote secretory lineage differentiation of ATOH1-null progenitors, suggesting that only ATOH1-expressing CRCs will be sensitive to these inhibitors (Kazanjian et al., 2010). We were also able to demonstrate that group 5 cell lines that lack ATOH1 expression did not respond to DBZ treatment.

Most CRC cases exhibit loss of ATOH1 due to deletions in the ATOH1 locus or methylation within the promoter region or coding sequence of the gene, highlighting its transcriptional silencing through genetic and epigenetic alterations (Bossuyt et al., 2009; Fu et al., 2020). We also confirmed epigenetic regulation of ATOH1 expression.

SPDEF is involved in the terminal differentiation of goblet cells. Noah et al. demonstrated that SPDEF is required to stimulate the expression of goblet cell-related

genes in LS174T cells when they are exposed to Notch-inhibitors (Noah et al., 2010). SPDEF depletion led to a massive increase in immature secretory progenitors in the intestine (Gregorieff et al., 2009; Kolev & Kaestner, 2023). Our qRT-PCR results revealed that there is an association between ATOH1 and SPDEF as knocking down ATOH1 led to reduced SPDEF expression, and vice versa. A study by Shroyer et al. detected a significant enrichment of SPDEF binding regions within ATOH1 target genes, suggesting their role as co-transcriptional regulators (Lo et al., 2017).

CDX homeodomain proteins were shown to control MUC2 expression (Mesquita et al., 2003). A study by Yamamoto et al. revealed CDX2-binding sites in the MUC2 promotor region (Yamamoto et al., 2003). However, CDX1 and CDX2 binding sites are very similar if not identical so that it is very difficult to distinguish their effects and, in many situations, when one is knocked out the other can most probably take its place functionally. Considering that the expression of CDX1 and CDX2 is regulated by methylation, a deficiency or reduced level of these factors can lead to impaired MUC2 expression (Wang et al., 2018; Wong et al., 2004).

We identified that methylation affects goblet cell differentiation in different ways via MUC2 or transcription factors promoters. Several reports indicated that demethylation of the MUC2 gene promoter enhanced its de novo expression in CRC cell lines (Andrianifahanana et al., 2006; Gratchev et al., 2001; Okudaira et al., 2010). Additionally, the proportion of analysed methylated sites was higher in non-mucinous carcinomas compared to mucinous carcinomas (Gratchev et al., 2001). For TFF3, Wang et al. found no significant association between promoter methylation and TFF3 expression (Yusufu et al., 2019). Overall, it seems most likely that the various implicated goblet cell controlling transcription factors, we have analysed, act in

combinations. It is then the disruption of these combinations by selection against goblet cell differentiation and MUC2 expression during tumour evolution that gives rise to the 5 different categories of MUC2 and TFF3 expression that have been identified in the CRC cell line panel.

CHAPTER 7

GENERAL DISCUSSION AND FUTURE OUTLOOK

CHAPTER 7: GENERAL DISCUSSION AND FUTURE OUTLOOK

Goblet cells are one of the secretory differentiated cells within the intestinal epithelium. The disruption of goblet cell differentiation has been shown to be associated with the more aggressive colorectal cancer (Conze et al., 2010; Gundamaraju & Chong, 2021; Kim & Ho, 2010). In addition to deregulation of cellular differentiation, hypersecretion of goblet cell product, MUC2 can also create an environment that allows cancer cells to escape chemotherapy and facilitates immune surveillance (Cantero-Recasens et al., 2022; Fujishima et al., 2012). Understanding the mechanisms involved in goblet cell differentiation might enable a more refined diagnosis of CRC and suggest novel approaches to treatment and prevention of CRC.

Our study reveals a new classification of CRC lines, and do so CRC, in the context of goblet cell differentiation and a possible novel subtype of CRCs. We categorised CRC cell lines into 4 groups by level of expression of MUC2 and TFF3 by both mRNA and immunofluorescence analysis, and a fifth category which was negative for the two goblet cell markers.

We originally hypothesised that group 1 cell lines, which express high levels of both MUC2 and TFF3, might be signet ring tumours. However, this is not the case due to the secretion of MUC2. They probably represent the well-known category of mucinous colorectal cancers. Most of the MUC2+ TFF3+ cell lines secrete both proteins with possibly just a small subset of group 3 that do not secrete.

One category displays notably high levels of TFF3 and no (or very low) MUC2 expression which has not been described elsewhere. It is important to investigate

group 4 cell lines that appear to be in the process of losing differentiation and potentially becoming more aggressive compared to differentiated types.

We extended our classification to cancer patients' tissue sections and validated the groups we identified in CRC cell lines. The immunofluorescence analysis was done with a commercial antibody against TFF3 and one of our antibodies (PR5D5) against MUC2. Though TFF3 and MUC2 completely overlap on normal colorectal tissue cells, and the majority of tissue sections showed heterogeneity in group types with respect to levels of MUC2 and TFF3 expression. This leads to the question of how these goblet cell differentiation groups relate to the usual pathology classifications and to clinical prognosis, with the eventual possibility of discovering what the underlying molecular changes are. The limitation of our study was the low number of slides used. Examining a much larger number of tissue sections allows us to validate the accuracy of our classification based on cell lines and explore the range of different associated prognoses, as well as to train artificial intelligence effectively. This might enable artificial intelligence to identify categories based only on H&E staining.

Our data agree with the widely reported primary dependence of MUC2 expression on the transcription factor ATOH1. To investigate what distinguishes the 4 main groups of cancers expressing MUC2 and/or TFF3, we compared the pairwise differences in their gene expression patterns. Several transcription factors, such as TOX3 and FOXA2 were significantly expressed in groups 1-4 versus group 5 which requires further investigation.

It is the high level TFF3 and completely negative/very low MUC2 expressing cancer cell lines (group 4) that we think are novel and most intriguing. The effects of 5-AZA-CdR treatment suggests that the absence of MUC2 expression in group 4 category

maybe due to methylation of the MUC2 promoter region. However, the result is confounded by the fact that HES1 increases expression with 5-AZA-CdR, which should have the opposite effect. We speculate that TFF3 may have a dual role in both the stimulation of growth in colonic tissues, especially their repair when damaged, and the maintenance of MUC2 mature structure.

We found no correlation between the groups of cancers we defined by MUC2 and TFF3 expression and enterocyte differentiation as measured by lumen formation. In particular, in contrast to enterocyte differentiation, the goblet cell differentiation does not seem to be dependent on ECM/Matrigel attachment.

Our observations fit in with selection in carcinogenesis, at least for adenocarcinomas, against differentiation perhaps especially to goblet cells. Is it possible that there is also selection especially against MUC2 expression, as in group 4, that might be associated with some advantage in TFF3 expression? Attacking TFF3 function could lead to new treatment approaches in CRC.

Goblet cells in cell lines provide a good model for further in-depth studies of differentiation. Clearly it is of interest to try and identify the driving genetic changes associated with each of our four main groups. ATOH1 downregulation, possibly by promoter methylation, could be one factor. However, our data indicate that changes in the expression of just one transcription factor are not enough. Several factors must be involved in any one transcription complex, and any one transcription factor is most probably involved in a variety of different transcription complexes whose functions might be quite unrelated. Overexpression or knocking down different combinations of transcription factors needs to be further investigated. Choosing the correct antibodies

is essential because some may exhibit varying activity on Western blot and immunofluorescence.

Single cell sequencing technology is quite efficient for determining intracell line variations, as well as differences between various CRC cell lines. We validated goblet cell associated genes through the scRNA-seq data of normal colon and LS180-COLO678 cell lines. The challenge with the sensitivity of single-cell analysis is that it helps define cell types but cannot generally explain how they arise. Conducting time-dependent experiments on various transcription factors can help us track their role in goblet cell differentiation.

REFERENCES

Aamann, L., Vestergaard, E. M., & Gronbaek, H. (2014). Trefoil factors in inflammatory bowel disease. *World J Gastroenterol*, *20*(12), 3223-3230. https://doi.org/10.3748/wjg.v20.i12.3223

Aasebø, K., Dragomir, A., Sundström, M., Mezheyeuski, A., Edqvist, P. H., Eide, G. E., Ponten, F., Pfeiffer, P., Glimelius, B., & Sorbye, H. (2020). CDX2: A Prognostic Marker in Metastatic Colorectal Cancer Defining a Better BRAF Mutated and a Worse KRAS Mutated Subgroup. *Front Oncol*, *10*, 8. https://doi.org/10.3389/fonc.2020.00008

Aihara, E., Engevik, K. A., & Montrose, M. H. (2017). Trefoil Factor Peptides and Gastrointestinal Function. *Annu Rev Physiol*, *79*, 357-380. https://doi.org/10.1146/annurev-physiol-021115-105447

Al-Khayal, K., Abdulla, M., Al-Obaid, O., Zubaidi, A., Vaali-Mohammed, M. A., Alsheikh, A., & Ahmad, R. (2016). Differential expression of mucins in Middle Eastern patients with colorectal cancer. *Oncol Lett*, *12*(1), 393-400. https://doi.org/10.3892/ol.2016.4672

Albert, T. K., Laubinger, W., Müller, S., Hanisch, F.-G., Kalinski, T., Meyer, F., & Hoffmann, W. (2010). Human Intestinal TFF3 Forms Disulfide-Linked Heteromers with the Mucus-Associated FCGBP Protein and Is Released by Hydrogen Sulfide. *Journal of Proteome Research*, *9*(6), 3108-3117. https://doi.org/10.1021/pr100020c

Allaire, J. M., Morampudi, V., Crowley, S. M., Stahl, M., Yu, H., Bhullar, K., Knodler, L. A., Bressler, B., Jacobson, K., & Vallance, B. A. (2018). Frontline defenders: goblet cell mediators dictate host-microbe interactions in the intestinal tract during health and disease. *Am J Physiol Gastrointest Liver Physiol*, *314*(3), G360-g377. https://doi.org/10.1152/ajpgi.00181.2017

Ambort, D., Johansson, M. E., Gustafsson, J. K., Nilsson, H. E., Ermund, A., Johansson, B. R., Koeck, P. J., Hebert, H., & Hansson, G. C. (2012). Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proc Natl Acad Sci U S A*, *109*(15), 5645-5650. https://doi.org/10.1073/pnas.1120269109

An, Y., Zhou, J., Lin, G., Wu, H., Cong, L., Li, Y., Qiu, X., & Shi, W. (2021). Clinicopathological and Molecular Characteristics of Colorectal Signet Ring Cell Carcinoma: A Review. *Pathol Oncol Res*, *27*, 1609859. https://doi.org/10.3389/pore.2021.1609859

Andrianifahanana, M., Moniaux, N., & Batra, S. K. (2006). Regulation of mucin expression: mechanistic aspects and implications for cancer and inflammatory diseases. *Biochim Biophys Acta*, *1765*(2), 189-222. https://doi.org/10.1016/j.bbcan.2006.01.002

Aragaki, M., Tsuchiya, K., Okamoto, R., Yoshioka, S., Nakamura, T., Sakamoto, N., Kanai, T., & Watanabe, M. (2008). Proteasomal degradation of Atoh1 by aberrant Wnt signaling maintains the undifferentiated state of colon cancer. *Biochem Biophys Res Commun*, *368*(4), 923-929. https://doi.org/10.1016/j.bbrc.2008.02.011

Aran, D., Sirota, M., & Butte, A. J. (2015). Systematic pan-cancer analysis of tumour purity. *Nat Commun*, *6*, 8971. https://doi.org/10.1038/ncomms9971

Arike, L., Holmén-Larsson, J., & Hansson, G. C. (2017). Intestinal Muc2 mucin O-glycosylation is affected by microbiota and regulated by differential expression of glycosyltranferases. *Glycobiology*, *27*(4), 318-328. https://doi.org/10.1093/glycob/cww134

Ashley, N., Yeung, T. M., & Bodmer, W. F. (2013). Stem cell differentiation and lumen formation in colorectal cancer cell lines and primary tumors. *Cancer Res*, *73*(18), 5798-5809. https://doi.org/10.1158/0008-5472.CAN-13-0454

Babyatsky, M., Lin, J., Yio, X., Chen, A., Zhang, J. Y., Zheng, Y., Twyman, C., Bao, X., Schwartz, M., Thung, S., Lawrence Werther, J., & Itzkowitz, S. (2009). Trefoil factor-3 expression in human colon cancer liver metastasis. *Clin Exp Metastasis*, *26*(2), 143-151. https://doi.org/10.1007/s10585-008-9224-9

Bae, J. M., Lee, T. H., Cho, N. Y., Kim, T. Y., & Kang, G. H. (2015). Loss of CDX2 expression is associated with poor prognosis in colorectal cancer patients. *World J Gastroenterol*, *21*(5), 1457-1467. https://doi.org/10.3748/wjg.v21.i5.1457

Bakaris, S., Cetinkaya, A., Ezberci, F., & Ekerbicer, H. (2008). Expression of homeodomain protein CDX2 in colorectal adenoma and adenocarcinoma. *Histol Histopathol*, *23*(9), 1043-1047. https://doi.org/10.14670/hh-23.1043

Bansil, R., & Turner, B. S. (2018). The biology of mucus: Composition, synthesis and organization. *Adv Drug Deliv Rev*, *124*, 3-15. https://doi.org/10.1016/j.addr.2017.09.023

Baran, B., Mert Ozupek, N., Yerli Tetik, N., Acar, E., Bekcioglu, O., & Baskin, Y. (2018). Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterology Res*, *11*(4), 264-273. https://doi.org/10.14740/gr1062w

Barriga, F. M., Montagni, E., Mana, M., Mendez-Lago, M., Hernando-Momblona, X., Sevillano, M., Guillaumet-Adkins, A., Rodriguez-Esteban, G., Buczacki, S. J. A., Gut, M., Heyn, H., Winton, D. J., Yilmaz, O. H., Attolini, C. S., Gut, I., & Batlle, E. (2017). Mex3a Marks a Slowly Dividing Subpopulation of Lgr5+ Intestinal Stem Cells. *Cell Stem Cell*, *20*(6), 801-816.e807. https://doi.org/10.1016/j.stem.2017.02.007

Becker, W. R., Nevins, S. A., Chen, D. C., Chiu, R., Horning, A. M., Guha, T. K., Laquindanum, R., Mills, M., Chaib, H., Ladabaum, U., Longacre, T., Shen, J., Esplin, E. D., Kundaje, A., Ford, J. M., Curtis, C., Snyder, M. P., & Greenleaf, W. J. (2022). Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nat Genet*, *54*(7), 985-995. https://doi.org/10.1038/s41588-022-01088-x

Belle, N. M., Ji, Y., Herbine, K., Wei, Y., Park, J., Zullo, K., Hung, L. Y., Srivatsa, S., Young, T., Oniskey, T., Pastore, C., Nieves, W., Somsouk, M., & Herbert, D. R. (2019). TFF3 interacts with LINGO2 to regulate EGFR activation for protection against colitis and gastrointestinal helminths. *Nat Commun*, *10*(1), 4408. https://doi.org/10.1038/s41467-019-12315-1

Benedix, F., Kube, R., Meyer, F., Schmidt, U., Gastinger, I., Lippert, H., & Group., C. R. C. P. T. S. (2010). Comparison of 17,641 patients with right- and left-sided colon cancer: Differences in epidemiology, perioperative course, histology,
and survival. *Diseases of the Colon & Rectum*, *53*(1), 57-64. https://doi.org/10.1007/DCR.0b013e3181c703a4

Bergstrom, J. H., Birchenough, G. M., Katona, G., Schroeder, B. O., Schutte, A., Ermund, A., Johansson, M. E., & Hansson, G. C. (2016). Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. *Proc Natl Acad Sci U S A*, *113*(48), 13833-13838. https://doi.org/10.1073/pnas.1611400113

Bergstrom, K. S., & Xia, L. (2013). Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology*, *23*(9), 1026-1037. https://doi.org/10.1093/glycob/cwt045

Beuling, E., Aronson, B. E., Tran, L. M., Stapleton, K. A., ter Horst, E. N., Vissers, L. A., Verzi, M. P., & Krasinski, S. D. (2012). GATA6 is required for proliferation, migration, secretory cell maturation, and gene expression in the mature mouse colon. *Mol Cell Biol*, *32*(17), 3392-3402. https://doi.org/10.1128/mcb.00070-12

Bhatia, R., Gautam, S. K., Cannon, A., Thompson, C., Hall, B. R., Aithal, A., Banerjee, K., Jain, M., Solheim, J. C., Kumar, S., & Batra, S. K. (2019). Cancer-associated mucins: role in immune modulation and metastasis. *Cancer Metastasis Rev*, *38*(1-2), 223-236. https://doi.org/10.1007/s10555-018-09775-0

Birchenough, G. M. H., Johansson, M. E., Gustafsson, J. K., Bergström, J. H., & Hansson, G. C. (2015). New developments in goblet cell mucus secretion and function. *Mucosal Immunology*, *8*(4), 712-719. https://doi.org/10.1038/mi.2015.32

Birchenough, G. M. H., Nyström, E. E. L., Johansson, M. E. V., & Hansson, G. C. (2016). A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion. *Science*, *352*(6293), 1535-1542. https://doi.org/doi:10.1126/science.aaf7419

Bodmer, W. F. (2006). Cancer genetics: colorectal cancer as a model. *J Hum Genet*, *51*(5), 391-396. https://doi.org/10.1007/s10038-006-0373-x
Bodmer, W. F., & Cavalli-Sforza, L. L. (1976). *Genetics, evolution and man*. W.H.Freeman & Co Ltd.
Boegh, M., & Nielsen, H. M. (2015). Mucus as a barrier to drug delivery – understanding and mimicking the barrier properties. *Basic Clin Pharmacol Toxicol*, *116*(3), 179-186. https://doi.org/10.1111/bcpt.12342
Bonhomme, C., Calon, A., Martin, E., Robine, S., Neuville, A., Kedinger, M., Domon-Dell, C., Duluc, I., & Freund, J. N. (2008). Cdx1, a dispensable homeobox gene for gut development with limited effect in intestinal cancer. *Oncogene*, *27*(32), 4497-4502. https://doi.org/10.1038/onc.2008.78
Bossuyt, W., Kazanjian, A., De Geest, N., Van Kelst, S., De Hertogh, G., Geboes, K., Boivin, G. P., Luciani, J., Fuks, F., Chuah, M., VandenDriessche, T., Marynen, P., Cools, J., Shroyer, N. F., & Hassan, B. A. (2009). Atonal homolog 1 is a tumor suppressor gene. *PLoS Biol*, *7*(2), e39. https://doi.org/10.1371/journal.pbio.1000039
Bourque, J., Opejin, A., Surnov, A., Iberg, C. A., Gross, C., Jain, R., Epstein, J. A., & Hawiger, D. (2021). Landscape of Hopx expression in cells of the immune system. *Heliyon*, *7*(11), e08311. https://doi.org/10.1016/j.heliyon.2021.e08311
Bu, X. D., Li, N., Tian, X. Q., Li, L., Wang, J. S., Yu, X. J., & Huang, P. L. (2010). Altered expression of MUC2 and MUC5AC in progression of colorectal carcinoma. *World J Gastroenterol*, *16*(32), 4089-4094. https://doi.org/10.3748/wjg.v16.i32.4089
Buechler, S. A., Stephens, M. T., Hummon, A. B., Ludwig, K., Cannon, E., Carter, T. C., Resnick, J., Gokmen-Polar, Y., & Badve, S. S. (2020). ColoType: a forty gene signature for consensus molecular subtyping of colorectal cancer tumors using whole-genome assay or targeted RNA-sequencing. *Sci Rep*, *10*(1), 12123. https://doi.org/10.1038/s41598-020-69083-y
Burclaff, J., Bliton, R. J., Breau, K. A., Ok, M. T., Gomez-Martinez, I., Ranek, J. S., Bhatt, A. P., Purvis, J. E., Woosley, J. T., & Magness, S. T. (2022). A Proximal-to-Distal Survey of Healthy Adult Human Small Intestine and Colon Epithelium by Single-Cell Transcriptomics. *Cell Mol Gastroenterol Hepatol*, *13*(5), 1554-1589. https://doi.org/10.1016/j.jcmgh.2022.02.007
Busch, M., & Dünker, N. (2015). Trefoil factor family peptides--friends or foes? *Biomol Concepts*, *6*(5-6), 343-359. https://doi.org/10.1515/bmc-2015-0020
Bustamante-Lopez, L. A., Nahas, S. C., Nahas, C. S. R., Pinto, R. A., Marques, C. F. S., & Cecconello, I. (2019). Is There a Difference between Right- Versus Left-Sided Colon Cancers? Does Side Make Any Difference in Long-Term Follow-Up? *Arq Bras Cir Dig*, *32*(4), e1479. https://doi.org/10.1590/0102-6720201900001e1479
Cantero-Recasens, G., Alonso-Marañón, J., Lobo-Jarne, T., Garrido, M., Iglesias, M., Espinosa, L., & Malhotra, V. (2022). Reversing chemorefraction in colorectal cancer cells by controlling mucin secretion. *Elife*, *11*. https://doi.org/10.7554/eLife.73926
Carulli, A. J., Keeley, T. M., Demitrack, E. S., Chung, J., Maillard, I., & Samuelson, L. C. (2015). Notch receptor regulation of intestinal stem cell homeostasis and crypt regeneration. *Dev Biol*, *402*(1), 98-108. https://doi.org/10.1016/j.ydbio.2015.03.012
Chan, C. W., Wong, N. A., Liu, Y., Bicknell, D., Turley, H., Hollins, L., Miller, C. J., Wilding, J. L., & Bodmer, W. F. (2009). Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *Proc Natl Acad Sci U S A*, *106*(6), 1936-1941. https://doi.org/10.1073/pnas.0812904106
Chen, R. M., Chiou, Y. S., Chong, Q. Y., Poh, H. M., Tan, T. Z., Zhang, M. Y., Ma, L., Zhu, T., Pandey, V., Basappa, Kumar, A. P., & Lobie, P. E. (2019). Pharmacological Inhibition of TFF3 Enhances Sensitivity of CMS4 Colorectal Carcinoma to 5-Fluorouracil through Inhibition of p44/42 MAPK. *Int J Mol Sci*, *20*(24). https://doi.org/10.3390/ijms20246215
Chen, X., Du, P., She, J., Cao, L., Li, Y., & Xia, H. (2016). Loss of ZG16 is regulated by miR-196a and contributes to stemness and progression of colorectal cancer. *Oncotarget*, *7*(52), 86695-86703. https://doi.org/10.18632/oncotarget.13435

Chen, Y., Ying, Y., Wang, M., Ma, C., Jia, M., Shi, L., Wang, S., Zheng, X., Chen, W., & Shu, X. S. (2023). A distal super-enhancer activates oncogenic ETS2 via recruiting MECOM in inflammatory bowel disease and colorectal cancer. *Cell Death Dis*, *14*(1), 8. https://doi.org/10.1038/s41419-022-05513-1

Chu, D., Zhang, Z., Zhou, Y., Wang, W., Li, Y., Zhang, H., Dong, G., Zhao, Q., & Ji, G. (2011). Notch1 and Notch2 have opposite prognostic effects on patients with colorectal cancer. *Ann Oncol*, *22*(11), 2440-2447. https://doi.org/10.1093/annonc/mdq776

Chu, D., Zheng, J., Wang, W., Zhao, Q., Li, Y., Li, J., Xie, H., Zhang, H., Dong, G., Xu, C., Li, M., Chen, D., & Ji, G. (2009). Notch2 expression is decreased in colorectal cancer and related to tumor differentiation status. *Ann Surg Oncol*, *16*(12), 3259-3266. https://doi.org/10.1245/s10434-009-0655-6

Conze, T., Carvalho, A. S., Landegren, U., Almeida, R., Reis, C. A., David, L., & Söderberg, O. (2010). MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. *Glycobiology*, *20*(2), 199-206. https://doi.org/10.1093/glycob/cwp161

Cui, H. Y., Wang, S. J., Song, F., Cheng, X., Nan, G., Zhao, Y., Qian, M. R., Chen, X., Li, J. Y., Liu, F. L., Zhu, Y. M., Tian, R. F., Wang, B., Wu, B., Zhang, Y., Sun, X. X., Guo, T., Yang, X. M., Zhang, H., . . . Chen, Z. N. (2021). CD147 receptor is essential for TFF3-mediated signaling regulating colorectal cancer progression. *Signal Transduct Target Ther*, *6*(1), 268. https://doi.org/10.1038/s41392-021-00677-2

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P. S., Rothenberg, M. E., Leyrat, A. A., Sim, S., Okamoto, J., Johnston, D. M., Qian, D., Zabala, M., Bueno, J., Neff, N. F., Wang, J., Shelton, A. A., Visser, B., Hisamori, S., Shimono, Y., van de Wetering, M., . . . Quake, S. R. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol*, *29*(12), 1120-1127. https://doi.org/10.1038/nbt.2038

Dame, M. K., Jiang, Y., Appelman, H. D., Copley, K. D., McClintock, S. D., Aslam, M. N., Attili, D., Elmunzer, B. J., Brenner, D. E., Varani, J., & Turgeon, D. K. (2014). Human colonic crypts in culture: segregation of immunochemical markers in normal versus adenoma-derived. *Lab Invest*, *94*(2), 222-234. https://doi.org/10.1038/labinvest.2013.145

Dang, L. H., Chen, F., Ying, C., Chun, S. Y., Knock, S. A., Appelman, H. D., & Dang, D. T. (2006). CDX2 has tumorigenic potential in the human colon cancer cell lines LOVO and SW48. *Oncogene*, *25*(15), 2264-2272. https://doi.org/10.1038/sj.onc.1209247

Demitrack, E. S., & Samuelson, L. C. (2016). Notch regulation of gastrointestinal stem cells. *J Physiol*, *594*(17), 4791-4803. https://doi.org/10.1113/jp271667

Dhanisha, S. S., Guruvayoorappan, C., Drishya, S., & Abeesh, P. (2018). Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. *Crit Rev Oncol Hematol*, *122*, 98-122. https://doi.org/10.1016/j.critrevonc.2017.12.006

Dieckow, J., Brandt, W., Hattermann, K., Schob, S., Schulze, U., Mentlein, R., Ackermann, P., Sel, S., & Paulsen, F. P. (2016). CXCR4 and CXCR7 Mediate TFF3-Induced Cell Migration Independently From the ERK1/2 Signaling Pathway. *Invest Ophthalmol Vis Sci*, *57*(1), 56-65. https://doi.org/10.1167/iovs.15-18129

Domingo, E., Ramamoorthy, R., Oukrif, D., Rosmarin, D., Presz, M., Wang, H., Pulker, H., Lockstone, H., Hveem, T., Cranston, T., Danielsen, H., Novelli, M., Davidson, B., Xu, Z. Z., Molloy, P., Johnstone, E., Holmes, C., Midgley, R., Kerr, D., . . . Tomlinson, I. (2013). Use of multivariate analysis to suggest a new molecular classification of colorectal cancer. *J Pathol*, *229*(3), 441-448. https://doi.org/10.1002/path.4139

Donovan, J. A., & Koretzky, G. A. (1993). CD45 and the immune response. *J Am Soc Nephrol*, *4*(4), 976-985. https://doi.org/10.1681/asn.V44976

Drummond, F. J., Sowden, J., Morrison, K., & Edwards, Y. H. (1998). Colon carbonic anhydrase 1: transactivation of gene expression by the homeodomain protein Cdx2. *FEBS Lett*, *423*(2), 218-222. https://doi.org/10.1016/s0014-5793(98)00103-3

Efstathiou, J. A., Noda, M., Rowan, A., Dixon, C., Chinery, R., Jawhari, A., Hattori, T., Wright, N. A., Bodmer, W. F., & Pignatelli, M. (1998). Intestinal trefoil factor controls the expression of the adenomatous polyposis coli-catenin and the E-cadherin-catenin complexes in human colon carcinoma cells. *Proc Natl Acad Sci U S A*, *95*(6), 3122-3127. https://doi.org/10.1073/pnas.95.6.3122

Ehrencrona, E., van der Post, S., Gallego, P., Recktenwald, C. V., Rodriguez-Pineiro, A. M., Garcia-Bonete, M. J., Trillo-Muyo, S., Backstrom, M., Hansson, G. C., & Johansson, M. E. V. (2021). The IgGFc-binding protein FCGBP is secreted with all GDPH sequences cleaved but maintained by interfragment disulfide bonds. *J Biol Chem*, *297*(1), 100871. https://doi.org/10.1016/j.jbc.2021.100871

Elphick, D. A., & Mahida, Y. R. (2005). Paneth cells: their role in innate immunity and inflammatory disease. *Gut*, *54*(12), 1802-1809. https://doi.org/10.1136/gut.2005.068601

Etienne-Mesmin, L., Chassaing, B., Desvaux, M., De Paepe, K., Gresse, R., Sauvaitre, T., Forano, E., de Wiele, T. V., Schüller, S., Juge, N., & Blanquet-Diot, S. (2019). Experimental models to study intestinal microbes-mucus interactions in health and disease. *FEMS Microbiol Rev*, *43*(5), 457-489. https://doi.org/10.1093/femsre/fuz013

Fearnhead, N. S., Wilding, J. L., & Bodmer, W. F. (2002). Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. *British Medical Bulletin*, *64*(1), 27-43. https://doi.org/10.1093/bmb/64.1.27

Fre, S., Hannezo, E., Sale, S., Huyghe, M., Lafkas, D., Kissel, H., Louvi, A., Greve, J., Louvard, D., & Artavanis-Tsakonas, S. (2011). Notch lineages and activity in intestinal stem cells determined by a new set of knock-in mice. *PLoS One*, *6*(10), e25785. https://doi.org/10.1371/journal.pone.0025785

Fu, Y., Yuan, S. S., Zhang, L. J., Ji, Z. L., & Quan, X. J. (2020). Atonal bHLH transcription factor 1 is an important factor for maintaining the balance of cell proliferation and differentiation in tumorigenesis. *Oncol Lett*, *20*(3), 2595-2605. https://doi.org/10.3892/ol.2020.11833

Fujishima, Y., Goi, T., Kimura, Y., Hirono, Y., Katayama, K., & Yamaguchi, A. (2012). MUC2 protein expression status is useful in assessing the effects of hyperthermic intraperitoneal chemotherapy for peritoneal dissemination of colon cancer. *Int J Oncol*, *40*(4), 960-964. https://doi.org/10.3892/ijo.2012.1334

Gao, X., Sedgwick, T., Shi, Y. B., & Evans, T. (1998). Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. *Mol Cell Biol*, *18*(5), 2901-2911. https://doi.org/10.1128/mcb.18.5.2901

Gao, X. H., Li, J., Gong, H. F., Yu, G. Y., Liu, P., Hao, L. Q., Liu, L. J., Bai, C. G., & Zhang, W. (2020). Comparison of Fresh Frozen Tissue With Formalin-Fixed Paraffin-Embedded Tissue for Mutation Analysis Using a Multi-Gene Panel in Patients With Colorectal Cancer. *Front Oncol*, *10*, 310. https://doi.org/10.3389/fonc.2020.00310

Gjorevski, N., & Ordonez-Moran, P. (2017). Intestinal Stem Cell Niche Insights Gathered from Both In Vivo and Novel In Vitro Models. *Stem Cells Int*, *2017*, 8387297. https://doi.org/10.1155/2017/8387297

Goodspeed, A., Heiser, L. M., Gray, J. W., & Costello, J. C. (2016). Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Mol Cancer Res*, *14*(1), 3-13. https://doi.org/10.1158/1541-7786.Mcr-15-0189

Grainger, S., Hryniuk, A., & Lohnes, D. (2013). Cdx1 and Cdx2 exhibit transcriptional specificity in the intestine. *PLoS One*, *8*(1), e54757. https://doi.org/10.1371/journal.pone.0054757

Gratchev, A., Siedow, A., Bumke-Vogt, C., Hummel, M., Foss, H. D., Hanski, M. L., Kobalz, U., Mann, B., Lammert, H., Mansmann, U., Stein, H., Riecken, E. O., & Hanski, C. (2001). Regulation of the intestinal mucin MUC2 gene expression in vivo: evidence for the role of promoter methylation. *Cancer Lett*, *168*(1), 71-80. https://doi.org/10.1016/s0304-3835(01)00498-0

Green, J. B., Timmcke, A. E., Mitchell, W. T., Hicks, T. C., Gathright, J. B., Jr., & Ray, J. E. (1993). Mucinous carcinoma--just another colon cancer? *Dis Colon Rectum*, *36*(1), 49-54. https://doi.org/10.1007/bf02050301

Gregorieff, A., Stange, D. E., Kujala, P., Begthel, H., van den Born, M., Korving, J., Peters, P. J., & Clevers, H. (2009). The ets-domain transcription factor Spdef promotes maturation of goblet and paneth cells in the intestinal epithelium. *Gastroenterology*, *137*(4), 1333-1345.e1331-1333. https://doi.org/10.1053/j.gastro.2009.06.044

Gronbaek, H., Vestergaard, E. M., Hey, H., Nielsen, J. N., & Nexo, E. (2006). Serum trefoil factors in patients with inflammatory bowel disease. *Digestion*, *74*(1), 33-39. https://doi.org/10.1159/000096591

Grondin, J. A., Kwon, Y. H., Far, P. M., Haq, S., & Khan, W. I. (2020). Mucins in Intestinal Mucosal Defense and Inflammation: Learning From Clinical and Experimental Studies. *Front Immunol*, *11*, 2054. https://doi.org/10.3389/fimmu.2020.02054

Gundamaraju, R., & Chong, W. C. (2021). Consequence of distinctive expression of MUC2 in colorectal cancers: How much is actually bad? *Biochim Biophys Acta Rev Cancer*, *1876*(1), 188579. https://doi.org/10.1016/j.bbcan.2021.188579

Guo, R. J., Funakoshi, S., Lee, H. H., Kong, J., & Lynch, J. P. (2010). The intestine-specific transcription factor Cdx2 inhibits beta-catenin/TCF transcriptional activity by disrupting the beta-catenin-TCF protein complex. *Carcinogenesis*, *31*(2), 159-166. https://doi.org/10.1093/carcin/bgp213

Gustafsson, J. K., Davis, J. E., Rappai, T., McDonald, K. G., Kulkarni, D. H., Knoop, K. A., Hogan, S. P., Fitzpatrick, J. A., Lencer, W. I., & Newberry, R. D. (2021). Intestinal goblet cells sample and deliver lumenal antigens by regulated endocytic uptake and transcytosis. *Elife*, *10*. https://doi.org/10.7554/eLife.67292

Gustafsson, J. K., Ermund, A., Johansson, M. E., Schütte, A., Hansson, G. C., & Sjövall, H. (2012). An ex vivo method for studying mucus formation, properties, and thickness in human colonic biopsies and mouse small and large intestinal explants. *Am J Physiol Gastrointest Liver Physiol*, *302*(4), G430-438. https://doi.org/10.1152/ajpgi.00405.2011

Gustafsson, J. K., & Johansson, M. E. V. (2022). The role of goblet cells and mucus in intestinal homeostasis. *Nat Rev Gastroenterol Hepatol*, *19*(12), 785-803. https://doi.org/10.1038/s41575-022-00675-x

Habowski, A. N., Flesher, J. L., Bates, J. M., Tsai, C. F., Martin, K., Zhao, R., Ganesan, A. K., Edwards, R. A., Shi, T., Wiley, H. S., Shi, Y., Hertel, K. J., & Waterman, M. L. (2020). Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Commun Biol*, *3*(1), 453. https://doi.org/10.1038/s42003-020-01181-z

Halm, D. R., & Halm, S. T. (2000). Secretagogue response of goblet cells and columnar cells in human colonic crypts. *Am J Physiol Cell Physiol*, *278*(1), C212-233. https://doi.org/10.1152/ajpcell.2000.278.1.C212

Hansson, G. C. (2019). Mucus and mucins in diseases of the intestinal and respiratory tracts. *J Intern Med*, *285*(5), 479-490. https://doi.org/10.1111/joim.12910

Hansson, G. C. (2020). Mucins and the Microbiome. *Annu Rev Biochem*, *89*, 769-793. https://doi.org/10.1146/annurev-biochem-011520-105053

Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*, *9*(1), 75. https://doi.org/10.1186/s13073-017-0467-4

Haveri, H., Westerholm-Ormio, M., Lindfors, K., Mäki, M., Savilahti, E., Andersson, L. C., & Heikinheimo, M. (2008). Transcription factors GATA-4 and GATA-6 in normal and neoplastic human gastrointestinal mucosa. *BMC Gastroenterol*, *8*, 9. https://doi.org/10.1186/1471-230x-8-9

Hermiston, M. L., Xu, Z., & Weiss, A. (2003). CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol*, *21*, 107-137. https://doi.org/10.1146/annurev.immunol.21.120601.140946

Hinoi, T., Tani, M., Lucas, P. C., Caca, K., Dunn, R. L., Macri, E., Loda, M., Appelman, H. D., Cho, K. R., & Fearon, E. R. (2001). Loss of CDX2 expression and microsatellite

instability are prominent features of large cell minimally differentiated carcinomas of the colon. *Am J Pathol*, *159*(6), 2239-2248. https://doi.org/10.1016/s0002-9440(10)63074-x
Hisamuddin, I. M., & Yang, V. W. (2006). Molecular Genetics of Colorectal Cancer: An Overview. *Curr Colorectal Cancer Rep*, *2*(2), 53-59. https://doi.org/10.1007/s11888-006-0002-2
Hollingsworth, M. A., & Swanson, B. J. (2004). Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer*, *4*(1), 45-60. https://doi.org/10.1038/nrc1251
Hossain, M. S., Karuniawati, H., Jairoun, A. A., Urbi, Z., Ooi, J., John, A., Lim, Y. C., Kibria, K. M. K., Mohiuddin, A. K. M., Ming, L. C., Goh, K. W., & Hadi, M. A. (2022). Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. *Cancers (Basel)*, *14*(7). https://doi.org/10.3390/cancers14071732
Hu, F. J., Li, Y. J., Zhang, L., Ji, D. B., Liu, X. Z., Chen, Y. J., Wang, L., & Wu, A. W. (2023). Single-cell profiling reveals differences between human classical adenocarcinoma and mucinous adenocarcinoma. *Commun Biol*, *6*(1), 85. https://doi.org/10.1038/s42003-023-04441-w
Huang, A., Yang, Y., Shi, J. Y., Li, Y. K., Xu, J. X., Cheng, Y., & Gu, J. (2021). Mucinous adenocarcinoma: A unique clinicopathological subtype in colorectal cancer. *World J Gastrointest Surg*, *13*(12), 1567-1583. https://doi.org/10.4240/wjgs.v13.i12.1567
Huang, C. Z., Yu, T., & Chen, Q. K. (2015). DNA Methylation Dynamics During Differentiation, Proliferation, and Tumorigenesis in the Intestinal Tract. *Stem Cells Dev*, *24*(23), 2733-2739. https://doi.org/10.1089/scd.2015.0235
Hull, R., Francies, F. Z., Oyomno, M., & Dlamini, Z. (2020). Colorectal Cancer Genetics, Incidence and Risk Factors: In Search for Targeted Therapies. *Cancer Manag Res*, *12*, 9869-9882. https://doi.org/10.2147/CMAR.S251223
Iacopetta, B. (2002). Are there two sides to colorectal cancer? *Int J Cancer*, *101*(5), 403-408. https://doi.org/10.1002/ijc.10635
Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., Aben, N., Goncalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P., van Dyk, E., Chang, H., de Silva, H., Heyn, H., Deng, X., Egan, R. K., Liu, Q., . . . Garnett, M. J. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, *166*(3), 740-754. https://doi.org/10.1016/j.cell.2016.06.017
Ishibashi, F., Shimizu, H., Nakata, T., Fujii, S., Suzuki, K., Kawamoto, A., Anzai, S., Kuno, R., Nagata, S., Ito, G., Murano, T., Mizutani, T., Oshima, S., Tsuchiya, K., Nakamura, T., Watanabe, M., & Okamoto, R. (2018). Contribution of ATOH1(+) Cells to the Homeostasis, Repair, and Tumorigenesis of the Colonic Epithelium. *Stem Cell Reports*, *10*(1), 27-42. https://doi.org/10.1016/j.stemcr.2017.11.006
Ito, G., Okamoto, R., Murano, T., Shimizu, H., Fujii, S., Nakata, T., Mizutani, T., Yui, S., Akiyama-Morio, J., Nemoto, Y., Okada, E., Araki, A., Ohtsuka, K., Tsuchiya, K., Nakamura, T., & Watanabe, M. (2013). Lineage-specific expression of bestrophin-2 and bestrophin-4 in human intestinal epithelial cells. *PLoS One*, *8*(11), e79693. https://doi.org/10.1371/journal.pone.0079693
Jarva, M. A., Lingford, J. P., John, A., Soler, N. M., Scott, N. E., & Goddard-Borger, E. D. (2020). Trefoil factors share a lectin activity that defines their role in mucus. *Nat Commun*, *11*(1), 2265. https://doi.org/10.1038/s41467-020-16223-7
Johansson, M. E., & Hansson, G. C. (2016). Immunological aspects of intestinal mucus and mucins. *Nat Rev Immunol*, *16*(10), 639-649. https://doi.org/10.1038/nri.2016.88
Johansson, M. E., Larsson, J. M., & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A*, *108 Suppl 1*(Suppl 1), 4659-4665. https://doi.org/10.1073/pnas.1006451107
Johansson, M. E., Sjovall, H., & Hansson, G. C. (2013). The gastrointestinal mucus system in health and disease. *Nat Rev Gastroenterol Hepatol*, *10*(6), 352-361. https://doi.org/10.1038/nrgastro.2013.35

Johansson, M. E. V., Thomsson, K. A., & Hansson, G. C. (2009). Proteomic Analyses of the Two Mucus Layers of the Colon Barrier Reveal That Their Main Component, the Muc2 Mucin, Is Strongly Bound to the Fcgbp Protein. *Journal of Proteome Research*, *8*(7), 3549-3557. https://doi.org/10.1021/pr9002504

Johnstone, C. N., White, S. J., Tebbutt, N. C., Clay, F. J., Ernst, M., Biggs, W. H., Viars, C. S., Czekay, S., Arden, K. C., & Heath, J. K. (2002). Analysis of the regulation of the A33 antigen gene reveals intestine-specific mechanisms of gene expression. *J Biol Chem*, *277*(37), 34531-34539. https://doi.org/10.1074/jbc.M204865200

Jung, P., Sommer, C., Barriga, F. M., Buczacki, S. J., Hernando-Momblona, X., Sevillano, M., Duran-Frigola, M., Aloy, P., Selbach, M., Winton, D. J., & Batlle, E. (2015). Isolation of Human Colon Stem Cells Using Surface Expression of PTK7. *Stem Cell Reports*, *5*(6), 979-987. https://doi.org/10.1016/j.stemcr.2015.10.003

Jung, T. H., Park, J. H., Jeon, W. M., & Han, K. S. (2015). Butyrate modulates bacterial adherence on LS174T human colorectal cells by stimulating mucin secretion and MAPK signaling pathway. *Nutr Res Pract*, *9*(4), 343-349. https://doi.org/10.4162/nrp.2015.9.4.343

Kachlik, D., Baca, V., & Stingl, J. (2010). The spatial arrangement of the human large intestinal wall blood circulation. *J Anat*, *216*(3), 335-343. https://doi.org/10.1111/j.1469-7580.2009.01199.x

Kahai., P., Mandiga., P., Wehrle., C. J., & Lobo., S. (2022). *Anatomy, Abdomen and Pelvis: Large Intestine* (S. [Internet], Ed.). Treasure Island (FL): StatPearls Publishing; 2023 Jan–. https://www.ncbi.nlm.nih.gov/books/NBK470577/

Kamphuis, J. B. J., Mercier-Bonin, M., Eutamene, H., & Theodorou, V. (2017). Mucus organisation is shaped by colonic content; a new view. *Sci Rep*, *7*(1), 8527. https://doi.org/10.1038/s41598-017-08938-3

Kanke, M., Kennedy Ng, M. M., Connelly, S., Singh, M., Schaner, M., Shanahan, M. T., Wolber, E. A., Beasley, C., Lian, G., Jain, A., Long, M. D., Barnes, E. L., Herfarth, H. H., Isaacs, K. L., Hansen, J. J., Kapadia, M., Guillem, J. G., Feschotte, C., Furey, T. S., . . . Sethupathy, P. (2022). Single-Cell Analysis Reveals Unexpected Cellular Changes and Transposon Expression Signatures in the Colonic Epithelium of Treatment-Naïve Adult Crohn's Disease Patients. *Cell Mol Gastroenterol Hepatol*, *13*(6), 1717-1740. https://doi.org/10.1016/j.jcmgh.2022.02.005

Kazanjian, A., Noah, T., Brown, D., Burkart, J., & Shroyer, N. F. (2010). Atonal Homolog 1 Is Required for Growth and Differentiation Effects of Notch/γ-Secretase Inhibitors on Normal and Cancerous Intestinal Epithelial Cells. *Gastroenterology*, *139*(3), 918-928.e916. https://doi.org/https://doi.org/10.1053/j.gastro.2010.05.081

Keller, M. S., Ezaki, T., Guo, R. J., & Lynch, J. P. (2004). Cdx1 or Cdx2 expression activates E-cadherin-mediated cell-cell adhesion and compaction in human COLO 205 cells. *Am J Physiol Gastrointest Liver Physiol*, *287*(1), G104-114. https://doi.org/10.1152/ajpgi.00484.2003

Kerr, D. J., Dunn, J. A., Langman, M. J., Smith, J. L., Midgley, R. S., Stanley, A., Stokes, J. C., Julier, P., Iveson, C., Duvvuri, R., & McConkey, C. C. (2007). Rofecoxib and cardiovascular adverse events in adjuvant treatment of colorectal cancer. *N Engl J Med*, *357*(4), 360-369. https://doi.org/10.1056/NEJMoa071841

Kim, J. J., & Khan, W. I. (2013). Goblet cells and mucins: role in innate defense in enteric infections. *Pathogens*, *2*(1), 55-70. https://doi.org/10.3390/pathogens2010055

Kim, Y. S., & Deng, G. (2008). Aberrant expression of carbohydrate antigens in cancer: the role of genetic and epigenetic regulation. *Gastroenterology*, *135*(1), 305-309. https://doi.org/10.1053/j.gastro.2008.06.013

Kim, Y. S., & Ho, S. B. (2010). Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Curr Gastroenterol Rep*, *12*(5), 319-330. https://doi.org/10.1007/s11894-010-0131-2

Kjellev, S. (2009). The trefoil factor family - small peptides with multiple functionalities. *Cell Mol Life Sci*, *66*(8), 1350-1369. https://doi.org/10.1007/s00018-008-8646-5

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., & Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187-1201. https://doi.org/10.1016/j.cell.2015.04.044

Kleppe, A., Skrede, O. J., De Raedt, S., Hveem, T. S., Askautrud, H. A., Jacobsen, J. E., Church, D. N., Nesbakken, A., Shepherd, N. A., Novelli, M., Kerr, R., Liestøl, K., Kerr, D. J., & Danielsen, H. E. (2022). A clinical decision support system optimising adjuvant chemotherapy for colorectal cancers by integrating deep learning and pathological staging markers: a development and validation study. *Lancet Oncol*, *23*(9), 1221-1232. https://doi.org/10.1016/s1470-2045(22)00391-6

Kleppe, A., Skrede, O. J., De Raedt, S., Liestøl, K., Kerr, D. J., & Danielsen, H. E. (2021). Designing deep learning studies in cancer diagnostics. *Nat Rev Cancer*, *21*(3), 199-211. https://doi.org/10.1038/s41568-020-00327-9

Kolev, H. M., & Kaestner, K. H. (2023). Mammalian Intestinal Development and Differentiation-The State of the Art. *Cell Mol Gastroenterol Hepatol*, *16*(5), 809-821. https://doi.org/10.1016/j.jcmgh.2023.07.011

Kong, S., Zhang, Y. H., & Zhang, W. (2018). Regulation of Intestinal Epithelial Cells Properties and Functions by Amino Acids. *Biomed Res Int*, *2018*, 2819154. https://doi.org/10.1155/2018/2819154

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., & Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. *Development*, *126*(9), 723-732.

Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., van de Velde, C. J., & Watanabe, T. (2015). Colorectal cancer. *Nat Rev Dis Primers*, *1*, 15065. https://doi.org/10.1038/nrdp.2015.65

Laudisi, F., Stolfi, C., Bevivino, G., Maresca, C., Franzè, E., Troncone, E., Lolli, E., Marafini, I., Pietrucci, D., Teofani, A., Di Grazia, A., Di Fusco, D., Colantoni, A., Ortenzi, A., Desideri, A., Monteleone, I., & Monteleone, G. (2022). GATA6 Deficiency Leads to Epithelial Barrier Dysfunction and Enhances Susceptibility to Gut Inflammation. *J Crohns Colitis*, *16*(2), 301-311. https://doi.org/10.1093/ecco-jcc/jjab145

Leow, C. C., Romero, M. S., Ross, S., Polakis, P., & Gao, W. Q. (2004). Hath1, down-regulated in colon adenocarcinomas, inhibits proliferation and tumorigenesis of colon cancer cells. *Cancer Res*, *64*(17), 6050-6057. https://doi.org/10.1158/0008-5472.Can-04-0290

Li, C., Zuo, D., Yin, L., Lin, Y., Li, C., Liu, T., & Wang, L. (2018). Prognostic Value of MUC2 Expression in Colorectal Cancer: A Systematic Review and Meta-Analysis. *Gastroenterol Res Pract*, *2018*, 6986870. https://doi.org/10.1155/2018/6986870

Li, J., Ma, X., Chakravarti, D., Shalapour, S., & DePinho, R. A. (2021). Genetic and biological hallmarks of colorectal cancer. *Genes Dev*, *35*(11-12), 787-820. https://doi.org/10.1101/gad.348226.120

Li, X., Hu, W., Zhou, J., Huang, Y., Peng, J., Yuan, Y., Yu, J., & Zheng, S. (2017). CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway. *Cell Commun Signal*, *15*(1), 38. https://doi.org/10.1186/s12964-017-0192-z

Liang, S. J., Li, X. G., & Wang, X. Q. (2019). Notch Signaling in Mammalian Intestinal Stem Cells: Determining Cell Fate and Maintaining Homeostasis. *Curr Stem Cell Res Ther*, *14*(7), 583-590. https://doi.org/10.2174/1574888x14666190429143734

Liu, H. (2017). *Goblet cell differentiation in human colorectal cancer cell lines* [DPhil thesis, University of Oxford (United Kingdom). https://ora.ox.ac.uk/objects/uuid:d907bbf2-76fc-409a-877c-529fc04ca042

Liu, Q., Niu, X., Li, Y., Zhang, J. R., Zhu, S. J., Yang, Q. Y., Zhang, W., & Gong, L. (2022). Role of the mucin-like glycoprotein FCGBP in mucosal immunity and cancer. *Front Immunol*, *13*, 863317. https://doi.org/10.3389/fimmu.2022.863317

Liu, T. C. (2021). *Colorectal cancer: The EMT transcription factor ZEB1, the interaction between myofibroblast and colorectal cancer cells and GMM analysis of cancer gene*

expression. [DPhil thesis, University of Oxford (United Kingdom). http://ora.ox.ac.uk/objects/uuid:b8ef534a-a9a7-4888-bbb2-1a50297bbec1
Liu, T. C., Kalugin, P. N., Wilding, J. L., & Bodmer, W. F. (2022). GMMchi: gene expression clustering using Gaussian mixture modeling. *BMC Bioinformatics*, *23*(1), 457. https://doi.org/10.1186/s12859-022-05006-0
Liu, Y., Yu, Z., Zhu, L., Ma, S., Luo, Y., Liang, H., Liu, Q., Chen, J., Guli, S., & Chen, X. (2023). Orchestration of MUC2 - The key regulatory target of gut barrier and homeostasis: A review. *Int J Biol Macromol*, *236*, 123862. https://doi.org/10.1016/j.ijbiomac.2023.123862
Lo, Y. H., Chung, E., Li, Z., Wan, Y. W., Mahe, M. M., Chen, M. S., Noah, T. K., Bell, K. N., Yalamanchili, H. K., Klisch, T. J., Liu, Z., Park, J. S., & Shroyer, N. F. (2017). Transcriptional Regulation by ATOH1 and its Target SPDEF in the Intestine. *Cell Mol Gastroenterol Hepatol*, *3*(1), 51-71. https://doi.org/10.1016/j.jcmgh.2016.10.001
Lueschow, S. R., & McElroy, S. J. (2020). The Paneth Cell: The Curator and Defender of the Immature Small Intestine. *Front Immunol*, *11*, 587. https://doi.org/10.3389/fimmu.2020.00587
Luo, C., Cen, S., Ding, G., & Wu, W. (2019). Mucinous colorectal adenocarcinoma: clinical pathology and treatment options. *Cancer Commun (Lond)*, *39*(1), 13. https://doi.org/10.1186/s40880-019-0361-0
Madsen, J., Nielsen, O., Tornøe, I., Thim, L., & Holmskov, U. (2007). Tissue localization of human trefoil factors 1, 2, and 3. *J Histochem Cytochem*, *55*(5), 505-513. https://doi.org/10.1369/jhc.6A7100.2007
Madsen, J., Sorensen, G. L., Nielsen, O., Tornøe, I., Thim, L., Fenger, C., Mollenhauer, J., & Holmskov, U. (2013). A variant form of the human deleted in malignant brain tumor 1 (DMBT1) gene shows increased expression in inflammatory bowel diseases and interacts with dimeric trefoil factor 3 (TFF3). *PLoS One*, *8*(5), e64441. https://doi.org/10.1371/journal.pone.0064441
Mashimo, H., Wu, D.-C., Podolsky, D. K., & Fishman, M. C. (1996). Impaired Defense of Intestinal Mucosa in Mice Lacking Intestinal Trefoil Factor. *Science*, *274*(5285), 262-265. https://doi.org/doi:10.1126/science.274.5285.262
Meng, H., Li, W., Boardman, L. A., & Wang, L. (2018). Loss of ZG16 is associated with molecular and clinicopathological phenotypes of colorectal cancer. *BMC Cancer*, *18*(1), 433. https://doi.org/10.1186/s12885-018-4337-2
Meran, L., Baulies, A., & Li, V. S. W. (2017). Intestinal Stem Cell Niche: The Extracellular Matrix and Cellular Components. *Stem Cells Int*, *2017*, 7970385. https://doi.org/10.1155/2017/7970385
Mesquita, P., Jonckheere, N., Almeida, R., Ducourouble, M. P., Serpa, J., Silva, E., Pigny, P., Silva, F. S., Reis, C., Silberg, D., Van Seuningen, I., & David, L. (2003). Human MUC2 mucin gene is transcriptionally regulated by Cdx homeodomain proteins in gastrointestinal carcinoma cell lines. *J Biol Chem*, *278*(51), 51549-51556. https://doi.org/10.1074/jbc.M309019200
Midgley, R. S., McConkey, C. C., Johnstone, E. C., Dunn, J. A., Smith, J. L., Grumett, S. A., Julier, P., Iveson, C., Yanagisawa, Y., Warren, B., Langman, M. J., & Kerr, D. J. (2010). Phase III randomized trial assessing rofecoxib in the adjuvant setting of colorectal cancer: final results of the VICTOR trial. *J Clin Oncol*, *28*(30), 4575-4580. https://doi.org/10.1200/jco.2010.29.6244
Milla, P. J. (2009). Advances in Understanding Colonic Function. *Journal of Pediatric Gastroenterology and Nutrition 48():p S43-S45*. https://doi.org/10.1097/MPG.0b013e3181a1171a
Miller, S. A., Ghobashi, A. H., & O'Hagan, H. M. (2021). Consensus molecular subtyping of colorectal cancers is influenced by goblet cell content. *Cancer Genet*, *254-255*, 34-39. https://doi.org/10.1016/j.cancergen.2021.01.009
Mirabelli, P., Coppola, L., & Salvatore, M. (2019). Cancer Cell Lines Are Useful Model Systems for Medical Research. *Cancers (Basel)*, *11*(8). https://doi.org/10.3390/cancers11081098

Mizoshita, T., Tsukamoto, T., Inada, K. I., Hirano, N., Tajika, M., Nakamura, T., Ban, H., & Tatematsu, M. (2007). Loss of MUC2 expression correlates with progression along the adenoma-carcinoma sequence pathway as well as de novo carcinogenesis in the colon. *Histol Histopathol*, *22*(3), 251-260. https://doi.org/10.14670/HH-22.251

Mouradov, D., Domingo, E., Gibbs, P., Jorissen, R. N., Li, S., Soo, P. Y., Lipton, L., Desai, J., Danielsen, H. E., Oukrif, D., Novelli, M., Yau, C., Holmes, C. C., Jones, I. T., McLaughlin, S., Molloy, P., Hawkins, N. J., Ward, R., Midgely, R., . . . Sieber, O. M. (2013). Survival in stage II/III colorectal cancer is independently predicted by chromosomal and microsatellite instability, but not by specific driver mutations. *Am J Gastroenterol*, *108*(11), 1785-1793. https://doi.org/10.1038/ajg.2013.292

Mouradov, D., Sloggett, C., Jorissen, R. N., Love, C. G., Li, S., Burgess, A. W., Arango, D., Strausberg, R. L., Buchanan, D., Wormald, S., O'Connor, L., Wilding, J. L., Bicknell, D., Tomlinson, I. P., Bodmer, W. F., Mariadason, J. M., & Sieber, O. M. (2014). Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res*, *74*(12), 3238-3247. https://doi.org/10.1158/0008-5472.Can-14-0013

Moutel, S., Beugnet, A., Schneider, A., Lombard, B., Loew, D., Amigorena, S., Perez, F., & Segura, E. (2020). Surface LSP-1 Is a Phenotypic Marker Distinguishing Human Classical versus Monocyte-Derived Dendritic Cells. *iScience*, *23*(4), 100987. https://doi.org/10.1016/j.isci.2020.100987

Múnera, J., Ceceña, G., Jedlicka, P., Wankell, M., & Oshima, R. G. (2011). Ets2 regulates colonic stem cells and sensitivity to tumorigenesis. *Stem Cells*, *29*(3), 430-439. https://doi.org/10.1002/stem.599

Nawa, T., Kato, J., Kawamoto, H., Okada, H., Yamamoto, H., Kohno, H., Endo, H., & Shiratori, Y. (2008). Differences between right- and left-sided colon cancer in patient characteristics, cancer morphology and histology. *J Gastroenterol Hepatol*, *23*(3), 418-423. https://doi.org/10.1111/j.1440-1746.2007.04923.x

Nguyen, L. H., Goel, A., & Chung, D. C. (2020). Pathways of Colorectal Carcinogenesis. *Gastroenterology*, *158*(2), 291-302. https://doi.org/10.1053/j.gastro.2019.08.059

Noah, T. K., Donahue, B., & Shroyer, N. F. (2011). Intestinal development and differentiation. *Exp Cell Res*, *317*(19), 2702-2710. https://doi.org/10.1016/j.yexcr.2011.09.006

Noah, T. K., Kazanjian, A., Whitsett, J., & Shroyer, N. F. (2010). SAM pointed domain ETS factor (SPDEF) regulates terminal differentiation and maturation of intestinal goblet cells. *Exp Cell Res*, *316*(3), 452-465. https://doi.org/10.1016/j.yexcr.2009.09.020

Nyström, E. E. L., Arike, L., Ehrencrona, E., Hansson, G. C., & Johansson, M. E. V. (2019). Calcium-activated chloride channel regulator 1 (CLCA1) forms non-covalent oligomers in colonic mucus and has mucin 2-processing properties. *J Biol Chem*, *294*(45), 17075-17089. https://doi.org/10.1074/jbc.RA119.009940

Nyström, E. E. L., Birchenough, G. M. H., van der Post, S., Arike, L., Gruber, A. D., Hansson, G. C., & Johansson, M. E. V. (2018). Calcium-activated chloride channel regulator 1 (CLCA1) controls mucus expansion in colon by proteolytic activity. *EBioMedicine*, *33*, 134-143. https://doi.org/10.1016/j.ebiom.2018.05.031

Okudaira, K., Kakar, S., Cun, L., Choi, E., Wu Decamillis, R., Miura, S., Sleisenger, M. H., Kim, Y. S., & Deng, G. (2010). MUC2 gene promoter methylation in mucinous and non-mucinous colorectal cancer tissues. *Int J Oncol*, *36*(4), 765-775. https://doi.org/10.3892/ijo_00000552

Okumura, R., Kurakawa, T., Nakano, T., Kayama, H., Kinoshita, M., Motooka, D., Gotoh, K., Kimura, T., Kamiyama, N., Kusu, T., Ueda, Y., Wu, H., Iijima, H., Barman, S., Osawa, H., Matsuno, H., Nishimura, J., Ohba, Y., Nakamura, S., . . . Takeda, K. (2016). Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. *Nature*, *532*(7597), 117-121. https://doi.org/10.1038/nature17406

Okumura, R., & Takeda, K. (2017). Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Exp Mol Med*, *49*(5), e338. https://doi.org/10.1038/emm.2017.20

Pandey, V., Zhang, X., Poh, H. M., Wang, B., Dukanya, D., Ma, L., Yin, Z., Bender, A., Periyasamy, G., Zhu, T., Rangappa, K. S., Basappa, B., & Lobie, P. E. (2022). Monomerization of Homodimeric Trefoil Factor 3 (TFF3) by an Aminonitrile Compound Inhibits TFF3-Dependent Cancer Cell Survival. *ACS Pharmacol Transl Sci*, *5*(9), 761-773. https://doi.org/10.1021/acsptsci.2c00044

Paone, P., & Cani, P. D. (2020). Mucus barrier, mucins and gut microbiota: the expected slimy partners? *Gut*, *69*(12), 2232-2243. https://doi.org/10.1136/gutjnl-2020-322260

Parikh, K., Antanaviciute, A., Fawkner-Corbett, D., Jagielowicz, M., Aulicino, A., Lagerholm, C., Davis, S., Kinchen, J., Chen, H. H., Alham, N. K., Ashley, N., Johnson, E., Hublitz, P., Bao, L., Lukomska, J., Andev, R. S., Björklund, E., Kessler, B. M., Fischer, R., . . . Simmons, A. (2019). Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature*, *567*(7746), 49-55. https://doi.org/10.1038/s41586-019-0992-y

Peignon, G., Durand, A., Cacheux, W., Ayrault, O., Terris, B., Laurent-Puig, P., Shroyer, N. F., Van Seuningen, I., Honjo, T., Perret, C., & Romagnolo, B. (2011). Complex interplay between β-catenin signalling and Notch effectors in intestinal tumorigenesis. *Gut*, *60*(2), 166-176. https://doi.org/10.1136/gut.2009.204719

Pellegrinet, L., Rodilla, V., Liu, Z., Chen, S., Koch, U., Espinosa, L., Kaestner, K. H., Kopan, R., Lewis, J., & Radtke, F. (2011). Dll1- and dll4-mediated notch signaling are required for homeostasis of intestinal stem cells. *Gastroenterology*, *140*(4), 1230-1240.e1231-1237. https://doi.org/10.1053/j.gastro.2011.01.005

Pendlebury, S., Duchesne, F., Reed, K. A., Smith, J. L., & Kerr, D. J. (2003). A trial of adjuvant therapy in colorectal cancer: the VICTOR trial. *Clin Colorectal Cancer*, *3*(1), 58-60. https://doi.org/10.3816/CCC.2003.n.013

Podolsky, D. K., Lynch-Devaney, K., Stow, J. L., Oates, P., Murgue, B., DeBeaumont, M., Sands, B. E., & Mahida, Y. R. (1993). Identification of human intestinal trefoil factor. Goblet cell-specific expression of a peptide targeted for apical secretion. *J Biol Chem*, *268*(9), 6694-6702.

Qi, F., Isaji, T., Duan, C., Yang, J., Wang, Y., Fukuda, T., & Gu, J. (2020). ST3GAL3, ST3GAL4, and ST3GAL6 differ in their regulation of biological functions via the specificities for the α2,3-sialylation of target proteins. *Faseb j*, *34*(1), 881-897. https://doi.org/10.1096/fj.201901793R

Ragulan, C., Eason, K., Fontana, E., Nyamundanda, G., Tarazona, N., Patil, Y., Poudel, P., Lawlor, R. T., Del Rio, M., Koo, S. L., Tan, W. S., Sclafani, F., Begum, R., Teixeira Mendes, L. S., Martineau, P., Scarpa, A., Cervantes, A., Tan, I. B., Cunningham, D., & Sadanandam, A. (2019). Analytical Validation of Multiplex Biomarker Assay to Stratify Colorectal Cancer into Molecular Subtypes. *Sci Rep*, *9*(1), 7665. https://doi.org/10.1038/s41598-019-43492-0

Rebuzzi, F., Ulivi, P., & Tedaldi, G. (2023). Genetic Predisposition to Colorectal Cancer: How Many and Which Genes to Test? *Int J Mol Sci*, *24*(3). https://doi.org/10.3390/ijms24032137

Reedijk, M., Odorcic, S., Zhang, H., Chetty, R., Tennert, C., Dickson, B. C., Lockwood, G., Gallinger, S., & Egan, S. E. (2008). Activation of Notch signaling in human colon adenocarcinoma. *Int J Oncol*, *33*(6), 1223-1229. https://doi.org/10.3892/ijo_00000112

Reynolds, I. S., O'Connell, E., Fichtner, M., McNamara, D. A., Kay, E. W., Prehn, J. H. M., Furney, S. J., & Burke, J. P. (2020). Mucinous adenocarcinoma is a pharmacogenomically distinct subtype of colorectal cancer. *Pharmacogenomics J*, *20*(3), 524-532. https://doi.org/10.1038/s41397-019-0137-6

Ribeirinho-Soares, S., Pádua, D., Amaral, A. L., Valentini, E., Azevedo, D., Marques, C., Barros, R., Macedo, F., Mesquita, P., & Almeida, R. (2021). Prognostic significance of MUC2, CDX2 and SOX2 in stage II colorectal cancer patients. *BMC Cancer*, *21*(1), 359. https://doi.org/10.1186/s12885-021-08070-6

Ribieras, S., Tomasetto, C., & Rio, M. C. (1998). The pS2/TFF1 trefoil factor, from basic research to clinical applications. *Biochim Biophys Acta*, *1378*(1), F61-77. https://doi.org/10.1016/s0304-419x(98)00016-x

Richman, P. I., & Bodmer, W. F. (1988). Control of differentiation in human colorectal carcinoma cell lines: epithelial-mesenchymal interactions. *J Pathol*, *156*(3), 197-211. https://doi.org/10.1002/path.1711560305

Rivat, C., Rodrigues, S., Bruyneel, E., Piétu, G., Robert, A., Redeuilh, G., Bracke, M., Gespach, C., & Attoub, S. (2005). Implication of STAT3 signaling in human colonic cancer cells during intestinal trefoil factor 3 (TFF3) - and vascular endothelial growth factor-mediated cellular invasion and tumor growth. *Cancer Res*, *65*(1), 195-202.

Robinson, B. S., Saeedi, B., Arthur, C. M., Owens, J., Naudin, C., Ahmed, N., Luo, L., Jones, R., Neish, A., & Stowell, S. R. (2020). Galectin-9 Is a Novel Regulator of Epithelial Restitution. *Am J Pathol*, *190*(8), 1657-1666. https://doi.org/10.1016/j.ajpath.2020.04.010

Rodilla, V., Villanueva, A., Obrador-Hevia, A., Robert-Moreno, À., Fernández-Majada, V., Grilli, A., López-Bigas, N., Bellora, N., Albà, M. M., Torres, F., Duñach, M., Sanjuan, X., Gonzalez, S., Gridley, T., Capella, G., Bigas, A., & Espinosa, L. (2009). Jagged1 is the pathological link between Wnt and Notch pathways in colorectal cancer. *Proceedings of the National Academy of Sciences*, *106*(15), 6315-6320. https://doi.org/doi:10.1073/pnas.0813221106

Rodriguez-Salas, N., Dominguez, G., Barderas, R., Mendiola, M., García-Albéniz, X., Maurel, J., & Batlle, J. F. (2017). Clinical relevance of colorectal cancer molecular subtypes. *Crit Rev Oncol Hematol*, *109*, 9-19. https://doi.org/10.1016/j.critrevonc.2016.11.007

Sadanandam, A., Lyssiotis, C. A., Homicsko, K., Collisson, E. A., Gibb, W. J., Wullschleger, S., Ostos, L. C., Lannon, W. A., Grotzinger, C., Del Rio, M., Lhermitte, B., Olshen, A. B., Wiedenmann, B., Cantley, L. C., Gray, J. W., & Hanahan, D. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med*, *19*(5), 619-625. https://doi.org/10.1038/nm.3175

Sawicki, T., Ruszkowska, M., Danielewicz, A., Niedzwiedzka, E., Arlukowicz, T., & Przybylowicz, K. E. (2021). A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms and Diagnosis. *Cancers (Basel)*, *13*(9). https://doi.org/10.3390/cancers13092025

Schaefer, K., Webb, N. E., Pang, M., Hernandez-Davies, J. E., Lee, K. P., Gonzalez, P., Douglass, M. V., Lee, B., & Baum, L. G. (2017). Galectin-9 binds to O-glycans on protein disulfide isomerase. *Glycobiology*, *27*(9), 878-887. https://doi.org/10.1093/glycob/cwx065

Schroeder, B. O. (2019). Fight them or feed them: how the intestinal mucus layer manages the gut microbiota. *Gastroenterol Rep (Oxf)*, *7*(1), 3-12. https://doi.org/10.1093/gastro/goy052

Shaik, J. P., Alanazi, I. O., Pathan, A. A. K., Parine, N. R., Almadi, M. A., Azzam, N. A., Aljebreen, A. M., Alharbi, O., Alanazi, M. S., & Khan, Z. (2020). Frequent Activation of Notch Signaling Pathway in Colorectal Cancers and Its Implication in Patient Survival Outcome. *J Oncol*, *2020*, 6768942. https://doi.org/10.1155/2020/6768942

Sheaffer, K. L., Kim, R., Aoki, R., Elliott, E. N., Schug, J., Burger, L., Schübeler, D., & Kaestner, K. H. (2014). DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes Dev*, *28*(6), 652-664. https://doi.org/10.1101/gad.230318.113

Sigurdsson, H. H., Kirch, J., & Lehr, C. M. (2013). Mucus as a barrier to lipophilic drugs. *Int J Pharm*, *453*(1), 56-64. https://doi.org/10.1016/j.ijpharm.2013.05.040

Silberg, D. G., Furth, E. E., Taylor, J. K., Schuck, T., Chiou, T., & Traber, P. G. (1997). CDX1 protein expression in normal, metaplastic, and neoplastic human alimentary tract epithelium. *Gastroenterology*, *113*(2), 478-486. https://doi.org/10.1053/gast.1997.v113.pm9247467

Silberg, D. G., Swain, G. P., Suh, E. R., & Traber, P. G. (2000). Cdx1 and cdx2 expression during intestinal development. *Gastroenterology*, *119*(4), 961-971. https://doi.org/10.1053/gast.2000.18142

Simmonds, N., Furman, M., Karanika, E., Phillips, A., & Bates, A. W. H. (2014). Paneth cell metaplasia in newly diagnosed inflammatory bowel disease in children. *BMC Gastroenterology*, *14*(1), 93. https://doi.org/10.1186/1471-230X-14-93

Siri, S., Zhao, Y., Maier, F., Pierce, D. M., & Feng, B. (2020). The Macro- and Micro-Mechanics of the Colon and Rectum I: Experimental Evidence. *Bioengineering (Basel)*, *7*(4). https://doi.org/10.3390/bioengineering7040130

Song, C., Chai, Z., Chen, S., Zhang, H., Zhang, X., & Zhou, Y. (2023). Intestinal mucus components and secretion mechanisms: what we do and do not know. *Exp Mol Med*, *55*(4), 681-691. https://doi.org/10.1038/s12276-023-00960-y

Sosnovski, K. E., Braun, T., Amir, A., Moshel, D., BenShoshan, M., VanDussen, K. L., Levhar, N., Abbas-Egbariya, H., Beider, K., Ben-Yishay, R., Asad Ali, S., Moore, S. R., Kugathasan, S., Abramovich, I., Glick Saar, E., Weiss, B., Barshack, I., Gottlieb, E., Geiger, T., . . . Haberman, Y. (2023). GATA6-AS1 Regulates Intestinal Epithelial Mitochondrial Functions, and its Reduced Expression is Linked to Intestinal Inflammation and Less Favourable Disease Course in Ulcerative Colitis. *J Crohns Colitis*, *17*(6), 960-971. https://doi.org/10.1093/ecco-jcc/jjad006

Stringer, E. J., Duluc, I., Saandi, T., Davidson, I., Bialecka, M., Sato, T., Barker, N., Clevers, H., Pritchard, C. A., Winton, D. J., Wright, N. A., Freund, J. N., Deschamps, J., & Beck, F. (2012). Cdx2 determines the fate of postnatal intestinal endoderm. *Development*, *139*(3), 465-474. https://doi.org/10.1242/dev.070722

Sugai, T., Habano, W., Jiao, Y. F., Tsukahara, M., Takeda, Y., Otsuka, K., & Nakamura, S. (2006). Analysis of molecular alterations in left- and right-sided colorectal carcinomas reveals distinct pathways of carcinogenesis: proposal for new molecular profile of colorectal carcinomas. *J Mol Diagn*, *8*(2), 193-201. https://doi.org/10.2353/jmoldx.2006.050052

Suh, E., Chen, L., Taylor, J., & Traber, P. G. (1994). A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol Cell Biol*, *14*(11), 7340-7351. https://doi.org/10.1128/mcb.14.11.7340-7351.1994

Tang, X., Huang, Y., Lei, J., Luo, H., & Zhu, X. (2019). The single-cell sequencing: new developments and medical applications. *Cell Biosci*, *9*, 53. https://doi.org/10.1186/s13578-019-0314-y

Taupin, D. R., Kinoshita, K., & Podolsky, D. K. (2000). Intestinal trefoil factor confers colonic epithelial resistance to apoptosis. *Proc Natl Acad Sci U S A*, *97*(2), 799-804. https://doi.org/10.1073/pnas.97.2.799

Ten Hoorn, S., de Back, T. R., Sommeijer, D. W., & Vermeulen, L. (2022). Clinical Value of Consensus Molecular Subtypes in Colorectal Cancer: A Systematic Review and Meta-Analysis. *J Natl Cancer Inst*, *114*(4), 503-516. https://doi.org/10.1093/jnci/djab106

Thomsen, E. R., Mich, J. K., Yao, Z., Hodge, R. D., Doyle, A. M., Jang, S., Shehata, S. I., Nelson, A. M., Shapovalova, N. V., Levi, B. P., & Ramanathan, S. (2016). Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat Methods*, *13*(1), 87-93. https://doi.org/10.1038/nmeth.3629

Tomic, G., Morrissey, E., Kozar, S., Ben-Moshe, S., Hoyle, A., Azzarelli, R., Kemp, R., Chilamakuri, C. S. R., Itzkovitz, S., Philpott, A., & Winton, D. J. (2018). Phospho-regulation of ATOH1 Is Required for Plasticity of Secretory Progenitors and Tissue Regeneration. *Cell Stem Cell*, *23*(3), 436-443.e437. https://doi.org/10.1016/j.stem.2018.07.002

Totaro, A., Castellan, M., Di Biagio, D., & Piccolo, S. (2018). Crosstalk between YAP/TAZ and Notch Signaling. *Trends Cell Biol*, *28*(7), 560-573. https://doi.org/10.1016/j.tcb.2018.03.001

Tsuchiya, K., Nakamura, T., Okamoto, R., Kanai, T., & Watanabe, M. (2007). Reciprocal targeting of Hath1 and beta-catenin by Wnt glycogen synthase kinase 3beta in

human colon cancer. *Gastroenterology*, *132*(1), 208-220. https://doi.org/10.1053/j.gastro.2006.10.031
Uchino, H., Kataoka, H., Itoh, H., Hamasuna, R., & Koono, M. (2000). Overexpression of intestinal trefoil factor in human colon carcinoma cells reduces cellular growth in vitro and in vivo. *Gastroenterology*, *118*(1), 60-69. https://doi.org/10.1016/s0016-5085(00)70414-8
Uchino, H., Kataoka, H., Itoh, H., Sekiya, R., Onitsuka, T., & Koono, M. (1999). Roles of intestinal trefoil factor (ITF) in human colorectal cancer: ITF suppresses the growth of colorectal carcinoma cells. *Hum Cell*, *12*(4), 181-188.
Ueo, T., Imayoshi, I., Kobayashi, T., Ohtsuka, T., Seno, H., Nakase, H., Chiba, T., & Kageyama, R. (2012). The role of Hes genes in intestinal development, homeostasis and tumor formation. *Development*, *139*(6), 1071-1082. https://doi.org/10.1242/dev.069070
Valenzuela, G., Canepa, J., Simonetti, C., Solo de Zaldivar, L., Marcelain, K., & Gonzalez-Montero, J. (2021). Consensus molecular subtypes of colorectal cancer in clinical practice: A translational approach. *World J Clin Oncol*, *12*(11), 1000-1008. https://doi.org/10.5306/wjco.v12.i11.1000
van der Flier, L. G., & Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol*, *71*, 241-260. https://doi.org/10.1146/annurev.physiol.010908.163145
van der Post, S., Jabbar, K. S., Birchenough, G., Arike, L., Akhtar, N., Sjovall, H., Johansson, M. E. V., & Hansson, G. C. (2019). Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. *Gut*, *68*(12), 2142-2151. https://doi.org/10.1136/gutjnl-2018-317571
van Es, J. H., van Gijn, M. E., Riccio, O., van den Born, M., Vooijs, M., Begthel, H., Cozijnsen, M., Robine, S., Winton, D. J., Radtke, F., & Clevers, H. (2005). Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature*, *435*(7044), 959-963. https://doi.org/10.1038/nature03659
VanDussen, K. L., Carulli, A. J., Keeley, T. M., Patel, S. R., Puthoff, B. J., Magness, S. T., Tran, I. T., Maillard, I., Siebel, C., Kolterud, A., Grosse, A. S., Gumucio, D. L., Ernst, S. A., Tsai, Y. H., Dempsey, P. J., & Samuelson, L. C. (2012). Notch signaling modulates proliferation and differentiation of intestinal crypt base columnar stem cells. *Development*, *139*(3), 488-497. https://doi.org/10.1242/dev.070763
VanDussen, K. L., & Samuelson, L. C. (2010). Mouse atonal homolog 1 directs intestinal progenitors to secretory cell rather than absorptive cell fate. *Dev Biol*, *346*(2), 215-223. https://doi.org/10.1016/j.ydbio.2010.07.026
Vichai, V., & Kirtikara, K. (2006). Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat Protoc*, *1*(3), 1112-1116. https://doi.org/10.1038/nprot.2006.179
Vinson, K. E., George, D. C., Fender, A. W., Bertrand, F. E., & Sigounas, G. (2016). The Notch pathway in colorectal cancer. *Int J Cancer*, *138*(8), 1835-1842. https://doi.org/10.1002/ijc.29800
Vocka, M., Langer, D., Petrtyl, J., Vockova, P., Hanus, T., Kalousova, M., Zima, T., & Petruzelka, L. (2015). Trefoil factor family (TFF) proteins as potential serum biomarkers in patients with metastatic colorectal cancer. *Neoplasma*, *62*(3), 470-477. https://doi.org/10.4149/neo_2015_056
Walker, E. M., Thompson, C. A., & Battle, M. A. (2014). GATA4 and GATA6 regulate intestinal epithelial cytodifferentiation during development. *Dev Biol*, *392*(2), 283-294. https://doi.org/10.1016/j.ydbio.2014.05.017
Wallaeys, C., Garcia-Gonzalez, N., & Libert, C. (2023). Paneth cells as the cornerstones of intestinal and organismal health: a primer. *EMBO Mol Med*, *15*(2), e16427. https://doi.org/10.15252/emmm.202216427
Wang, Y., Li, Z., Li, W., Liu, S., & Han, B. (2018). Methylation of CDX2 gene promoter in the prediction of treatment efficacy in colorectal cancer. *Oncol Lett*, *16*(1), 195-198. https://doi.org/10.3892/ol.2018.8670

Wang, Y., Song, W., Wang, J., Wang, T., Xiong, X., Qi, Z., Fu, W., Yang, X., & Chen, Y. G. (2020). Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med*, *217*(2). https://doi.org/10.1084/jem.20191130

Ward, T. M., Cauley, C. E., Stafford, C. E., Goldstone, R. N., Bordeianou, L. G., Kunitake, H., Berger, D. L., & Ricciardi, R. (2022). Tumour genotypes account for survival differences in right- and left-sided colon cancers. *Colorectal Dis*, *24*(5), 601-610. https://doi.org/10.1111/codi.16060

Washabau, R. J., & Day, M., J. (2013). Large Intestine. In *Canine and Feline Gastroenterology* (pp. 729-777). Elsevier Inc. https://doi.org/10.1016/B978-1-4160-3661-6.00058-4

Wilding, J. L., & Bodmer, W. F. (2014). Cancer cell lines for drug discovery and development. *Cancer Res*, *74*(9), 2377-2384. https://doi.org/10.1158/0008-5472.CAN-13-2971

Wilding, J. L., McGowan, S., Liu, Y., & Bodmer, W. F. (2010). Replication error deficient and proficient colorectal cancer gene expression differences caused by 3'UTR polyT sequence deletions. *Proc Natl Acad Sci U S A*, *107*(49), 21058-21063. https://doi.org/10.1073/pnas.1015604107

Witek, M. E., Nielsen, K., Walters, R., Hyslop, T., Palazzo, J., Schulz, S., & Waldman, S. A. (2005). The putative tumor suppressor Cdx2 is overexpressed by human colorectal adenocarcinomas. *Clin Cancer Res*, *11*(24 Pt 1), 8549-8556. https://doi.org/10.1158/1078-0432.Ccr-05-1624

Wong, N. A., Britton, M. P., Choi, G. S., Stanton, T. K., Bicknell, D. C., Wilding, J. L., & Bodmer, W. F. (2004). Loss of CDX1 expression in colorectal carcinoma: promoter methylation, mutation, and loss of heterozygosity analyses of 37 cell lines. *Proc Natl Acad Sci U S A*, *101*(2), 574-579. https://doi.org/10.1073/pnas.0307190101

Yamagishi, H., Kuroda, H., Imai, Y., & Hiraishi, H. (2016). Molecular pathogenesis of sporadic colorectal cancers. *Chin J Cancer*, *35*, 4. https://doi.org/10.1186/s40880-015-0066-y

Yamamoto, H., Bai, Y. Q., & Yuasa, Y. (2003). Homeodomain protein CDX2 regulates goblet-specific MUC2 gene expression. *Biochem Biophys Res Commun*, *300*(4), 813-818. https://doi.org/10.1016/s0006-291x(02)02935-2

Yang, Q., Bermingham, N. A., Finegold, M. J., & Zoghbi, H. Y. (2001). Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science*, *294*(5549), 2155-2158. https://doi.org/10.1126/science.1065718

Yang, R. Y., Rabinovich, G. A., & Liu, F. T. (2008). Galectins: structure, function and therapeutic potential. *Expert Rev Mol Med*, *10*, e17. https://doi.org/10.1017/s1462399408000719

Yang, S., & Yu, M. (2021). Role of Goblet Cells in Intestinal Barrier and Mucosal Immunity. *J Inflamm Res*, *14*, 3171-3183. https://doi.org/10.2147/JIR.S318327

Yang, X., Zhang, L., Song, X., He, W., Zhang, D., Lu, Q., Wu, J., Wu, C., & Jiang, J. (2018). MicroRNA-613 promotes colon cancer cell proliferation, invasion and migration by targeting ATOH1. *Biochem Biophys Res Commun*, *504*(4), 827-833. https://doi.org/10.1016/j.bbrc.2018.09.054

Yang, Y., Lin, Z., Lin, Q., Bei, W., & Guo, J. (2022). Pathological and therapeutic roles of bioactive peptide trefoil factor 3 in diverse diseases: recent progress and perspective. *Cell Death Dis*, *13*(1), 62. https://doi.org/10.1038/s41419-022-04504-6

Yeung, T. M., Gandhi, S. C., & Bodmer, W. F. (2011). Hypoxia and lineage specification of cell line-derived colorectal cancer stem cells. *Proc Natl Acad Sci U S A*, *108*(11), 4382-4387. https://doi.org/10.1073/pnas.1014519107

Yeung, T. M., Gandhi, S. C., Wilding, J. L., Muschel, R., & Bodmer, W. F. (2010). Cancer stem cells from colorectal cancer-derived cell lines. *Proc Natl Acad Sci U S A*, *107*(8), 3722-3727. https://doi.org/10.1073/pnas.0915135107

Young, M. D., & Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*, *9*(12). https://doi.org/10.1093/gigascience/giaa151

Yu, C., Yu, J., Yao, X., Wu, W. K., Lu, Y., Tang, S., Li, X., Bao, L., Li, X., Hou, Y., Wu, R., Jian, M., Chen, R., Zhang, F., Xu, L., Fan, F., He, J., Liang, Q., Wang, H., . . . Wang, J. (2014). Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing. *Cell Res*, *24*(6), 701-712. https://doi.org/10.1038/cr.2014.43

Yuan, Z., Zhao, Z., Hu, H., Zhu, Y., Zhang, W., Tang, Q., Huang, R., Gao, F., Zou, C., Wang, G., & Wang, X. (2021). IgG Fc Binding Protein (FCGBP) is Down-Regulated in Metastatic Lesions and Predicts Survival in Metastatic Colorectal Cancer Patients. *Onco Targets Ther*, *14*, 967-977. https://doi.org/10.2147/OTT.S285171

Yusufu, A., Shayimu, P., Tuerdi, R., Fang, C., Wang, F., & Wang, H. (2019). TFF3 and TFF1 expression levels are elevated in colorectal cancer and promote the malignant behavior of colon cancer by activating the EMT process. *Int J Oncol*, *55*(4), 789-804. https://doi.org/10.3892/ijo.2019.4854

Yusup, A., Huji, B., Fang, C., Wang, F., Dadihan, T., Wang, H. J., & Upur, H. (2017). Expression of trefoil factors and TWIST1 in colorectal cancer and their correlation with metastatic potential and prognosis. *World J Gastroenterol*, *23*(1), 110-120. https://doi.org/10.3748/wjg.v23.i1.110

Zecchini, V., Domaschenz, R., Winton, D., & Jones, P. (2005). Notch signaling regulates the differentiation of post-mitotic intestinal epithelial cells. *Genes Dev*, *19*(14), 1686-1691. https://doi.org/10.1101/gad.341705

Zhang, M., & Wu, C. (2020). The relationship between intestinal goblet cells and the immune response. *Biosci Rep*, *40*(10). https://doi.org/10.1042/BSR20201471

Zhang, Y., Sun, L., Lei, C., Li, W., Han, J., Zhang, J., & Zhang, Y. (2022). A Sweet Warning: Mucin-Type O-Glycans in Cancer. *Cells*, *11*(22). https://doi.org/10.3390/cells11223666

Zheng, X., Tsuchiya, K., Okamoto, R., Iwasaki, M., Kano, Y., Sakamoto, N., Nakamura, T., & Watanabe, M. (2011). Suppression of hath1 gene expression directly regulated by hes1 via notch signaling is associated with goblet cell depletion in ulcerative colitis. *Inflamm Bowel Dis*, *17*(11), 2251-2260. https://doi.org/10.1002/ibd.21611

Zhou, B., Lin, W., Long, Y., Yang, Y., Zhang, H., Wu, K., & Chu, Q. (2022). Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduction and Targeted Therapy*, *7*(1), 95. https://doi.org/10.1038/s41392-022-00934-y

Zhu, D. H., Niu, B. L., Du, H. M., Ren, K., Sun, J. M., & Gong, J. P. (2012). Hath1 inhibits proliferation of colon cancer cells probably through up-regulating expression of Muc2 and p27 and down-regulating expression of cyclin D1. *Asian Pac J Cancer Prev*, *13*(12), 6349-6355. https://doi.org/10.7314/apjcp.2012.13.12.6349

Zhu, J., Deane, N. G., Lewis, K. B., Padmanabhan, C., Washington, M. K., Ciombor, K. K., Timmers, C., Goldberg, R. M., Beauchamp, R. D., & Chen, X. (2016). Evaluation of frozen tissue-derived prognostic gene expression signatures in FFPE colorectal cancer samples. *Sci Rep*, *6*, 33273. https://doi.org/10.1038/srep33273

Zhuang, Q., Shen, A., Liu, L., Wu, M., Shen, Z., Liu, H., Cheng, Y., Lin, X., Wu, X., Lin, W., Li, J., Han, Y., Chen, X., Chen, Q., & Peng, J. (2021). Prognostic and immunological roles of Fc fragment of IgG binding protein in colorectal cancer. *Oncol Lett*, *22*(1), 526. https://doi.org/10.3892/ol.2021.12787