An Exploration of Metachronous Colorectal Lesion Predictors from a Bowel Screening Population

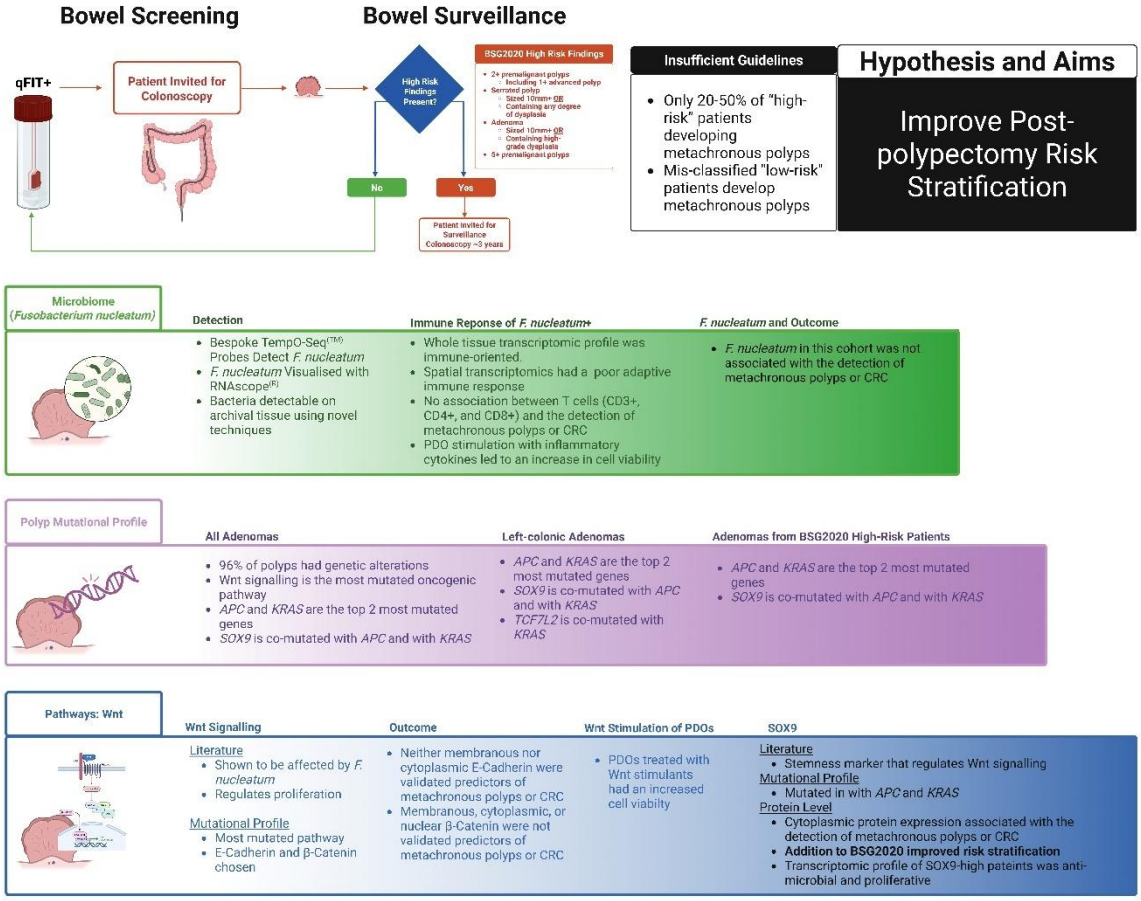
Sara Samir Foad Al-Badran
BSc. (Hons), MSc.

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

School of Cancer Sciences,
College of Medical, Veterinary, and Life Sciences,
University of Glasgow

September 2025

Graphical Abstract



Abstract

The rising incidence and mortality rates of colorectal cancer (CRC) continue to impact patients and burden healthcare systems worldwide. Multiple factors contribute to increasing CRC risk and, while modifiable ones such as improving diet, and smoking cessation are within an individual's control, non-modifiable factors such as age and a genetic propensity for CRC remain major causes of the disease. This led to many countries adopting national screening programs to promote the early detection and prevention of CRC. Many modalities can be used for screening, but several countries (including the United Kingdom (UK)) begin with a quantitative faecal immunochemical test (qFIT) and, should a positive result be found, a colonoscopy is recommended. Not only do screening colonoscopies detect CRC but also find pre-malignant polyps from which most CRC is derived. These are removed opportunistically as a mode of prevention; however the presence of polyps is associated with ongoing future risk of polyps and CRC, leading to the need for surveillance. In 2020, the British Society for Gastroenterology published updated guidelines (BSG2020 Guidelines) for the surveillance of patients who underwent a screening colonoscopy. They define a "high-risk" patient as one whose colonoscopy revealed two or more polyps where at least one was advanced, a serrated polyp larger than 1cm or with dysplasia, an adenoma larger than 1cm or with high-grade dysplasia, or five or more polyps. According to BSG2020 Guidelines, "high-risk" patients are recommended to undergo a second colonoscopy for surveillance of metachronous polyps or CRC within three years of the screening colonoscopy. However, with only 20-50% of "high-risk" patients developing metachronous polyps, coupled with the endoscopy services backlog in the UK, these risk stratification guidelines are not sufficient. This thesis aimed to identify a suitable tool to improve the current BSG2020 Guidelines for risk stratification within a bowel screening population. Initially, *Fusobacterium nucleatum*, was investigated with relation to the detection of metachronous polyps or CRC, through detection by a novel technique, TempO-Seq™. Numerous evidence of the involvement of *F. nucleatum* in CRC is available, implicating it in supporting tumorigenesis and metastasis. *F. nucleatum*+ patients revealed an immune-oriented whole tissue transcriptomic profile, however, spatial transcriptomics revealed a poor adaptive immune response in those patients. Similarly, there was no association

between lymphocytic markers (CD3+ T cells, CD4+ T cells, and CD8+ T cells) and the detection of metachronous polyps or CRC in patients with a known *F. nucleatum* status, although the stimulation of patient-derived organoids with inflammatory cytokines led to an increase in cell viability. Ultimately, *F. nucleatum* in this cohort was not associated with the detection of metachronous polyps or CRC. However, mutational analysis of adenomas from these patients revealed mutations in multiple Wnt pathway-associated genes. At the protein level, E-Cadherin and ß-Catenin were not associated with the detection of metachronous polyps or CRC. However, the high expression stemness marker and Wnt regulator SOX9 in the cytoplasm was associated with the detection of metachronous polyps or CRC. Patients with high SOX9 protein expression demonstrated a proliferative transcriptomic profile, with multiple anti-microbial pathways being positively enriched. Finally, the addition of SOX9 protein expression to current BSG2020 Guidelines significantly improved risk stratification. This research provides a basis for the use of additional stratification methods to improve the rate of prediction of metachronous polyps or CRC, using simple well-established methods like immunohistochemistry.

Table of Contents

List of Figures

List of Tables

List of Accompanying Material

Selected Poster Presentations:

SOX9 Expression in Colorectal Adenomas Improves Surveillance Colonoscopy Risk Stratification in a Bowel Screening Population. **Sara Samir Foad Al-Badran**, Christopher Bigley, Mark Johnstone, Aula Ammar, Alexander Winton, Jennifer Hay, Jean Quinn, Jakub Jawny, Ditte Andersen, Natalie Fisher, Philip D. Dunne, Noori Maka, Gerard Lynch, Stephen McSorley, and Joanne Edwards on behalf of the INCISE Collaborative. 4th International Conference on Cancer Prevention, Heidelberg, Germany 2024.

Improving Surveillance Colonoscopy Stratification Using SOX9 Expression in Colorectal Adenomas. **Sara Samir Foad Al-Badran**, Christopher Bigley, Mark Johnstone, Aula Ammar, Alexander Winton, Jennifer Hay, Jean Quinn, Jakub Jawny, Ditte Andersen, Natalie Fisher, Philip Dunne, Noori Maka, Gerard Lynch, Stephen McSorley, Joanne Edwards. Inaugural WntUK Meeting, The Crick Institute, London, UK 2024.

TempO-Seq™ Detects Fusobacterium nucleatum in Formalin-fixed Paraffin-embedded Patient Tissue from a Bowel Screening Population. **Sara Samir Foad Al-Badran**, Mark Johnstone, Jennifer Hay, Jakub Jawny, Noori Maka, William Rooney, Ditte Andersen, Gerard Lynch, Stephen McSorley, and Joanne Edwards. Best Poster Session: 35th European Congress of Pathology, Dublin, Ireland 2023.

Oral Presentations:

Exploring the Impact of Fusobacterium nucleatum on Colorectal Polyps from Patients within the Scottish Bowel Screening Programme. **Sara Samir Foad Al-Badran**, Stephen McSorley, Mark Johnstone, Noori Maka, Jennifer Hay, Jakub Jawny, Ditte Andersen, Natalie Fisher, Philip Dunne, Gerard Lynch, and Joanne Edwards. BACR 65th Anniversary Meeting, Edinburgh, UK 2024

SOX9 Expression in Colorectal Adenomas Improves Surveillance Colonoscopy Risk Stratification in a Bowel Screening Population. **Sara Samir Foad Al-Badran**, Christopher Bigley, Mark Johnstone, Aula Ammar, Alexander Winton, Jennifer Hay, Jean Quinn, Jakub Jawny, Ditte Andersen, Natalie Fisher, Philip D. Dunne, Noori Maka, Gerard Lynch, Stephen McSorley, and Joanne Edwards on behalf of the INCISE Collaborative. Rapid Fire Talk: 4th International Conference on Cancer Prevention, Heidelberg, Germany 2024.

Publications:

Preprint, under revision at *Scientific Reports:* SOX9 Expression in Colorectal Adenomas Improves Surveillance Colonoscopy Risk Stratification in a Bowel Screening Population. **Sara Samir Foad Al-Badran**, Christopher Bigley, Mark Johnstone, Aula Ammar, Alexander Winton, Jennifer Hay, Jean Quinn, Jakub Jawny, Ditte Andersen, Natalie Fisher, Philip D. Dunne, Noori Maka, Gerard Lynch, Stephen McSorley, and Joanne Edwards on behalf of the INCISE Collaborative.
medRxiv 2024.06.28.24309576; doi: https://doi.org/10.1101/2024.06.28.24309576

Awards:

Medical Research Scotland Annual Researcher Development Event 4th Year Oral Presentation Award 2025– *Runner-up*

Acknowledgements

There isn’t a word powerful enough to convey how lucky I am to have had Prof. Joanne Edwards as a primary supervisor. Joanne's kindness, patience, personal and professional support, and incredible scientific acumen allowed me to get to a place I always dreamt of being: on the other side of a PhD project. I also cannot thank enough my secondary supervisor, Dr. Stephen McSorley, for his support, input, and for showing me how to look at my research from a different point of view. To my third supervisor Dr. Tomoko Iwata, thank you for the stimulating discussions and ideas, and to my industry supervisor Dr. Ditte Andersen, you have my utmost gratitude as so much of this work would not have been possible without your teachings and guidance. Dr. Jean Quinn, you helped me through times I never thought would pass. I will always be in your debt. Dr. Gerard Lynch, your discussions and suggestions over the last four years have improved my work so much, so thanks for all the meetings.

I would like to extend my deepest thanks to the team at BioClavis, especially Paul Capewell, Samantha Campbell, and William Rooney for all their help. A huge thank you to all my friends at the Glasgow Tissue Research Facility, especially Dr. Hannah Morgan, for always going above and beyond. Thanks to Dr. Christopher Bigley for all his digital pathology support. INCISE, I wouldn’t be here without you. Thank you to everyone who has made this project what it is.

I would be amiss if I did not thank Assya, Lelo, and Natalie for holding my hand while I entered the world of coding. It might seem small to you, but it meant the world to me. Emma, you are truly the best. Thank you for taking me with you into the amazing realm of public engagement, I would have never expanded my horizons like that if it wasn’t for you. Thank you for your support while I was writing and thank you for making this thesis ten times better with all your suggestions.

A PhD is not an easy undertaking, but everyone who has been a part of M19 in the past 4 years has turned it into an adventure. Aula, Amna, and Gift, thank you for your support, help, and friendship. Guang-Yu, thanks for always checking my math and being there for me. Kathryn, you are a role model if I have ever seen one, and I fear no more the heat of the sun because of you. C, through two

degrees, your help and friendship has meant the world. Molly, you are a ray of sunshine and a smart cookie, I am better for being your friend. Lewis: never change; I am so glad your MSc project led you here. Nam, thank you for always brightening my day. Mark Johnstone, your friendship, support, and amazing stories are things I will never forget.

A special (and absurd) thank you to my Pod 3 (party pod) companions: Molly, Lewis and Yoana. Ashley, you made those afternoons in TC2 something to look forward to. To the Molerat Workforce (Tom, Emma, Kathryn, C, Molly, Rory, and Lewis), being a fellow molerat is an honour and a privilege. Thank you to all the members of the Roxburgh and Jamieson labs for making the second floor the warm and welcoming environment it is.

To my friends spotted across the globe: Salma Almekhlafi, Vania Teoh, Azza Fazar, and Sarah Shabbir, thank you for not letting those thousands of miles get between us.

I would also like to thank Simi and the Cafe Ø team for giving us all something to look forward to with their incredible lunches. To all my entertainment outlets (especially the Captain's Pod crew, and The Josh Johnson Show), thanks for keeping me going through the less exciting parts.

To my Glasgow family, Uncle Abdulla, Auntie Farah, Hiba, Noor, and Karam, thank you for all the love. Nothing I have ever done would have been possible without my grandparents, aunts, uncles, and cousins. Thank you all for the prayers and for believing in me.

To my perfect parents: Dr. Ferdous Salih Alsalman and Dr. Samir Foad Al-Badran, you raised me to be capable of anything and gave me the best of everything. Nothing I ever do to repay you will be enough. I hope I made you proud.

This work was supported by Medical Research Scotland, to whom I am grateful. I would also like to thank the NHS Endowment Grant for funding all the spatial transcriptomics work.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Sara Samir Foad Al-Badran

Definitions/Abbreviations

Abbreviations	Definition	Abbreviations	Definition	Abbreviations	Definition
5-FU	5-Fluorouracil	Gal-GalNAc	D-galactose-b(1-3)-N-acetyl-D-galactosamine	ON	Overnight
95% CIs	95% confidence intervals	GC Content	Guanine-Cytosine Content	pAdj	Adjusted P-value
APCs	Antigen-presenting Cells	gFOBT	Guaiac Faecal Occult Blood Test	PBS	Phosphate-buffered Saline
APS	Ammonium Persulfate	GI	Gastrointestinal	PCA	Principle Component Analysis
BH	Benjamini-Hochberg	GIA	Genomic Innovation Alliance	PDOs	Patient-derived Organoids
BRR	Brain Total RNA	GO	Gene Ontology	PFA	Paraformaldehyde
BSA	Bovine Serum Albumin	GSEA	Gene Set Enrichment Analysis	PHTS	PTEN-hamartoma Tumour Syndrome
BSG	British Society for Gastroenterology	GTRF	Glasgow Tissue Research Facility	PJS	Peutz-Jeghers Syndrome
BSP	Bowel Screening Programme	H&E	Haematoxylin & Eosin	PPAP	Polymerase Proofreading-associated Polyposis
CAF	Cancer-associated Fibroblast	HCl	Hydrochloric Acid	PT	Pre-treatment
CIMP	CpG Island Methylator Phenotype	HEIR	Heat-induced Antigen Retrieval	qFIT	Quantitative Faecal Immunochemical Test
CMMRD	Constitutional Mis-Match Repair Deficiency Syndrome	HRs	Hazard Ratios	ROCK	Rho-associated Protein Kinase
CMS	Consensus Molecular Subtypes	Hs	*Homo sapiens*	ROS	Reactive Oxygen Species
CPM	Counts per Million	H-score	Histoscore	rpm	Rounds per Minute
CRC	Colorectal Cancer	Hs-TBP	*Homo sapiens* TATA-Box Binding Protein	RT	Room Temperature
CRP	C Reactive Protein	Hs-UBC	*Homo sapiens* Ubiquitin C	sDNA-FIT	Stool DNA Faecal Immunochemical Test
DAB	3,3'-Diaminobenzidine	IHC	Immunohistochemistry	SDS	Sodium Dodecyl Sulfate
DapB	*Bacillus subtilis* Dihydrodipicolinate Reductase	INCISE	Integrated Technologies for Improved Polyp Surveillance	SMI	Spatial Molecular Imager
DCC	Deleted in Colorectal Cancer	JPS	Juvenile Polyposis Syndrome	SOX9	SRY-box Transcription Factor 9
DDO	Down-stream Detector Oligo	KEGG	Kyoto Encyclopedia of Genes and Genomes	SPS	Serrated Polyposis Syndrome
DEGs	Differentially Expressed Genes	KRAS	Kirsten Rat Sarcoma Virus	TBS	Tris-buffered Saline
DepMap	Dependency Map	LPS	Lipopolysaccharide	TCR	T cell Receptor
DMSO	Dimethyl Sulfoxide	LS	Lynch Syndrome	TEMED	N,N,N',N'-Tetramethylethylenediamine
DOs	Detector Oligos	MAP	MUTYH-associated Polyposis	TempO-Seq™	Templated Oligo Assay with Sequence Readout
DTT	1.4-Dithiothreitol	MHC	Major Histocompatibility Complex	TIGIT	T cell immunoglobulin and ITIM domain
EB-16S-rRNA	Eubacteria 16S rRNA	MPS	Mixed Polyposis Syndrome	Tm	Melting Temperature
EDTA	Diaminoethanetetra-acetic Acid Disodium Salt Dihydrate	MSI	Microsatellite Instable	TMA	Tissue Microarray
ESGE	European Society of Gastrointestinal Endoscopy	MSigDb	Molecular Signature Database	TNM	Tumour, Node, Metastasis
FAP	Familial Adenomatous Polyposis	MSS	Microsatellite Stable	UDO	Up-stream Detector Oligo
FBS	Foetal Bovine Serum	MW	Molecular Weight	UK	United Kingdom
FCCTX	Familial Colorectal Cancer Type X	NaCl	Sodium Chloride	UMAP	Uniform Manifold Approximation and Projection
FFPE	Formalin-fixed Paraffin-embedded	NaOH	Sodium Hydroxide	URR	Reference Total RNA
FIT	Faecal Immunochemical Test	NATT	NaCl-Tris-Tween Buffer	US	United States
Fn	*F. nucleatum, Fusobacterium nucleatum*	NES	Normalised Enrichment Scores	USPSTF	US Preventive Services Taskforce
FOLFIRI	Folinic acid/Fluorouracil/Irinotecan	NETs	Neutrophil Extracellular Traps	WB	Western Blot
FOLFOX	Folinic acid/Fluorouracil/Oxaliplatin	NK	Natural Killer		
FOV	Fields of View	NSCs	No-sample Controls		

Chapter 1 Introduction

1.1 Gastrointestinal Tract

Beginning with the mouth and terminating with the anus, the human gastrointestinal (GI; Figure 1.1) tract is a nine-metre-long tube through which food is consumed, processed, digested, and eliminated (Ogobuiro *et al.*, 2023). The mouth contains saliva, which consists of several substances including water, mucin (to aid in swallowing), amylase (to initiate starch digestion), and microorganisms native to the oral cavity. The salivary components and the teeth contribute to the early stages of digestion, both mechanical (chewing) and enzymatic (Pedersen *et al.*, 2018) which results in the production of a bolus. The bolus then passes through the oesophagus by peristalsis, into the stomach. Most digestion takes place in the highly acidic environment of the stomach (pH~2), aided by muscular contractions and the release of enzymes such as pepsin and lipases, which turn the food bolus into chyme (Singh and Gallier, 2014). Chyme moves through the small intestine where up to 90% of water and nutrients are processed by secreted enzymes from GI accessory organs (pancreatic secretions and bile) and absorbed into the bloodstream by microvilli (Capuano, 2017). Any remaining nutrients and water are then absorbed in the large bowel (colon), and indigestible materials are compacted into faeces which are eliminated by defecation through the rectum and anus (Bruneau and Bruneau, 2017). This thesis focuses on the large bowel.

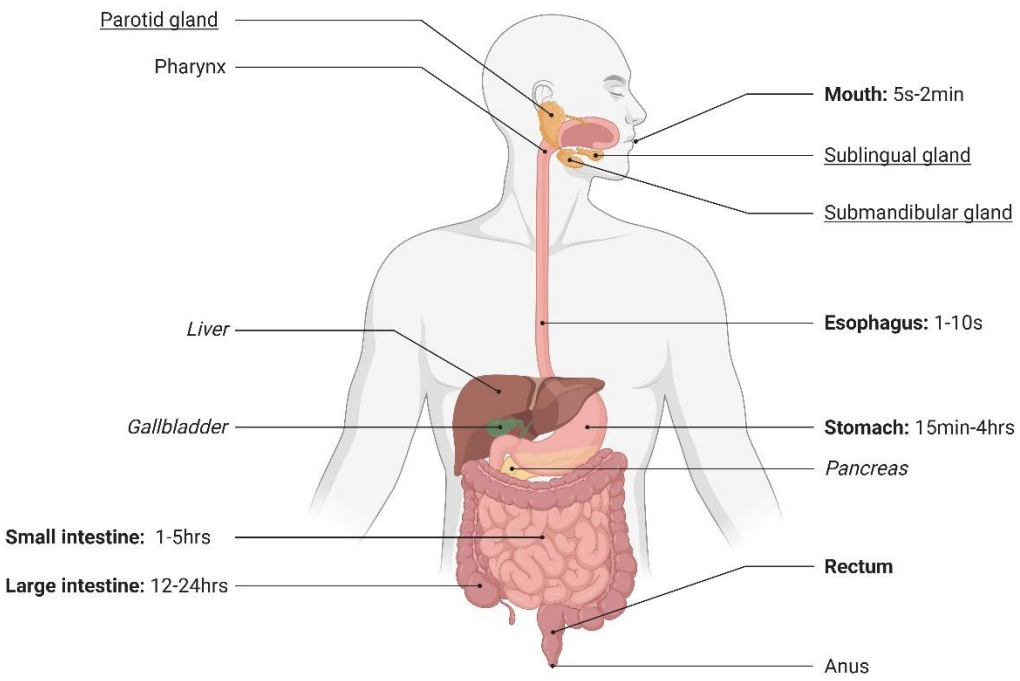


Figure 1.1. Schematic of the Human Digestive Tract with Transient Times.
Organs in bold are primary digestive system organs where food materials transit. Organs in italics are accessory digestive organs that aid in digestion but fundamentally belong to other systems. Underlined organs are salivary glands within the mouth. Figure created with BioRender.

1.1.1 Bowel Anatomy and Function

The colon is made up of four layers: the inner most mucosa, the submucosa, the muscular layer, and the serosa. The muscular layer is composed of two layers perpendicular to each other, which contribute to the muscular contractions required for colonic motility (Azzouz and Sharma, 2023). Colonic motility is crucial to the function of the colon, which is to move indigestible materials towards the rectum for elimination. The proximal parts of the colon, the ascending and transverse colon, absorb any remaining water and vitamins, beginning the formation of stool. This is then stored in the descending colon, until contractions from the sigmoid colon move it into the rectum prior to defecation (Azzouz and Sharma, 2023).

1.2 Colorectal Cancer

1.2.1 Epidemiology

Among the many diseases that can affect the bowel, cancer has been at the forefront of medical research. Colorectal cancer (CRC) is a disease characterised

by uncontrolled growth of cells of the colon and the rectum. It is the third most common cancer and has the second highest cancer-mortality worldwide with 903,859 reported deaths in 2022 (Bray *et al.*, 2024). Despite improvements in diagnosis and treatment, it is predicted that incidence will increase by 63% by 2040, with a 73% increase in mortality (Morgan *et al.*, 2023). In the United Kingdom (UK), 44,063 cases were reported between 2017 and 2018, with 16,808 deaths (CRUK, 2023). The incidence of CRC in the UK is higher in males (56%), older people (43%), and people from white ethnic groups (90%; CRUK, 2023; Delon et al., 2022).

1.2.2 Aetiology

Multiple factors contribute to the incidence of CRC. CRC can be caused by germline mutations passed through hereditary syndromes, or through acquired genetic events, termed sporadic CRC, with risk factors contributing to the likelihood of CRC development in either scenario.

1.2.2.1 Hereditary CRC

It has been reported that 20% of patients diagnosed with CRC before the age of 50 carry germline mutations (Stoffel *et al.*, 2018). Hereditary CRC can be divided into polyposis syndromes and non-polyposis syndromes. The first group includes adenomatous polyposis syndromes such as familial adenomatous polyposis (FAP), which is a dominant disorder with an 80-100% risk of CRC if left unmanaged and is mainly attributed to mutations in *APC* (Samadder *et al.*, 2019), and *MUTYH*-associated polyposis (MAP), caused by *MUYTH* mutations that involve the base excision repair pathway (Patel *et al.*, 2021). Less common adenomatous polyposis syndromes such as polymerase proofreading-associated polyposis (PPAP), and constitutional mis-match repair deficiency syndrome (CMMRD) can confer a higher risk of early onset CRC (Palles *et al.*, 2012; Aronson *et al.*, 2022). Another subgroup of polyposis syndromes are the hamartomatous polyposis syndromes, such as Peutz-Jeghers Syndrome (PJS), PTEN-hamartoma tumour syndrome (PHTS), and juvenile polyposis syndrome (JPS). PJS is characterised by hyperpigmentation in the form of distinct “spots” on the lips, mouth, and finger/toe tips (Gilad *et al.*, 2019). Both PJS and PHTS are linked to manifestation in non-GI organs (Yehia *et al.*, 2025), while JPS is limited to

manifestation within the GI tract (Cohen *et al.*, 2019). Serrated polyposis syndrome (SPS) and mixed polyposis syndrome (MPS) are characterised by non-adenomatous polyps and polyps of a mixed histology, respectively (Jaeger *et al.*, 2012; Dekker *et al.*, 2020).

Non-polyposis syndromes include Lynch syndrome (LS) which involves an MMR-deficient profile (Moreira *et al.*, 2012), and familial colorectal cancer type X (FCCTX) which differs from LS through its MMR-proficient profile (Xu *et al.*, 2020).

1.2.2.2 Sporadic CRC

Any CRC that is caused by a somatic mutation is considered sporadic. As mutations are a key step in the development of CRC, sporadic CRCs can be categorised by their mutational burden. Hypermutated CRCs (16%) are MMR-deficient and have a higher rate of occurrence in the right colon. Non-hypermutated CRCs (84%) are MMR-proficient and may carry mutations in many different genes (Muzny *et al.*, 2012). Figure 1.2 represents the mutation rate of important genes within each group as identified by the Cancer Genome Atlas Network.

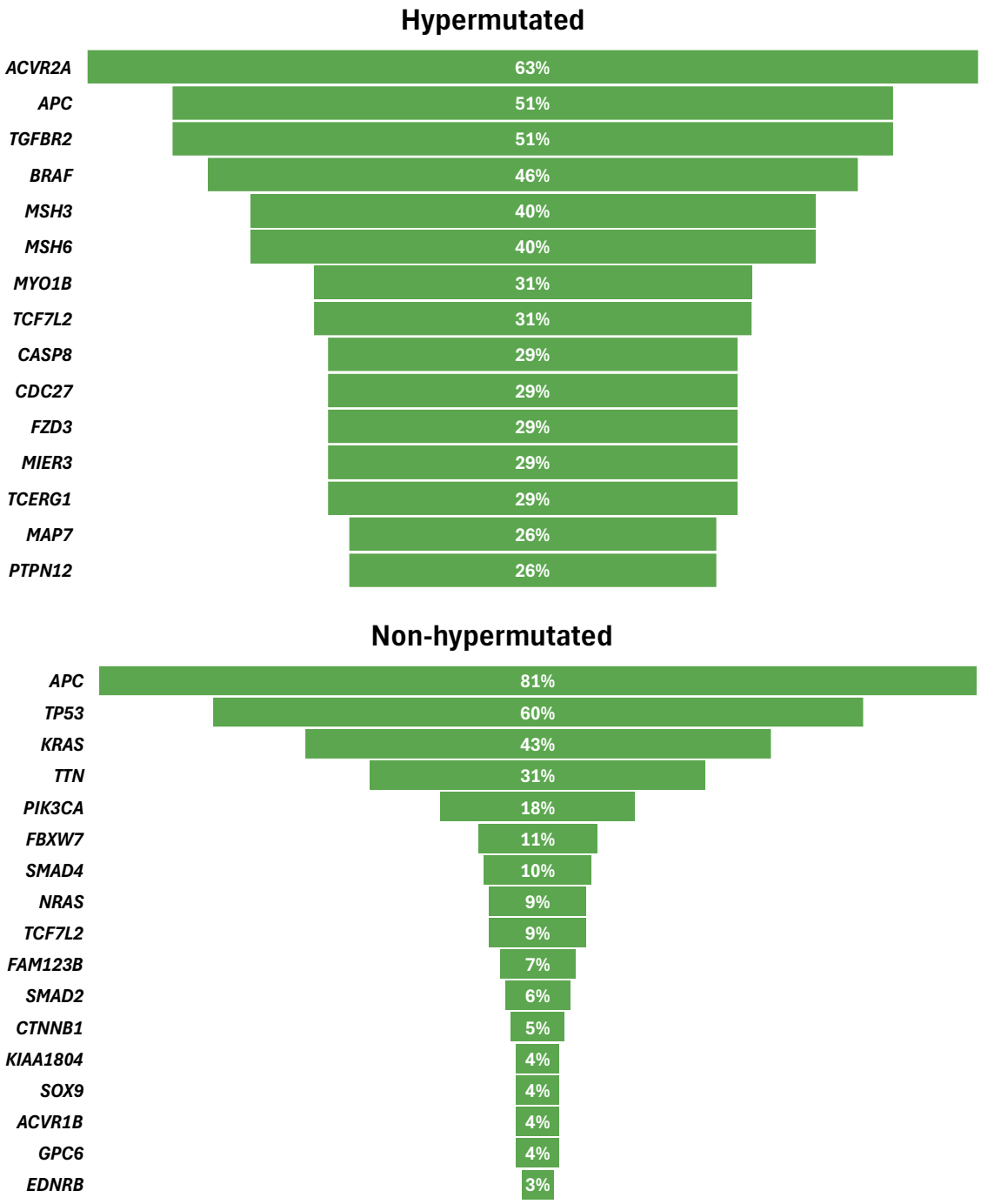


Figure 1.2. Common Mutations in Colorectal Cancer (CRC) Patients Reported by the Cancer Genome Atlas Network.
(Muzny *et al.*, 2012).

1.2.2.3 Risk Factors of CRC

Many lifestyle choices that contribute to general poor health have been shown to be risk factors for CRC. A diet high in red meats and alcohol and low in fibre is an established contributor to the development of CRC (Chan *et al.*, 2011; Sawicki *et al.*, 2021). Smoking has been shown to not only increase CRC

incidence in a dose-dependent manner but is also associated with the development of cancers with a poorer prognosis (Liang, Chen and Giovannucci, 2009; Carr *et al.*, 2018). Excess weight and a sedentary lifestyle were also found to be potential causes of reduced CRC survival (Fedirko *et al.*, 2014). Although increasing age (a non-modifiable risk factor) has long been considered a CRC risk factor, it is important to note the increasing incidence of CRC onset in individuals below 50 years of age (Vuik *et al.*, 2019).

1.2.3 CRC Sidedness

The location of CRC is significant in terms of prognosis and treatment. Clinically, right-sided tumours usually arise in older females (Haggar and Boushey, 2009), from sessile polyps and serrated polyps, and metastasise to the peritoneum (Benedix *et al.*, 2010). These tumours were shown to have a poorer prognosis with a higher risk of recurrence (Park *et al.*, 2015). Left-sided tumours are more predominant in younger males (Segev, Kalady and Church, 2018), arise from tubular and villous polyps, and metastasise to the liver and lungs (Benedix *et al.*, 2010). The areas of the transverse colon that are categorised as right or left are shown in Figure 1.3.

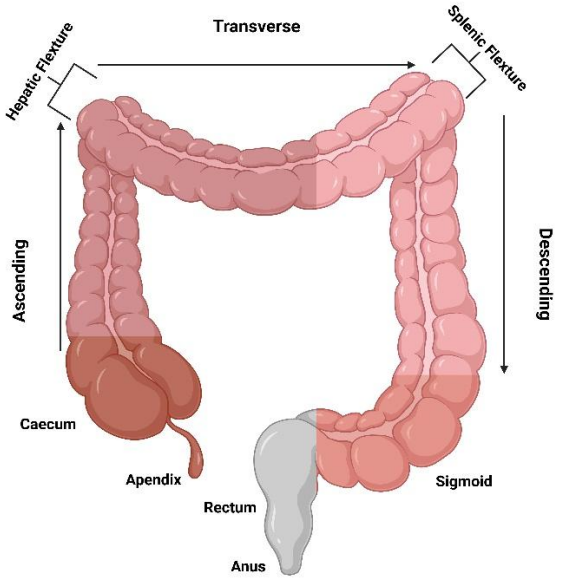


Figure 1.3. Anatomy of the Colon.
The caecum and ascending colon, the 90° hepatic flexure, and majority of the transverse colon represent the areas where right-sided colon cancer arises. The remainder of the transverse colon, the 90° splenic flexure, and the descending colon represent areas where left-sided colon cancer arises. Colour variations denote different parts of the colon. The rectum is coloured grey. Figure created with BioRender.

1.2.4 CRC Prognosis and Clinical Presentation

1.2.4.1 Staging

Upon diagnosis of CRC, the extent of advancement must be determined to inform prognosis and treatment options. This process, known as staging, depends on identifying the level of cancer growth. A few systems are available for this; the most commonly used in current clinical practice is the tumour, node, metastasis (TNM) staging system (Table 1.1). T values are used to describe the extent of invasion of the solid tumour through the layers of the primary organ. The higher the value the more advanced the tumour is. The N values represent how many regional lymph nodes have cancer cells within them and are used to measure regional spread. Finally, M values describe whether the primary tumour has spread to other organs and established itself in these distant sites. Generally, these stages coincide with the age of the tumour, i.e., the earlier the detection occurs, the lower the stage and the better the prognosis (Li *et al.*, 2016; Rosen and Sapra, 2023).

1.2.4.2 Symptoms

One of the main problems presented by CRC is its asymptomatic nature in the early stages, leading to late diagnosis. Most symptoms experienced by patients are due to the larger mass blocking the bowel, leading to blood in the stool, abdominal pain and vomiting, and in more advanced cases, the presence of a palpable mass in the abdomen (Burnett-Hartman *et al.*, 2021). This leads to diagnosis occurring later in the development of the tumour, which greatly affects prognosis.

Table 1.1. TNM Classification for CRC Staging.

T (Tumour)		N (Lymph Node)		M (Metastasis)			TNM		
Stage	**Clinical Presentation**	**Stage**	**Clinical Presentation**	**Stage**		**Clinical Presentation**	**Stage**	**Classification**	**Prognosis**
T0	No tumour	**N0**	No regional node involvement	**M0**		No metastasis to distant sites	**0**	Carcinoma *in situ* Tis, N0, M0	-
T1	Invasion to the submucosa	**N1**	Involvement of 1-3 regional nodes	**M1**	**A**	Metastasis to 1 area	**I**	Localised cancer T1, N0, M0 T2, N0, M0	5-year survival: 90%
T2	Invasion to the muscularis propria	**N2**	Involvement of 4-6 regional nodes		**B**	Metastasis to 2+ areas	**II**	Locally advanced cancer, early T1, N1, M0 T2, N1, M0	5-year survival: 84%
T3	Invasion to the subserosa	**N3**	Involvement of 7+ regional nodes		**C**	Metastasis to the peritoneum	**III**	Locally advanced cancer, late T1-4, N2-3, M0	5-year survival: 65%
T4	Invasion beyond the layers of the colon to adjacent structures	**Nx**	Not assessable				**IV**	Metastatic cancer T1-4, N1-3, M1	5-year survival: 11%
Tis	Carcinoma *in situ*								
Tx	Not assessable								

1.2.4.3 CRC and the Immune Response

The immune response has been increasingly investigated in the past decade as both a supporter of cancer growth and an inhibitor to cancer progression in a context-dependent manner. The ability of cancer to evade the immune system and immune cells promoting tumour development has been considered an emerging hallmark of cancer and an enabling characteristic (Hanahan and Weinberg, 2011).

In fact, systemic inflammation has been linked to CRC prognosis, with acute phase inflammation markers like C Reactive Protein (CRP) and serum albumin, as well as blood-count measurements like the neutrophil-lymphocyte ratio being associated with the survival of CRC patients undergoing surgery (Woo, Kim and Kim, 2015; Li, Zhao and Zheng, 2019).

However, this is more complicated locally, with the various immune residents of the tumour microenvironment leading to different effects. The adaptive immune response is comprised of both T cells and B cells (Cooper, 2015). T cell activation is depicted in Figure 1.4. T cells are activated by extracellular signals from antigen-presenting cells (APCs) via the T cell receptor (TCR), which lacks a functional intracellular domain. This requires a signal transduction complex to deliver the extracellular signals to the cytoplasm of the immune cells (Menon *et al.*, 2023). CD3 is comprised of multiple subunits which interact with the TCR to receive signals and to successfully deliver them into the cell, highlighting the importance of functional CD3 for the activation of T cells (Dong *et al.*, 2019).

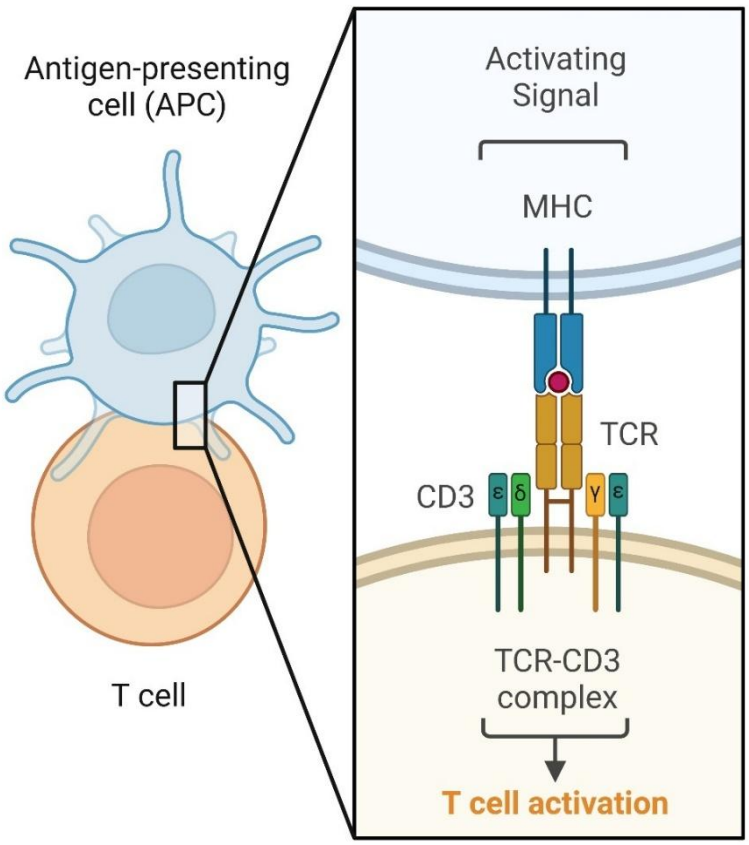


Figure 1.4. T Cell Activation.
Antigen-presenting cells signal through the major histocompatibility complex (MHC), and this is recognised by the TCR-CD3 complex on T cells. The signal travels through the TCR to CD3, which can deliver it to the cytoplasm of the lymphocyte due to its longer intracellular domain. This leads to activation of the T cell. Figure created with BioRender.

It has long been established that activated T cells are classified by their function, which is determined by which MHC class the activating signal originated from, and which T cell co-receptor detected it: CD4, or CD8. CD4+ T cells are activated by MHC-II, and are known as “helper T cells”, while CD8+ T cells are activated by MHC-I, and are known as “cytotoxic T cells” (Meuer, Schlossman and Reinherz, 1982).

The immune response in CRC is multifaceted, and T cells play a major role. CD4+ T cells help activate and regulate CD8+ T cells, interact with other immune cells such as natural killer cells involved in the anti-tumour immune response, and inhibit angiogenesis. CD8+ T cells are major contributors to the anti-tumour response and are key to cancer immunotherapy (Raskov *et al.*, 2020). However, the continued activation of these cells can lead to T cell exhaustion which has been associated with cancer progression (Xie *et al.*, 2023).

On the other hand, immune cells like macrophages can have a dual role in CRC. For instance, macrophages polarised to the M1 type have a strong antigen-presenting capacity, promoting CD4+ T cell recruitment and function (Huo *et al.*, 2023). However, macrophages polarised to the M2 type are more abundant in CRC tissue have been associated with poor prognosis (Huang *et al.*, 2024). While stage at diagnosis has long been considered the most important prognostic factor for CRC, the role of the immune response cannot be overlooked, despite its complexity.

1.2.5 CRC Management

1.2.5.1 Surgery

Surgical interventions, like the removal of the tumour mass or the more radical removal of a section of the bowel, are among the most common and effective CRC treatments, and although advancements have made surgery safer, post-operative complications remain a concern (Krieg *et al.*, 2024). Thus, prior to surgery chemotherapy and radiotherapy are often used to downstage the tumour prior to excision, particularly in locally advanced rectal cancer and increasingly in colonic cancer. Surgery carries a risk of reduced quality of life, such as the possible need for a stoma, or adverse effects on sexual performance, in addition to limited movement during extended recovery (Świątkowski *et al.*, 2022).

1.2.5.2 Chemotherapy

Chemotherapy is a systemic approach to cancer treatment that uses toxic agents to kill rapidly dividing cells. It is a direct outcome of observed effects of mustard gas on lymphatic tissue and bone marrow in the 1900s (Anand *et al.*, 2023). Decades later, regimens consisting of multiple agents (such as FOLFOX: folinic acid/fluorouracil/oxaliplatin and FOLFIRI: folinic acid/fluorouracil/irinotecan), are used pre-operatively (neoadjuvant therapy) and/or post-operatively (adjuvant therapy) with good outcomes (Arredondo *et al.*, 2013). However, the systemic and cytotoxic nature of chemotherapy leads to a host of side effects ranging from hair loss and fatigue to vomiting and taste changes that greatly reduce quality of life and require supporting management (Anand *et al.*, 2023).

1.2.5.3 Radiotherapy

Radiotherapy is a local approach to cancer treatment that uses a directed beam of ionising radiation to destroy the DNA of rapidly dividing cells (Karaca and Kırlı Bölükbaş, 2025). New modalities, such as intensity-modulated and volumetric modulated arc radiotherapy, allow the delivery of high doses to the tumour while significantly decreasing off-target delivery to healthy adjacent tissue (Dröge *et al.*, 2015). However, side-effects due to toxicity continue to occur with only symptomatic relief treatment options available to patients (De Ruysscher *et al.*, 2019).

1.2.5.4 Targeted Therapy

To maximise therapeutic potential and minimise off-target effects, targeted therapies have become the focus of cancer research in recent years. Immunotherapy is at the forefront of this area of research, with approaches such as the use of monoclonal antibodies and immune checkpoint inhibitors being used to treat CRC. For example, the monoclonal antibodies bevacizumab and cetuximab were both approved in 2004 for clinical use to target angiogenic molecules such as VEGF and growth factors like EGFR, respectively (Kasi *et al.*, 2023; Z. L. Liu *et al.*, 2023), and nivolumab was approved in 2025 for CRC in its capacity to block PD-1 in immune cells (Rajan *et al.*, 2016).

1.3 CRC Development

The majority of CRC, particularly sporadic disease, originates from benign overgrowths of cells known as polyps (Simon, 2016). These polyps increase in size and malignancy over time, with the chance of them developing into CRC increasing during a 7–15-year period, if left unattended (Huck and Bohl, 2016; Rutter *et al.*, 2025). Polyps come in a variety of shapes and histologies, and these differences can be important prognostic factors.

1.3.1 Polyp Types

1.3.1.1 Morphological

There are two main polyp morphology types: pedunculated and sessile. Pedunculated and sub-pedunculated polyps are characterised by their

"mushroom"-like appearance, where a stalk bridges the polyp body to the colonic mucosa (Ciocalteu *et al.*, 2018). Sessile polyps resemble flat "bumps" on the colonic mucosa and are often covered with a mucus cap or can be entirely flat; thus, they are harder to identify during visualisation procedures such as colonoscopy (Obuch, Pigott and Ahnen, 2015; Moy *et al.*, 2018). These descriptions were detailed by the Paris Classification which proposed a working framework for superficial lesion classification through endoscopy (Participants in the Paris Workshop, 2003), and are shown in Figure 1.5.

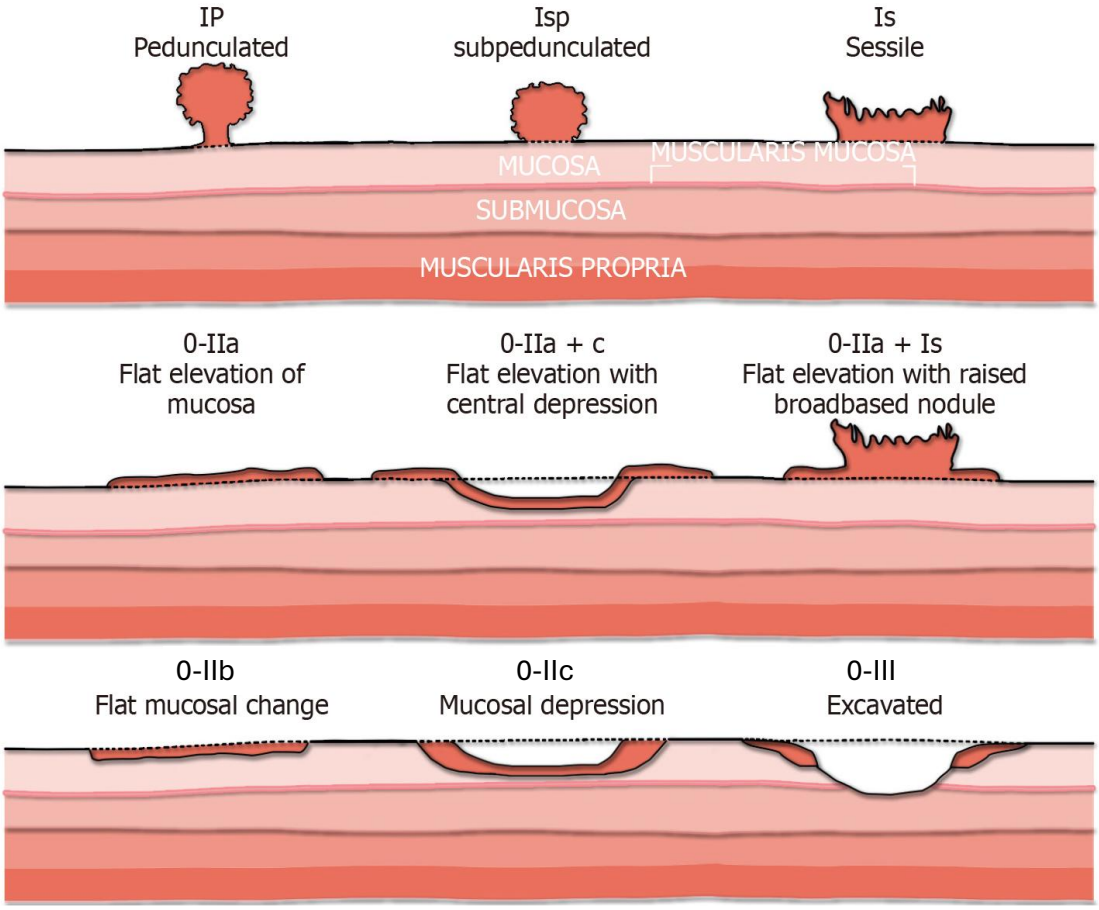


Figure 1.5. Diagram of the Type 0 Variant of the Paris Classification of Neoplastic Lesions. Polypoid (0-I): Ip refers to pedunculated lesions characterised by a stalk. Isp refers to sub-pedunculated lesions characterised by a broader attachment to the bowel lining than that of true pedunculated lesions. Is refers to protruding sessile lesions which have no stalk but are connected to the bowel lining by a wide base. **Non-polypoid (0-II):** 0-II refers to flat lesions that may be elevated. 0-IIa+c refers to lesions with flatly elevated sides and a depressed centre, and 0-IIa+Is refers to flatly elevated lesions with a protruding central nodule of sessile characteristics. Non-elevated lesions include 0-IIb, which is flat, 0-IIc, which is depressed, and 0-III which is an excavated lesion that could surpass the mucosa. Adapted from (Participants in the Paris Workshop, 2003; Mathews, Draganov and Yang, 2021).

1.3.1.2 Histological

There are four main polyp types that can be sub-categorised based on their malignant potential: neoplastic polyps carry the highest chance of malignancy, while truly hyperplastic polyps, hamartomatous polyps, and inflammatory polyps are considered to be non-neoplastic. Neoplastic polyps, which include the conventional adenomas and serrated lesions occur through two important, but different pathways discussed below.

Conventional adenomas are the most commonly occurring polyp type and are sub-divided into different histological sub-groups, all with varying degrees of dysplasia. The histology of traditional adenomas depends on their villous content. Along with the size of the polyp, villous histology (resembling finger-like projections) confers a higher risk of malignant transformation (Witold *et al.*, 2018). Tubular adenomas have the appearance of branched glands and less than 20% villous features. They are most common in adults and carry the least malignant potential. With a mostly conserved crypt shape, their nuclei are enlarged and display hyperchromatic staining upon haematoxylin & eosin (H&E) staining, as well as a loss of polarity (Galuppini *et al.*, 2021). Villous adenomas have over 75% villous features and make up less than 10% of all adenomas. They have deep crypts with a dysplastic epithelium (Shussman and Wexner, 2014). Tubulovillous adenomas have between 25-75% villous features and make up less than 25% of all adenomas (Shussman and Wexner, 2014). Sessile serrated polyps, another type of neoplastic polyp, present as flat with a saw-like appearance, and dilated crypts. These were thought to have no malignant potential but have since been linked to various mutations and reported to lead to CRC (Obuch, Pigott and Ahnen, 2015).

The non-neoplastic polyps do not carry malignant potential on their own. Hyperplastic polyps are sometimes considered serrated lesions, although they are not dysplastic (Monreal-Robles *et al.*, 2021). Hamartomatous polyps are made up of a collection of cells and tissues native to the area where the polyp is found (Cauchin, Touchefeu and Matysiak-Budnik, 2015), and although they are benign, they are commonly found in hereditary CRC syndromes discussed in 1.2.2.1. Inflammatory polyps are considered pseudo-polyps, although their appearance can be difficult to distinguish from true polyps. They are not

strongly linked to the development of CRC as they are mainly associated with other inflammatory bowel diseases (Ashktorab *et al.*, 2020) like inflammatory bowel disease and ulcerative colitis. Recent meta-analysis (Candel *et al.*, 2025) of the literature on the association between inflammatory polys and CRC posits that it is more accurate to describe any associations between inflammatory polyps and CRC as a surrogate for inflamed state of the bowel in inflammatory bowel disease and ulcerative colitis patients as the evidence supporting the malignant transformation of inflammatory polyps in CRC is extremely scarce (Politis *et al.*, 2017).

1.3.2 Adenoma-Carcinoma Sequence

Following the transition from normal tissue to benign polyp, a series of steps lead to the development of a cancer. In 1990, Fearon and Vogelstein described what is now known as the adenoma-carcinoma sequence, a series of mutational events that lead to a benign polyp becoming malignant (Figure 1.6) accounting for ~85% of CRCs. This model was based on the notion that epithelial adenomas begin in a small pocket of cells and rise by clonal expansion (Ponder and Wilkinson, 1986). The model described sequential mutations in oncogenes, tumour suppressor genes, and loss of alleles that contribute to progression to malignancy. The initial step is the transition from normal epithelium to hyperproliferating epithelium through a mutation/loss of a gene on the 5q chromosome, *APC*. The association of *APC* mutations with colorectal disease has been linked to inherited CRC syndromes (1.2.2.1). However, research continues to place *APC* at the centre of various crucial cellular processes. For example, the role of *APC* in mitosis was the subject of investigation, with its ability to bind to microtubules and the kinetochore in order to participate in the spindle assembly checkpoint becoming clear, when silencing it led to error-prone mitotic progression (Dikovskaya *et al.*, 2007). The connection between *APC* and apoptosis gained further attention when apoptotic cells in *APC*-deficient CRC cell lines increased when *APC* was restored (Morin, Vogelstein and Kinzler, 1996; Cristofaro *et al.*, 2015). *APC* restoration was also shown to be directly responsible for regaining differentiation in tumours (Dow *et al.*, 2015). However, the most important role of *APC* is its involvement in Wnt signalling. This highly conserved pathway and its functions regulating cellular development originated through the independent discovery of the embryonic functions of the *int-1*

mouse oncogene and its homologue in *Drosophila melanogaster*, *Wg* (Nusse *et al.*, 1984; Rijsewijk *et al.*, 1987). Since then, the cell proliferation functions of this pathway have become well-studied and targeted for therapeutic potential (Liu *et al.*, 2013; Srivastava and Srivastava, 2019). The main mechanism of this pathway is the regulation of ß-Catenin levels, which are regulated by the ß-Catenin destruction complex. *APC* is a major component of this complex, which regulates the levels of free cytoplasmic ß-Catenin and controls its translocation to the nucleus and initiation of transcription of cellular processes such as proliferation and survival (Parker and Neufeld, 2020). This pathway is of particular interest and will be further described in later sections of this thesis.

The next step in the Fearon and Vogelstein model is a mutation in the 12p chromosome resident Kirsten rat sarcoma virus (*KRAS*) gene. Members of the RAS family were described as binary switches by Milburn et al. (1990) when they successfully visualised RAS proteins. *KRAS* is "off" when bound to a GDP molecule but is subsequently turned "on" when GDP is exchanged for GTP. It is an imbalance of these two states that leads to an overactivation of *KRAS* and to the oncogenic effects it imposes (Johnson, Burkhart and Haigis, 2022). *KRAS* is involved in multiple pathways that, when activated, lead to cellular proliferation and survival, for example the RAF-MEK-ERK pathway and the PI3K pathway (H. J. Kim *et al.*, 2021). *KRAS* mutations are of therapeutic value in CRC, with common mutations such as $KRAS^{G12C}$ and $KRAS^{G12D}$ acting as barriers to effective response with anti-EGFR based treatments and representing desirable targets for drug development (Singhal *et al.*, 2024; Li, Zhao and Li, 2025).

The next step is the loss of heterozygosity at chromosome 18q, which was initially attributed to deleted in colorectal cancer (*DCC*), now known as netrin receptor DCC. As its function in axon guidance and neural function became clear (Jasmin *et al.*, 2020), a neighbouring gene on the same chromosomal location emerged as a possible replacement. The tumour-supressing function of *DPC4* was elucidated in pancreatic cancer (Hahn *et al.*, 1996) and later in CRC (Thiagalingam *et al.*, 1996). Eventually, it was recognised as a member of the SMAD transcription family and is now known as *SMAD4*. SMAD4 plays an integral role in TGFß signalling, where it forms a complex with other SMAD molecules, that translocates to the nucleus to aid in transcription of downstream targets (Du *et al.*, 2018). The role of the functional pathway in general and of *SMAD4* in

specific in arresting the cell cycle in early tumorigenesis was elucidated by Hunt et al. (1998).

The last step in the sequence is the loss of *TP53* on chromosome 17p. Encoding the "guardian of the genome" p53, the involvement of wild-type *TP53* in cell cycle regulation and DNA repair is well known. Mutations affecting either the structural integrity of p53 or its transcriptional ability lead to a loss in its ability to regulate cellular functions including cell division and apoptosis (Ludwig, Bates and Vousden, 1996; Marei *et al.*, 2021).

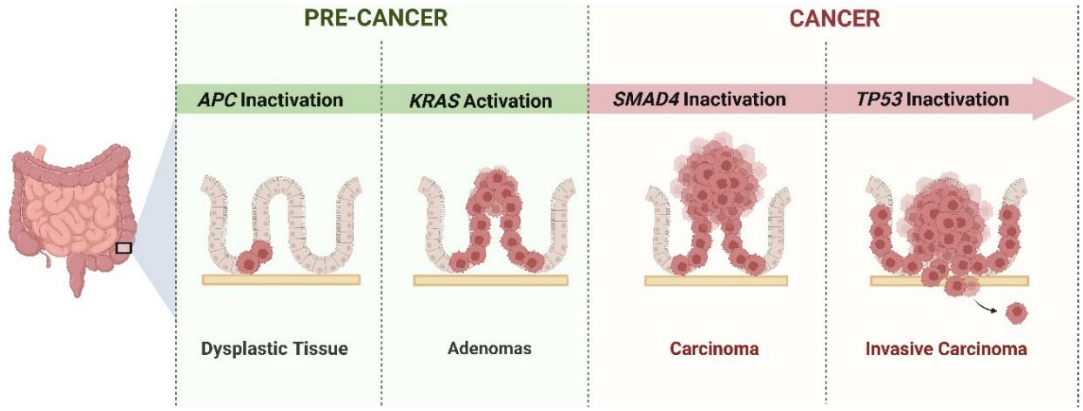


Figure 1.6. Mutations Acquired Through Adenoma-Carcinoma Progression.
The accumulation of loss-of-function of tumour suppressor gene APC and the gain-of-function of oncogene KRAS are important in the transformation from dysplastic tissue to adenomas. Loss-of-function of other tumour suppressor genes such as SMAD4 and TP53 have been cited as important steps for the progression of carcinoma. Figure created with BioRender.

1.3.3 Sessile Serrated Polyps Pathway

The sessile serrated polyp pathway contributes to ~15% of CRCs and involves a different set of mutations to the adenoma-carcinoma sequence, occurring in varying sequences. For sessile serrated polyps, two *BRAF*^{V600E} mutation-mediated pathways are generally accepted. *BRAF* activating mutations lead to an overactivation of the MAPK pathway that is present only in serrated polyps (Kambara *et al.*, 2004). This leads to a short-lived proliferative surge that should end in cellular senescence through the activity of proteins such as p16INK4A (Kriegl *et al.*, 2011). However, cellular senescence can by silenced through methylation (X. Chen *et al.*, 2025), leading to a serrated polyp that progresses into a microsatellite stable (MSS) tumour. Alternatively, methylation could affect the MMR gene *MLH1*, which leads to high-grade dysplasia and a

microsatellite instable (MSI) CpG island methylator phenotype (CIMP; Fennell et al., 2018). As these polyps represent the minority of developing neoplastic lesions, this thesis focused on the conventional adenomas described in 1.3.2.

1.3.4 Polyp Progression into Consensus Molecular Subtypes of CRC

The differences between cancers arising from adenomas and serrated polyps do not end in the differing histology and mutational profile afforded by either polyp, but there is also a difference in the eventual molecular subtype under which the established CRC will be classified. There are four consensus molecular subtypes (CMS) for CRC that are stratified by underlying biology: CMS1 (MSI and Immune Signature), CMS2 (Canonical), CMS3 (Metabolic), and CMS4 (Mesenchymal). Around 14% of CRC fall under CMS1, 37% under CMS2, 13% under CMS3, and 23% under CMS4 leaving the remaining 13% of CRC within a “non-consensus group” (Guinney *et al.*, 2015).

Current research suggests that the adenoma-carcinoma pathway gives rise to CMS2 and CMS3 cancers, which are characterised by WNT pathway activation and MYC-mediated cell proliferation and metabolic deregulation, while CRC from serrated polyps is likely to be classified as CMS1 and CMS4, the immune-rich and stroma-rich subtypes with the worst prognosis (Dunne and Arends, 2024).

1.4 CRC Screening

As previously discussed, the prognosis for CRC greatly depends on the stage at diagnosis, leading to early detection becoming an indispensable tool in reducing CRC mortality. In fact, the results of a recent randomised trial run between multiple countries revealed that screening led to a significant reduction in CRC deaths within a 10-year period (Bretthauer *et al.*, 2022). This led to both the development of various screening tests and the launch of population screening programmes worldwide.

1.4.1 Testing Modalities for CRC Screening

CRC screening methods include non-invasive baseline tests as well as more invasive visualisation methods to assess the bowel. Stool-based tests are a

common first step in CRC screening, with their main objective being to detect blood and/or abnormal cells within the stool. The guaiac faecal occult blood test (gFOBT) was one of the earliest CRC screening tests offered. It was developed in the 1970s and involved placing a stool sample on guaiac paper that detects haemoglobin through a colour change on the addition of H_2O_2 (Greegor, 1967). Although easy to perform at home, gFOBT began to face scrutiny due to its low sensitivity for lower haemoglobin concentrations (Allison, Feldman and Tekawa, 1990) and issues with haemoglobin stability due to environmental factors such as temperature and prolonged time between sample collection and testing (Grazzini *et al.*, 2010; Van Roon *et al.*, 2012). This led to the adoption of faecal immunochemical tests (FITs), which use human haemoglobin-specific antibodies to detect blood in the stool in both a quantitative and qualitative capacity (Carroll, Seaman and Halloran, 2014). The quantitative results achieved by FIT (qFIT) are a major advantage, as healthcare systems can implement a variety of positivity thresholds (Wilschut *et al.*, 2011). This, in addition to the ease of sample collection and the ability to automate the analysis, has led to FIT becoming the preferred test in many healthcare systems.

When visualisation of the bowel is required, endoscopic examination is the gold standard. Modern multi-functional endoscopes are composed of a lighting unit, a camera, an air/irrigation unit, and a functional unit such as a snare, encapsulated within a tube (Katzir, 2003). Depending on how much of the bowel is to be visualised, a sigmoidoscopy or a colonoscopy might be performed, with the former visualising the rectum and the descending colon, while the latter also covers the transverse and the ascending colon (Juul *et al.*, 2024). Regardless, both endoscopic modes require major preparation by the patients, which are summarised in Figure 1.7. In addition, these invasive procedures carry a risk of complications such as bowel perforation and bleeding which might present immediately or sometime after the procedure (Rai and Mishra, 2017).

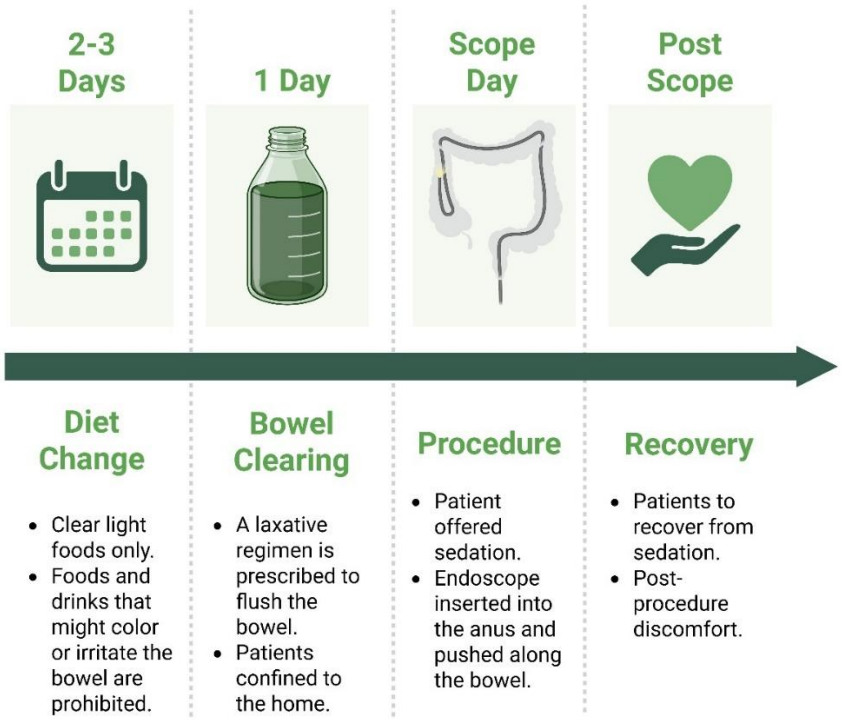


Figure 1.7. Patient Preparation for Colonoscopy.
Figure created with BioRender.

With the high incidence of CRC, and the different treatment options available, healthcare costs associated with the management of CRC are increasing with advancing disease. In 2018, CRC-related costs in European Union member states and the UK, Norway, Switzerland, and Iceland added up to €18 billion (Klimeck *et al.*, 2023). As early-stage disease is asymptomatic and has a better prognosis, programs aiming to screen the population for CRC have been adopted worldwide using the described screening methods among others.

1.4.2 United Kingdom

The aim of CRC screening programmes is to detect CRC precursor lesions or early-stage CRC to improve prognosis and increase cost-effectiveness in healthcare. In the UK, bowel screening begins with an at-home qFIT sample collection kit, starting at the age of 50 (or 60 years in Northern Ireland). Patients with a positive qFIT are invited to an index colonoscopy, during which any precursor lesions (polyps - 1.3) found are removed in a process called polypectomy. If participants return a negative qFIT, or a colonoscopy with no CRC-related findings, they are discharged back into the bowel screening

programme (BSP) where they will retake the qFIT every two years (Bowel Cancer Screening, 2025).

1.4.3 International Programmes

In the United States (US), the US Preventive Services Taskforce (USPSTF) recommends CRC screening begins at age 45 to identify cases developing in younger patients. Additionally, they recommend stopping screening at age 75 to reduce unfavourable outcomes for individuals with co-morbidities or a reduced life expectancy. The recommended testing modalities include annual stool-based tests such as gFOBT or FIT, or a stool DNA FIT (sDNA-FIT) every one to three years to detect abnormal colonic cells. The USPTSF also recommends a CT-colonography (also known as a virtual colonoscopy) to detect bowel abnormalities every five years, a flexible sigmoidoscopy every five to 10 years, and a colonoscopy every 10 years (Davidson *et al.*, 2021).

Although the incidence of CRC in people younger than 50 years old has been observed in Europe, the European Society of Gastrointestinal Endoscopy (ESGE) continues to recommend an age range of 50-75 years old for CRC screening, with the recommendation including a FIT followed by a colonoscopy depending on local resources (Săftoiu *et al.*, 2020).

1.5 Post-polypectomy Surveillance

Following from the screening guidelines described in 1.3, the presence and removal of a polyp drive a post-polypectomy surveillance pipeline aimed to stratify patients at higher risk of developing metachronous polyps or CRC who might warrant more follow-up. There has been evidence that risk stratification of adenomas decreases CRC mortality (Løberg *et al.*, 2014), and polypectomy patients who underwent surveillance procedures had significantly lower rates of CRC (Atkin *et al.*, 2017). In fact, post-screening surveillance can help circumvent issues with interval cancers arising from missed polyps due to poor endoscopy techniques (Corley *et al.*, 2014). Indeed, a high-quality index colonoscopy is the basis of accurate post-screening surveillance, as it allows for the differentiation between a true metachronous lesion and a missed index lesion. A high-quality index colonoscopy covers the entire bowel (including the caecum), ensures the

bowel is sufficiently prepped to allow for the detection of polyps less than 5mm in size, is performed by an endoscopist with a high adenoma detection rate, and involves the complete removal of detected polyps (Gupta *et al.*, 2020).

1.5.1 United Kingdom

In 2002 (four years before the introduction of the bowel cancer screening program in the UK), recommendations for post-polypectomy surveillance were released by the British Society for Gastroenterology (BSG). Taking into consideration the number of polyps found at the baseline colonoscopy, and their size (small polyps were less than 1cm), three risk groups were described alongside follow-up colonoscopy recommendations. Low risk patients (those with 1-2 small adenomas) were no longer followed-up with or were followed-up with after 5 years if they had other risk factors like co-morbidities or a family history of cancer. The intermediate risk group was defined as those with 3-4 small adenomas or those with at least 1 large adenoma and they were surveilled at 3 years. Finally, the high risk group was defined as patients with 5 or more small adenomas or 3 or more adenoma with at least one large adenoma, and they required surveillance after one year. Low risk patients who had no adenomas at follow-up and intermediate risk patients who had two negative follow-ups were discharged from surveillance, while high risk patients without new high risk adenomas had an extended follow up time of 3 years, and high risk patients with high risk findings at follow-up continued with annual visits (Atkin and Saunders, 2002). Since then, revisions were published to include guidelines for patients with other risk factors like family history and bowel co-morbidities (Cairns *et al.*, 2010), and patients with serrated polyps (East *et al.*, 2017).

In 2020, an updated set of guidelines was published by the BSG, taking into account factors like the bowel cancer screening program and polyp dysplasia and histology. These new guidelines included a set of “high risk findings”, that if present, require a surveillance colonoscopy to be performed within 3 years of the screening scope. These findings are defined as multiple premalignant polyps where at least one of them is advanced, a polyp with serrated histology measuring over 10mm or containing dysplastic features, a conventional adenoma measuring over 10mm or containing highly dysplastic features, or five or more premalignant polyps (Figure 1.8; Rutter et al., 2020).

These risk factors have been associated with a higher risk of metachronous polyps or CRC. High-grade dysplasia in the index polyp was significantly associated with advanced lesions or CRC upon surveillance (Fairley *et al.*, 2014) and those with large polyps with high-grade dysplasia had the highest chance of another lesion within 3 years (Facciorusso *et al.*, 2016).

Similarly, polyps measuring 10mm or more were found to have a higher odds ratio for the detection of more advanced lesions and of CRC upon follow-up (Fairley *et al.*, 2014), and the chance of high-grade dysplasia being present increased with increasing polyp size (Pickhardt *et al.*, 2010).

Finally, a study found that increasing number of polyps at the index colonoscopy (3 or more polyps) almost doubled the odds ratio of an advanced metachronous lesion and almost quadrupled it for CRC (Fairley *et al.*, 2014), while another reported higher incidence rate of advanced adenomas upon surveillance for patients with multiple advanced adenomas (Laish *et al.*, 2017).

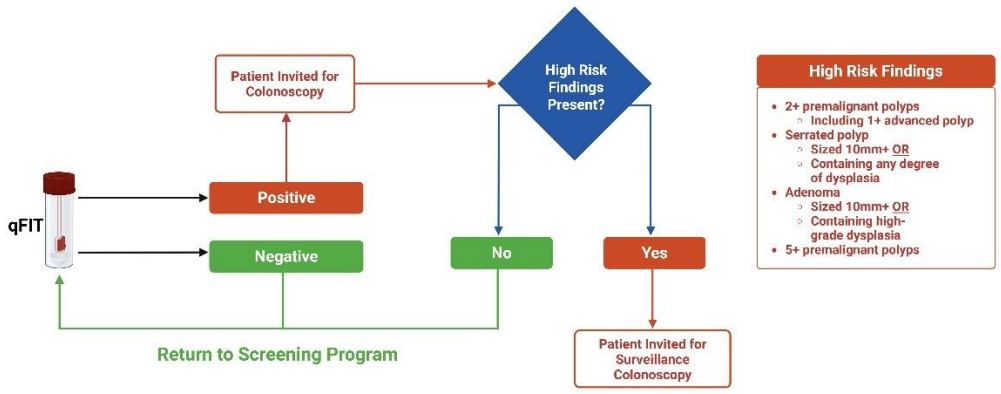


Figure 1.8. Summary of CRC Screening and Surveillance Guidelines in the United Kingdom. qFIT = Quantitative Faecal Immunochemical Test. Figure created with BioRender.

1.5.2 International Surveillance Guidelines

Outside the UK, post-polypectomy surveillance recommendations do not differ too much from those issued by the BSG. The US Multi-Society Task Force define "high risk adenomas" identified within a colonoscopy as any adenoma ≥10mm in size, any adenoma with a degree of villous histology (including tubulovillous histology), any adenoma with high-grade dysplasia, or 3 or more adenomas.

Patients with such findings are recommended to undergo a surveillance colonoscopy at the 3-year interval (Gupta *et al.*, 2020).

The ESGE also recommends a surveillance colonoscopy after 3 years for patients who had a complete resection of an adenoma measuring ≥10mm or an adenoma with high-grade dysplasia, or 5 or more adenomas, or a serrated polyp larger than 10mm or with any degree of dysplasia (Hassan *et al.*, 2020).

However, guidelines in Asian countries like Korea differ in the strength of these recommendations. For instance, although adenomas ≥10mm in size, and more than 3-4 adenomas fall under the 3-year interval for a surveillance colonoscopy, these are recommended only conditionally (i.e. depending on the individual clinical situation). This is the same for patients with serrated lesions (with or without dysplasia). Korean guidelines only strongly recommend a surveillance colonoscopy after 3 years in cases where villous histology is present (including tubulovillous adenomas) and in adenomas with high-grade dysplasia (Kim *et al.*, 2022).

1.5.3 Problems with Current Surveillance Methods

The post-polypectomy surveillance aims to monitor patients who are likely to develop metachronous polyps or CRC in the future. However, the metrics used to assign “high-risk” status are insufficient, as it has been reported that between 20-50% of post-polypectomy patients develop metachronous polyps (Hao *et al.*, 2020). While this is evidence to warrant surveillance colonoscopies, the current guidelines continue to expose the patients who were stratified as “high-risk” but who never develop metachronous polyps or CRC to the risk and discomfort of surveillance colonoscopies. Conversely, patients stratified as “low-risk” but do develop metachronous polyps or CRC are at a disadvantage as they are often missed. In fact, a recent study examining the effectiveness of the BSG2020. An audit was performed on the INCISE bowel screening patient cohort and applied the new 2020 guidelines, and this revealed that not only did this risk stratification not differentiate between advanced and non-advanced metachronous lesion risk, but also found that 39% of high-risk patients did not have a metachronous lesion while 13% of low-risk patients had a metachronous lesion or CRC (Johnstone *et al.*, 2023). This is not supportive of the overarching

goal of cancer screening and early detection, which is to prevent the development of malignant tumours to decrease the incidence and mortality rates of CRC.

1.5.3.1 Endoscopy Backlog

In addition to the increased risk to patients, the burden that surveillance colonoscopies place on healthcare resources in the UK is too large to ignore. Prior to the COVID-19 pandemic, there was an increase of 12-15% in demand for endoscopy services in the UK in 2019, with 25% of service providers failing to meet urgent cancer waiting targets (Ravindran *et al.*, 2021). When COVID-19 restrictions came into effect in March 2020, 91% of services across the UK reported a referral backlog of 2-6 months (Catlow *et al.*, 2021), and in January-April 2021, only 13% of endoscopy services in the UK met their surveillance waiting times (Ravindran *et al.*, 2022). This backlog was attributed to different reasons, some long-standing such as limited endoscopy staff, physical capacity for patient intake, and shortage of equipment, while others were temporary, such as the enforced downtime between procedures in effect during the pandemic (Catlow *et al.*, 2021; Ravindran *et al.*, 2021, 2022).

It has become clear that not only is the 3-year period post index colonoscopy would become arbitrary with the rising demand, but also that with half of high-risk patients not seeing benefits of the surveillance colonoscopies, while the mis-stratified low-risk patients being missed from surveillance, these resources are better used for other purposes.

1.6 Human Microbiome

The human microbiome is the accepted term for the trillions of bacteria, archaea, and fungi that inhabit the human body. These microorganisms have been linked to human health for centuries; therefore, the Human Microbiome Project was launched to further understand their role (Turnbaugh *et al.*, 2007). The balance of the various microbial populations is of the utmost importance, and the loss of this balance, known as dysbiosis, has been linked to a variety of diseases. For instance, a marked difference in the microbiome of type-2 diabetic patients was identified in comparison to non-diabetics (Larsen *et al.*, 2010) and

microbial biomarkers were shown to be relevant in liver cirrhosis (Qin *et al.*, 2014).

Although the human microbiome refers to all microbes, it is imperative to differentiate the species that are local to one area of the body and invasive to another. This gives rise to the concept of anatomical microbiomes, which have been shown to modulate the process of disease. A study on SARS-CoV-2 patients revealed that dysbiosis in the nasal cavity was associated with a higher viral RNA load (Rhoades *et al.*, 2021), while site-specific dysbiosis was associated with atopic dermatitis (Bjerre *et al.*, 2021).

1.6.1 The Gut Microbiome

Another anatomical microbiome is the gut microbiome. Under healthy conditions, it has been shown to support normal human function. For example, genomic analysis of bacterial species commonly found in the gut revealed that the microbiome is an essential source of vitamin B production (Magnúsdóttir *et al.*, 2015).

However, the role of the gut microbiome in CRC development has been increasingly studied, with different bacteria being involved in different disease stages (Mizutani, Yamada and Yachida, 2020). It has been agreed between experts that the mechanism of involvement of the microbiome in CRC development is multifaceted, and includes genotoxicity, genomic integration, driving inflammation, modulating host immunity, and affecting metabolism (Scott *et al.*, 2019). Some bacteria that have been shown to be involved in CRC include *Escherichia coli*, the pks+ strains of which have been shown to produce the DNA alkylating agent colibactin (Wilson *et al.*, 2019), which induces a mutational signature in 3D cultured cells similar to that of CRC patients (Pleguezuelos-Manzano *et al.*, 2020). Another established species in CRC development is *Bacteroides fragilis* which promotes chemoresistance (Ding *et al.*, 2025). A multi-model study on *Parvimonas micra* showed an enrichment of the bacteria in stool samples and tissue biopsies from CRC patients and was an independent prognostic factor of poor survival (L. Zhao *et al.*, 2022). The same study highlighted that *P. micra* was also associated with a higher tumour burden in mice and was linked to an up-regulation in cell proliferation genes (L. Zhao *et*

al., 2022). On the other hand, *Ruminococcus spp.* and other producers of the anti-tumorigenic acid butyrate, were less abundant in CRC patient stool samples when compared to stool samples of healthy individuals (Weir *et al.*, 2013), highlighting the role of dysbiosis in disease formation.

There are also bacterial species that are emerging in the study of CRC, i.e. those which were thought to be commensals or more involved in other diseases such as *Fusobacterium nucleatum*.

1.6.2 *Fusobacterium nucleatum*

The oral microbiome exists in a rapidly changing and harsh environment, characterised by constant mechanical force changes (during events such as chewing), fluctuating pH, temperatures, and aerobic levels. This leads to the production of biofilms by the oral bacteria as a survival mechanism. A biofilm is defined as a collection of species that form a single entity within a matrix of water and secreted polymers such as cellulose (Serra, Richter and Hengge, 2013). Many bacterial species take residence in the biofilm, and as such the biofilm has been implicated in many oral diseases.

One such resident is a Gram-negative, rod shaped, anaerobic bacterium known as *F. nucleatum*. Long thought to be a commensal bacterium, *F. nucleatum* began emerging as a pathogen when it was regularly found in diseased samples and was subsequently linked to periodontitis (Saygun *et al.*, 2011) and gingivitis (Liu *et al.*, 2014) in the oral cavity. Interestingly, it was also found at distant sites in the body such as the amniotic fluid and cord blood from neonatal septic pregnancies (Wang *et al.*, 2013), atherosclerosis plaques (Figuero *et al.*, 2011), and blood from patients suffering from Alzheimer's disease (Sparks Stein *et al.*, 2012). *F. nucleatum* was also detected is CRC tissue (Castellarin *et al.*, 2012), and a study reported that not only was *F. nucleatum* found in CRC, but also that the strains in oral cavity samples matched those from the CRC tissue of the same patients (Komiya *et al.*, 2019). The question of how an oral microbe is found in these sites leads to a consideration of the link between gum health and systemic health. The most accepted route of transmission is through inflamed, bleeding gums, which allows the bacteria to enter the bloodstream and disseminate throughout the body (Abed *et al.*, 2020).

1.6.2.1 *F. nucleatum* in CRC

The relationship between CRC and *F. nucleatum* was first suggested with the detection of *F. nucleatum* DNA in CRC (Kostic *et al.*, 2012). *F. nucleatum* has also been linked to poor cancer-specific survival, as well as poor prognostic factors such as right-sided cancers, poorly differentiated cancers, and high CIMP status (Mima, Nishihara, *et al.*, 2016). In addition, *F. nucleatum* was found in CRC liver metastasis and was associated with the cancer cells at distant sites (Bullman *et al.*, 2017).

F. nucleatum has been shown to self-enrich through modulation of the tumour microenvironment. Mouse model and cell line work demonstrated how *F. nucleatum* uses its lectin Fap2, to bind to D-galactose-b(1-3)-N-acetyl-D-galactosamine (Gal-GalNAc), a polysaccharide found on CRC cells, to attach and enrich itself in CRC (Abed et al., 2016; Figure 1.9). Furthermore, *F. nucleatum* can support an immunosuppressive environment. Fap2 allows *F. nucleatum* to modify and inhibit the anti-tumour immune response by binding to T cell immunoglobulin and ITIM domain (TIGIT), an inhibitory immune cell surface receptor (Gur et al., 2015; Figure 1.9). Additionally, there is evidence of recruitment and selective expansion of immune inhibitory myeloid-derived suppressor cells through the bacterial release of short-chain fatty acids (Kostic *et al.*, 2013).

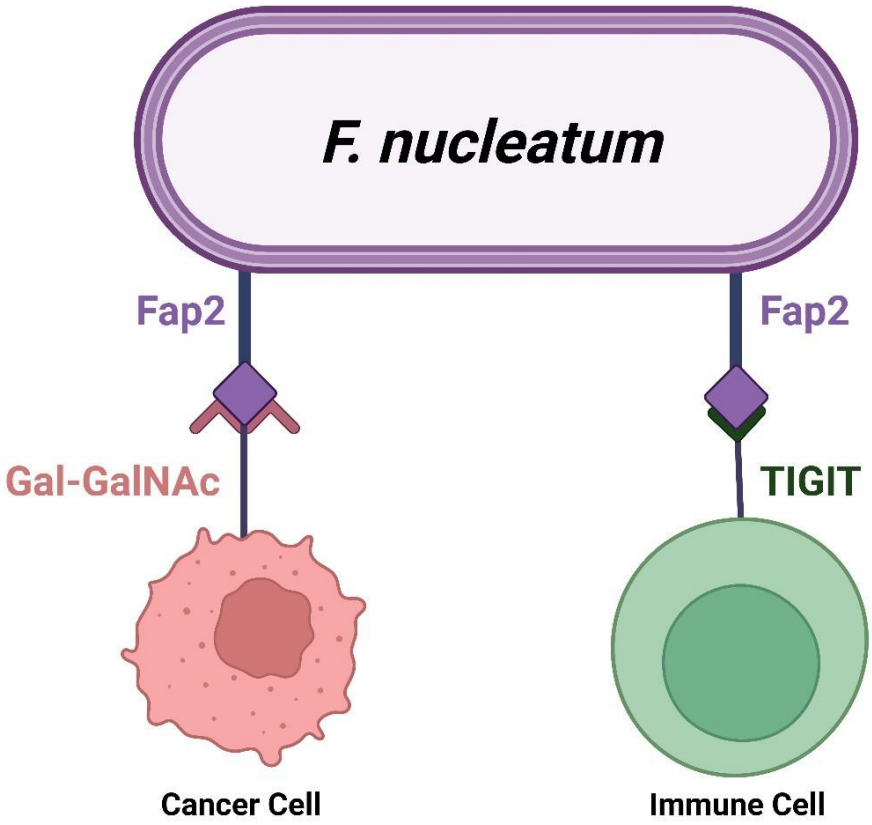


Figure 1.9. *F. nucleatum* Interacts with Cancer and Immune Cells Using Fap2.
Bacterial lectin Fap2 allows *F. nucleatum* to enrich itself in CRC by binding the bacteria to cancer cell surface polysaccharide Gal-GalNAc. Fap2 also allows for bacteria-mediated immunomodulation of CRC as it can bind to immune-inhibitory cell surface receptor TIGIT, leading to a suppression of the anti-tumour immune response. Gal-GalNAc: D-galactose-b(1-3)-N-acetyl-D-galactosamine. TIGIT: T cell immunoglobulin and ITIM domain.
Figure created with BioRender.

Besides its ability to enrich itself, *F. nucleatum* can manipulate cells to generate a more proliferative, cancer-supporting environment. For example, it has been shown to induce Wnt/ß-Catenin signalling and gene transcription by interrupting the E-Cadherin/ß-Catenin adhesion complex. When a Wnt molecule activates its receptor, a cascade of events is initiated that leads to increased levels of ß-Catenin and the eventual transcription of downstream genes. *F. nucleatum* was shown to mimic this cascade by using its FadAc adhesion proteins to bind to E-Cadherin. E-Cadherin binds α- and ß-Catenin as part of the catenin/cadherin complex for epithelial cell adhesion (Kim *et al.*, 2019). The FadAc interaction phosphorylates E-Cadherin, which leads to the release of its bound ß-Catenin. This increase in cytosolic ß-Catenin levels leads to the same downstream effects of the canonical Wnt signalling pathway (Figure 1.10; Rubinstein et al., 2013).

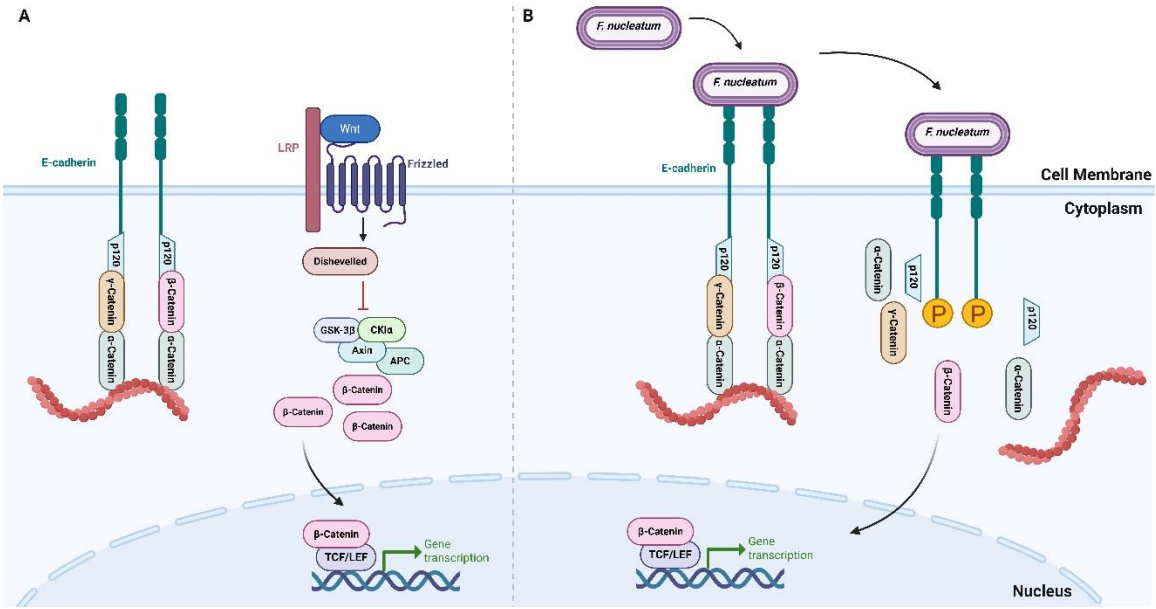


Figure 1.10. *F. nucleatum* Can Mimic Wnt-mediated Signalling.
[A: Left] E-Cadherin binds β-Catenin in a chain with p120, α-Catenin, γ-Catenin, and actin filaments (depicted as red wavy molecules) as part of the cell-cell adhesion complex. **[A: Right]** When Wnt binds to LRP and Frizzled, the β-Catenin destruction complex (APC, GSK, Axin, CKIα) becomes disrupted by dishevelled, allowing β-Catenin levels to rise in the cytosol. β-Catenin then translocates to the nucleus, where it binds to TCF/LEF transcription factor, and aids in transcription of oncogenes involved in cell proliferation like MYC. **[B]** F. nucleatum binds to cell surface E-Cadherin, through its FadAc molecules. This internalises and phosphorylates E-Cadherin, leading to the release of β-Catenin to the cytosol. β-Catenin is then free to translocate into the nucleus, bind to TCF/LEF transcription factors, and begin the transcription process of downstream oncogenes like MYC. Figure created with BioRender.

Furthermore, *F. nucleatum* has been shown to activate the NF-κB pathway. As a Gram-negative bacterium, *F. nucleatum* presents lipopolysaccharide (LPS) on its surface. LPS can bind to cell surface Toll-like receptor 4 (TLR4) leading to the recruitment of the adapter proteins MyD88 and TIRAP, which activate IRAKs 1/4 and TRAF6. Alternatively, LPS-mediated TLR4 activation could also lead to the recruitment of adaptors TRIF and TRAM, activating RIP1 and TRAF6. Regardless of which adaptor complex is recruited, TAK1 is activated and subsequently so is IKK, which phosphorylates and degrades the NF-κB inhibitor IκB. The NF-κB complex can now translocate to the nucleus, where it aids in the transcription of oncogenic mIR-21 (Figure 1.11; Yang et al., 2016).

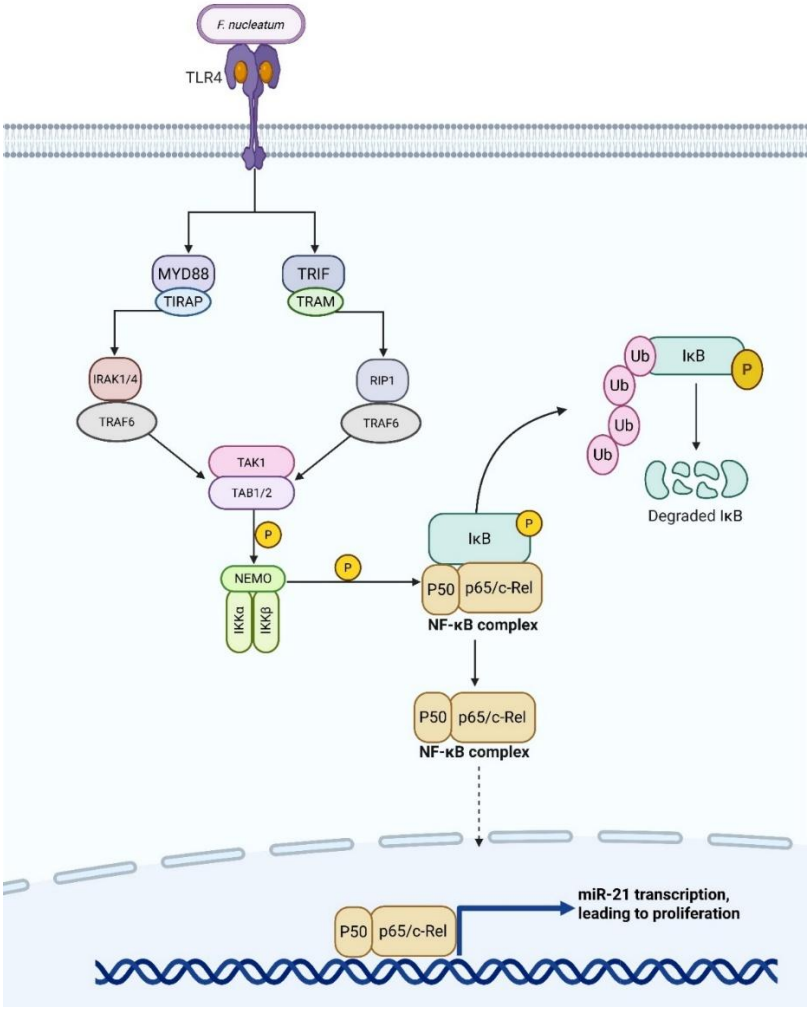


Figure 1.11. *F. nucleatum* Can Activate NFκB Signalling.
LPS mediated activation of NF-κB signalling leads to cellular proliferation. LPS recruits adaptor proteins like MyD88 and TIRAP by binding to cell surface TLR4. This leads to IRAK1/4 and TRAF6 activation. TLR4 activation by LPS could also recruit different adaptor proteins like TRIF and TRAM which activate RIPS1 and TRAF6. Both pathways lead to TAK1 and IKK activation. IKK phosphorylates and degrades NF-κB inhibitor IκB, allowing the NF-κB complex to translocate from the cytoplasm to the nucleus where it aids in transcription of target genes. LPS: Lipopolysaccharide. NF-κB: Nuclear factor kappa-light-chain-enhancer of activated B cells. MyD88: Myeloid Differentiation primary response 88. TIRAP: Toll/IL-1 receptor domain containing adaptor protein. TLR4: Toll-like Receptor 4. IRAK1/4: Interleukin-1 Receptor-Associated Kinase 1/4. TRAF6: Tumor Necrosis Factor Receptor-Associated Factor 6. RIPS1: Receptor Interacting Protein Kinase 1. TRIF: TIR-domain-containing adapter-inducing interferon-β. TAK1: Transforming Growth Factor-β-Activated Kinase 1. IκB: Inhibitor of κB. Ub: Ubiquitin. P: Phosphate. Figure created with BioRender.

The effects of *F. nucleatum* on CRC have led to a large body of research, however, it seems that this has been confined to established malignant tumours. Investigation into the effects of *F. nucleatum* on adenoma tissue is sparse (Kostic *et al.*, 2013; McCoy *et al.*, 2013; Flanagan *et al.*, 2014; Yu *et al.*, 2016), which raises the question of how early the microbiome is involved in CRC development.

1.7 Aims and Hypothesis

With the impact of CRC-directed healthcare continuing to rise, and with current post-screening surveillance practices being insufficient, novel avenues of improving bowel screening were required. As little is known on the effects of the microbiome on colorectal polyps, this could potentially be an area of interest. To that end, this thesis aimed to investigate whether the microbiome of index polyps detected during bowel screening is predictive of metachronous polyps or CRC development. It further aimed to investigate other avenues for the prediction of metachronous polyps, identified within intrinsic characteristics of the polyp though mutational analysis and protein level studies. To achieve this, the following main objectives were proposed:

1. Identify suitable methods of detecting and visualising *F. nucleatum* in archival formalin-fixed paraffin-embedded (FFPE) tissue from the largest polyp taken from screening patients within the Scottish BSP.

2. Examine the spatial environment around *F. nucleatum* to ascertain any bacterial-mediated differences in pathway regulation.

3. Investigate the immune response within FFPE tissue from screening patients within the Scottish BSP, in relation to the presence of *F. nucleatum*.

4. Explore the mutational landscape of bowel screening patient polyp tissue to identify genes of interest and investigate any correlation between the presence of bacteria and mutations.

5. Investigate the protein expression levels of identified targets of interest (E-Cadherin, ß-Catenin, and SOX9), assessing their association with clinical factors, and identifying their potential as predictive biomarkers for the detection of metachronous polyps or CRC.

5. Determine whether the growth of polyps is influenced by pathways of interest by stimulating these pathways in 3D FAP patient-derived organoids.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 INCISE Tissue Cohort

The Integrated Technologies for Improved Polyp Surveillance (INCISE) cohort consisted of 2642 bowel screening patients from the Greater Glasgow and Clyde area (2009-2016) who on index colonoscopy, had a minimum of one polyp removed and received follow-up colonoscopies as part of the Scottish Surveillance programme for 6 months - 6 years after the index colonoscopy (Johnstone *et al.*, 2021). Pathological information on each excised polyp was obtained from histopathology reports. These included but were not limited to, number of polyps present at index and follow-up colonoscopies, polyp histology (whether a polyp presented adenoma or serrated features), polyp morphology (characterised by the presence or absence of villous features), polyp location (right/left colon, or rectum), polyp size ($<$10mm$\leq$), and presence of advanced polyps. From these clinical variables, the BSG2020 Guidelines were determined and risk category recorded for each patient. Patient status at follow-up surveillance colonoscopy was dependent on the presence or absence of a metachronous polyp(s) or CRC and was determined using electronic endoscopy reporting software (Ver. 2.5, Unisoft GI Reporting Software, Unisoft Medical Systems, UK) and Telepath Laboratory Information Management System electronic pathology database. All analysis was done on the index polyp from each patient, defined as the largest polyp in patients with multiple polyps or as the only polyp in patients with single polyps.

2.1.1.1 *F. nucleatum* Subset

The *F. nucleatum* cohort consisted of full section tissue from 198 INCISE patients that were randomly selected from the larger INCISE cohort. 68% of patients were male (n=134), 95% were of screening age (n=188), and 73% of patients presented with polyps in the left colon (n=145). 19% of patients had polyps with high-grade dysplasia (n=37), 49% of patients presented with 2-4 polyps (n=96), 94% had polyps larger than 10mm (n=187), and 55% of patients had polyps with tubulovillous features (n=108). 63% of patients were found to be high-risk according to current BSG2020 Guidelines (n= 125) and 55% of patients (n=109) had metachronous polyps or CRC at follow-up (Figure 2.1).

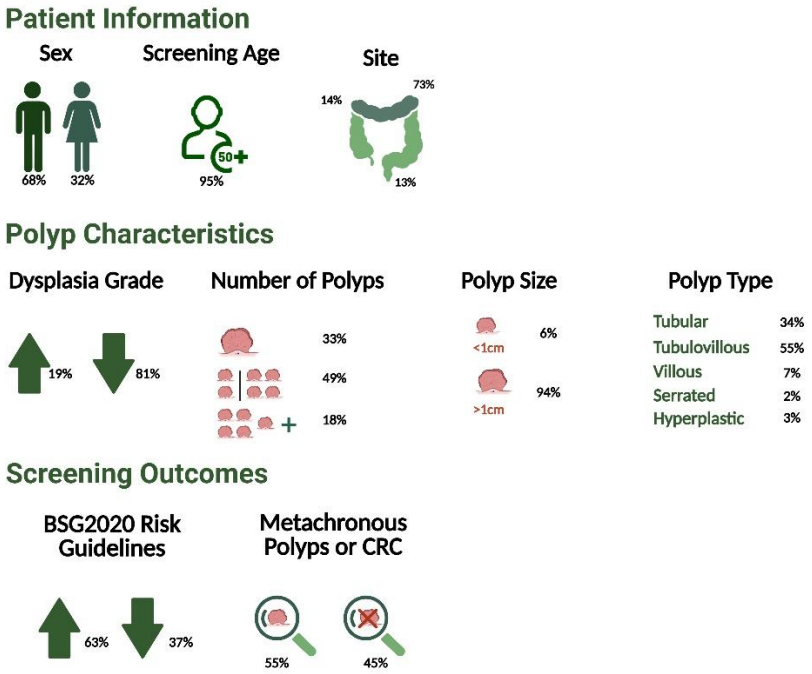


Figure 2.1. *F. nucleatum* Cohort Description.
Cohort description of various variables pertaining to patients, polyps, and screening outcomes. Arrows for "Dysplasia Grade" indicate high-grade dysplasia (upward pointing arrow) and low-grade dysplasia (downward pointing arrow). Arrows for "BSG2020 Risk Guidelines" indicate high risk (upward pointing arrow) and low risk (downward pointing arrow). N=198. Screening age is at 50 years. Figure created with BioRender.

2.1.1.2 Mutational Subset

The mutational subset consisted of index adenomas from 722 INCISE patients selected based on tissue availability (tissue block must have allowed for 10µm sections). 74% of patients were male (n=535), 96% were of screening age (n=690), and 71% of patients presented with polyps in the left colon (n=511). 16% of patients had polyps with high-grade dysplasia (n=112), 51% of patients presented with 2-4 polyps (n=365), 92% has polyps larger than 10mm (n=668), and 52% of patients had polyps with tubulovillous features (n=372). 65% of patients were found to be high-risk according to current BSG2020 Guidelines (n=469) and 59% of patients (n=472) had metachronous polyps or CRC at follow-up (Figure 2.2).

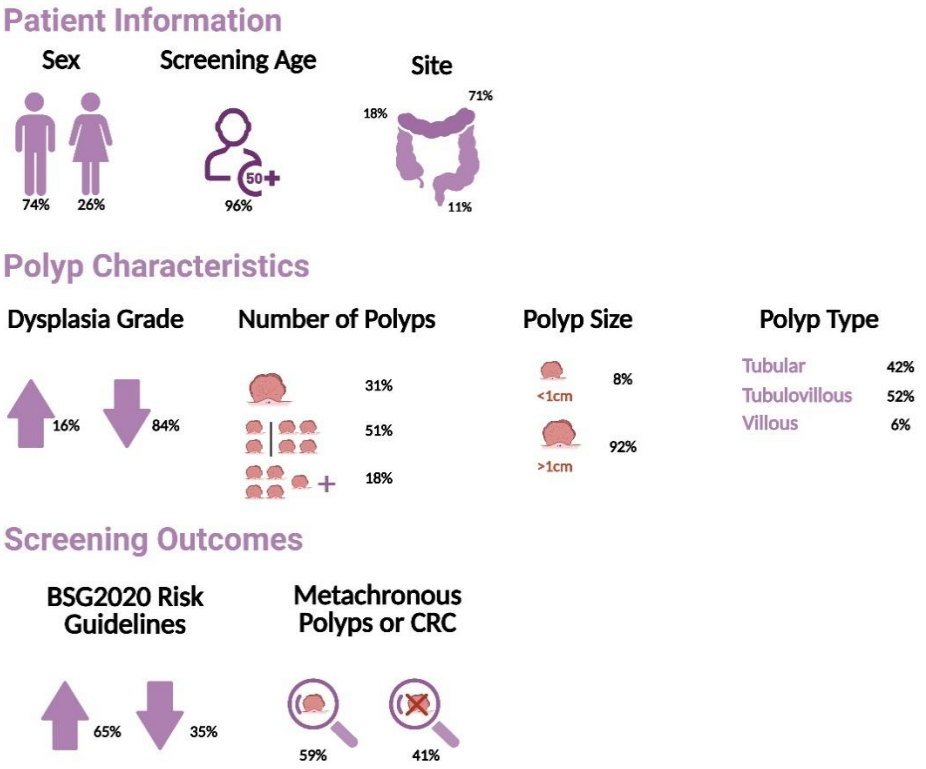


Figure 2.2. Mutational Cohort Description.
Cohort description of various variables pertaining to patients, polyps, and screening outcomes. Arrows for "Dysplasia Grade" indicate high-grade dysplasia (upward pointing arrow) and low-grade dysplasia (downward pointing arrow). Arrows for "BSG2020 Risk Guidelines" indicate high risk (upward pointing arrow) and low risk (downward pointing arrow). N=722. Screening age is at 50 years. Figure created with BioRender.

2.1.1.3 INCISE TMA Cohort: Construction

The tissue cohort was characterised using full sections before cases were selected for the construction of a tissue microarray (TMA). The INCISE TMA cohort was constructed from 1258 patients selected based on the size of the polyp allowing multiple cores to be obtained (this biases the TMA to patients with larger polyps). Each patient was represented by four cores 0.6mm in diameter. A pathologist identified areas of luminal and basal epithelium within each polyp, and 2 of the 4 cores were taken from the luminal epithelium, and 2 from the basal epithelium. This was to account for tissue heterogeneity and facilitate tissue assessment in both luminal and basal regions within the same polyp. A diagram explaining the mapping of the TMA slide is shown in Figure 2.3.

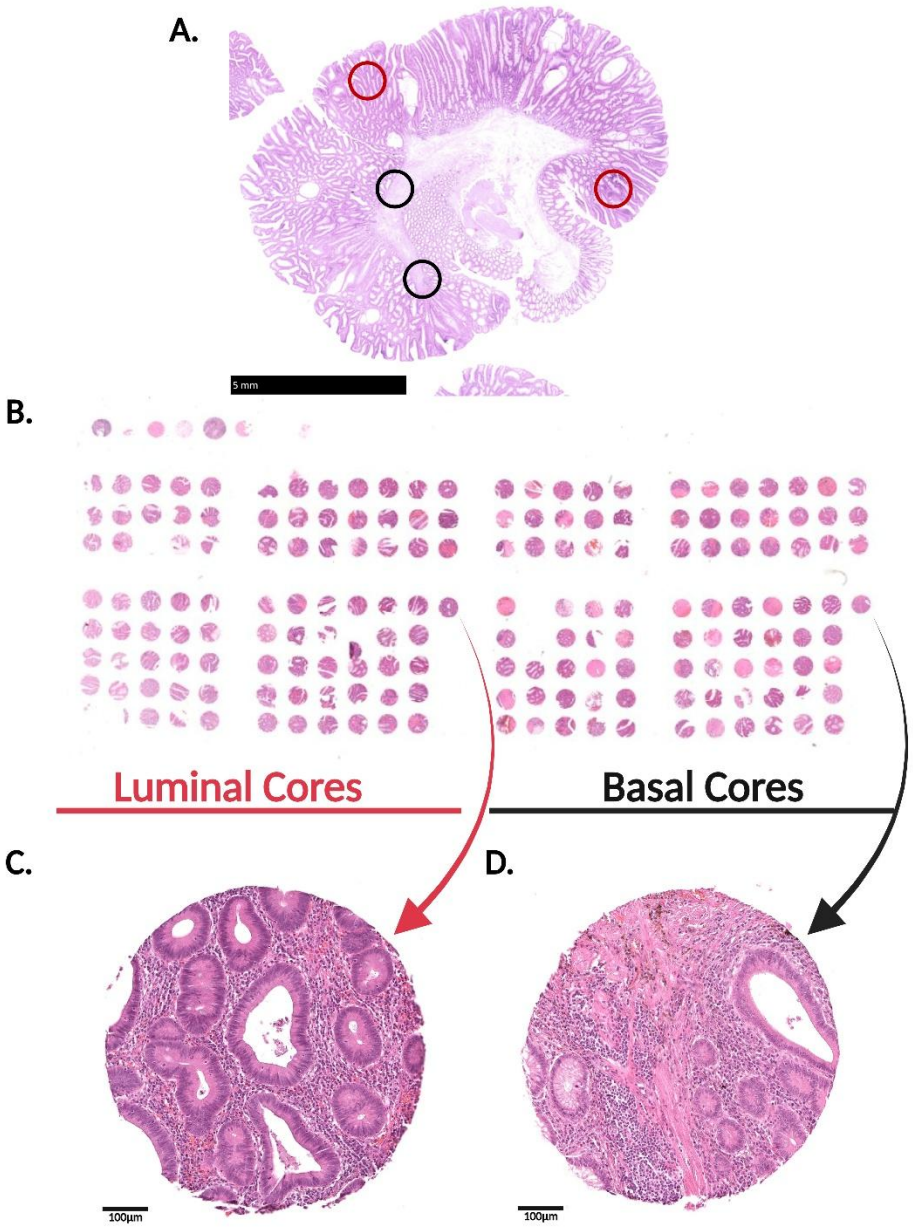


Figure 2.3. TMA Construction.
[A] H&E image of a sample polyp indicating areas designated as luminal (red circles) and basal (black) circle. **[B]** H&E TMA slide where luminal cores are on the left, and basal cores of the same patients at the same location are on the right. **[C]** H&E of luminal polyp core. **[D]** H&E of basal polyp core. C and D are of the same patient.

In order to build prediction models that are generalisable and consistent, and in adherence to the REMARK criteria, cases within this cohort were divided into two separate training and test datasets through a process of computational randomisation.

2.1.1.3.1 Training Dataset

The training dataset of the TMA cohort comprised of 869 patients (70%). Due to low numbers, patients with polyps 10mm in size or less (n=38) and patients with non-adenoma polyps (n=13) were excluded from analysis. Following these exclusion criteria, the training dataset and matched TMA consisted of 818 patients with matched tissue and data.

71% of patients were male (n=584), 96% were of screening age (n=785), and 73% of patients presented with polyps in the left colon (n=n=595). 16% of patients had polyps with high-grade dysplasia (n=132), 54% of patients presented with 2-4 polyps (n=440), and 48% of patients had polyps with tubulovillous features (n=397). 66% of patients were found to be high-risk according to current BSG2020 Guidelines (n= 538) and 56% of patients (n=455) had metachronous polyps or CRC at follow-up (Figure 2.4).

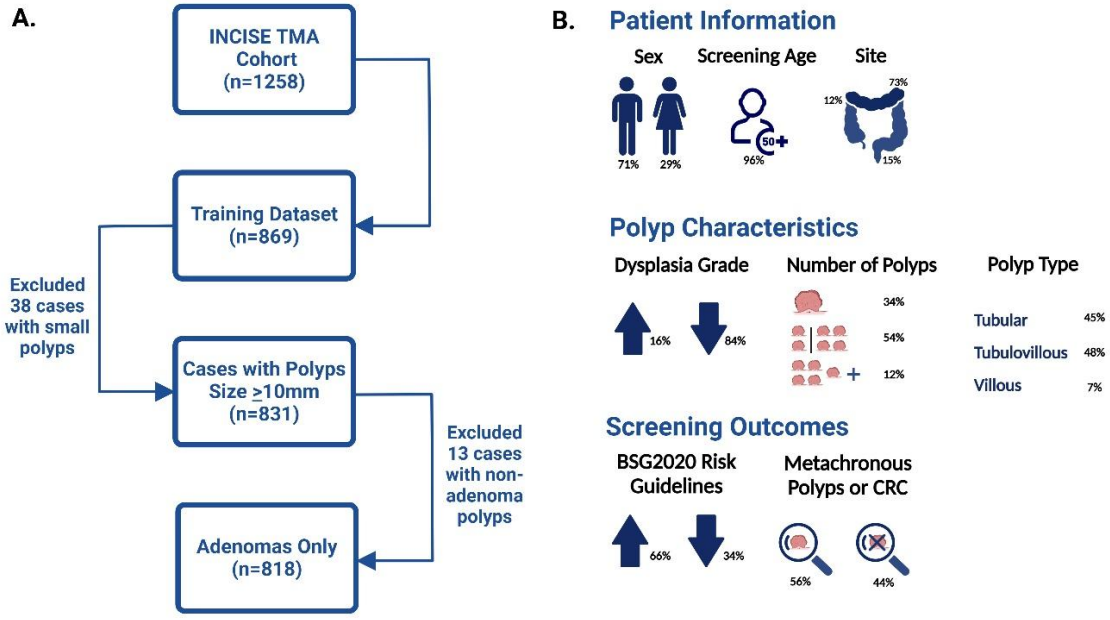


Figure 2.4. Training Dataset Description.
[A] CONSORT diagram of cases included and excluded in the training dataset. **[B]** Cohort description of various variable pertaining to patients, polyps, and screening outcomes. Arrows for "Dysplasia Grade" indicate high-grade dysplasia (upward pointing arrow) and low-grade dysplasia (downward pointing arrow). Arrows for "BSG2020 Risk Guidelines" indicate high risk (upward pointing arrow) and low risk (downward pointing arrow). Screening age is at 50 years. Figure created with BioRender.

2.1.1.3.2 Test Dataset

The test dataset of the TMA cohort comprised of 389 patients (30%). As with the training dataset, patients with polyps 10mm in size or less (n=18) and patients

with non-adenoma polyps (n=13) were excluded from analysis. Following this exclusion criteria, 358 patients with matched data and tissue formed the test dataset and TMA cohort.

75% of patients were male (n=268), 95% were of screening age (n=341), and 73% had left colonic polyps (n=259). 16% had polyps with high grade dysplasia (n=58), 54% presented with 2-4 polyps (n=194), and 55% had polyps with tubulovillous features (n=198). 71% of patients were found to be high-risk according to current BSG2020 Guidelines (n=253) and 57% of patients (n=205) had metachronous polyps or CRC at follow-up (Figure 2.5).

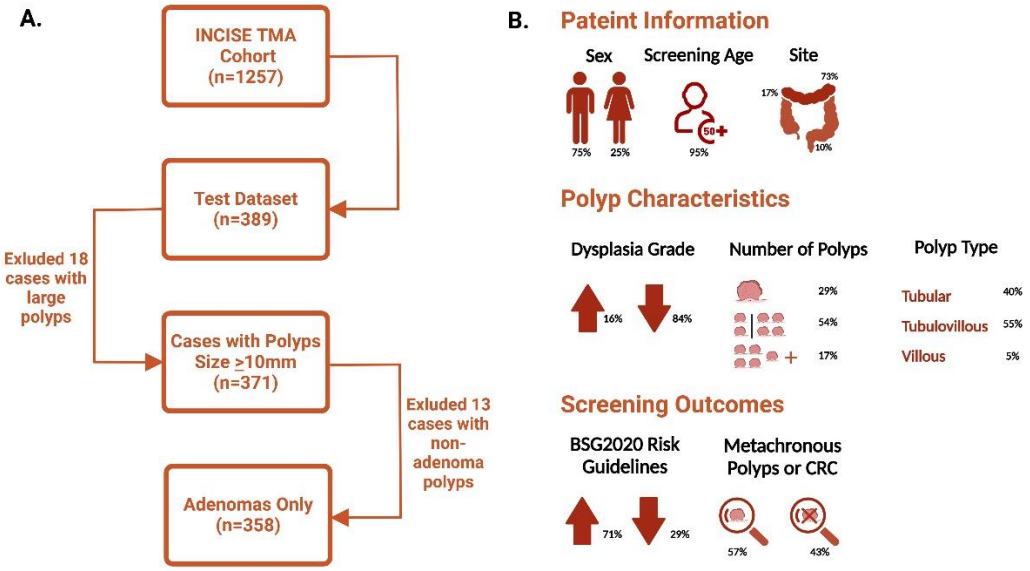


Figure 2.5. Test Dataset Description.
[A] CONSORT diagram of cases included and excluded in the test dataset. **[B]** Cohort description of various variable pertaining to patients, polyps, and screening outcomes. Arrows for "Dysplasia Grade" indicate high-grade dysplasia (upward pointing arrow) and low-grade dysplasia (downward pointing arrow). Arrows for "BSG2020 Risk Guidelines" indicate high risk (upward pointing arrow) and low risk (downward pointing arrow). Screening age is at 50 years. Figure created with BioRender.

2.1.2 Ethical Approval Statement

Ethical approval was obtained from the West of Scotland Research Ethics Committee (22/WS/0020). Patient information is held within the Glasgow and Clyde Safe Haven (GSH20CO002) and was anonymised before use. The protocols used and described in this study were performed in accordance with the REMARK guidelines.

2.1.3 Antibodies

Table 2.1. Information on Antibodies Used

Primary Antibodies								
Target	**Isotype**	**Clonality**	**Clone**	**Manufacturer**	**Cat. / Ref.**	**Host Species**	**Species Cross-reactivity**	**Application**
SOX9	IgG	Polyclonal	-	EMD Millipore	AB5535	Rabbit	Human/Mouse /Rat	IHC WB
CD3*	IgG	Monoclonal	LN10	Novacastra	NCL-L-CD3-565	Mouse	Human	IHC
CD4*	-	Monoclonal	4B12	Vector	VP-C319	Mouse	Human	IHC
CD8*	IgG	Monoclonal	C8/144B	Dako	M7103	Mouse	Human	IHC
IgG1κ	IgG1κ	Monoclonal	MG1-45	BioLegend	401402	Mouse	Human/Mouse /Rat	IHC (Isotype Control)
ß-actin	IgG	Monoclonal	8H10D10	Cell Signalling Technologies	3700S	Mouse	Human/Mouse /Rabbit/Monkey /Dog	WB (Loading Control)
ß-Catenin	IgG1κ	Monoclonal	ß-Catenin-1	Dako	M3539	Mouse	Human	IHC WB
ß-tubulin	IgG	Polyclonal	-	abcam	AB21058	Rabbit	Human/Mouse /Rat	WB (Loading Control)
E-Cadherin	IgG	Monoclonal	36/E-Cadherin (RUO)	BD Biosciences	610182	Mouse	Human/Mouse /Rat/Dog	WB IHC
c-Myc	IgG	Monoclonal	Y69	abcam	AB32072	Rabbit	Mouse/Rat /Human	WB

Secondary Antibodies					
Product Name	**Isotype**	**Manufacturer**	**Cat. / Ref.**	**Host Species**	**Species Cross-reactivity**
Anti-Rabbit HRP-linked Antibody	IgG	Cell Signaling Technologies	7074S	Goat	Rabbit
Anti-mouse HRP-linked Antibody	IgG	Cell Signaling Technologies	7076S	Horse	Mouse
Peroxidase-conjugated Anti-Rabbit	IgG	Jackson ImmunoResearch Laboratories	111-035-144	Goat	Rabbit
Peroxidase-conjugated Anti-Mouse	IgG	Jackson ImmunoResearch Laboratories	715-035-150	Donkey	Mouse

IHC: Immunohistochemistry. WB: Western Blot.

* Denotes diagnostic-grade antibodies.

2.1.4 Chemicals

Table 2.2. Information on General Chemicals Used

Name	Full Product Name	Cat. / Ref.	Manufacturer	Origin
Acrylamide	Rotiphorese®Gel 30	3029.1	Carl Roth GmbH	Germany
Alcohol	Absolute Ethanol	20821.330	VWR Chemicals	UK
APS	Ammonium Persulfate	17874	ThermoScientific	UK
Bromophenol	Bromophenol Blue	B0126-25G	Sigma-Aldrich	USA
Citric Acid	Citric Acid	27109	Sigma-Aldrich	USA
DMSO	Dimethyl Sulfoxide	BP231-1	Fisher	UK
DTT	1.4-Dithiothreitol	38225822	Roche	Mannheim, Germany
EDTA	Diaminoethanetetra-acetic acid disodium salt dihydrate	D/0700/60	Fisher	UK
Glycerol	Glycerol	G5516-1L	Sigma-Aldrich	USA
Glycine	Glycine	G7126-5KG	Sigma-Aldrich	USA
H_2O_2	Hydrogen Peroxide	23619.297	VWR Chemicals	UK
HCl	Hydrochloric Acid S.G.	H/1200/PB08	Fisher	UK
Methanol	Methanol, ≥99.8%	M/4056/17	Fisher	UK
NaOH	Sodium Hydroxide	S/4920/53	Fisher	UK
$Na_4P_2O_7$	Sodium Pyrophosphate Tetrabasic	P8010-500G	Sigma-Aldrich	USA
NaCl	Sodium Chloride	27810.295	VWR Chemicals	UK
SDS	Sodium Dodecyl Sulfate	L4509-500G	Sigma-Aldrich	USA
TEMED	N,N,N',N'-Tetramethylethylenediamine	T9281-50ML	Sigma-Aldrich	USA
Tris	Tris Base	BP152-1	Fisher	UK
Tri-sodium citrate	Tri-Sodium Citrate Dihydrate	S/3320/53	Fisher	UK
Triton	Triton® X-100	T9284-100ML	Sigma-Aldrich	USA
Tween	Tween® 20	P1379-500ML	Sigma-Aldrich	USA
Xylene	Xylene	X/0250/PB17	Fisher	UK

Table 2.3. Compounds for Stimulation Experiments

Drug Name	Manufacturer	Cat. / Ref.	Stock Concentration	Solubility
Wnt Stimulants				
BML-284 (Wnt Agonist 1)	Selleckchem	S8178	10mM	DMSO
AZD2858		S7253		
Laduviglusib (CHIR-99021)		S1263		
IM-12		S7566		
Cytokines				
LIGHT/TNSF14	R&D Systems	664-LI-025/CF	100µg/mL	PBS
Lymphotoxin		884-LY-025/CF	20µg/mL	
IL-6	Thermo Fisher Scientific	200-06-20UG	20µg/mL	
TNF-α	RayBiotech	230-00243	20µg/mL	
Control				
5-Fluorouracil	Selleckchem	S1209	10mM	DMSO

PBS: Phosphate-buffered Saline.

2.2 Methods

2.2.1 Testing Specificity of Research-grade Antibodies

Antibodies for each target for use in immunohistochemistry (IHC) experiments were purchased after a literature review was undertaken to determine the most appropriate antibody to use. When a diagnostic-grade antibody was available, that product was purchased. As diagnostic-grade antibodies are robust and well validated, specificity and sensitivity validation was performed for research-grade antibodies only. This was undertaken using western blotting (WB) and IHC on cell pellets.

2.2.1.1 Selection of Low and High Expressing Cell Lines *in silico*

Antibody specificity was performed to validate all research grade antibodies. Cell lines that express low levels of the target protein, and others that express high levels of the target protein were identified *in silico* using the Dependency Map (DepMap; (Tsherniak *et al.*, 2017)) package on RStudio. Target expression by cell linage was first obtained to reveal the best tissue types to use. This was then followed by the expression of the target in cell lines derived from the selected tissue type. Breast and colorectal cell lines were used due to their availability within our research group.

2.2.1.2 Cell Culture

Unless specified otherwise, all reagents used in cell culture were manufactured by Gibco (UK), and all plasticware was manufactured by Corning (UK).

2.2.1.2.1 Bringing Up Cells from Frozen

Two cell lines, MCF-7 and DLD-1 were identified through DepMap. Aliquots were thawed on ice, and transferred to a 15mL tube, and 10mL DMEM(1X) + GlutaMAX-I (Ref. 31966-047) for MCF-7 and RPMI Medium 1640 (1X) (Ref. 21875-034) for DLD-1 with 10% foetal bovine serum (FBS; Ref. 10270-106) added. The tubes were centrifuged at 1200rpm for 3.5 minutes at room temperature (RT). The supernatant was discarded, and the cells were resuspended in 5mL media before transferring to a T25 culture flask. Flasks were incubated at 37°C in 5% CO_2 until they reached 80% confluency. Media was aspirated and replaced as required.

2.2.1.2.2 Passaging

Confluent T25 flasks were washed in 5mL Phosphate-buffered Saline (PBS, Ref. 14190-094) and passaged by incubating in 2mL 0.05% Trypsin-EDTA (Ref. 25300-054) for approximately 3-5 minutes at 37℃, or until the cells were no longer adhering to the flask. Cells were split into T75 flasks at a 1:3 cell to media ratio. They were subsequently incubated at 37℃ in 5% CO_2 until they reached 80% confluency. Media was aspirated and replaced as required.

2.2.1.3 Cell Lysates Production

80% confluent T75 flasks were passaged by treating with 0.05% trypsin for 5 minutes and neutralised with cell culture medium containing 10% FBS. A hemocytometer was used to count the cells, which were seeded into 6-well plates at a density of $2.5x10^5$ cells/well and incubated for 48 hours until ~80% confluent. Lysis of cells was carried out chemically by the addition of Pierce RIPA Buffer (Ref. 89901, ThermoScientific) with PhosSTOP EASYpack phosphatase inhibitor (Ref. 04906845001, Roche) and cOmplete Tablets EDTA-free EASYpack protease inhibitor (Ref.04693132001, Roche), and mechanically by shearing using a 21G needle. Cells were then incubated on ice for 30 minutes, then centrifuged at 4℃ for 15 minutes at 15,000rpm. The supernatant was removed into a new microcentrifuge tube and lysates were either used immediately or stored at -20℃. Lysates were prepared for use by adding Laemmli's Buffer (4x Sample Buffer: 252mM Tris-HCl, 8mM $Na_4P_2O_7$, 15.5mM EDTA, 40% glycerol, 277mM, 0.417mM bromophenol blue; with 518mM DTT and dH_2O).

2.2.1.4 SDS-PAGE and Western Blotting

Unless stated otherwise, all equipment used was manufactured by BioRad (USA).

2.2.1.4.1 Hand Casting Gels

A 1mm spacer plate was assembled with a front plate into a casting frame and inserted into a casting stand. The assembly was filled with dH_2O to check for leakage, before 10% resolving gel solution (10% acrylamide, 0.05% APS, 0.05% TEMED, 375mM Tris with 0.1% SDS, and H_2O), was added. Any bubbles were removed by the addition of 0.1% SDS. Once the resolving gel was set (approximately 20 minutes at RT), the SDS was removed, and the stacking gel

(3% acrylamide, 0.1% APS, 0.13% TEMED, 125mM Tris with 0.1% SDS, and H_2O) was added and a 10-well comb was inserted. The stacking gel was allowed to set (approximately 20 minutes at RT), and the gels were either used directly or stored in running buffer at 4℃ for no longer than 5 days.

2.2.1.4.2 Running Buffer Preparation

Tris/Glycine/SDS running buffer (25mM Tris,192mM glycine, 3mM SDS) was prepared in 2L batches and stirred until clear.

2.2.1.4.3 Gel Electrophoresis

The combs were removed from the gels and either 2 gels or 1 gel and 1 buffer dam were secured in a leak-free gel holder and placed in a Mini Trans-Blot Electrophoretic tank (Cat. 1703930). The tank was filled halfway with running buffer, ensuring the gel holder was topped-up and overflowing to facilitate electric current conduction. The wells were washed prior to sample loading by pipetting running buffer in them. 3µL Precision Plus Protein Dual Xtra Standards (Cat. 1610377) was added, followed by 15µL of protein lysates. The tank lid was secured, and the gels were allowed to run at 115V for 90 minutes, or until the loading buffer has reached the bottom of the gel.

2.2.1.4.4 Transfer Buffer Preparation

Tris/Glycine transfer buffer (25mM Tris-base, 192mM glycine, 20% methanol) was prepared in 2L batches and stirred until clear.

2.2.1.4.5 Wet Protein Transfer

Prior to completion of the gel run, components of the protein transfer sandwich shown in Figure 2.6 were soaked in transfer buffer. The gel was removed from the tank and placed on the nitrocellulose membrane as shown in Figure 2.6. The sandwich was inserted into a blot frame inside the tank. An ice cooling unit was inserted in the tank to absorb excess heat. The tank was then filled to the top with transfer buffer, the lid secured, and the proteins were transferred at 300mA for 110 minutes.

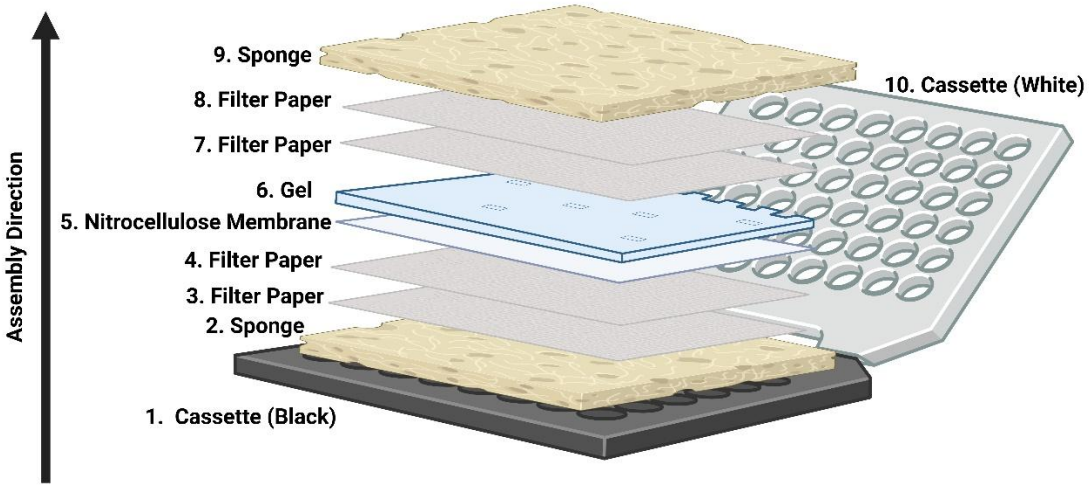


Figure 2.6. Protein Transfer Sandwich Assembly.
2 sponges and 4 Whatman 3mm filter paper (Cat. 3030-917, Cytiva) ensure a tight sandwich for a successful transfer of proteins from the gel to 1 piece of Amersham™ Protan™ 0.45µm nitrocellulose blotting membrane (Ref. 10600002, Cytiva). Bubbles are smoothed out with a roller after the addition of each layer. Figure created with BioRender.

2.2.1.4.6 Blocking of Non-specific Binding

Once the transfer was completed, the nitrocellulose membrane was blocked in 3% w/v bovine serum albumin (BSA; Cat. 5217, Tocris) in NaCl-Tris-Tween buffer (NATT; 150mM NaCl, 20mM Tris, 0.03% Tween with pH adjusted to 7.4 using HCl) for 2 hours at RT on a platform shaking at 16rpm.

2.2.1.4.7 Incubation with Primary Antibodies

Primary antibodies were diluted in 0.3% w/v BSA in NATT at the concentrations shown in Table 2.4. Membranes were incubated in primary antibodies overnight (ON) at 4°C with agitation. The membranes were then removed from the primary antibody solution and washed (3x for 15 minutes) in NATT on a slowly shaking platform. Primary antibodies were stored at -20°C for reuse.

2.2.1.4.8 Incubation with Secondary Antibodies

Species-matched secondary antibodies were diluted in 0.3% w/v BSA in NATT with anti-ladder at the concentrations shown in Table 2.4. Washed membranes were incubated in secondary antibodies for 90 minutes at RT on a shaking platform, before washing (3x for 15 minutes) in NATT with agitation.

Table 2.4. Conditions of Antibodies for Western Blots

Antibody	Concentration	Incubation Conditions
Primary		
SOX9	1:1000	Overnight at 4°C with agitation
E-Cadherin	1:5000	Overnight at 4°C with agitation
ß-Catenin	1:1000	Overnight at 4°C with agitation
c-Myc	1:1000	Overnight at 4°C with agitation
ß-actin	1:20,000	Overnight at 4°C with agitation
ß-tubulin (HRP-Conjugated)	1:5000	90 minutes at RT with agitation
Secondary		
Anti-Rabbit HRP-linked Antibody	1:5000	90 minutes at RT with agitation
Anti-mouse HRP-linked Antibody	1:10,000	90 minutes at RT with agitation
Peroxidase-conjugated Anti-Rabbit	1:5000	90 minutes at RT with agitation
Peroxidase-conjugated Anti-Mouse	1:5000	90 minutes at RT with agitation

2.2.1.4.9 Imaging

Following the final wash, membranes were gently dried of excess NATT, and incubated in ECL (Ref. 32106, ThermoScientific, UK) for 5 minutes, and inserted into Li-Cor Odyssey Fc 2800 imaging system. Images were allowed to develop for 2 minutes and saved as ".tiff".

2.2.1.4.10 Re-probing with Other Antibodies

Membranes were kept at 4°C in NATT for re-probing, allowing blotting of loading controls on the same membrane. Membranes were washed three times for 15 minutes in NATT, then stripped of bound antibodies using Re-Blot Plus Strong Solution (Cat. 2504, EMD Millipore) diluted in dH_2O for 15 minutes with agitation. The membranes were then washed in NATT (3x for 15 minutes) and the new antibody (diluted to the desired concentration (Table 2.4, in 0.3% w/v BSA in NATT) was added to the membrane. Depending on the loading control used, the

incubation was either 90 minutes at RT or overnight at 4℃, with agitation. The remainder of the steps are as previously described.

2.2.1.4.11 Protein Quantification

Images were analysed using ImageJ (Ver. 1.54d, NIH), and band densitometry was quantified using a signal-to-noise ratio detection method. Finally, Prism (Ver. 10.0.3, GraphPad) was used for statistical analysis and graph generation.

2.2.1.5 Cell Pellets

Cell pellets for antibody validation experiments were made by Edinburgh Pathology Histology Research Service using untreated MCF-7 and DLD-1 cells and supplied by Dr. Zeanap Mabrouk. In brief, flasks of untreated cultured cells at 80% confluency were trypsinised and moved to a 15mL falcon tube and topped up with media. A pellet was achieved by centrifugation at 1,200rpm for 3 minutes at RT and removing the supernatant. The pellet was washed in PBS and spun at 1,200rpm for 3 minutes. The washed pellet was then mixed with 4% paraformaldehyde (PFA) and moved to an Eppendorf tube; were it was fixed for 20 minutes at RT. This was followed by washing twice in PBS and spinning at 1,200rpm for 3 minutes each time. The dried pellet was then mixed with 1% agarose and left to sit in a cassette ON at 4℃. Then, it was dehydrated through a series of increasing alcohols (15 minutes in 50% ethanol, 15 minutes in 75% ethanol, 15 minutes in 100% ethanol, twice) and cleared in Histo-Clear II (twice for 10 minutes each). The pellet was then embedded in paraffin wax and allowed to cool for 1hr. Sections were cut and baked at 50℃ ON. Staining was carried out for each marker as detailed in 2.2.2.

2.2.2 Immunohistochemistry

IHC was used to stain the cell pellets for antibody validation, as well as INCISE TMA Cohort for SOX9, E-Cadherin, ß-Catenin, CD3, CD4, and CD8. SOX9 and E-Cadherin staining was performed by SSFA. ß-Catenin staining was performed by Dr. Mark Johnstone. CD3 and CD4 staining was performed by Marjolien de Koning, and CD8 by Lewis Irvine, under supervision of SSFA.

Table 2.5. Primary Antibody Optimised Conditions for IHC

Antibody	Concentration	HIER	Final Staining Method
SOX9	1:4000	pH9	Automated
E-Cadherin	1:10,000	pH6	Manual and Automated
ß-Catenin	1:600	pH6	Manual
CD3*	1:100	pH9	Semi-automated
CD4*	1:100	pH9	Automated
CD8*	1:400	pH6	Automated
IgG1κ (Isotype control)	Matched to primary conditions		

HEIR: Heat-induced Epitope Retrieval.

*Denotes diagnostic-grade antibodies.

2.2.2.1 Manual Staining

Manual IHC follows a 2-day protocol with an overnight incubation in primary antibodies. Each run included a negative no-antibody control. The first run included a matched isotype control. It was used to optimise staining conditions for each primary antibody.

2.2.2.1.1 Slide Requisition and Preparation

Patient tissue was requested from the NHS at Greater Glasgow and Clyde Tissue Biorepository and cut by the Glasgow Tissue Research Facility (GTRF). Full sections were cut at a thickness of 4µm, and TMAs were sectioned at a thickness of 2.5µm. After sectioning tissue, slides were baked at 37℃ and stored at 4℃ until used. TMAs slides were baked for 1 hour at 60℃ before staining to reduce core loss during the staining process.

2.2.2.1.2 Slide Dewaxing and Tissue Rehydration

Slides were twice submerged in HistoClear-II (Order No. HS-202, National Diagnostics) in 3-minute intervals for dewaxing. This was followed by rehydration in a decreasing alcohol gradient (twice for 2 minutes in 100% ethanol, once for 2 minutes in 90% ethanol, and once for 2 minutes in 70% ethanol). Finally, they were submerged in running tap water for 10 minutes.

2.2.2.1.3 Antigen Retrieval

Heat-induced antigen retrieval (HEIR) was employed to recover antigens blocked during tissue formalin-fixation. This was done using a citrate buffer at pH6 (9.3mM tri-sodium citrate and 1.77mM citric acid) or Tris/EDTA buffer at pH9 (4.5mM Tris and 0.73mM EDTA), depending on the antibody's optimised conditions (Table 2.5). Buffer was heated in an unlidded microwavable pressure cooker for 14 minutes. Then, slides were placed in the buffer. The pressure cooker was assembled, and pressure was allowed to build for approximately 3 minutes, after which the slides were heated under pressure for 5 minutes. Slides were cooled in the buffer for 30 minutes before being washed in running water for at least 10 minutes. For cell pellets, this procedure was unchanged except for the reduction in time from 5 minutes to 2.5 minutes.

2.2.2.1.4 Blocking of Endogenous Peroxidase Activity

To reduce non-specific background staining, endogenous peroxidases were blocked in fresh 3% H_2O_2 for 15 minutes at RT. Slides were then washed in running water for 10 minutes.

2.2.2.1.5 Blocking of Non-Specific Off-Target Binding

The tissue was circled using a hydrophobic pen (Cat. H-4000, Vector Laboratories) to ensure aqueous reagents applied to the slide would cover the tissue and not run off. To reduce off-target binding, tissue was blocked by incubation in 10% casein (Ref. SP-5020, Vector Laboratories) for 30 minutes at RT in a dark humidified chamber.

2.2.2.1.6 Incubation in Primary Antibody

Primary antibodies were prepared at the concentrations shown in Table 2.5 using antibody diluent (Ref. S0809, Dako). Blocking solution was blotted off and approximately 200µL of primary antibody solution was added to cover the tissue. Slides were incubated ON at 4°C in a dark humidified chamber.

2.2.2.1.7 Incubation in Secondary Antibody

Primary antibody was blotted off and slides were washed (2x for 5 minutes) in Tris-buffered Saline (TBS; 309mM Tris, 936mM NaCl at pH7.5), then incubated

for 30 minutes in a species-appropriate HRP-linked secondary antibody (Ref. MP-7500, Vector Laboratories) at RT in a dark humidified chamber.

2.2.2.1.8 Visualization with DAB Chromogen Substrate

Slides were washed in TBS (2x for 5 minutes), then incubated in 3,3'-Diaminobenzidine (DAB; Ref. SK-4105, Vector Laboratories) for 5 minutes at RT in a dark humidified chamber. Slides were washed to remove excess DAB in running water for 10 minutes.

2.2.2.1.9 Counterstaining

Slides were counterstained in Haematoxylin Gill III (Ref. 3801540E, Surgipath) for 1 minute, washed in running water for 2 minutes, and dipped in 0.5% acid alcohol (Ref. 3803650E, Surgipath) three times to remove excess haematoxylin. Slides were washed in running water for 2 minutes and submerged in Scott's Tap Water Substitute (Ref. 3802901E, Surgipath) for 2 minutes to develop the blue counterstain. Slides were washed in running water for 2 minutes.

2.2.2.1.10 Dehydration and Mounting

Tissue was dehydrated through an increasing alcohol gradient (once in 70% ethanol for 2 minutes, once in 90% ethanol for 2 minutes, and twice in 100% ethanol for 2 minutes each) and cleared in HistoClear-II twice, for 2 minutes each. Slides were mounted with coverslips using Pertex mounting medium (Ref. 00801-EX, HistoLab), which was allowed to dry for a minimum of 1 hour.

2.2.2.1.11 IHC Quality Check

All mounted slides were viewed under a brightfield microscope to assess staining quality and ensure the absence of staining artefacts and bubbles, which could negatively impact the scanning process. Any slides that failed this step had their coverslips gently removed (by submerging in xylene overnight at RT), placed in HistoClear-II and remounted. All mounted slides were cleaned of excess mounting medium using 100% ethanol and gently scraping off the excess using a 23G needle.

2.2.2.2 Scanning and Visualisation

Slides were sent to GTRF for scanning using the S60 NanoZoomer (Hamamatsu, Japan) scanner at x20 magnification and were uploaded to NZConnect (Ver. 1.1.0, Hamamatsu) for visualisation.

2.2.2.3 Semi-automated Staining

Unless stated otherwise, all materials used were manufactured by Epredia (UK). TMA slides were baked as outlined in 2.2.2.1.1. Slides were placed in the pre-treatment (PT) Module, which was filled with Dewax and HIER Buffer (Cat. TA-999-DHBH) of the appropriate pH. All dewaxing, rehydration, and antigen retrieval steps were performed automatically in the PT module. Endogenous peroxidases were blocked in Ultravision Hydrogen Peroxide Block (Cat. TA-125-H202Q) for 10 minutes at RT, followed by blocking of non-specific binding in Ultravision Protein Block (Cat. TA-125-PBQ). The slides were incubated in CD3 antibody (Table 2.5) diluted in Antibody Diluent OP Quanto (Cat. TA-125-ADQ) for 30 minutes at RT. The signal was amplified by incubating the slides for 10 minutes at RT in Primary Antibody Amplifier (Cat. TL-125-QPB), followed by Horse Radish Peroxidase Polymer Quanto (Cat. TL-125-QPH) for 10 minutes at RT. DAB Quanto (Cat. TA-125-QHDX) was added for 5 minutes to visualise the tissue. Counterstaining, dehydrating, mounting, and scanning steps were carry out as described in 2.2.2.1.92.2.2.1.10.

2.2.2.4 Automated Staining

2.2.2.4.1 Study Set-up

Slides were entered into a newly added study in the Leica Bond software and marked as “Study Tissue”, “Negative Control”, or “Positive Control” accordingly. Coded stickers were printed and placed on the corresponding slides.

2.2.2.4.2 Preparation of Staining Kit and Primary Antibody

A BOND Polymer Refine Detection Kit (Ref. DS9800, Biosystems) was used, the components of which are listed in Table 2.6. Primary antibodies were prepared as described in Table 2.5.

Table 2.6. Components of BOND Polymer Refine Detection

Reagent	Manufacturer	Reference
Peroxide Block	Leica Biosystems	26157670
Post Primary		26160412
Polymer		26161245
DAB Part 1		26162141
DAB Part B		26159342
DAB Part B		26159341
Haematoxylin		26158336

2.2.2.4.3 Automated Staining

An automated staining protocol was created on the Bond RX Fully Automated Research Stainer (Cat. 21.2821, Leica), that was comparable to that described in 2.2.2.1.

2.2.2.4.4 Dehydration and Mounting

Dehydration and mounting were done manually. Slides were placed in water, and then in 70% ethanol for 2 minutes, followed by 2 times in 100% ethanol for 2 minutes each, and finally in xylene twice for 2 minutes each. Mounting was as described in 2.2.2.1.10.

2.2.2.4.5 Visualisation and Scanning

Slides were checked for debris and artefacts as outlined in 2.2.2.1.11. They were subsequently scanned and visualised as described in 2.2.2.2.

2.2.3 Staining Assessment

2.2.3.1 Manual Histo-Scoring for Epithelial Cells

Cell staining was scored manually by compartment in epithelial cells (nuclear staining, cytoplasmic staining, and membranous staining, as available) and weighted histoscores were calculated using the equation below (Kirkegaard *et al.*, 2006):

$$(0 \times negative\ staining\ percentage) + (1\ \times\ weak\ staining\ percentage) + (2 \times moderate\ staining\ percentage) + (3 \times strong\ staining\ percentage)$$

This produced a range of scores between 0-300, where 0 is no staining and 300 is strong staining.

2.2.3.2 Digital Pathology for Epithelial Cells

Given the subjectivity of manual assessment of IHC, digital pathology platforms were chosen to generate weighted histoscores. QuPath (Bankhead *et al.*, 2017) is an open-source platform that was used to assess nuclear and cytoplasmic staining in the absence of membrane staining.

2.2.3.2.1 Image Set-up

Images in ".ndpi" format were imported into QuPath, and image type was set to "Brightfield (DAB)" to indicate the type of microscopy and chromogen used for target visualisation.

2.2.3.2.2 TMA Slide Dearraying

Using the "TMA Dearrayer" function, a grid was formed on TMA slides with coordinates for the location of each core. This grid was designed to detect cores 0.6mm in diameter, and set to match the slide's map, with 20 columns and 8 rows. It was manually modified to accurately assign missing cores, to exclude poor quality cores (Figure 2.7), and to include out of position cores and cores of different sizes.

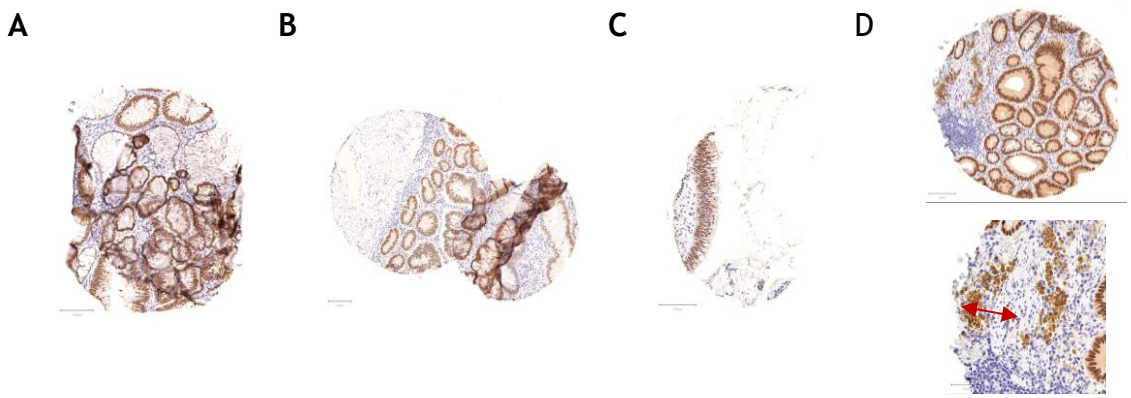


Figure 2.7. Selected Examples of Cores Excluded Due to Poor Quality.
[A] Two cores overlapped and folded together. **[B]** One core (right) heavily folded on itself and out of position. **[C]** Poor quality core with very little scorable epithelium. **[D - Top]** Core with hemosiderin (old red blood cells) artefacts and **[D - Bottom]** Hemosiderin artefact expanded (red arrows). Hemosiderin cores are excluded as they will skew the intensity scoring. Scale bars=100μm.

2.2.3.2.3 Stain Vector Estimation

A core exhibiting as many staining intensities as possible and including background, was selected as a model to estimate stain vectors. This involves a pre-processing step to increase the staining intensity using an RBG visualiser.

2.2.3.2.4 Cell Detection

A watershed cell detection step was used to build objects to represent cells. This allowed for cell expansion to be calculated from nuclear parameters, including but not exclusive to minimum and maximum areas to be considered as nuclei, and a sigma value for gaussian filters to reduce noise.

2.2.3.2.5 Tissue Classification

Tissue classes were annotated (mainly the epithelium and lamina propria) using a set brush tool on most cores to account for tissue differences between different patients.

2.2.3.2.6 Staining Intensity Selection

To generate a histoscore (H-score), 3 thresholds were manually selected from the DAB channel to represent pixels with weak, moderate, and strong DAB values (Figure 2.8).

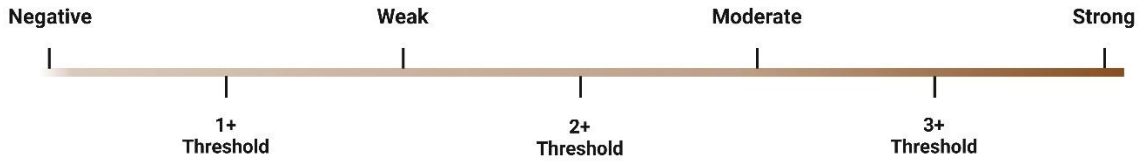


Figure 2.8. Three Intensity Thresholds on DAB Staining Spectrum.
Any values below the 1+ threshold will be very weak to negative. Any values between the 1+ and 2+ thresholds will be weak. Any values between the 2+ and 3+ thresholds will be moderate. Any values above the 3+ threshold will be strong. Figure created with BioRender.

2.2.3.2.7 Classifier Construction

A random trees classifier was built from the cell detection, tissue classification, and selected staining intensities to allow for the generation of a histoscore that represents each core. The values were changed for each cellular compartment to generate different values.

2.2.3.3 Histo-Score Reliability Validation

To ensure consistency and reliability of staining assessment by digital pathology, manual scoring was performed by a single observer (Sara Al-Badran) blinded to clinical outcome for 10% of the cores. This was then compared to QuPath-generated scores and an intraclass correlation coefficient (ICC) was calculated. ICCs above 0.7 were considered good.

2.2.3.4 Point Scoring for Immune Cells

Immune cell scoring was done by Marjolein de Koning, Arwen Shirmohammad Damavandi, and Lewis Irvine as part of honours and Masters projects, supervised by Sara Al-Badran. In brief, positive immune cells were counted in both the epithelium and lamina propria. A ratio percentage of positive cells was determined using the equation below to account for differences in amount of tissue present in each core:

$$\frac{Positive\ Cells\ in\ Epithelium + Positive\ Cells\ in\ Lamina\ Propria}{Number\ of\ Objects\ Detected - Number\ of\ Objects\ Considered\ Artefacts} \times 100$$

This was normalised to the differences in areas designated as epithelium and lamina propria in each core, and a final score was generated using the below equation:

$$\frac{Positive\ Cells\ Ratio\ Percentage}{Number\ of\ Cells\ in\ Lamina\ Propria / Number\ of\ Cells\ in\ Epithelium}$$

2.2.4 RNA Preparation

2.2.4.1 Human RNA

qPCR Reference Total RNA (URR; Cat. 636690) and human Brain Total RNA (BRR; Cat. 636530) were all purchased from Takara (Japan). URR was sourced from a collection of human adult tissue from both male and female donors. BRR was sourced from human brains of 8 individuals of Asian and Caucasian descent who were between the ages of 21-82. Both URR and BRR were isolated using the modified guanidinium thiocyanate method. These were diluted in water to a concentration of 100ng/µL, then aliquoted and stored at -80°C until required.

2.2.4.2 *F. nucleatum* RNA Extraction

F. nucleatum (ATCC 10953) cultures were kindly provided by Prof. Gordon Ramage (University of Glasgow). RNA extraction was done using RNeasy® Mini Kit (Cat. 74104, Qiagen) with no deviation from the manufacturer's instruction. In brief, *F. nucleatum* cells were harvested into microcentrifuge tubes in an anaerobic chamber, and centrifuged at 1300rpm for 2 minutes at RT. The pellet

was resuspended in 70% ethanol and 700µL was transferred to a spin column with a collection tube. This was centrifuged at a minimum of 8,000g for 15 seconds and the supernatant was discarded. This spin step was repeated three times with the addition of kit-supplied RW1 Buffer, and RPE Buffer the second and third time. The spin column was placed in a new collection tube and centrifuged at 20,000rpm for 1 minute to dry the membrane. The spin column was moved to a new collection tube and 30µL of RNase-free water was added to the spin column, which was centrifuged at 8,000g for 1 minute. This yielded 60µL of RNA at a concentration of 10ng/µL.

2.2.4.3 Titrations of Bacterial RNA

Human URR was prepared to a final concentration of 10ng/µL by adding 12µL of URR to 108µL of UltraPure water. This was used to prepare 10-fold serial dilutions of bacterial RNA to human RNA, as shown in Table 2.7.

Table 2.7. 10-Fold Serial Dilutions of Bacterial RNA in Human RNA.

Dilution	*F. nucleatum RNA*	Human URR
Fn1:Hs1	6µL purified RNA at 10ng/µL	6µL at 10ng/µL
Fn1:Hs10	2µL purified RNA at 10ng/µL	18µL at 10ng/µL
Fn1:Hs100	2 µL from Fn1:Hs10 at 10ng/µL	18µL at 10ng/µL
Fn1:Hs1,000	2 µL from Fn1:Hs100 at 10ng/µL	18µL at 10ng/µL
Fn1:Hs10,000	2 µL from Fn1:Hs1,000 at 10ng/µL	18µL at 10ng/µL
Fn1:Hs100,000	2 µL from Fn1:Hs10,000 at 10ng/µL	18µL at 10ng/µL
Fn1:Hs1,000,000	2 µL from Fn1:Hs100,000 at 10ng/µL	18µL at 10ng/µL

URR: Universal Reference RNA. Fn: *F. nucleatum*. Hs: *Homo sapiens*

2.2.5 TempO-Seq™

2.2.5.1 Bacterial Probe Design

F. nucleatum subsp. *Polymorphum* (ATCC10953) 16S ribosomal RNA (NCBI Ref. NR_117288.1) was used to compare against prospective probe sequences. Sequences found in the literature (e.g (Yang *et al.*, 2016; Nishimura *et al.*, 2021)) were used as starting point for the design of the *F. nucleatum* probes. To identify the potential target regions and specificity to *F. nucleatum*, the probe sequences were run on the SILVA ribosomal RNA database project platform (Quast *et al.*,

2013; Yilmaz *et al.*, 2014). Probe sequences were entered in the 5'-3' direction, and 0 mismatches were allowed before the TestProbe3.0 function was run. A table of matched regions was exported as a ".csv" file and was checked for compatibility with organisms and number of mismatches. Reverse complementary sequences to the probes were obtained from a reverse complement online tool (Stothard, 2000). These were checked against the 16S region for *F. nucleatum* to ensure target binding and to check the sequences flanking the target. Sequences targeted by probes had to be at least 50 base pairs long.

2.2.5.2 Bacterial Probe *in silico* Quality Check

The eight potential Fusobacterium probe sequences were checked for the following quality factors:

- Specificity to organism: as presented on the SILVA ribosomal RNA database
- GC content: must be between 40-60% to ensure probe stability. This was tested using Eurofins Genomics online tool (Eurofins, no date).
- Melting temperature (Tm): must be between 60-70°C. This was established by entering the sequence into the Eurofins Genomics online tool (Eurofins, no date).
- Homopolymers: preferably none, as they lead to structural bending. However, if present they must be less than 6-mers in length as more would lead to less reliable sequencing. This was tested by manual examination of the probe sequences.
- Self-annealing capacity and hairpin structure formation: checked to ensure the structural integrity of the probes, using Eurofins Genomics online tool (Eurofins, no date), and the MaxExpect function of the RNAStructure tool (Reuter and Mathews, 2010).

A summary of the of the probe sequences and quality controls parameters is shown in Table 2.8.

Table 2.8. Parameters of *F. nucleatum* Probes Quality Check

	Probe ID	Probe Name	Direction	Sequences (5'-3')	Organism Specificity (SILVA)	GC Content (%)	Tm (°C)	Homopolymers	Self-annealing Capacity (Bases)	Hairpin Probability (%)
Probe 1	Fn16S_113694	Green	DDO	TCTAGTTACACAGTTTCCAACGCAA	*Fusobacterium*	40	59.7	No	4	<50
			UDO	TACAGAGTTGAGCCCTGCATTTTCA	*Fusobacterium*	44	61.3	Yes TTTT	4	70 - 80
Probe 2	Fn16S_113695	Indigo	DDO	ACGAATTTCACCTCTACACTTGTAG	*Fusobacterium*	40	59.7	No	10	80 - 90
			UDO	TTCCGCTTACCTCTCCAGTACTCTA	*Fusobacterium* and uncultured rumen bacterium	48	63	No	8	<50
Probe 3	Fn16S_113696	Magenta	DDO	TACCTCTCCAGTACTCTAGTTACAC	*Fusobacterium*	44	61.3	No	12	<50
			UDO	AGTTTCCAACGCAATACAGAGTTGA	*Fusobacterium*	40	59.7	No	12	80 - 90
Probe 4	Fn16S_113697	Cyan	DDO	GTCCCCGTCAATTCCTTTGAGTTTC	Rice and marine metagenome	48	63	Yes CCCC	8	60 - 70
			UDO	ATACTTGCGTGCGTACTCCCCAGGC	Uncultured *Clostridiales bacterium*	60	67.9	Yes CCCC	6	50 - 60
Probe 5	Fn16S_113698	Gold	DDO	ACCTCTCCAACTAGCTAATGGGACG	*Fusobacterium* and uncultured *Treponema sp.*	54	64.6	No	14	<50
			UDO	CAAAGCTCTCTCACAGCGCATGTAG	*Fusobacterium* and uncultured organism	52	64.6	No	8	50 - 60
Probe 6	NA	NA	DDO	TCCCCGTCAATTCCTTTGAGTTTCA	Rice and marine metagenome	44	61.3	Yes CCCC	8	60 - 70
			UDO	TACTTGCGTGCGTACTCCCCAGGCG	Uncultured bacteria	64	69.5	Yes CCCC	6	90 - 95
Probe 7	NA	NA	DDO	GCCGTTACCTCTCCAACTAGCTAAT	*Fusobacterium* and others	48	63°C	No	8	<50
			UDO	GGGACGCAAAGCTCTCTCACAGCGC	*Fusobacterium*	64	69.5	No	4	60 - 70
Probe 8	NA	NA	DDO	TAGTTACACAGTTTCCAACGCAATA	*Fusobacterium*	36	58.1	No	4	<50
			UDO	CAGAGTTGAGCCCTGCATTTTCACA	*Fusobacterium*	48	63	Yes TTTT	6	<50

Probe directions and sequences are included. Probes specific to *Fusobacterium* are those with a GC content between 40-60%, melting temperatures between 60-70°C, a low self-annealing capacity, homopolymers less than 6-mers in length (preferably none), and a hairpin probability <90%. Specific probes that were sent for manufacture were assigned Probe IDs. Failed probes that were not manufactured were not assigned IDs (NA). Probes were given color-coded names for simplicity. Probe: pair of detector oligos, DO: Detector Oligo, DDO: Down-stream Detector Oligo, UDO: Up-stream Detector Oligo, GC Content: Guanine-Cytosine Content, and Tm: Melting Temperature.

2.2.5.3 Probe Manufacture

Human and bacterial probes that passed *in silico* quality check were ordered for manufacture by IDT through their LabReady service at a concentration of 100μM in IDTE buffer at pH8 (Table 2.9).

Table 2.9. Identification Guide of *F. nucleatum* Probes

Probe ID	Probe Name
Fn16S_113694	Green
Fn16S_113695	Indigo
Fn16S_113696	Magenta
Fn16S_113697	Cyan
Fn16S_113698	Gold

2.2.5.4 Human Probe Design and *in silico* Quality Check

Human probes were selected as internal assay controls. These included SOX9, E-Cadherin, ß-Catenin, COX2, and p53. Probe sequences were already available on BioClavis databases and were tested in silico using the UCSC BLAT tool (Kent, 2002) for target specificity in human genomes.

2.2.5.5 Probe Pooling

Detector oligo (DO) pools of the human probes and bacterial probes were prepared as outlined in Table 2.10 and Table 2.11. Each down-stream DO (DDO) and upstream DO (UDO) of a probe pair was added at the same volume and concentration.

Table 2.10. Components, Volumes, and Concentrations of Primary Detector Oligo Pools

DO Pool	Components	Components Volume & Concentration		Notes
		Each Oligo	Oligo Pool (Total)	
DO1	DDO_Fn16S_113694	5µL at 100µM	50µL of oligos at 10µM	DO Pool 1 includes all the Fuso probes only.
	UDO_ Fn16S_113694	5µL at 100µM		
	DDO_ Fn16S_113695	5µL at 100µM		
	UDO_ Fn16S_113695	5µL at 100µM		
	DDO_ Fn16S_113696	5µL at 100µM		
	UDO_ Fn16S_113696	5µL at 100µM		
	DDO_ Fn16S_113697	5µL at 100µM		
	UDO_ Fn16S_113697	5µL at 100µM		
	DDO_ Fn16S_113698	5µL at 100µM		
	UDO_ Fn16S_113698	5µL at 100µM		
DO2	DDO_SOX9_22075	5µL at 100µM	100µL of oligos at 5µM	DO Pool 2 includes all the human probes and all the Fuso probes.
	UDO_SOX9_22075	5µL at 100µM		
	DDO_COX2_1514	5µL at 100µM		
	UDO_COX2_1514	5µL at 100µM		
	DDO_TP53_7287	5µL at 100µM		
	UDO_TP53_7287	5µL at 100µM		
	DDO_CDH1_1186	5µL at 100µM		
	UDO_CDH1_1186	5µL at 100µM		
	DDO_CTNNB1_1633	5µL at 100µM		
	UDO_CTNNB1_1633	5µL at 100µM		
	DDO_Fn16S_113694	5µL at 100µM		
	UDO_ Fn16S_113694	5µL at 100µM		
	DDO_ Fn16S_113695	5µL at 100µM		
	UDO_ Fn16S_113695	5µL at 100µM		
	DDO_ Fn16S_113696	5µL at 100µM		
	UDO_ Fn16S_113696	5µL at 100µM		
	DDO_ Fn16S_113697	5µL at 100µM		
	UDO_ Fn16S_113697	5µL at 100µM		
	DDO_ Fn16S_113698	5µL at 100µM		
	UDO_ Fn16S_113698	5µL at 100µM		
DO3	DDO_SOX9_22075	5µL at 100µM	80µL of oligos at 6.25µM	DO Pool 3 is comprised of the same elements of DO Pool 2 at a different final concentration.
	UDO_SOX9_22075	5µL at 100µM		
	DDO_COX2_1514	5µL at 100µM		
	UDO_COX2_1514	5µL at 100µM		
	DDO_TP53_7287	5µL at 100µM		
	UDO_TP53_7287	5µL at 100µM		
	DDO_CDH1_1186	5µL at 100µM		
	UDO_CDH1_1186	5µL at 100µM		
	DDO_CTNNB1_1633	5µL at 100µM		
	UDO_CTNNB1_1633	5µL at 100µM		
	DDO_ Fn16S_113696	5µL at 100µM		
	UDO_ Fn16S_113696	5µL at 100µM		
	DDO_ Fn16S_113697	5µL at 100µM		
	UDO_ Fn16S_113697	5µL at 100µM		
	DDO_ Fn16S_113698	5µL at 100µM		
	UDO_ Fn16S_113698	5µL at 100µM		
DO4	DDO_SOX9_22075	5µL at 100µM	70µL of oligos at 7.14µM	DO Pool 4 excludes COX2 but includes the remaining human probes and bacterial probes. This pool was used in different titrations with DO5 to make up DOs 6-10.
	UDO_SOX9_22075	5µL at 100µM		
	DDO_TP53_7287	5µL at 100µM		
	UDO_TP53_7287	5µL at 100µM		
	DDO_CDH1_1186	5µL at 100µM		
	UDO_CDH1_1186	5µL at 100µM		
	DDO_CTNNB1_1633	5µL at 100µM		
	UDO_CTNNB1_1633	5µL at 100µM		
	DDO_ Fn16S_113696	5µL at 100µM		
	UDO_ Fn16S_113696	5µL at 100µM		
	DDO_ Fn16S_113697	5µL at 100µM		
	UDO_ Fn16S_113697	5µL at 100µM		
	DDO_ Fn16S_113698	5µL at 100µM		
	UDO_ Fn16S_113698	5µL at 100µM		
DO5	DDO_COX2_1514	5µL at 100µM	10µL at 50µM	DO Pool 5 includes COX2 only. It was used with DO4 in different titrations to make up DOs 6-10.
	UDO_COX2_1514	5µL at 100µM		

DO: Detector Oligo, UDO: Upstream detector oligo, DDO: Downstream detector oligo

Table 2.11. Composition of DO Pools as Used Experimentally

DO Pool	Stock Concentration	Final Concentration	Composition	
			DOs	TE Buffer
DO1	10μM	30nM	3μL of 10μM DO1	997μL
DO1	30nM	0.3nM	5μL of 30nM DO1	495μL
DO2	5μM	30nM	3μL of 5μM DO2	497μL
DO2	30nM	0.3nM	5μL of 30nM DO2	495μL
DO3	6.25μM	30nM	3μL of 6.25μM	622μL
DO3	30nM	0.3nM	5μL of 30nM	495μL
DO4	7.14μM	300nM	5μL of 7.14μM DO4	114.05μL
DO4	300nM	30nM	20μL of 30nM DO4	180μL
DO5	50μM	300nM	5μL of 50μM DO5	828.33μL
DO5	300nM	30nM	20μL of 300nM DO5	180μL
DO5	30nM	3nM	20μL of 30nM DO5	180μL
DO6	30nM DO4 30nM DO5	3nM DO4 3nM DO5	10μL of DO4 30nM 10μL of DO5 at 30nM	80μL
DO7	30nM DO4 3nM DO5	3nM DO4 0.3nM DO5	10μL DO4 at 30nM 10μL DO5 at 3nM	80μL
DO8	30nM DO4 3nM DO5	3nM DO4 0.09nM DO5	10μL DO4 at 30nM 3μL DO5 at 3nM	87μL
DO9	300nM DO4 3nM DO5	3nM DO4 0.03nM DO5	5μL of 300nM DO4 5μL of 3nM DO5	490μL
DO10	300nM DO4 3nM DO5	3nM DO4 0.009nM DO5	5μL 300nM DO4 1.5μL 3nM DO5	493.5μL

2.2.5.6 Bulk FFPE Tissue Lysate Preparation

Healthy colon tissue sections (Product code: HP-311) were obtained from Amsbio (UK). INCISE tissue sections were prepared by GTRF. All FFPE tissue sections were 4μm thick, the tissue area was roughly estimated by overlaying the tissue with a millimetre paper. The full FFPE tissue was scraped off and placed in 1x FFPE Lysis Buffer (BioSpyder), using 2μL Lysis Buffer/1mm^2. Molecular biology-grade mineral oil (Thermofisher) was added, and the tissue was deparaffinised by incubating at 95℃ for 5 min. Further processing included a 30-minute incubation at 37℃ in 2.5% FFPE Protease (BioSpyder), followed by mechanical homogenization by trituration. Lysates were then stored at -80℃ until required. This was performed by BioClavis staff.

2.2.5.7 TempO-Seq™ Assay

Unless specified otherwise, all reagents used in Templated Oligo Assay with Sequence Readout (TempO-Seq™) were supplied by BioSpyder (USA).

2.2.5.7.1 Sample Preparation

All experiments included BRR and URR controls, at least one dilution of *F. nucleatum* in human RNA, and no sample controls (NSCs) in the same 96-well plate, all run in duplicate.

2.2.5.7.2 DO Hybridization

A hybridisation mix containing 2x Annealing Buffer and DO pool was prepared as described in 2.2.5.4. To prevent evaporation during lengthy heated incubations, 6µL of mineral oil were added to each well of a labelled assay plate. Then 2µL of Hybridisation Mix was added to each sample well, followed by 2µL of samples lysed in 2x FFPE Lysis Buffer to the appropriately labelled wells. The plates were sealed, centrifuged at 1000rpm for 1 minute at RT, and placed in a T100 thermal cycler (BioRad). Hybridisation was performed by incubating assay plates at 70℃ for 10 minutes. The temperature was then reduced to 45℃ by lowering at a rate of 0.5℃/minute, then held at 45℃ for 16 hours. Plates were then cooled to 25℃.

2.2.5.7.3 Excess DO Digestion

To remove unhybridised DOs, a nuclease mix containing 10x Nuclease Buffer, a Nuclease Enzyme, and UltraPure water was prepared, and 24µL added to each well. The plate was sealed, centrifuged at 1000rpm for 1 minute at RT, and placed in a thermocycler, where it was incubated at 37℃ for 90 minutes.

2.2.5.7.4 DO Ligation

As DOs target adjacent sequences for hybridisation, this allows them to be ligated. A ligation mix consisting of 10x Ligation Buffer, Ligase Enzyme, and UltraPure water was prepared and 24µL added to each well. The plate was sealed, centrifuged at 1000rpm for 1 minute at RT, and placed in a thermocycler, where it was incubated at 37°C for 1 hour, and then at 80℃ for 15 minutes.

2.2.5.7.5 PCR Amplification

Once ligation was complete, 10µL of ligated product was added to PCR plates containing a proprietary PCR Pre-Mix including indexed primers. The plate was sealed and centrifuged at 1000rpm for 1 minute at RT. The PCR reaction was

programmed as follows: 37℃ for 10 minutes, then 95℃ for 1 minute, followed by 35 cycles of 95℃ for 10 seconds, then 65℃ for 30 seconds, then 68℃ for 30 seconds. The optical thermal cycler read the plate after each cycle to measure the amplified product. Finally, the plate was incubated at 68℃ for 2 minutes and held at 25℃. This was stored at -20℃ or used right away. The amplification was inspected after each run prior to sequencing; a positive amplification curve indicates an amplifiable product that can be sequenced.

2.2.5.8 Library Pooling and Purification

20µL of barcoded PCR products was added to a flip-spin reservoir and purified using the Macherey-Nagel NucleoSpin Gel and PCR Cleanup Kit (Cat. 740609.50, Macherey-Nagel) and quantified by Qubit using the dsDNA kit (Cat. Q32851, ThermoFisher Scientific).

2.2.5.9 Sequencing

This step was performed on barcoded samples pooled into a single library and run on the Illumina MiniSeq High-Output flowcell. Sequencing reads were demultiplexed by BCL2FASTQ software (Illumina, USA).

2.2.5.10 Analysis of TempO-Seq™ Read Counts

Before analysis, sequencing reads in ".FASTQ" format were converted to simple read count tables using BioClavis' automated pipeline. Sums of read counts from each probe for each sample were averaged, and then normalized by dividing the read count value by the product of the sum of read counts for each probe and the average of probe sums of all samples as shown in the equation below:

$$\frac{Read\ Count\ Value\ of\ Each\ Probe}{(Sum\ of\ Read\ Counts\ for\ Each\ Probe\ \times Average\ of\ Probe\ Sums\ of\ All\ Samples)}$$

Thresholding of *F. nucleatum* negative or positive status was determined by the *F. nucleatum* probe signal on the human reference RNA (BRR and URR). The average plus 1 standard deviation for each *F. nucleatum* probe for those samples was calculated, and any signal from the *F. nucleatum* probes on the INCISE polyp lysates above the thresholds was considered an *F. nucleatum* positive sample for that probe. This analysis was done on MS Excel (Ver. 2301, Microsoft, USA).

2.2.6 RNAscope®

Unless specifically stated otherwise, all reagents used for RNAscope® were manufactured by ACD-biotechne (UK).

2.2.6.1 Case Selection

Patients who were included in this study were selected from the subset of patients that underwent TempO-Seq™ using the *F. nucleatum* probes as described in 2.2.5, that were deemed *F. nucleatum*+.

2.2.6.2 Slide Requisition and Preparation

Patient tissue was requested from NHS Greater Glasgow and Clyde Tissue and cut by GTRF, no more than 2 weeks prior to use. Slides were stored at 4°C until used.

2.2.6.3 RNAscope® Run

Components of an RNAscope® 2.5 LSx Reagent Kit – BROWN (Table 2.12) and probes (

Table 2.13) were transferred to Open BOND containers (Ref. OP309700, Leica Biosystems). Slides were registered as outlined in 2.2.2.4.1. RNAscope® was run overnight following a protocol designed and registered by an ACD technician. The slides were dehydrated and mounted with coverslips as outlined in 2.2.2.1.10, apart from the use of xylene instead of HistoClear-II.

Table 2.12. Components of RNAscope® 2.5 LSx Reagent Kit – BROWN (Ref. 322700)

Reagent	Reference
RNAscope® 2.5 LSx Rinse	23936795
RNAscope® 2.5 LSx Rinse	23936793
RNAscope® LSx Hematoxylin	23937036
RNAscope® LSx DAB Part 1	23937819
RNAscope® LSx DAB Part B	23937211
RNAscope® LSx Bluing	23937309
RNAscope® LSx H_2O_2	23937475
RNAscope® LSx AMP 1 DAB	23937511
RNAscope® LSx AMP 2 DAB	23936461
RNAscope® LSx AMP 3 DAB	23937730
RNAscope® LSx AMP 4 DAB	23936050
RNAscope® LSx AMP 5 DAB	23936344
RNAscope® LSx AMP 6 DAB	23936162
RNAscope® LSx Protease	23936584

Table 2.13. Summary of RNAscope® Probes

Probe Target	Probe Purpose	Reagent	Reference
Homo sapiens Ubiquitin C	Detect all RNA	RNAscope® 2.5 LS Positive Control Probe – Hs-UBC	312028
Eubacteria 16S rRNA	Detect bacterial RNA	RNAscope® 2.5 LS Probe – EB-16S-rRNA	464468
F. nucleatum 23S rRNA	Detect *F. nucleatum* RNA	RNAscope® 2.5 LS Probe – B-Fusobacterium-23S-1zz	486408
Homo sapiens TATA-Box Binding Protein	Positive Control	RNAscope® 2.5 LS Probe – Hs-TBP	314298
Bacillus subtilis dihydrodipicolinate reductase	Negative Control	RNAscope® 2.5 LS Negative Control Probe – DapB	312038

2.2.6.4 Visualisation and Scanning

Slides were checked under a brightfield microscope for artefacts and scanned for visualisation as outlined in 2.2.2.4.5.

2.2.7 DNA Preparation

DNA was prepared by Jakub Jawny and Scott Murray of GTRF, for the INCISE collaborative.

2.2.7.1 Preparation of FFPE Tissue Curls

FFPE blocks were placed into a microtome and serially sectioned. Two 10µm curls were taken and stored in a 2mL DNA/RNA-free tube. Curls were stored in 4℃ until used.

2.2.7.2 Preprocessing of FFPE Samples from Curls

Curls were centrifuged briefly before the addition of 300µL mineral oil and vortexed. They were then heated at 80℃ for 2 minutes and allowed to cool at RT. A master mix containing 224µL lysis buffer, 25µL proteinase K, and 1µL blue dye (per reaction) was prepared, and 250µL/sample was added and vortexed. This was followed by centrifugation at 10,000g for 20 seconds. The contents were then transferred to Eppendorf tubes and heated at 56℃ for 30 minutes. This was followed by a 4-hour incubation at 80℃. Then the samples were allowed to cool at RT for 5 minutes. The aqueous phase of the mixture was then mixed with 10µL of RNase A, and the samples were incubated for 5 minutes at

RT, then centrifuged at full speed for 5 minutes. The aqueous phase was then promptly transferred to the prepared Maxwell® FFPE Cartridge.

2.2.7.3 Instrumental Run and Quantitation

The deck tray was set-up with the cartridges for use and the seals removed, and plungers positioned into the cartridge wells. Elution tubes were fitted for each cartridge and 50µL nuclease-free water was added to each elution tube. The on-screen instructions for the Maxwell® RSC DNA FFPE Kit experiment were followed. After the run was completed, the elution tubes (containing extracted DNA) were retrieved and allowed to come to RT for quantitation, which was performed using a calibrated Qubit™ 4 Fluorochrome.

2.2.7.4 Storage

Samples were stored for long-term use at -70℃ or below.

2.2.7.5 DNA Sequencing

DNA sequencing was performed by the Genomic Innovation Alliance (GIA, UK). The SureSelect XT2 HS2 kit (Part NumbeR: G9983D, Agilent, USA) was used for sample sequencing on the NovaSeq 6000 (Illumina, UK). The SureSelect CancerPlus panel (Design ID: S3225252, Agilent, USA) was used to enrich regions of interest. Resulting library quality and quantity was determined by TapeStation D1000 ScreenTape (Cat. 5-67-5582, Agilent, USA).

2.2.7.6 Next Generation Sequencing Analysis Pipeline

This pipeline and all steps within it were developed and performed by GIA. In brief, sequencing data were processed, and single nucleotide variation files were generated using the HOLMES pipeline (V1.3., V1.3.1). Bcl2fastq conversion software (V2.19.1.403, V2.20.0.422 on C++; Gerstung et al. 2012, 2014) was used to convert NovaSeq 6000-generated ".cblc" raw data files to FASTQ files and aligned using Burrows-Wheeler Alignment (V0.7.15 as a C program; Li and Durbin 2009). deepSNV/Shearwater (V1.22/V1.1.0, V1.22.0.5/ V1.1.0 as an R package/R wrapper script;Gerstung et al., 2012; Gerstung, Papaemmanuil and Campbell, 2014) was used for calling single nucleotide variants, while Pindel (V0.2.5b8-ww1 as a standalone application; Ye et al. 2009) was used for large

insertions/deletions. Once called, variants were annotated using CAVA (V1.2.2.ww1, V1.2.2.ww5 on Python; Münz et al., 2015). Discordant grouping was used to identify structural variation breakpoints using BRASS (V5.3.3-ww10 on C++), and copy-numbers were called using geneCN (V2.1 on Perl and on R).

2.2.7.7 Filtering Strategy

The HOLMES pipeline was run in the tumour-only mode on a subset of the INCISE cohort, against a panel of 21 normal tonsil samples. This allowed for germline variants and artefacts to be removed from the INCISE test samples, which generated a purely somatic output.

2.2.8 Patient-Derived Organoids

FAP patient-derived organoids (PDOs) were obtained from Prof. Farhat Din (University of Edinburgh) through Prof. Seth Coffelt (University of Glasgow).

2.2.8.1 Bringing up from Frozen

Thawed organoids were moved to a 15mL centrifuge tube and topped up with 5mL wash media (DMEM with 10% FBS and Primocin), and spun at 1000rpm for 3 minutes at RT. The supernatant was carefully discarded taking care not to disrupt the PDO pellet. The PDOs were resuspended in 60-120µL (depending on pellet size) of Matrigel® Matrix Basement Membrane (Ref. 354230, Corning). 20µL of PDO-Matrigel mix were placed in pre-heated 12-well plate with an average of 3 domes per well, incubated upside down for a minimum of 30 minutes at 37℃ to solidify, before being righted and 800µL of Supplemented IntestiCult Media (Cat.100-0190 Stem Cell Technologies) with Y-27632 rho-associated protein kinase (ROCK) inhibitor (Cat. 72304, Stem Cell Technologies) and Primocin (Cat. ant-pm-05, InvivoGen) added. They were checked every 48 hours and media was changed, or they were passaged as required.

2.2.8.2 Passaging of PDOs

PDOs in Matrigel domes had the old media removed and were washed in cold PBS. They were then mechanically disrupted using a Pasteur pipette, moved to a 15mL centrifuge tube, and topped up with PBS. The PDOs were centrifuged at 800rpm for 3 minutes at RT. The PBS was carefully aspirated to avoid disrupting

the pellet. The PDO pellet was vigorously resuspended in 150µL of wash media to break up the organoids into a single-cell suspension. 5ml of wash media added before the tube was centrifuged at 1000rpm for 3 minutes at RT. The excess media was carefully aspirated, and the PDOs resuspended in 60-120µL (depending on pellet size) of Matrigel® Matrix Basement. 20µL of PDO-Matrigel mix were placed in pre-heated 12-well plate with an average of 3 domes per well, incubated upside down for a minimum of 30 minutes at 37℃ to solidify, before being righted and 800µL of Supplemented IntestiCult Media with ROCK inhibitor and Primocin added. They were checked every 48 hours and media was changed, or they were passaged as required.

2.2.8.3 Seeding of PDOs for Pathway Stimulation

Media from wells with confluent PDO domes was aspirated, and the domes incubated with 1mL Cell Recovery Solution (Ref. 354253, Corning), for 1 hour at 37℃ and 5% CO_2. Domes were then mechanically disrupted using a Pasteur pipette and moved to 15mL tubes to be spun at 1000rpm for 3 minutes at RT. The supernatant was aspirated and the PDOs were washed in PBS before being spun at 1000rpm for 3 minutes at RT. After aspirating the supernatant, the PDO pellet was resuspended in 150µL wash media to loosely break up the organoids. The PDOs were topped up with 5mL wash media and spun at 1000rpm for 3 minutes at RT. The supernatant was removed, and the pellet was resuspended in 100-400µL wash media (depending on pellet size). Cells/mL were counted using 10µL of PDO cell suspension loaded into a CellDrop (DeNovix, Delaware, USA). PDO cells were seeded at $5x10^5$ cells in Matrigel in 5µL dots per well in 96-well pre-warmed plates. Plates were flipped and incubated at 37℃ and 5% CO_2 to allow domes to form, before 100µL of Supplemented Intesticult media was added to each well.

2.2.9 Cellular Treatment Studies

2.2.9.1 Compound Selection and Preparation

2.2.9.1.1 Cytokines Preparation

The selection of suitable compounds was previously done within the group by Dr. Kathryn Pennel, Dr. Guang-Yu Lian, and Molly McKenzie. LIGHT, TNF-α,

Lymphotoxin, and IL-6 were all prepared to stock concentrations, and 5-FU was included as positive assay control at 10μM as described in Table 2.3.

2.2.9.1.2 Wnt Signalling Pathway Stimulants Preparation

A literature search for compounds that successfully stimulate Wnt signalling was carried out, and 4 compounds (Table 2.3) were purchased from SelleckChem. Compounds were prepared to a stock concentration of 10mM using DMSO as per the manufacturer's instructions. 5-FU was included as positive assay control at 10μM.

2.2.9.2 Compound Concentration Testing on 2D Cell Lines

2.2.9.2.1 Cytokines

The optimum concentrations to affect cellular changes were investigated by Dr. Kathryn Pennel, Dr. Guang-Yu Lian, and Molly McKenzie. All cytokines were used at 1:1000.

2.2.9.2.2 Wnt Stimulants

In order to discern the best Wnt stimulant at the optimum concentration for use in PDOs, a series of concentrations for each compound were tested on 3 different CRC cell lines with increased cell viability as an outcome. HT29, HCT116, and DLD-1 cells were grown to 80% confluency, trypsinsed as described in 2.2.1.2, and seeded at 10,000 cells/well in 96-well plates. Cells were allowed to adhere in growth media for 24 hours at 37℃ and 5% CO_2. The media was aspirated and replaced with CHIR-99201 (at 10μM, 3μM, and 1μM), BML-284 (at 10μM, 3μM, 1μMm, and 0.5μM), AZD2858 (at 3μM, 1μM, 0.5μM, and 0.1μM), and IM-12 (at 10μM, 3μM, and 1μM) for 72 hours at 37℃ 5% CO_2. All experiments included a media-only control, a 1% DMSO vehicle control, and a 5-FU positive control. All conditions were done in technical triplicates, and the experiment was done on biological triplicates.

2.2.9.3 Cell Viability of 2D Cell Lines

To check whether Wnt-stimulation affected 2D CRC cell line growth, media/vehicle/compound were aspirated from each well and replaced with 100μL of fresh growth media, to which 10μL WST-1 reagent was added. Plates

were gently tapped to mix the well contents, and the plates were incubated at 37°C and 5% CO_2 for 90 minutes. The GloMax® Discover (Serial Number: 9700100424, Promega, USA), was used to shake the plate for 30 seconds, followed by absorbance reading at 450nm. Readings from triplicate wells for each condition were averaged and normalised to vehicle control. Normalised results for each condition from biological triplicates were averaged and plotted using GraphPad Prism10 (Ver.10.6.0 (890) for Windows, GraphPad Software, Boston, Massachusetts USA).

2.2.9.4 Treatment of 3D PDOs

2.2.9.4.1 Treatment with Cytokines

Twenty-four hours after PDOs were seeded (as described in 2.2.8.3), old media was carefully aspirated and 100µL of Supplemented IntestiCult media, 5-FU positive control, LIGHT, TNF-α, Lymphotoxin, and IL-6 made up in Supplemented IntestiCult media (1:1000) were added to triplicate wells for each condition. These were incubated for 72 hours at 37°C and 5% CO_2. A vehicle control was not included because all the cytokines were made up in PBS/dH_2O.

2.2.9.4.2 Stimulation of Wnt Signalling

The most effective compounds were chosen based on the 2D cell line viability results, and their most effective concentrations were used to stimulate PDOs. 24 hours after PDOs were seeded (as described in 2.2.8.3), old media was carefully aspirated and 100µL of Supplemented IntestiCult media, 1% DMSO vehicle control, 5-FU positive control, CHIR-99201 (3µM and 1µM), and AZD2858 (1µM and 0.1µM) made up in Supplemented IntestiCult media were added to triplicate wells for each condition. These were incubated for 72 hours at 37°C and 5% CO_2.

2.2.9.5 Detection of Treatment Effects on 3D PDOs

Following the 72-hour incubation period, PDO domes across all control and experimental conditions were imaged using a brightfield microscope (Axio Vert.A1, Carl Zeiss Microscopy, Germany) and accompanying Zen Blue software (Zen 2012 (blue edition), Carl Zeiss Microscopy, Germany). Cell viability was measured via metabolic activity using the CellTiter-Blue® assay (Cat.G8080, Promega, USA). 20µL of CellTiter-Blue® was added to the wells, and the plates

were incubated for 4 hours at 37℃ and 5% CO_2. The GloMax® Discover, was used to shake the plate for 30 seconds, followed by fluorescence reading at 560nm. Readings from triplicate wells for each condition were averaged and normalised to vehicle control. Normalised results for each condition from biological triplicates were averaged and plotted using GraphPad Prism. This was done for both Wnt stimulants and cytokines.

2.2.10 Statistical Analysis for IHC

Statistical analysis for IHC was performed on RStudio (R Core Team, 2021) and SPSS (Ver. 27, IBM). Scores (H-scores and point counts) were dichotomised into low and high based on thresholds generated using *Survminer* (Ver.0.4.9; Kassambara, Kosinski and Biecek, 2021) and *Maxstat* (Ver. 0.7-25; Hothorn, 2017) packages. Unless stated otherwise, the generated thresholds were used. Where suitable, thresholds were generated from patients within the training dataset and subsequently applied to the test dataset. 1-survival analysis used Kaplan-Meier curves to demonstrate patients getting metachronous polyps or CRC as time progressed. X^2 analysis was used to reveal any relationships between protein expression levels and clinical factors. Cox regression analysis was used to produce hazard ratios (HRs) with 95% confidence intervals (95% CIs), and to determine which variables were independent prognostic factors. Histograms were used to visualise the distribution of the frequency of scores, and violin plots were used to visualise the distribution of scores across different groups.

2.2.11 Mutational Analysis

All analysis was performed on RStudio (R Core Team, 2021; 2023.12.0+369; MA, USA) unless otherwise specified. All analysis was performed by the author unless otherwise specified.

Mutational analysis was performed using the *Maftools* (Mayakonda *et al.*, 2018) (Ver. 2.18.0) package under the BiocManager Repository (Morgan and Ramos, 2023; Ver. 1.30.22). Oncoplots were generated using the “oncoplot” function. Pairwise Fisher’s Exact test was performed to establish co-occurring and mutually exclusive mutated gene pairs using the “somaticInteractions” function.

Lollipop plots were generated using the MutationMapper tool on cBioPortal (Cerami *et al.*, 2012; Gao *et al.*, 2013; de Bruijn *et al.*, 2023). All statistical significance for mutational analysis was set at an adjusted p-value (pAdj) of <0.05.

2.2.12 Bulk Transcriptomics Analysis

All transcriptomics data originated from TempO-Seq™ whole transcriptome profiling experiments on the INCISE cohort as described in 2.2.5.7, using the Human Whole Transcriptome v2.0 panel of 19703 genes. Initial quality control and preprocessing was carried out by Dr. Natalie Fisher and Dr. Philip Dunne at Queen's University Belfast as part of the INCISE Collaborative. In brief, probe variance across all samples was calculated using the *sapply* function. Probes with a variance less than the 25th percentile were removed. Samples with a low read count (< 95% CI [<2.28x106 counts]) were identified with *stats* (Ver. 4.2.3) and removed, as well as samples which did not match batch information. The counts were batch corrected using *ComBat_seq* (sva Ver. 3.46.0; Zhang, Parmigiani and Johnson, 2020). Probe IDs were mapped to gene symbols with duplicated genes collapsed using *MaxMeans* (Langfelder and Horvath, 2008, 2012; *WGCNA* Ver. 1.72-1), resulting in 14,993 genes. Normalised counts were generated using quantile normalisation and log2 +1 transformed.

2.2.12.1 Principle Component Analysis

Principle Component Analysis (PCA) was conducted using the prcomp() function under R's stats package (Ver 4.3.3) on normalised expression data.

2.2.12.2 Differential Gene Expression

Differentially expressed genes (DEGs) were analysed using *DESeq2* (Love, Huber and Anders, 2014; Ver. 1.42.1), which allowed for the extraction of Log2 Fold-Change, p-value, and pAdj (Wald-statistic).

2.2.12.3 Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was used to compare gene expression within pre-defined biological pathways between two groups, using differential expression data containing $\log_2$ fold change generated from *DESeq2* (2.2.12.2).

GSEA was performed using *ClusterProfiler* (Ver.3.18; Yu et al., 2012) for the *Homo sapiens* organism annotation. Three different ontologies/pathway databases were explored: Gene Ontology (GO; Ashburner et al., 2000; Consortium et al., 2023), Molecular Signature Database (MSigDb; Subramanian et al., 2005; Liberzon et al., 2011, 2015), and the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2024).

2.2.12.4 Visualisation

All data was visualised using *ggplot2* (Ver. 3.5.0; Wickham, 2016). geom_point() was used to visualise PCA, and volcano plots to display DEGs. geom_violin() with geom_box() were used to generate violin plots overlaid with box plots to visualize the expression patterns of individual DEGs between two groups. *ComplexHeatmap* (Ver. 2.21.1; Gu, 2022) was used to generate heatmaps to investigate any patterns of expression of DEGs between two groups. For GSEA, dot plots were generated using dotplot() under the *DOSE* package (Ver. 3.28.2; Yu et al., 2015) to show activated and suppressed gene sets based on gene ratio and pAdj. Bar plots were generated using geom_col to show statistically significant upregulated and downregulated gene sets ranked by normalised enrichment scores (NES). UpSet plots were used to show leading edge genes shared across enriched gene sets using the ComplexUpset package (Ver. 1.3.3; (Lex *et al.*, 2014).

2.2.13 Spatial Transcriptomics on CosMx™

2.2.13.1 Patient Selection

Patients included in this study (n=10) were the same patients who were included in the RNAscope® study described in 2.2.6.

2.2.13.2 TMA Construction

A TMA consisting of 30 cores was constructed. For each patient, two *F. nucleatum*+ cores and one *F. nucleatum*- core were included. *F. nucleatum*+ cores were selected with the bacterial location (visualised by RNAscope®) serving as the centroid for each core (or as close as possible).

2.2.13.3 CosMx™ Experiment

CosMx™ was performed by Claire Kennedy Dietrich from the Jamieson Spatial Lab as detailed below. Unless stated otherwise all reagents are from NanoString.

The FFPE TMA slide was sectioned at 5µm and placed on a Leica Bond+ slide. The slide was dried at 37℃ ON followed by baking at 60℃ ON. The slide was then loaded onto the Lecia Bond automated system for deparaffinisation, followed by HIER at 100℃ for 15min in a citrate buffer (pH6). This was followed by an incubation in proteinase K (3µg/mL) for 20min at 37℃.

Slides were then incubated in 0.001% fiducial solution diluted in 2XSSCT for 5min at RT, and washed sequentially in PBS for 1 min, 10% NBF for 1min, and NBF Stop buffer for 5min repeated. The slides were then incubated in NHS-Acetate buffer for 15min at RT, followed by two washes in 2XSSC for 5min. The probe panel comprising of 6200 genes (including 20 negative control probes) was denatured at 95 degrees for 2min before immediately being cooled on ice for at least 1min. The probes were then mixed with buffer R, RNase inhibitor and DEPC water, and applied to the slide. Hybridisation was carried out for 17 hours at 37℃ in a hybridisation oven ON.

The following day the slide was washed twice in a solution of 50% formamide and 50% 4XSSC in a water bath set to 37℃ for 25min. This was followed by two washes in 2XSSC for 2min each, before incubating the slide in Dapi for 15min at RT in the dark. The slide was then washed in PBS for 5min before the standard NanoString antibody panel containing cell membrane marker CD298/B2M, epithelium marker PanCK, immune marker CD45, and macrophage marker CD68 was prepared with blocking buffer and the slide was incubated for 1 hour at RT in the dark. The slide was then washed three times in PBS for 5min each, before the flow cell was affixed to the slide and 2XSSC was added into the flow cell. The slide was then processed on the CosMx™ Spatial Molecular Imager (SMI), with 178 fields of view (FOV) selected to ensure the instrument ran within the appropriate time scale.

The CosMx™ Instrument then performs 16 imaging cycles, during each cycle 1 out of 4 possible types of reporter fluorophores hybridise to the readout domain of

the target probes. After hybridisation an image is acquired followed by cleavage and removal of the reporter probes. This process is then repeated across all cycles to capture the complete transcriptomic information. The acquired images are then decoded into RNA transcripts which can be mapped into a corresponding segmented cell. The raw files are subsequently exported from the AtoMx® platform for analysis in R.

2.2.13.4 CosMx™ Data Analysis

CosMx™ data was quality checked, normalised, underwent dimensionality reduction, integrated, clustered, and bulk-analysed by John Cole from the MVLS Shared Research Facilities. Dr. Assya Legrini assisted with neighbourhood analysis and pathway analysis.

In brief, the dataset was segmented under default settings using AtomMx® and exported as flat text files. The dataset was loaded into R using *Seurat* (Ver. 5.1.0; Satija et al., 2015; Stuart et al., 2019; Hao et al., 2021, 2024) and cells assigned to the appropriate samples and conditions based on FOV. The relationship between per cell feature counts, number of features, and sample was assessed using *FeatureScatter, VlnPlot*, and *SpatialFeaturePlot* functions. Data normalisation was performed using SCTransform (Ver. 0.4.2; Hafemeister and Satija, 2019) and PCA.

Clustering was performed using *FindNeighbors* and *FindClusters* functions under default settings and a uniform manifold approximation and projection (UMAP) generated to show a clear per sample batch effect. The PCA was integrated on sample using *Harmony* (Ver. 3.8; Korsunsky et al., 2019), under default settings.

To identify cell types per cluster *FindAllMarkers* function specifying only.pos = TRUE, min.pct = 0.25 and logfc.threshold = 0.25. *SingleR* (Ver. 2.1; Aran et al., 2019) with the Human Primary Cell Atlas database was used as a prior, with clusters manually curated based on the markers in Table 3.3.

For direct differential analysis between *F. nucleatum*+ and *F. nucleatum*-, the data was collapsed into pseudo counts on sample plus cell type. Genes with > 10 reads on average per cell type, and pseudo samples with > 100k reads were

retained. For each cell type, the data was corrected for library size using *CombatSeq* (Ver. 4.5; Leek et al., 2012), then pairwise differential expression was calculated using *DESeq2* (Ver. 1.48.1) under default settings. The results were visualised using *Searchlight* (Ver. 2.0.3) specifying pAdj < 0.05, absolute log2fold > 1 for significance, and using the String (Ver. 11.5; (Szklarczyk *et al.*, 2022)) and TRRUST (Ver. 2.0; Han et al., 2015) databases.

For the spatial analysis a custom python script was written to take each core and mark concentric circles, of distance 1 FOV, from either the core midpoint (*F. nucleatum*-) or the bacteria (*F. nucleatum*+). Circles were used to create bands labelled 1-9, with 1 being the innermost and 9 the outermost, and cells assigned to bands. Next, pseudo counts were calculated by sample and band, corrected using *ComBat_Seq* by library size, and normalised using *DESeq2*. For each gene an ANNOVA was performed, using the model expression ~ distance + group, where: expression = the normalised pseudo bulk expression, distance = band number, group = *F. nucleatum*-/*F. nucleatum*+. The p-values were adjusted using Benjamini-Hochberg (BH), and significance set at pAdj<0.05. The resultant significant genes showed differential expression between *F. nucleatum*-/ *F. nucleatum*+ across the radius of the cores.

Neighbourhood analysis was performed using *CellularNeighborhoods* (Hu *et al.*, 2024) and *mClust* (Ver.), where the neighbourhood size was set to a maximum of 50µm. Pathway analysis was performed using *decoupleR* (Ver. 3.21; Badia-I-Mompel et al., 2022) and *OmniPath* (Ver. 3.21; Türei et al., 2021).

Chapter 3 Detection of *F. nucleatum* and Its Effects on Colorectal Polyps

3.1 Introduction

F. nucleatum is an anaerobic Gram-negative rod bacterium mainly present in the oral cavity that was considered to be a commensal bacterium but has since been shown to be an opportunistic pathogen (Han, 2015). It has been implicated in not only dental diseases (Yang *et al.*, 2014) but also in conditions ranging from poor pregnancy outcomes (Han *et al.*, 2004) to cerebral aneurysms (Pyysalo *et al.*, 2013). In recent years, *F. nucleatum* has been linked with colorectal cancer. Small-scale sequencing studies have found that *Fusobacterium* species were enriched not only in CRC tissue (Marchesi *et al.*, 2011; Kostic *et al.*, 2012), but also in colorectal adenoma tissue and the stool samples of individuals with colorectal adenomas or CRC, when compared to matched normal samples (Kostic *et al.*, 2013). Higher *F. nucleatum* levels have been associated with poor cancer-specific survival, poor prognostic factors including right-sided cancers, poorly differentiated cancers, and high CIMP status (Mima, Nishihara, *et al.*, 2016). *F. nucleatum* was also found in CRC liver metastases and was associated with the presence of tumour cells at distant sites (Bullman *et al.*, 2017).

As previously described in this thesis (1.6.2) *F. nucleatum* is able to enrich itself within the tumour microenvironment through its Fap2 molecule, which also allows it to supress the anti-tumour immune response (Figure 1.9). Furthermore, the mechanisms of *F. nucleatum*-mediated tumour support, such as the triggering of NF-κB signalling through the binding of bacterial LPS to the TLR4 receptor (Figure 1.11) and the mimicking of Wnt signalling through the bacterial binding to E-Cadherin by *F. nucleatum*'s FadAc molecule (Figure 1.10) were discussed.

The role of *F. nucleatum* in CRC progression is becoming clearer through mechanistic studies, and although involvement with colorectal adenomas has been reported (Kostic *et al.*, 2013; McCoy *et al.*, 2013; Flanagan *et al.*, 2014; Yu *et al.*, 2016), its role in metachronous adenoma development is not well understood. This chapter will test the hypothesis that the presence of *F. nucleatum* informs metachronous polyp or CRC risk. To that end, bespoke probes for a novel sequencing technique (TempO-Seq™) were designed and used to detect *F. nucleatum* in INCISE polyp tissue, and RNAscope® was employed to visualise the bacteria. The mutational profile of the *F. nucleatum* positive (*F.*

nucleatum+) and *F. nucleatum* negative (*F. nucleatum*-) patients will be explored, and so will the transcriptomic differences between these two patient groups. Additionally, the adaptive immune response will be investigated through the scope of *F. nucleatum* presence, and the effects of growth on adenomas from immune-related cytokines will be test. Furthermore, high-plex imaging of RNA transcripts on a single cell level will be used to explore the spatial transcriptome of the areas where *F. nucleatum* was detected, through the CosMx™ SMI platform. Finally, the associations between the presence of *F. nucleatum* and clinical outcomes will be investigated.

3.2 Detection of *F. nucleatum* by TempO-Seq™

TempO-Seq™ is a targeted sequencing assay that targets specific sequences in genes but does this for every gene. Traditional RNA-Seq is difficult to scale to high throughput environments due to its cost and sensitivity-limiting RNA extraction and cDNA synthesis steps. TempO-Seq™ foregoes the need for purified RNA by targeting RNA available in tissue lysates. Two detector oligos (DOs) are designed to hybridise to sequences in proximity to each other, which can then be ligated together. Any unhybridised or excess DOs are enzymatically eliminated, and amplification of the DOs is carried out to add sequencing adaptors. While TempO-Seq™ cannot be used as a discovery method – as the DOs are designed for known targets - it provides high levels of specificity, as it eliminates positional bias and background noise, distinguishes homologous genes, and yields a simple readout (Yeakley *et al.*, 2017).

3.2.1 *In Assay* Probe Quality Check

3.2.1.1 Target Detection of Human and Bacterial Probes

Probes for *F. nucleatum* were designed and tested *in silico* (2.2.5.1-2.2.5.2) to ensure they accurately detected their targets. Human probes were available through established panels supplied by BioClavis Ltd. and were included as internal assay controls. TempO-Seq™ was performed on purified human reference RNA, purified *F. nucleatum* RNA, and a mixture of purified human reference RNA and purified *F. nucleatum* RNA. As shown in Figure 3.1A, the signal from the human probes was mostly detected in the human reference RNA samples, with the COX2 probe dominating the majority of the reads. Thus, the signal from the other human

probes was compressed. However, adjusting the y-axis allowed the signal from the other human probes to become visible. While the signal from COX2 remained the strongest, ß-catenin was second strongest, followed by SOX9, E-Cadherin, and p53 sequentially in the human reference RNA samples (Figure 3.1B).

The *F. nucleatum* probes all generated a signal in the purified bacteria-human RNA mixed samples, with the magenta probe consistently providing the strongest signal, followed by cyan, indigo, green, and gold probes. This pattern was maintained as the ratio of *F. nucleatum* RNA increased and human RNA decreased, and in multiple replicates of *F. nucleatum*-only RNA (Figure 3.1A). The signal from the COX2 probe was also detected in the bacteria-human RNA mixed samples (Figure 3.1B), with the signal levels gradually decreasing with decreasing human RNA concentrations. In the *F. nucleatum* only RNA samples, no human probe signal was detected, suggesting that the human probes are specific to human genes. Of note, the indigo probe represented most of the reads in sample FnRNA8. However, this is a negligible observation as the other seven *F. nucleatum* only RNA samples do not show a similar trend.

Four of the five bacterial probes generated no signal in the human brain reference RNA (BRR) and human universal RNA (URR) samples (2.2.4.1). However, the cyan probe produced a very weak signal in the URR sample. Despite all five *F. nucleatum* probes generating detectable signals, further inspection of the probes' sequence targets revealed that the green and indigo probes target sequences overlap with each other, hence only the magenta, cyan, and gold probes were taken forward for further analysis.

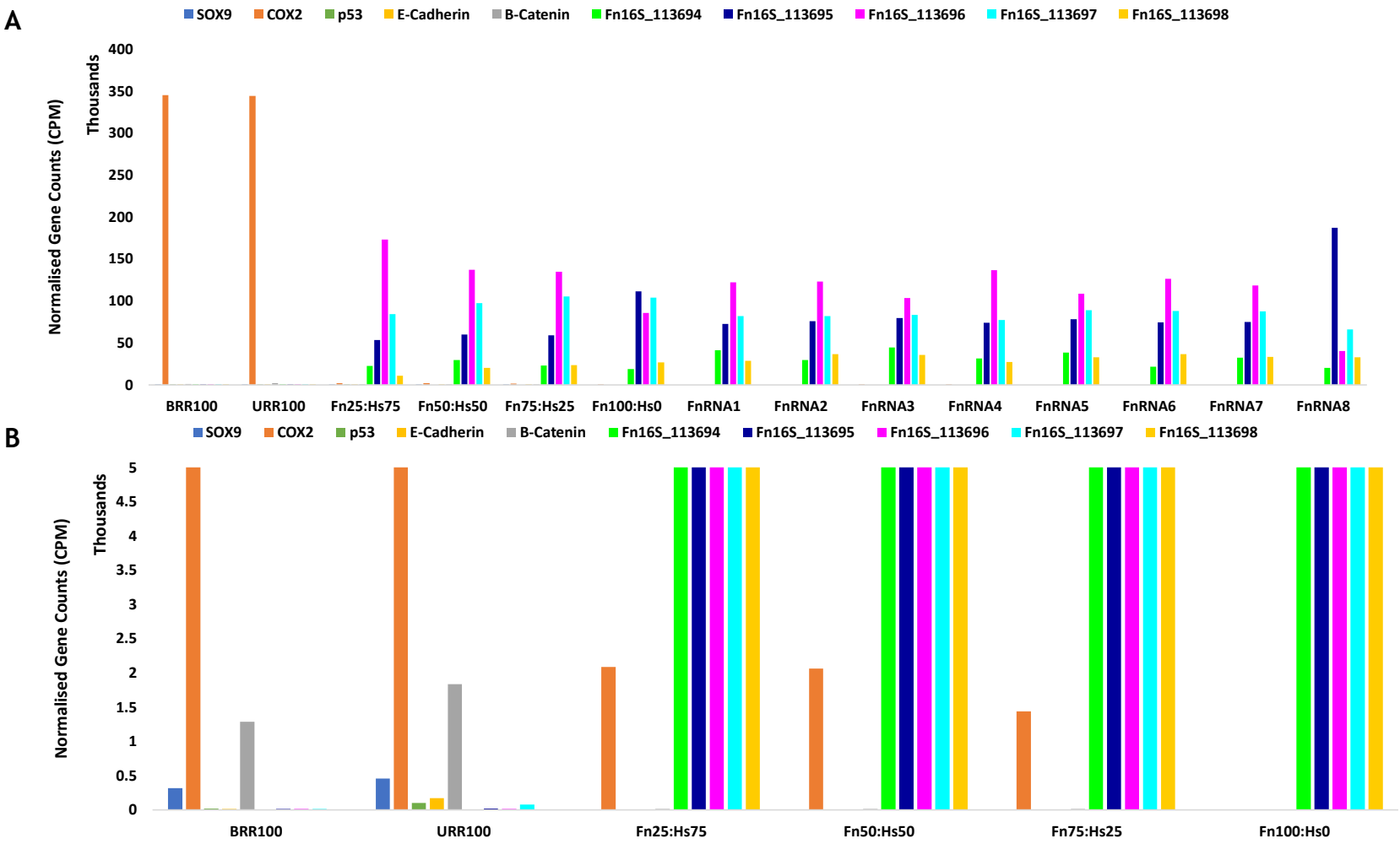


Figure 3.1. *In Vitro* Quality Check of Human Probes and *F. nucleatum* Probes.
[A] Normalised signal of TempO-Seq™ probes.**[B]** Normalised signal of TempO-Seq™ probes, with adjusted y-axis showing human reference RNA samples and bacteria-human RNA mixed samples only. Human Probes: SOX9, COX2, p53, E-Cadherin, β-Catenin. *F. nucleatum* probes: Green, Indigo, Magenta, Cyan, Gold. BRR: Brain Reference RNA. URR: Universal Reference RNA. CPM: Counts per Million. Fn: *Fusobacterium nucleatum*. Hs: *Homo sapiens*. FnRNA1-8: RNA extracted from different 8 *F. nucleatum* colonies. Data was obtained from duplicate runs.

3.2.1.2 Probe Specificity Tests in Bacterial and Human RNA

Once the *F. nucleatum* probes were shown to be specific to their target, they were tested to see how they perform against a high background of human RNA. 10-fold serial dilutions of *F. nucleatum* RNA in human URR RNA were prepared as in Table 2.7. Human reference RNA and pure *F. nucleatum* RNA controls were included. As shown in Figure 3.2A, the bacterial probes were able to detect their targets in dilutions as high as $1:1x10^4$ (bacteria to human). Only human probes carried a detectable signal in the human reference RNA samples, while only bacterial probes were detectable in the pure *F. nucleatum* RNA samples. COX2 remained the most prominent human signal, and it continued to compress signals from other probes. However, when adjusting the y-axis (Figure 3.2B), signal from the magenta and cyan bacterial probes were detected in dilutions up to $1:10x10^6$. This demonstrated that the *F. nucleatum* probes could produce a signal against high human RNA background.

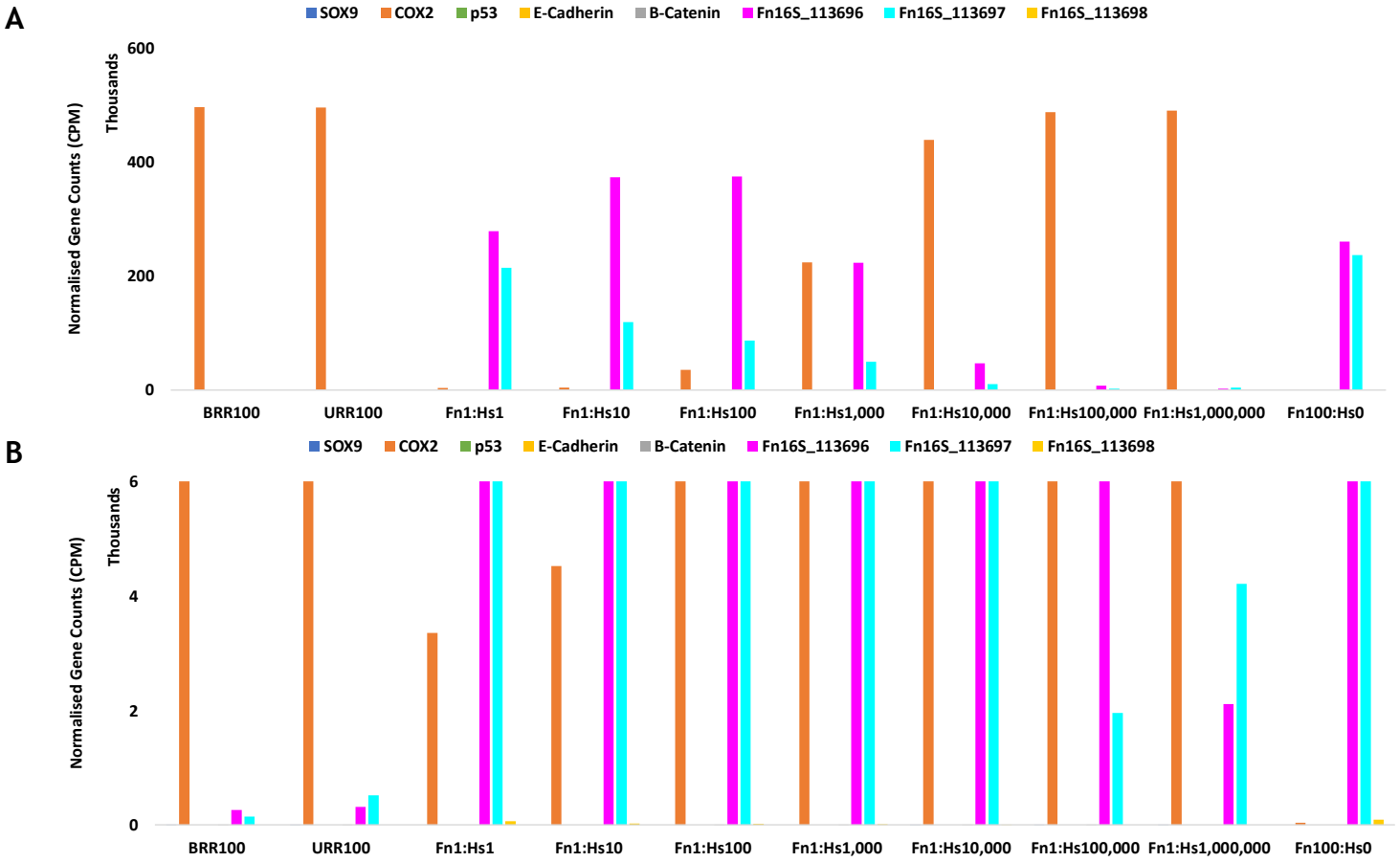


Figure 3.2. Bacterial Probes are Sensitive to their Targets in a High Background Environment.
Serial dilutions were performed in technical duplicates and were presented here as averages. **[A]** Two bacterial probes (magenta and cyan) were clearly detectable in bacteria-to-human RNA dilutions of up to 1:10,000. Only the COX2 human probe was detectable to varying levels in all samples including human RNA. **[B]** After adjusting the y-axis, the magenta and cyan bacterial probes were detectable in the 1:106 bacteria-to-human RNA dilutions. The gold bacterial probe had a very low signal in the 1:1 and 1:10 bacteria-to-human RNA dilutions, and in the *F. nucleatum* RNA control. BRR: Brain Reference RNA. URR: Universal Reference RNA. CPM: Counts per Million. Fn: *Fusobacterium nucleatum*. Hs: *Homo sapiens*.

3.2.1.3 Probe Sensitivity Testing on Human Tissue Lysates

The previous experiments demonstrated that the designed probes were able to detect their targets in a purified RNA environment. As TempO-Seq™ is a method designed to work on FFPE tissue, lysates from healthy colon samples and lysates from 15 INCISE patients were tested alongside the serial dilution of bacterial-to-human RNA described in 3.2.1.2.

As shown in Figure 3.3A, the signal from the COX2 human probe continued to dominate most of the reads, while the 10-fold serial dilution controls exhibited similar results to those shown in Figure 3.2, where the magenta and cyan probes could detect their targets in bacteria-to-human RNA dilutions as high as $1:1x10^4$. To allow the signal from the other probes to be visible, Figure 3.3B displays the same data zoomed in to the lowest parts of the y-axis. This revealed a visible signal from the magenta and cyan probes in the normal colon samples, and in one INCISE sample.

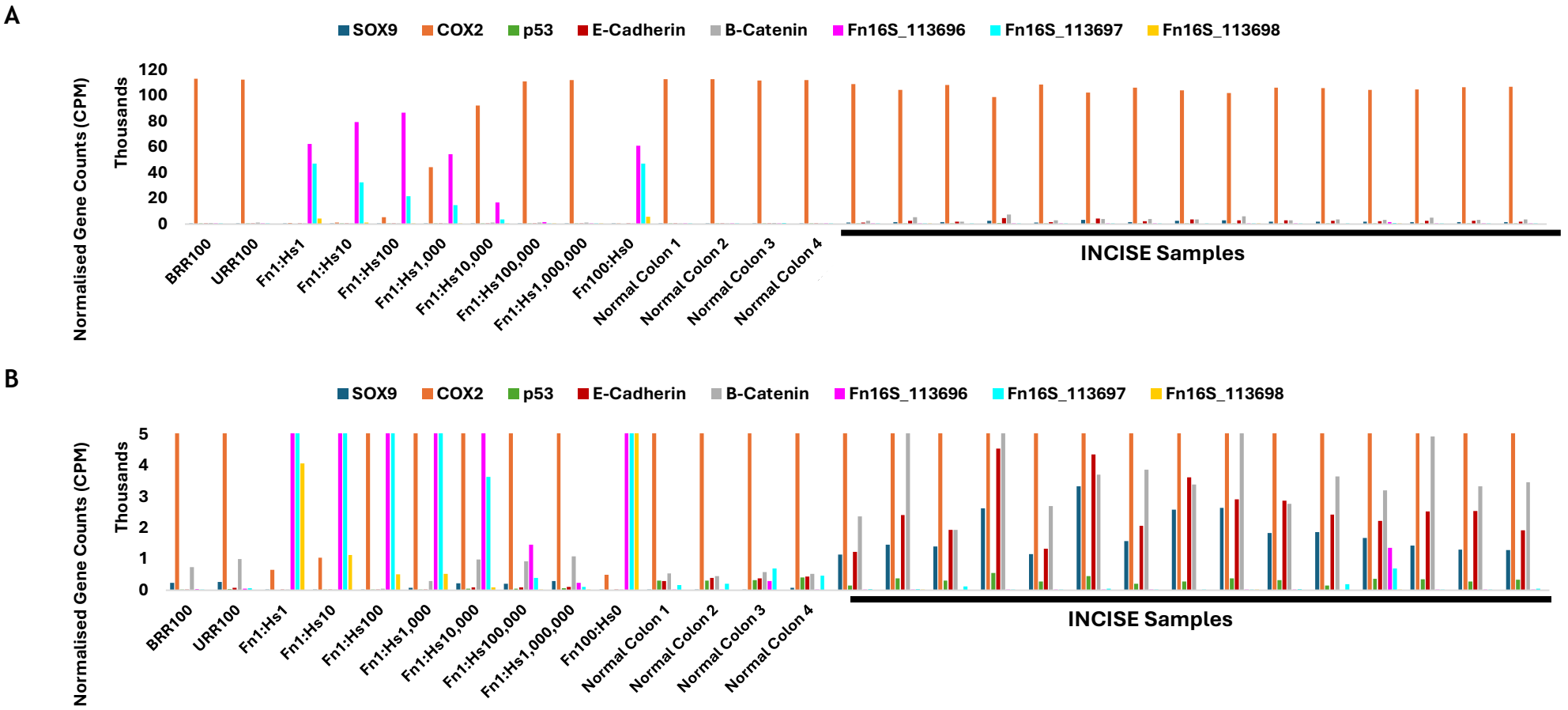


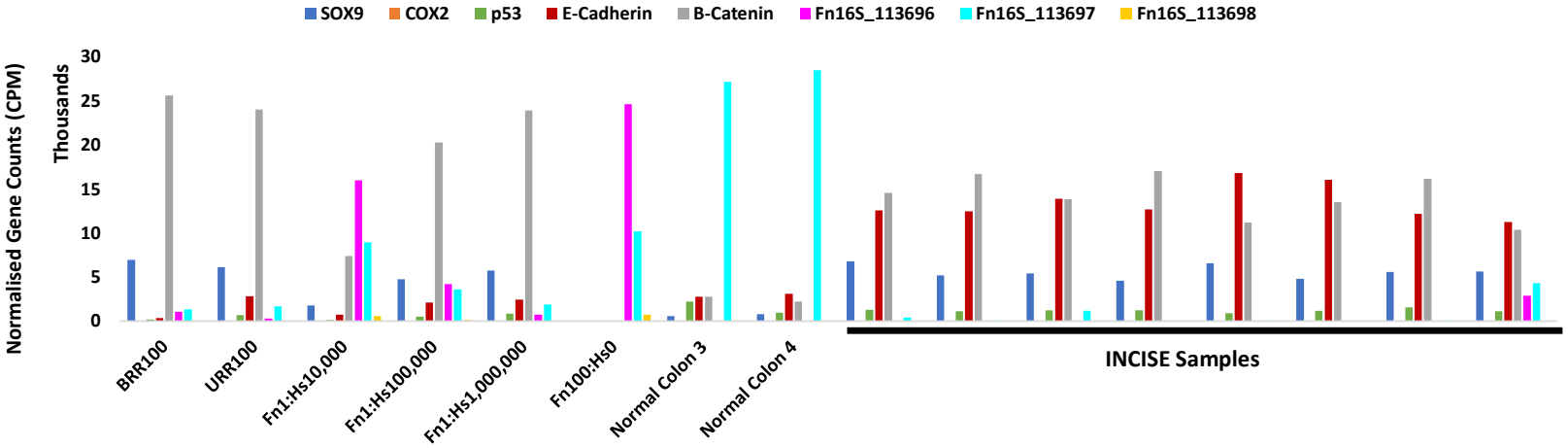
Figure 3.3. TempO-Seq™ Probes Can Detect their Targets in a Biological Setting.
[A] Lysates prepared from normal human colon FFPE tissue samples were included alongside lysates prepared from colonic adenoma FFPE tissue from INCISE patients. 10-fold serial dilutions of bacteria-to-human RNA and human reference RNA were included. Signal from COX2 took up the majority of the reads. **[B]** After adjusting the y-axis, signal from the *F. nucleatum* probes was more visible in the normal colon samples and in some INCISE samples (n=15). BRR100 and URR100 were run in technical duplicates. All other data points were obtained from a single sample run. BRR: Brain Reference RNA. URR: Universal Reference RNA. CPM: Counts per Million. Fn: *Fusobacterium nucleatum*. Hs: *Homo sapiens*.

3.2.1.4 Optimization of Probe Concentrations in Detector Oligo Pools

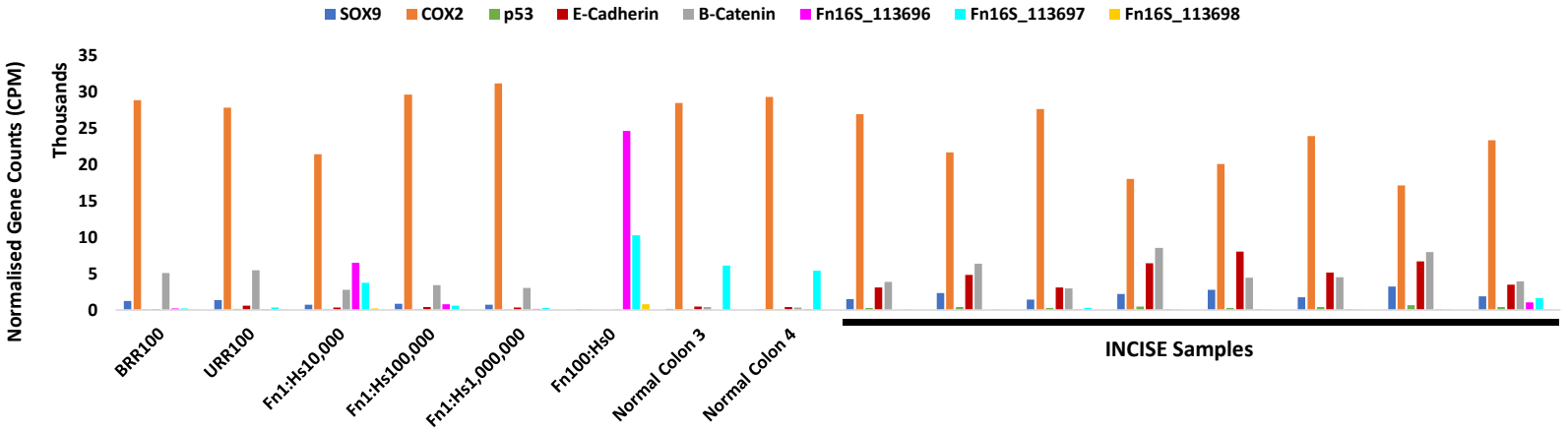
As outlined previously, COX2 signal repeatedly overpowered the signals from the other human probes, as well as the bacterial probes. To allow signals from the remainder of the probes to be displayed more clearly, the COX2 probe had to be diluted in the detector oligo (DO) pool. Dilutions (as described in Table 2.11) were made of 0.03nM, 0.09nM, and 0.009nM in 3nM of other probes, as well as removal from the DO pool altogether. Samples used to test the effect of each dilution on the overall signal included human reference RNA, purified bacterial RNA, bacterial RNA diluted in human RNA, healthy colon tissue lysates, and INCISE polyp tissue lysates.

When the COX2 probe was removed (Figure 3.4A), signals from other probes were amplified and became clearer. β-Catenin became the most prominent signal in samples with human RNA and all trends observed in Figure 3.3 not only remained but became more prominent. When COX2 was present in the DO pool at 0.09nM (other probes at 3nM), it regained dominance (Figure 3.4B). When the COX2 probe concentration was lowered to 0.03nM (Figure 3.4C), the signal was visible but not dominant. Finally, the lowest concentration of 0.009nM (Figure 3.4D) allowed other signals to appear more prominent, however it appeared to be too low to represent the true signal pattern of COX2. Since the inclusion of COX2 at lower concentrations allowed for other signals to be amplified and become clearer, the complete removal of COX2 was deemed unnecessary. COX2 at 1% of the DO pool (0.03nM in 3nM of other human and bacterial probes) was used for future experiments.

A


SOX9
COX2
p53
E-Cadherin
B-Catenin
Fn16S_113696
Fn16S_113697
Fn16S_113698
Normalised Gene Counts (CPM)
Thousands
30
25
20
15
10
5
0
BRR100
URR100
Fn1:Hs10,000
Fn1:Hs100,000
Fn1:Hs1,000,000
Fn100:Hs0
Normal Colon 3
Normal Colon 4
INCISE Samples

B


SOX9
COX2
p53
E-Cadherin
B-Catenin
Fn16S_113696
Fn16S_113697
Fn16S_113698
Normalised Gene Counts (CPM)
Thousands
35
30
25
20
15
10
5
0
BRR100
URR100
Fn1:Hs10,000
Fn1:Hs100,000
Fn1:Hs1,000,000
Fn100:Hs0
Normal Colon 3
Normal Colon 4
INCISE Samples

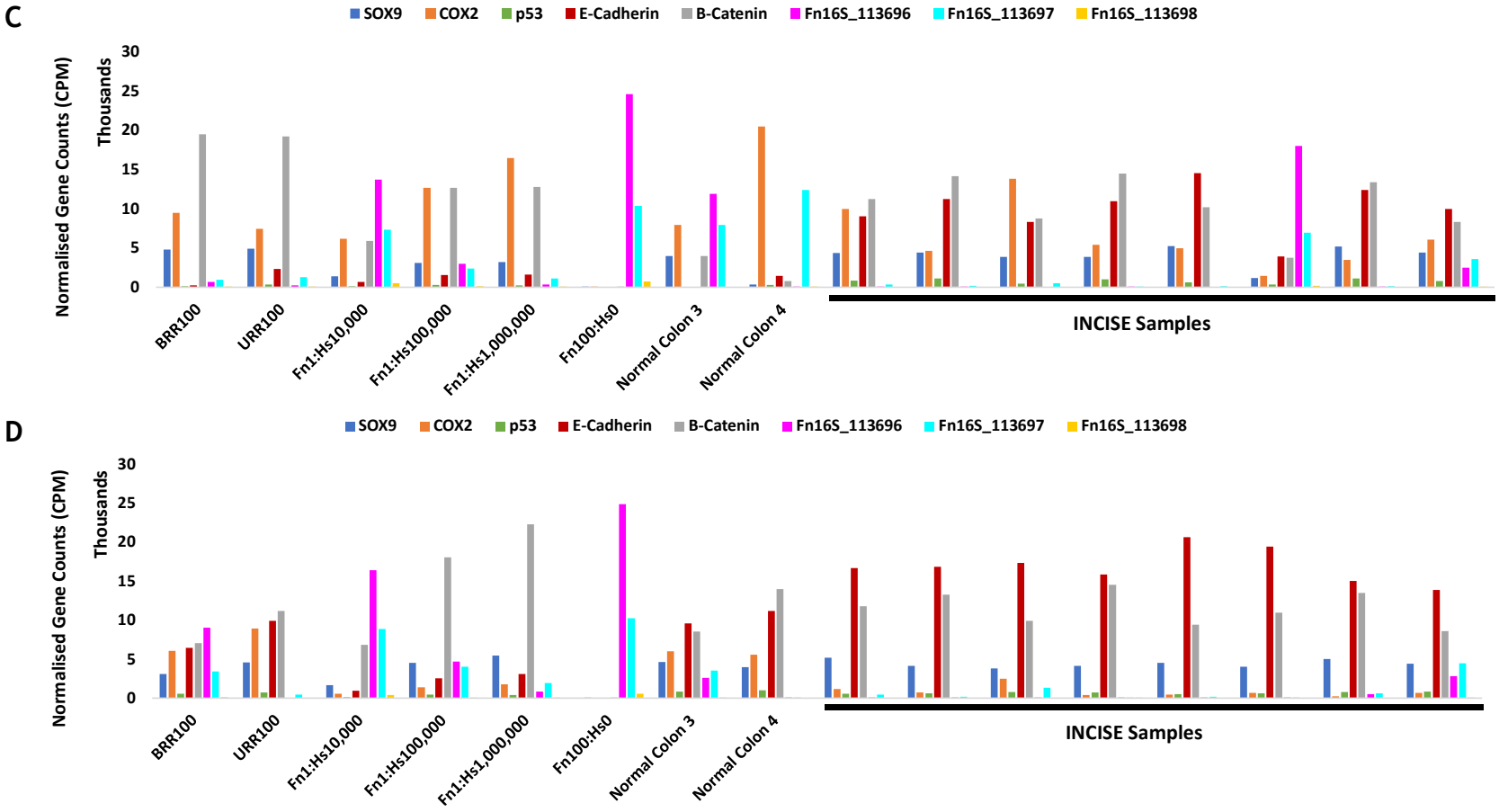


Figure 3.4. Optimisation of COX2 Probe Concentration in DO Pool.
In order to allow other signals to be visible, the COX2 probe within the DO pool was **[A]** completely removed (0nM COX2 to 3nM other probes), **[B]** reduced to 3% of the DO pool (0.09nM COX2 to 3nM other probes), **[C]** reduced to 1% of the DO pool (0.03nM COX2 to 3nM other probes), and **[D]** reduced to 0.3% of the DO pool (0.009nM COX2 to 3nM other probes). Controls (BRR100 through Fn100:Hs0) were run in duplicates. All other data points were obtained from a single sample run. BRR: Brain Reference RNA. URR: Universal Reference RNA. CPM: Counts per Million. Fn: *Fusobacterium nucleatum*. Hs: *Homo sapiens*. INCISE Samples n=8.

3.2.2 TempO-Seq™ on INCISE Lysates

After validating the designed probes' specificity and sensitivity to their targets, and after optimization of the DO pool, TempO-Seq™ was performed on 200 INCISE lysates. Of these, one failed the PCR step, and one had no clinical data, thus both were removed from the analysis, leaving 198 valid samples.

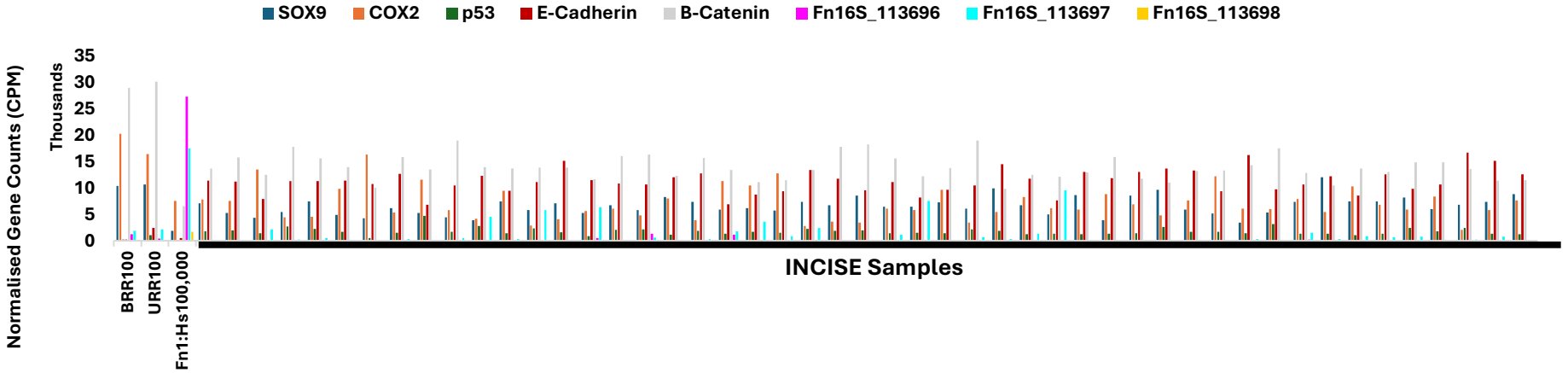
As shown in Figure 3.5, all human probes were detectable, to varying degrees, suggesting the TempO-Seq™ run was successful. The three *F. nucleatum* probes were detectable in the bacterial-human RNA control and displayed a strong signal in some of the INCISE lysates. To determine which samples were to be considered positive and negative for *F. nucleatum* (*F. nucleatum*+ and *F. nucleatum*-, respectively), thresholds (Table 3.1) were calculated for each bacterial probe from the averaged read counts plus the standard deviation of the human reference RNA controls (2.2.5.10).

Table 3.1. Summary of Thresholds for F. nucleatum TempO-Seq™ Probes

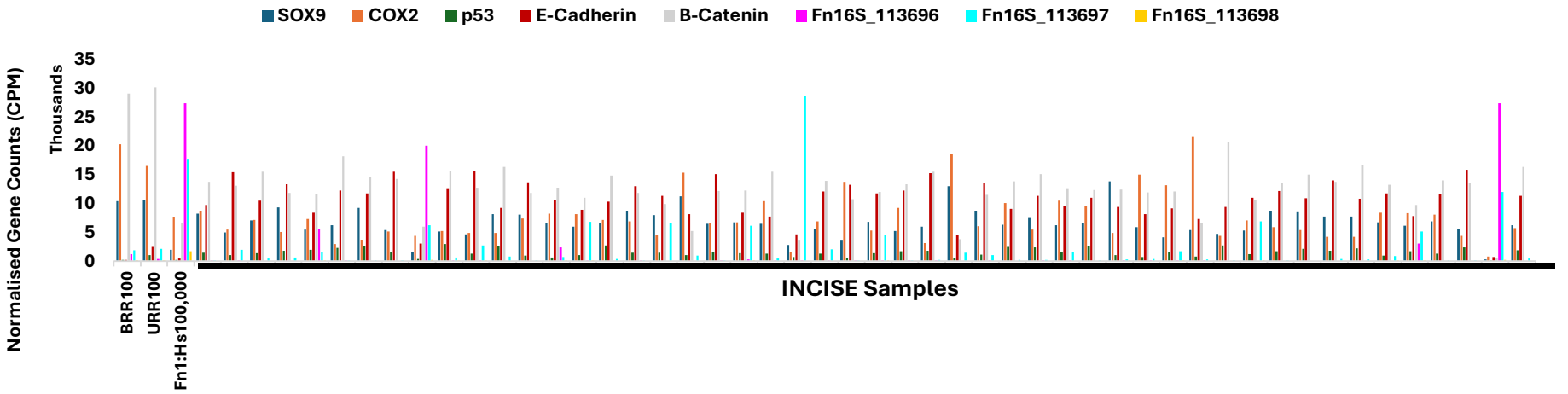
	Magenta Probe	Cyan Probe	Gold Probe
Threshold	1572.237	3104.139	39.9422

In total, 9% (n=17) of INCISE samples were above the threshold for the magenta probe, 24% (n=47) were above the threshold for the cyan probe, and 11% (n=21) were above the threshold for the gold probe (Figure 3.6). Samples that were above the threshold for any two probes were considered *F. nucleatum*+. 7% (n=13) of samples were above the threshold for both the magenta and cyan probes, 9% (n=18) of samples were above the threshold for both the cyan and gold probes, and 6% (n=11) of samples were above the threshold for both the magenta and gold probes. 21% (n=42) of samples were positive for any two probes. 5% (n=10) of samples were above the threshold for all three probes.

A


SOX9
COX2
p53
E-Cadherin
B-Catenin
Fn16S_113696
Fn16S_113697
Fn16S_113698
Normalised Gene Counts (CPM)
Thousands
35
30
25
20
15
10
5
0
BRR100
URR100
Fn1:Hs100,000
INCISE Samples

B


SOX9
COX2
p53
E-Cadherin
B-Catenin
Fn16S_113696
Fn16S_113697
Fn16S_113698
Normalised Gene Counts (CPM)
Thousands
35
30
25
20
15
10
5
0
BRR100
URR100
Fn1:Hs100,000
INCISE Samples

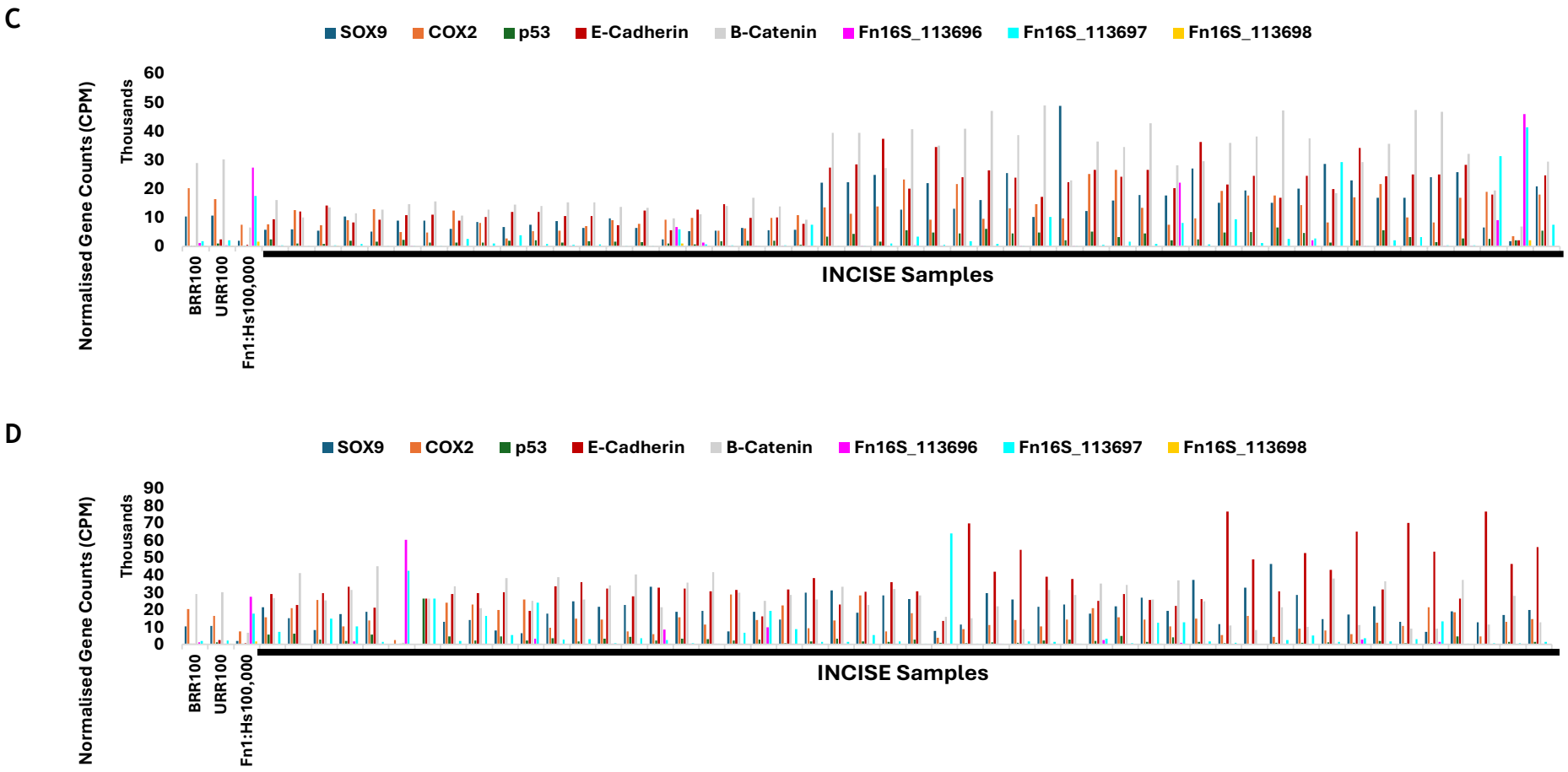
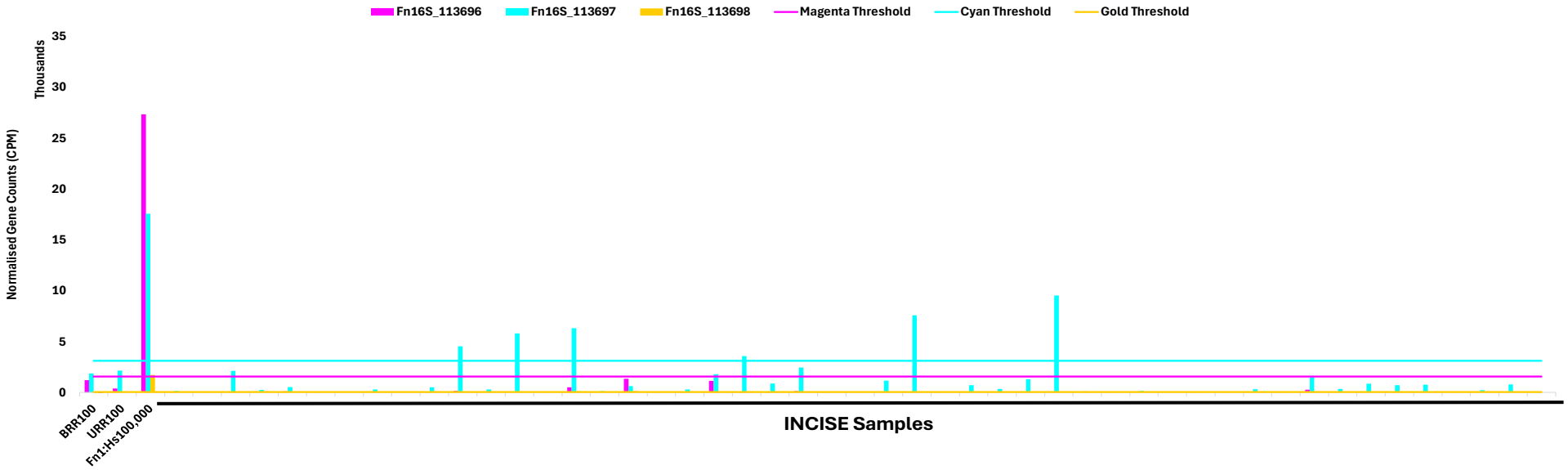
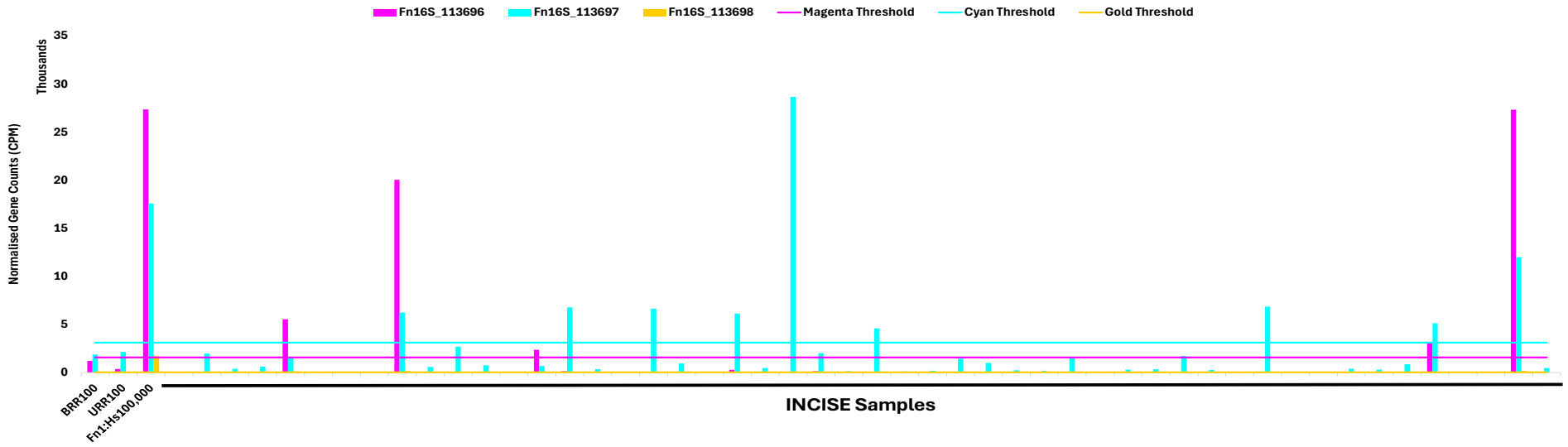


Figure 3.5. TempO-Seq™ on INCISE Samples.
Human probes and *F. nucleatum* probes were successfully used for TempO-Seq™ on 198 INCISE samples. Due to the large number of samples, a maximum of 50 samples is shown per graph. **[A]** Samples 1-50 with BRR100, URR100, and Fn1:Hs100,000 controls. **[B]** Samples 51-101 with BRR100, URR100, and Fn1:Hs100,000 controls. **[C]** Samples 102-151 with BRR100, URR100, and Fn1:Hs100,000 controls. **[D]** Samples 152-198 with BRR100, URR100, and Fn1:Hs100,000 controls. BRR: Brain Reference RNA. URR: Universal Reference RNA. CPM: Counts per Million. Fn: *Fusobacterium nucleatum*. Hs: *Homo sapiens*. BRR100, URR100, and Fn1:Hs100,000 were run in technical duplicates. INCISE Samples (n=198) were run once.

A


Fn16S_113696
Fn16S_113697
Fn16S_113698
Magenta Threshold
Cyan Threshold
Gold Threshold
Thousands
Normalised Gene Counts (CPM)
35
30
25
20
15
10
5
0
BRR100
URR100
Fn1:Hs100,000
INCISE Samples

B


Fn16S_113696
Fn16S_113697
Fn16S_113698
Magenta Threshold
Cyan Threshold
Gold Threshold
Thousands
Normalised Gene Counts (CPM)
35
30
25
20
15
10
5
0
BRR100
URR100
Fn1:Hs100,000
INCISE Samples

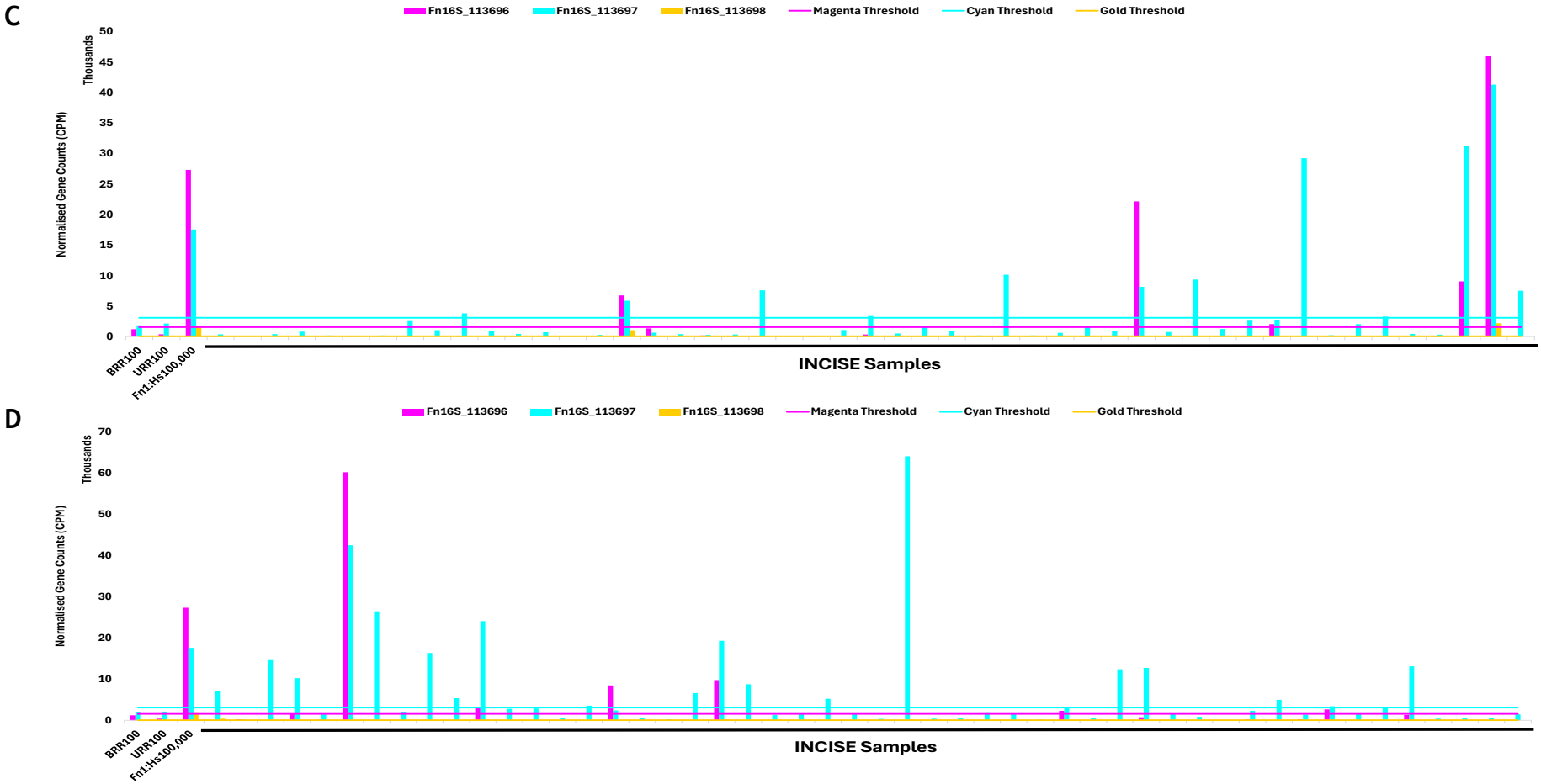


Figure 3.6. TempO-Seq™ *F. nucleatum* Probes against Thresholds.
Due to the large number of INCISE samples (n=198), a maximum of 50 samples is shown per graph. Samples crossing the horizontal threshold indicators were considered positive for that probe. **[A]** Samples 1-50 with BRR100, URR100, and Fn1:Hs100,000 controls. **[B]** Samples 51-101 with BRR100, URR100, and Fn1:Hs100,000 controls. **[C]** Samples 102-151 with BRR100, URR100, and Fn1:Hs100,000 controls. **[D]** Samples 152-198 with BRR100, URR100, and Fn1:Hs100,000 controls. BRR: Brain Reference RNA. URR: Universal Reference RNA. CPM: Counts per Million. Fn: *Fusobacterium nucleatum*. Hs: *Homo sapiens*. BRR100, URR100, and Fn1:Hs100,000 were run in technical duplicates. INCISE Samples were run once.

3.3 Validation of TempO-Seq™ by RNAscope®

An RNA *in situ* hybridisation technique, RNAscope® was designed for use on FFPE tissue, to detect mRNA transcripts either chromogenically for single-plex applications, or fluorescently for multi-plex purposes. RNAscope® uses a proprietary double Z probe technology, allowing the probes to bind to multiple RNA targets. Only paired Z probes can bind the signal amplification molecules, increasing the specificity of the assay, and decreasing background off-target signal (Wang *et al.*, 2012).

To validate the bacterial detection with TempO-Seq™, full section slides from ten INCISE patients included in the TempO-Seq™ experiment that were *F. nucleatum*+ were used for RNAscope® as described in 2.2.6, and *in assay* positive and negative controls included in the run (Figure 3.7D). To ensure the validity of the signal used, three probes were used for each sample, as shown in Figure 3.7. The HS-UBC universal probe was used to detect any available RNA to guarantee the presence and quality of the RNA (Figure 3.7A). A 16S bacterial RNA probe was used to detect RNA from any bacteria (Figure 3.7B), and finally, a probe specific to 23S RNA from *F. nucleatum* was used to detect any *F. nucleatum* RNA on the polyp tissue (Figure 3.7C). As expected, the more universal the probe was, the clearer the signal and the higher the density of detected RNA. As the probes became more target specific, the signal became weaker and the density decreased significantly, indicating decreased levels of bacteria in general, and specifically of *F. nucleatum*.

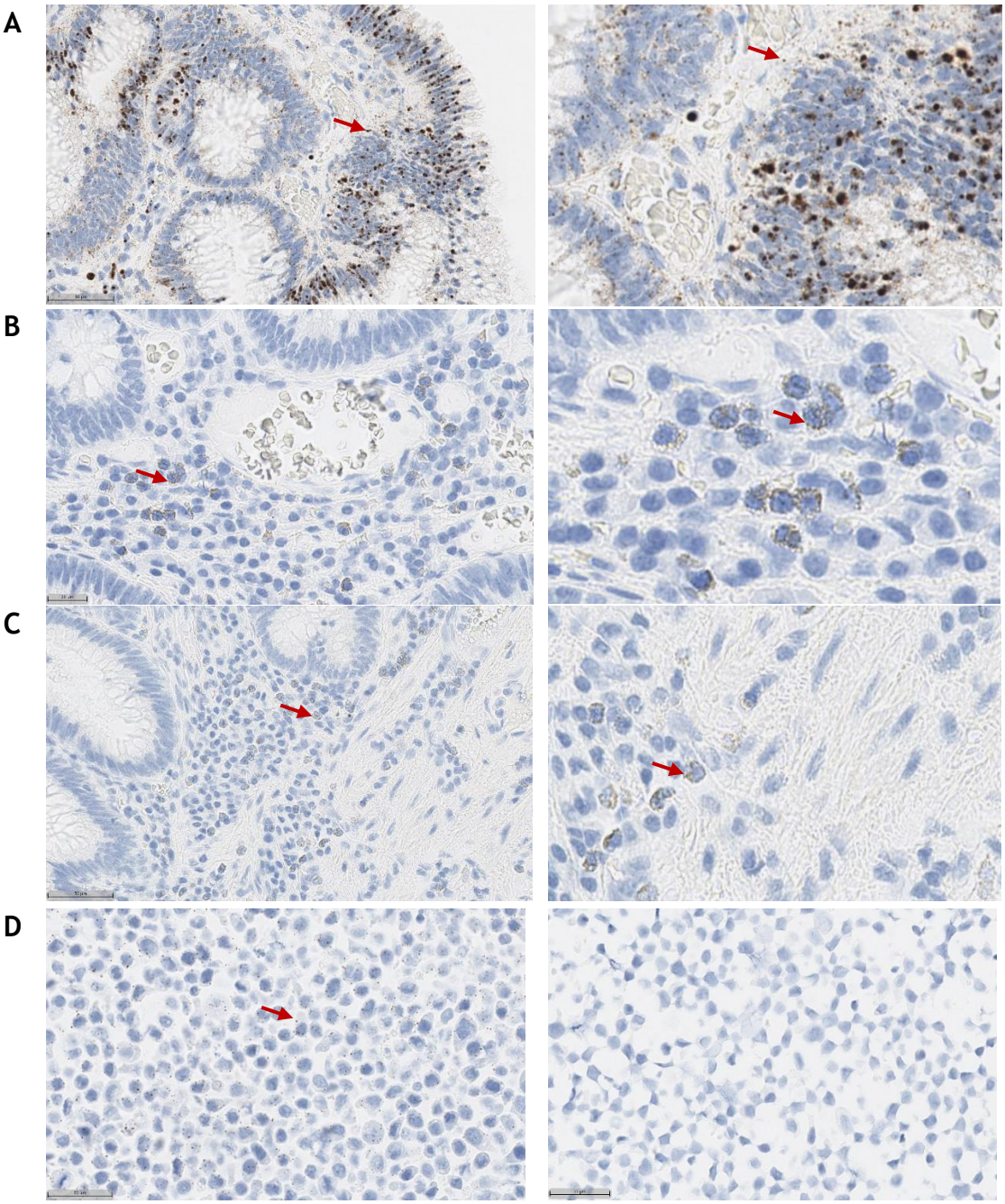


Figure 3.7. Representative Images of RNAscope® on an INCISE Patient.
[A: Left] HS-UBC universal probe for chromogenic RNAscope® . **[A: Right]** Zoomed in area from A: Left. **[B: Left]** 16S RNA bacterial probe for chromogenic RNAscope® . **[B: Right]** Zoomed in area from B: Left. **[C: Left]** 23S RNA *F. nucleatum* probe for chromogenic RNAscope® . **[C: Right]** Zoomed in area from C: Left. **[D: Left]** Positive Control: HS-TBP probe for chromogenic RNAscope® on HeLa cell pellets (supplied by ACD, a biotechne company). **[D: Right]** Negative Control: DapB probe for chromogenic RNAscope® on HeLa cell pellets (supplied by ACD, a biotechne company). **[General Key]:** Brown dots are positive RNAscope® signal. Larger brown dots indicate mRNA aggregates. Red arrows point at positive RNAscope® signal. While only 1 INCISE patient is represented here, this was done for 10 *F. nucleatum*+ INCISE patients.

The weak signal and low density of the 16S and *F. nucleatum* probes made digital signal quantification extremely difficult. Although attempts were made using specialist software HALO (Indica Labs) and Visiopharm®, the brown signal of the RNAscope® probes was too weak and the positive RNA transcripts too small and too few to be differentiated from the background negative tissue. While this unfortunately meant that further quantitative analysis was not possible, the number of cells associated with *F. nucleatum* was visually determined. A cutoff of 16 or less *F.* nucleatum associated cells was qualitatively used to dichotomise the RNAscope® patients to validate the TempO-Seq™ groupings. Of the five patients who had less than 16 *F. nucleatum*-associated cells by RNAscope®, three were considered *F. nucleatum*- by TempO-Seq™, and of the five patients who had more than 16 *F. nucleatum*-associated cells by RNAscope®, four were considered *F. nucleatum*+ by TempO-Seq™.

3.4 Mutational Profile of *F. nucleatum* Patient Subset

Following the detection of *F. nucleatum*, the mutational profile of the *F. nucleatum*+ and *F. nucleatum*- patients assigned by TempO-Seq™ was explored to reveal any underlying mutations. Of the 198 patients with *F. nucleatum* status, mutational data was available for 120. The top 10 most mutated genes were identified, and somatic interaction analysis was conducted to reveal any co-occurring or mutually exclusive gene pairs.

3.4.1 *F. nucleatum* Negative

Of the *F. nucleatum*- patients, 99% (n=108) had mutations (Figure 3.8A). *APC* (91%) and *KRAS* (41%) were the top two most mutated genes with "multi-hit" and "missense" mutations, respectively. *HLA-A* (30%, third) and *ZFHX3* (23%, fourth) had mainly "missense" mutations. "Frameshift" mutations (insertions and deletions) were the most common alterations in *SOX9* (22%), the fifth most mutated gene in this group. Mutated in 21% of patients in this subset, *ARID1A* (sixth) was mostly mutated by "frameshift insertions", while *HLA-B* (seventh), mostly had "missense" alterations. Both *FBXW7* (eighth) and *TCF7L2* (ninth) were mutated in 18% of the *F. nucleatum*- patients. While "missense" mutations were most common for *FBXW7*, *TCF7L2* was equally affected by both "missense" mutations and "frameshift insertions". *KMT2C* was mutated in 17% of patients mostly through "missense" mutations. While *HLA-B* co-mutated with *KRAS* at an event ratio of 44% (pAdj=0.029), no other gene pairs were mutating in a co-occurring or mutually exclusive manner (Figure 3.8C, Table 3.2).

3.4.2 *F. nucleatum* Positive

Although 100% of *F. nucleatum*+ patients had mutations, the number of patients in this subset is very low (n=11). This led to some different mutational trends than previously described. Although *APC* is still the top mutated gene (73%) with mainly multi-hit events, *KRAS* is no longer among the top ten most mutated genes (Figure 3.8B). *HLA-A* (second), *HLA-C* (third), and *KMT2C* (fourth), were all mutated in 36% of patients. Both *HLA-A* and *KMT2C* had mainly "multi-hit" mutation events, while "missense" mutations were more common in *HLA-C*. The rest of the top 10 most mutated genes were mutated in 27% of the patients within this group. Protein kinase anchoring and signalling complex assembly

protein, *AKAP9* (fifth), and DNA repair protein, *FANCD2* (sixth), both had mostly "missense" mutations. *FAT1* (seventh), *MGA* (eighth), and *ZFHX3* (tenth) did not have a dominant mutation type, while *MSH3* (ninth) had mostly "multi-hit" events. Somatic interactions analysis revealed no gene pairs that were co-mutating or mutually exclusive in their mutation pattern (Figure 3.8D, Table 3.2).

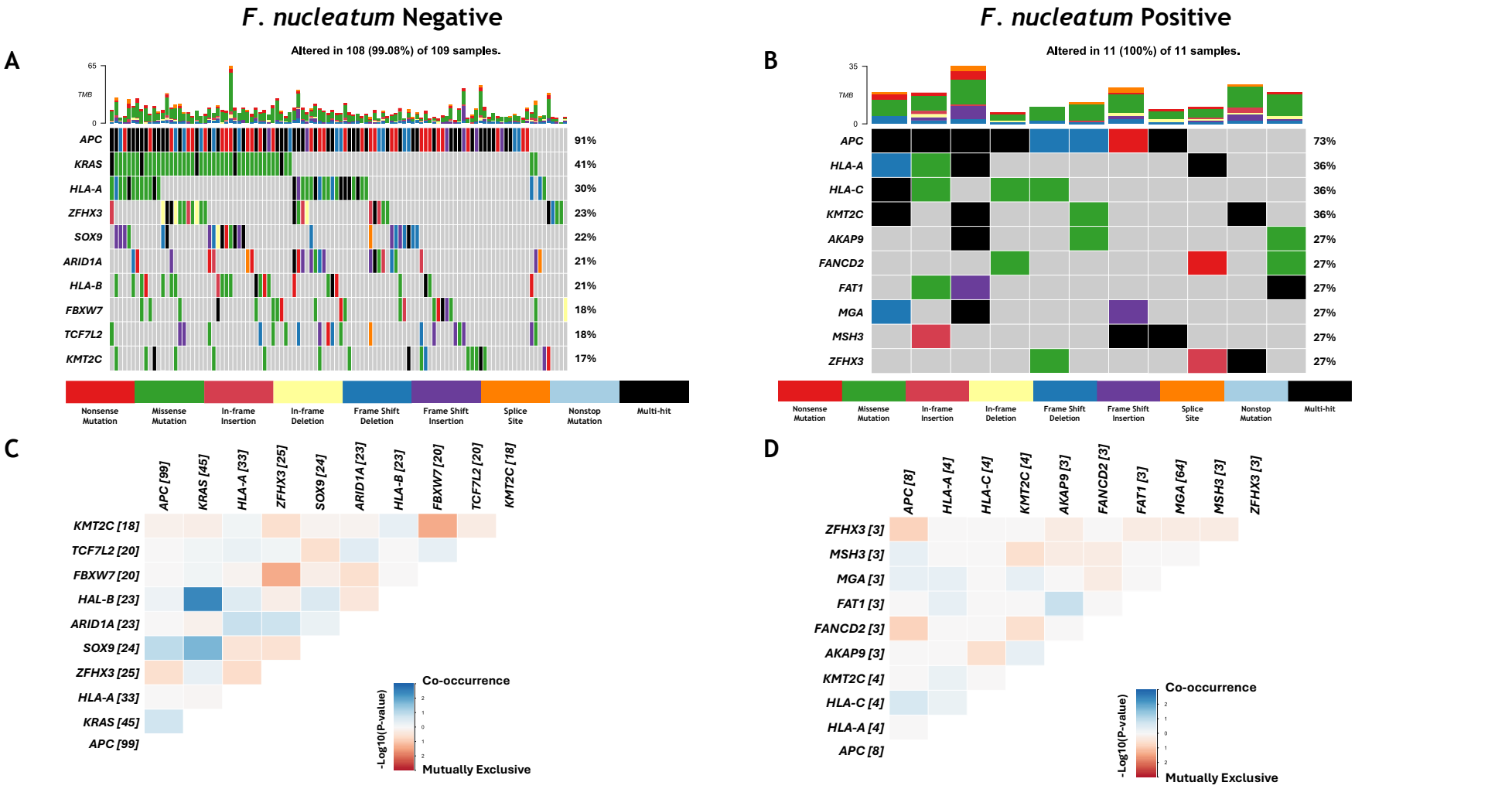


Figure 3.8. Comparison of the Mutational Landscape of Adenomas by *F. nucleatum* Status.
Oncoplots depicting the top 10 most mutated genes in **[A]** *F. nucleatum*- patients, and **[B]** *F. nucleatum*+ patients. Genes are ranked by the percentage of patients in which the gene is affected. “Multi-hit” denotes genes with multiple mutations in a sample. Somatic interactions depicting co-occurring or mutually exclusive gene pairs as statistically true events in **[C]** *F. nucleatum*- patients, and **[D]** *F. nucleatum*+ patients. N=120.

Table 3.2. Top Five Gene Pair Mutations for Adenomas Stratified by *F. nucleatum* Status

Rank	Gene Pair	pAdj	Significant	Event	Event Ratio
***F. nucleatum* Negative**					
1	*HLA-B/KRAS*	0.0300	Yes	Co-occurrence	16/36
2	*KRAS/SOX9*	0.1452	No	Co-occurrence	15/39
3	*FBXW7/ZFHX3*	0.2206	No	Mutually Exclusive	1/43
4	*FBXW7/KMT2C*	0.2206	No	Mutually Exclusive	0/38
5	*APC/SOX9*	0.5668	No	Co-occurrence	24/75
***F. nucleatum* Positive**					
1	*APC/FANCD2*	0.9470	No	Mutually Exclusive	1/9
2	*APC/ZFHX3*	0.9470	No	Mutually Exclusive	1/9
3	*AKAP9/FAT1*	0.9470	No	Co-occurrence	2/2
4	*APC/HLA-C*	0.9818	No	Co-occurrence	4/4
5	*AKAP9/HLA-C*	0.9818	No	Mutually Exclusive	0/7

Gene pairs are ranked by most significant pAdj. From total n=120.

3.5 Bulk Transcriptomics Analysis of *F. nucleatum*+ Patients

Although numbers were low for the *F. nucleatum*+ patients for which mutational data was available, the mutations seen in immune related genes like *HLA-A* and *HLA-B* were of interest, and so the transcriptomic profile of these patients was of interest to investigate the biological processes underlying the host-bacterial interaction.

Of the 198 patients with TempO-Seq™-based *F. nucleatum* status, bulk transcriptomic data was available for 190 of them, all of which were included in the analysis. The patients were grouped into *F. nucleatum*- (n=170) and *F. nucleatum*+ (n=20) as described in 3.2.2. RStudio was used for all subsequent analysis, and the *F. nucleatum*+ group was compared to the *F. nucleatum*- group.

3.5.1 Principal Component Analysis and Differential Gene Expression

To assess the transcriptomic differences between the *F. nucleatum*- and *F. nucleatum*+ patients, PCA was used. PCA did not show distinct data clusters between the two groups (Figure 3.9A). When applying the cut-off criteria of

pAdj<0.05 and Log2 Fold-Change of <-0.5 and >0.5, DESeq2 analysis revealed 29 upregulated genes and 38 downregulated genes for a total of 67 differentially expressed genes, in the *F. nucleatum*+ group. The top 10 of each are displayed in Figure 3.9B.

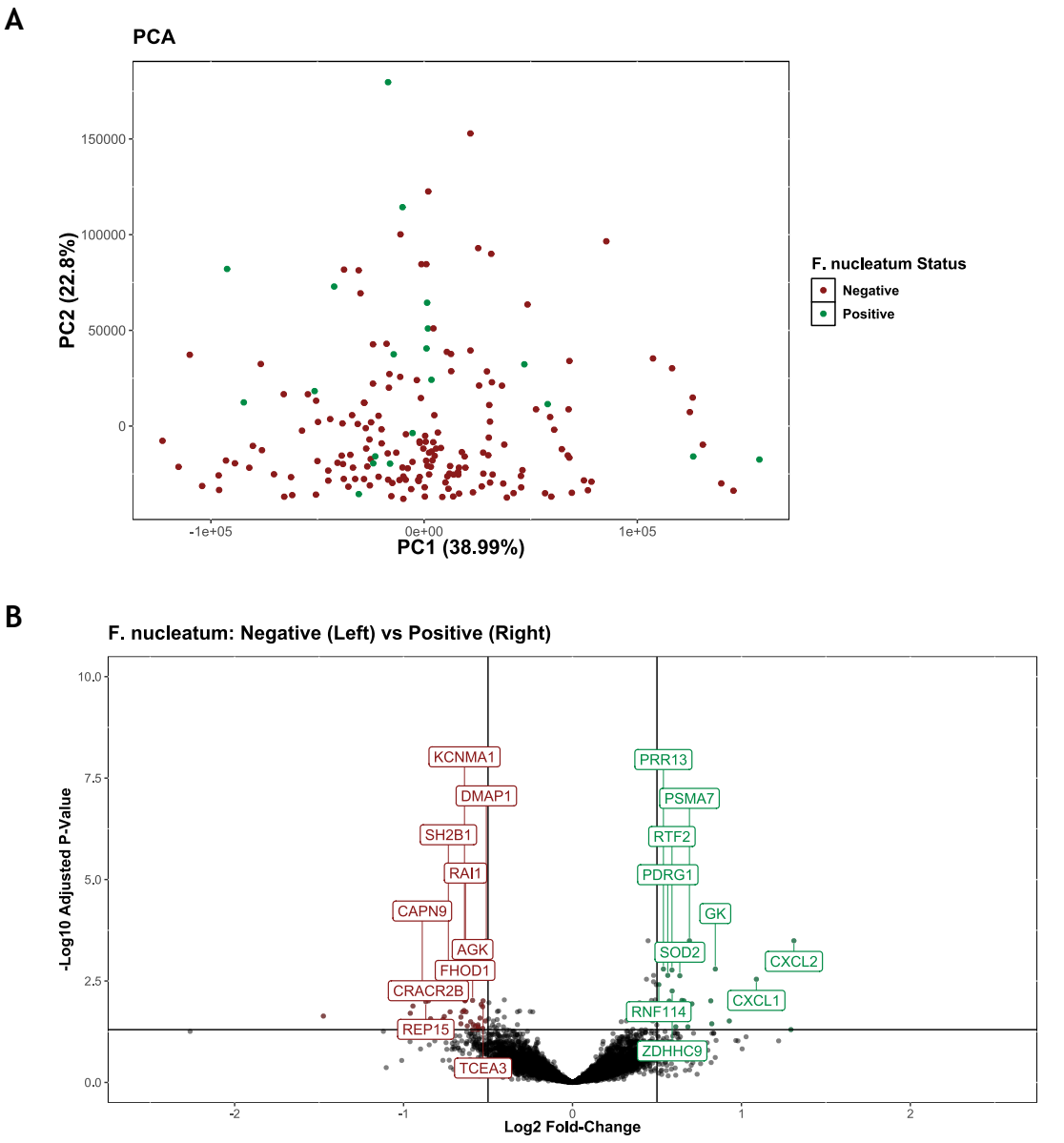


Figure 3.9. Transcriptomic Analysis Comparing *F. nucleatum*+ to *F. nucleatum*- Status in INCISE Polyp Patients.
[A] Principal Component Analysis (PCA) of *F. nucleatum* status groups. **[B]** Volcano plot of differentially expressed genes (DEGs). The *F. nucleatum*+ group (n=20) was compared to the *F. nucleatum*- group (n=170). The top 10 most downregulated genes (red) and the top 10 most upregulated genes (green) in the *F. nucleatum*+ group are shown in boxes. Log2 Fold-change <-0.5 and >0.5. pAdj <0.05. Total n=190.

To visualise the expression pattern of each DEG across the two groups, a heatmap was generated (Figure 3.10A), revealing a pattern consistent with the regulation direction observed in the volcano plot, which was verified by a

heatmap focussing only on the top 10 upregulated and downregulated genes (Figure 3.10B).

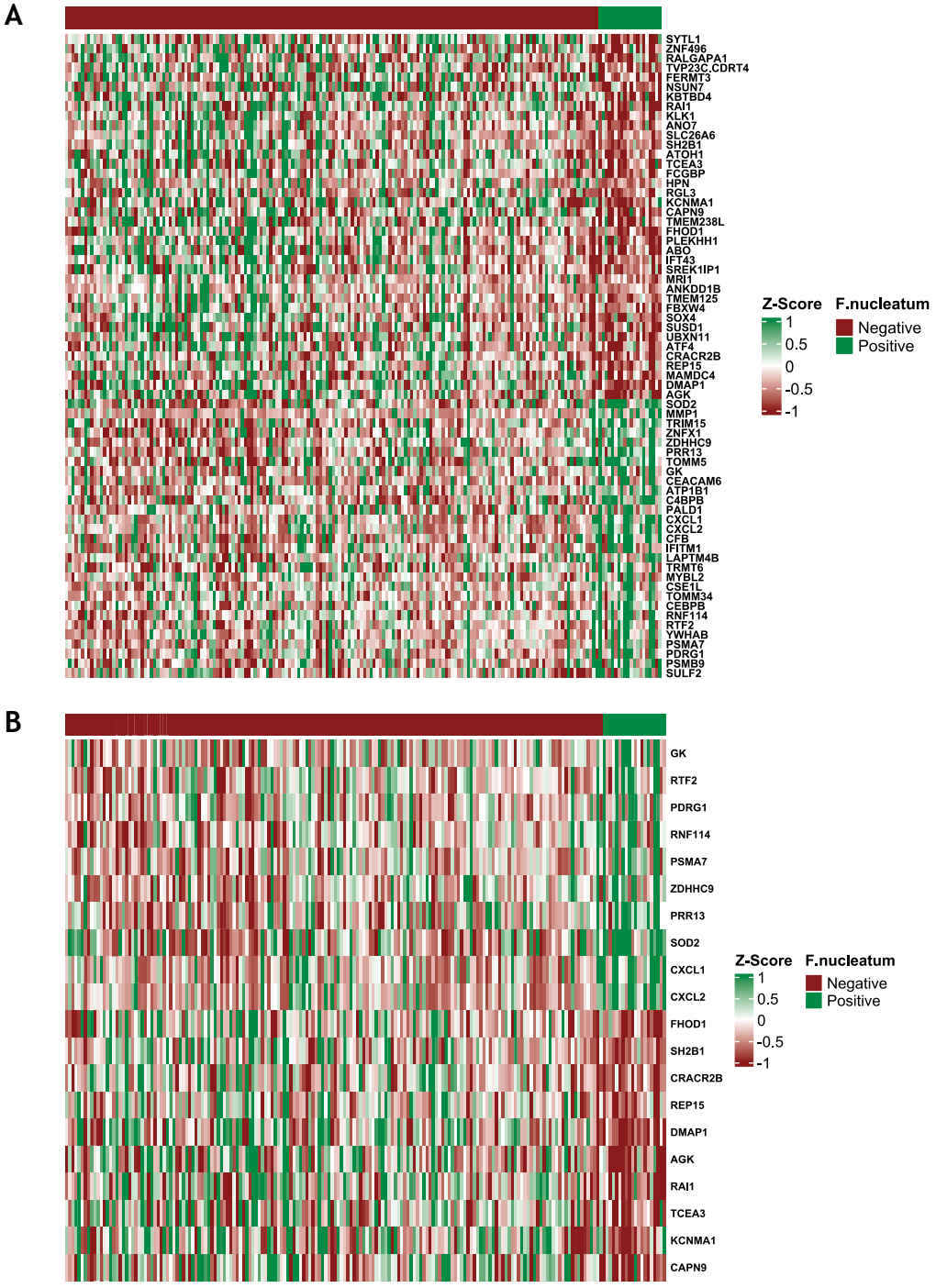


Figure 3.10. Heatmaps Visualising Gene Expression Across the *F. nucleatum*+ and *F. nucleatum*- Groups.
[A] Heatmap of all 67 DEGs. **[B]** Heatmap of top 10 most upregulated and top 10 most downregulated genes. Total n=190, *F. nucleatum*+ group n=20, and *F. nucleatum*- group n=170.

Violin plots were generated to display the expression pattern of individual genes between *F. nucleatum* status groups. As seen in Figure 3.11, the expression pattern remained consistent.

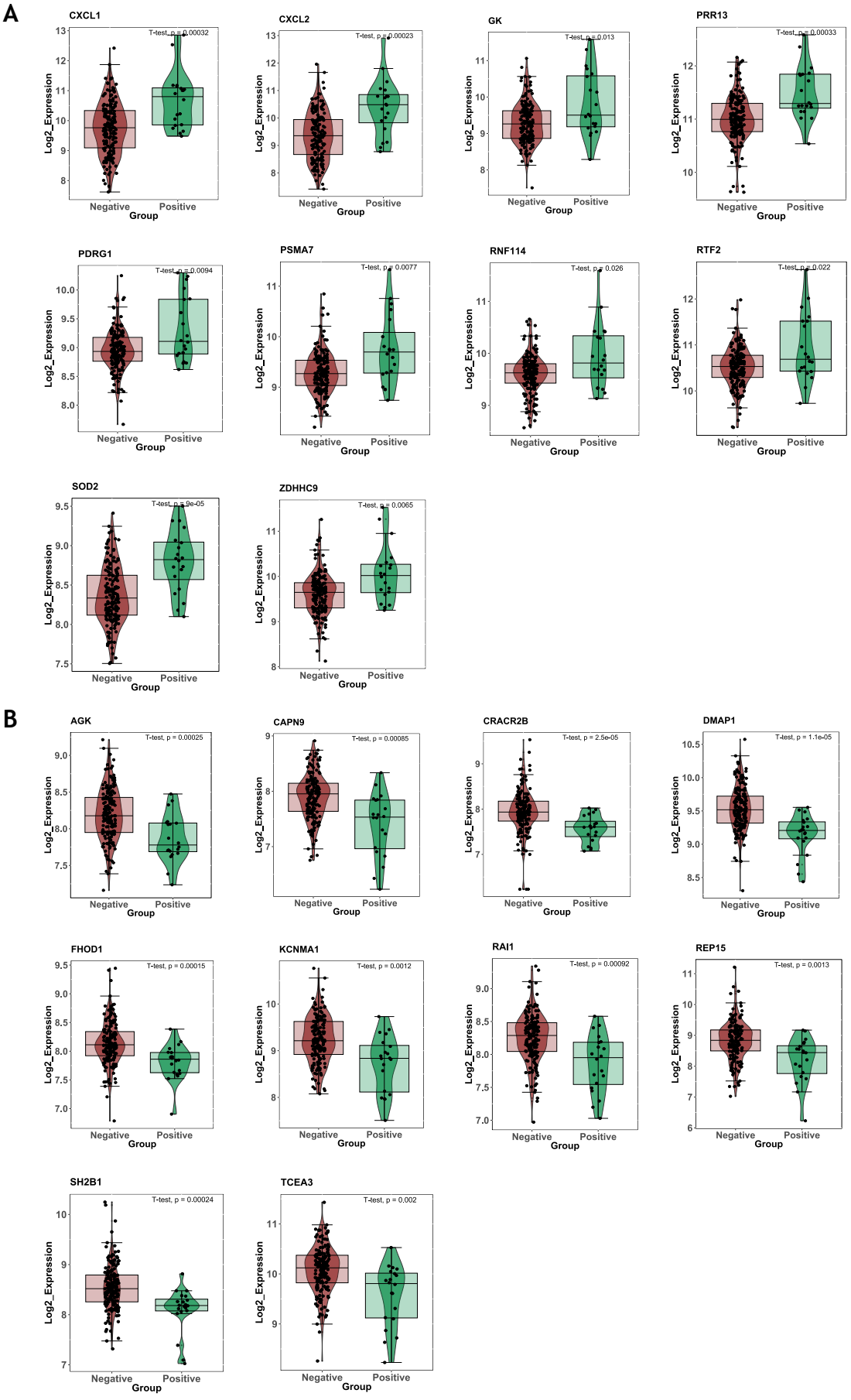


Figure 3.11. Violin Plots of Top 20 DEGs.
[A] The top 10 most upregulated genes in the *F. nucleatum*+ group. **[B]** The top 10 most downregulated genes in the *F. nucleatum*+ group. Total n=190, *F. nucleatum*+ group n=20, and *F. nucleatum*- group n=170.

3.5.2 Gene Set Enrichment Analysis

To understand the molecular signatures enriched within the *F. nucleatum*+ samples, pairwise GSEA was performed using three publicly available biological resources: GO, MSigDb, and the KEGG. The following results all had a pAdj<0.05 and a normalised enrichment score (NES) <-1 or >1.

3.5.2.1 Gene Ontology

73 GO terms were enriched. A dot plot (Figure 3.12A) displayed the top 10 activated and suppressed GO terms in the *F. nucleatum*+ group. Four terms were suppressed: “sperm ejaculation”, “neurotransmitter uptake”, “monoamine transport”, and “neurotransmitter transport”. Activated terms that suggest the innate immune response such as “neutrophil migration”, “granulocyte migration”, “leukocyte chemotaxis”, “myeloid leukocyte migration”, and “leukocyte migration” were among the top 10. Other activated terms such as “cytokine activity” and “chemotaxis” suggest a more comprehensive immune response signature. Finally, a microbial-related term “response to bacterium” was also among the top 10 activated terms.

A bar plot (Figure 3.12B) of the top 10 positively and negatively enriched GO terms, ordered by the NES shows the same negatively enriched terms as those suppressed in the dot plot. Additional immunity-related terms are positively enriched, such as “regulation of T-helper 1 type immune response”, “response to chemokine”, and “cellular response to chemokine”. Further analysis revealed multiple positively enriched terms that meet the cut-off criteria but were ranked below the top 10, that confer a signature of microbial infection recognition and response (Figure 3.12C).

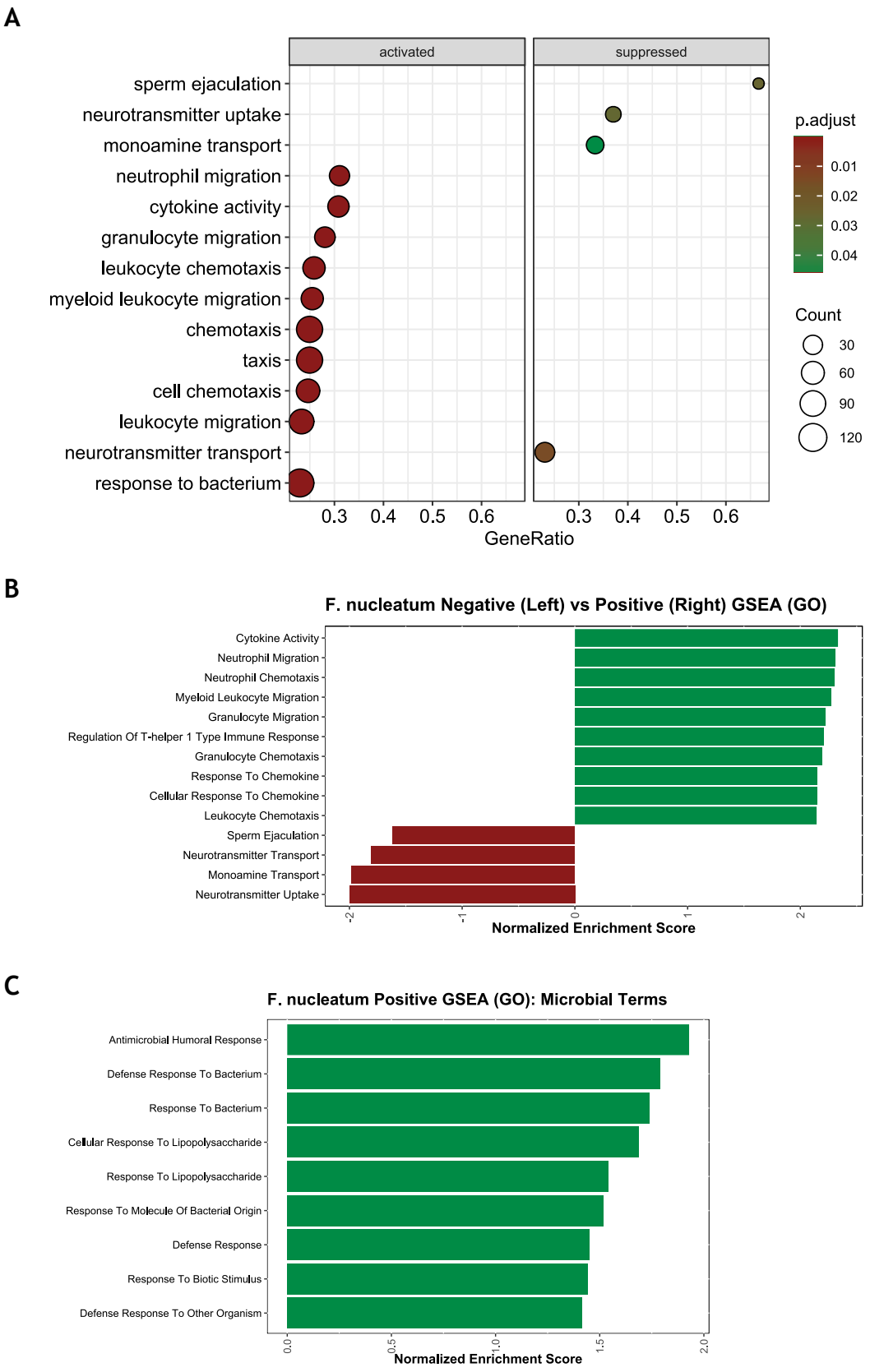


Figure 3.12. GO GSEA Summary.
[A] Dot plot of top 10 activated (left) and supressed (right) GO terms ranked by gene ratio. The x-axis shows the calculated gene ratio of input genes corresponding to the gene set in each term. Dot size refers to gene counts while colour refers to pAdj. **[B]** Bar plot of top 10 positively enriched (green) and top 10 negatively enriched (red) GO terms ranked by NES. **[C]** Bar plot of statistically significant positively enriched microbial GO terms. These were ranked lower than 10 and did not appear in B. The *F. nucleatum*+ group was compared to the *F. nucleatum*- group. Total n=190, *F. nucleatum*+ group n=20, and *F. nucleatum*- group n=170.

As many of the enriched terms have similar functions, leading edge genes were common across them. UpSet plots were produced to visualise the interconnectivity between the enriched terms. Among the ten positively enriched terms, two genes were shared, *CCL19* and *XCL1*. Five genes, *CCR7*, *CXCR1*, *IL1B*, *IL23A*, and *RIPOR2* were common between 8 terms, and one gene, *TGFB2,* was shared across seven terms. As all of these terms pertain to immune functions, it was not surprising that fifteen genes were shared across nine terms (Figure 3.13).

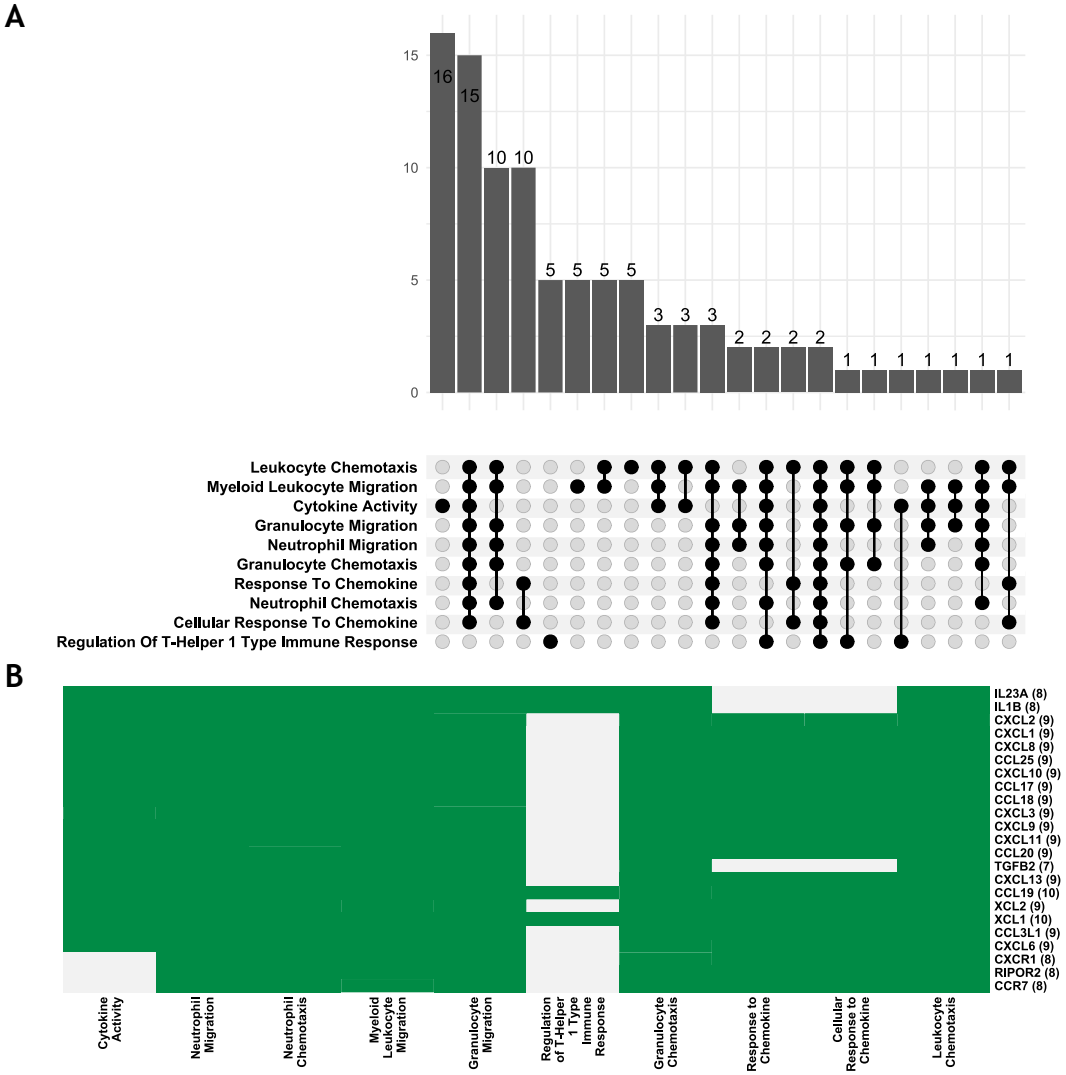


Figure 3.13. Overlapping Leading Edge Genes in Positively Enriched GO Terms. **[A]** UpSet plot showing number of overlapping genes as bars across intersecting GO terms (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least seven GO terms.

One gene was shared across all four negatively enriched terms (Figure 3.14A), *SLC6A4*, while three genes, *SLC22A1*, *GPM6B*, and *SLC6A2* were shared across three terms pertaining to neurotransmitter receptor functions (Figure 3.14B).

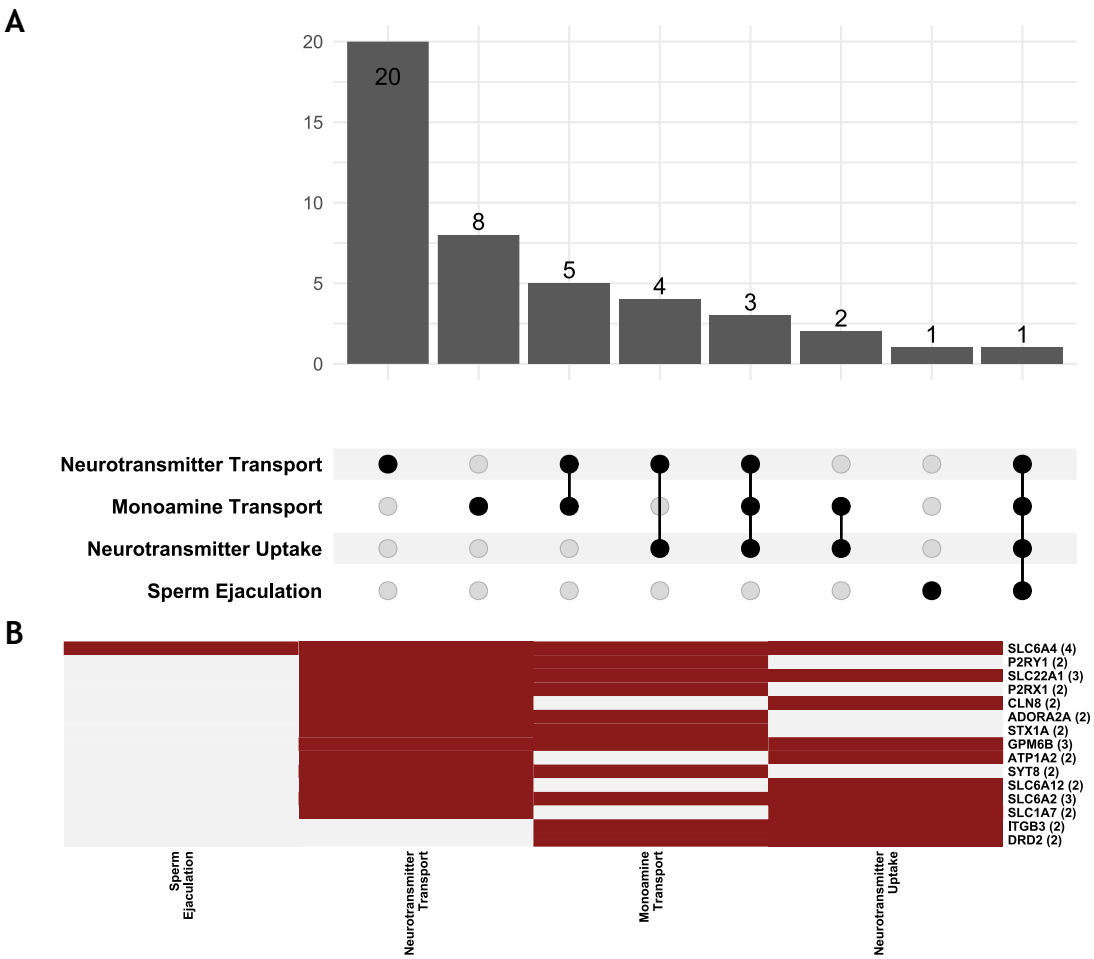


Figure 3.14. Overlapping Leading Edge Genes in Negatively Enriched GO Terms. **[A]** UpSet plot showing number of overlapping genes as bars across intersecting GO terms (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least two GO terms.

The nine positively enriched microbial terms shared three genes, *CTSG*, *CXCL13*, and *LTF*. Of the 16 genes common between eight terms, eight were members of the CXC chemokine family, three were interleukins (*IL10*, *IL1B*, and *IL6*), as well as *NOS2*, *RIPK2*, *S100A9*, *SERPINE1*, and *SLC11A1*. Of the 16 genes shared between seven of the microbial terms, non-immune genes such as *ARID5A*, *IRAK3*, *MMP3*, and *WNT5A* stand out (Figure 3.15).

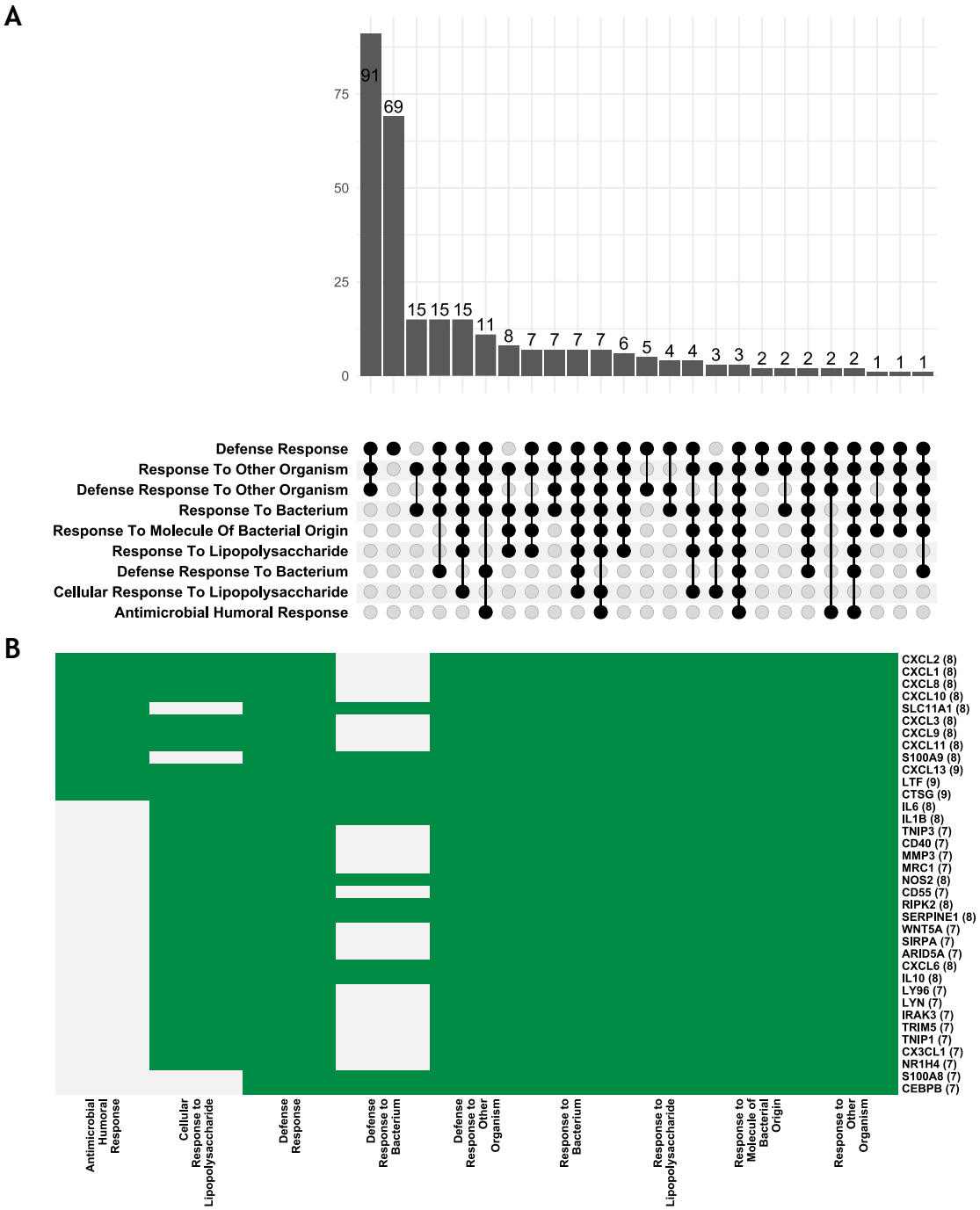


Figure 3.15 Overlapping Leading Edge Genes in Positively Enriched Microbial GO Terms.
[A] UpSet plot showing number of overlapping genes as bars across intersecting GO terms (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least seven GO terms.

GSEA enrichment score plots, that represent the position of enriched genes within GO terms of interest, reveal that the positively enriched terms involved in immune response are highly ranked (Figure 3.16A), while genes from negatively enriched terms, especially “sperm ejaculation”, were ranked lower (Figure 3.16B). Similarly, genes from the selected microbial terms that were positively enriched were all ranked highly (Figure 3.17).

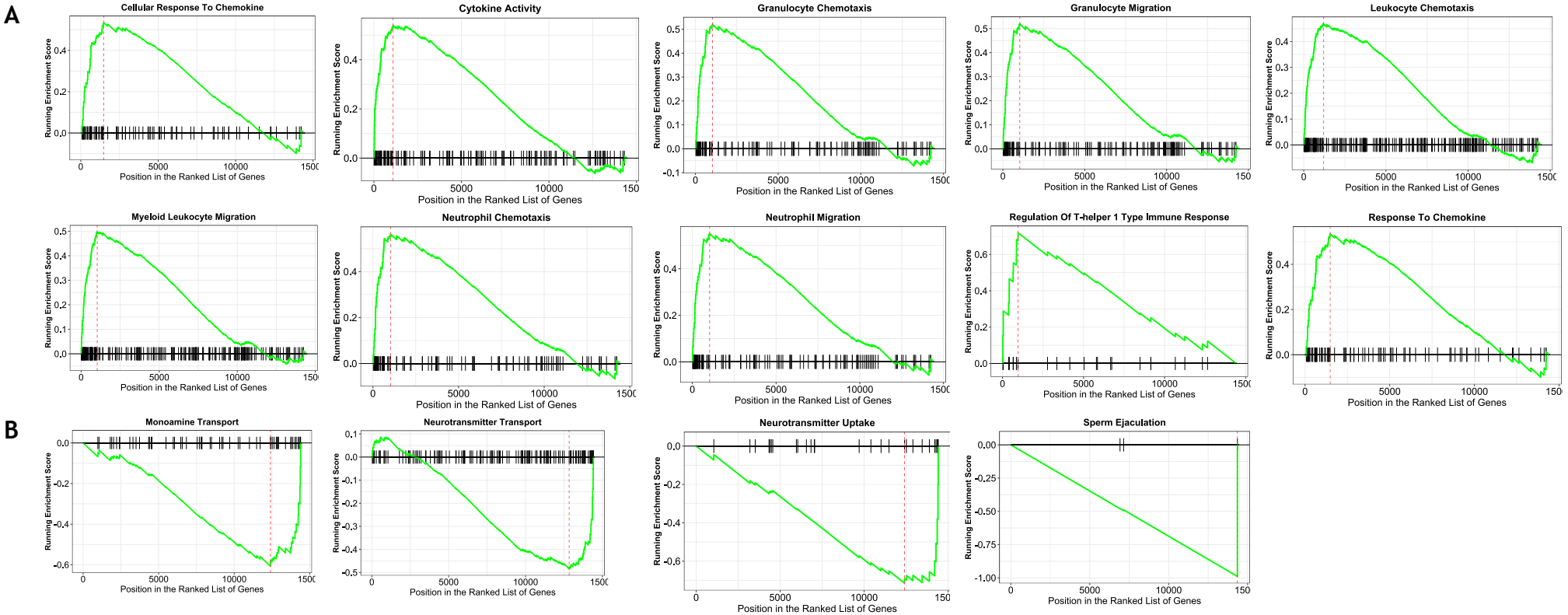


Figure 3.16. GSEA Plots of Top Enriched GO Terms.
[A] Positively enriched GO Terms. **[B]** Negatively enriched GO terms. The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The vertical red line represents the point where gene enrichment is most positive or negative.

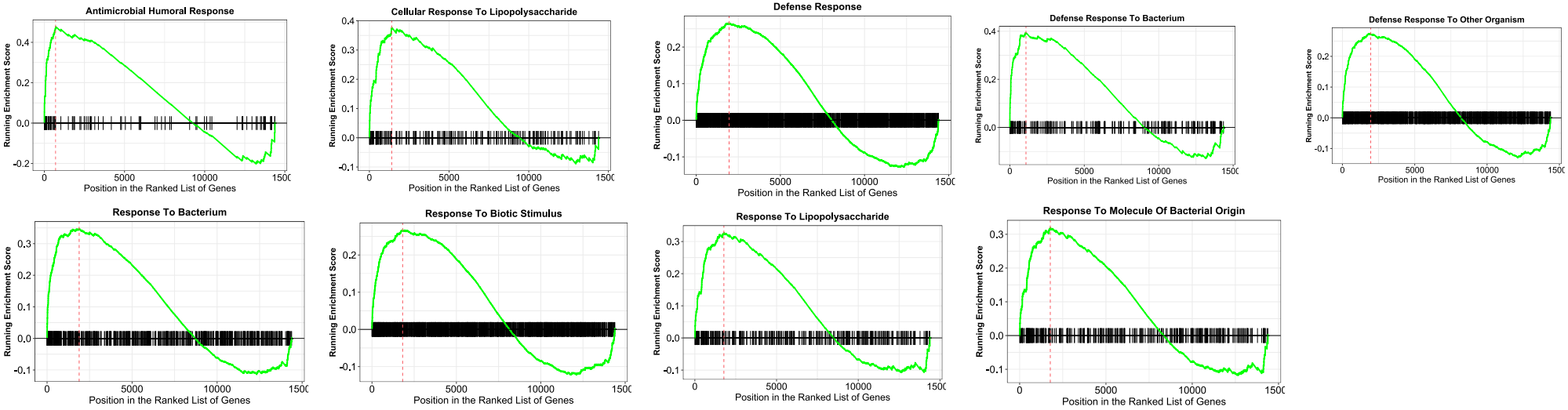


Figure 3.17. GSEA Plots of Positively Enriched Microbial GO Terms.
The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The vertical red line represents the point where gene enrichment is most positive or negative.

3.5.2.2 Molecular Signature Database

MSigDb offers gene sets from multiple organisms. The human gene set collection Hallmarks was analysed. Twenty-two gene sets had a pAdj<0.05. Figure 3.18A shows a dot plot of the top ten activated and suppressed gene sets. Of the top ten activated gene sets, six were categorised as “immune”: “Interferon Alpha Response”, “Inflammatory Response”, “Allograft Rejection”, “Interferon Gamma Response”, “Complement”, and “IL6 JAK STAT3 Signalling” and three were categorised as “signalling”: “IL2 JAK STAT5 Signalling”, “KRAS Signalling Up”, and “TNFA Signalling via NF-KB”. “UV Response Up” was categorised as “DNA Damage”. The one suppressed gene set, “KRAS Signalling Down”, was categorised as “signalling”.

The bar plot in Figure 3.18B shows many of the same activated gene sets to be positively enriched, with the addition of “pathway” category entry, “Reactive Oxygen Species Pathway”, and “proliferation” category entry, “MYC Targets V1”. “KRAS Signalling Down” was the only negatively enriched gene set.

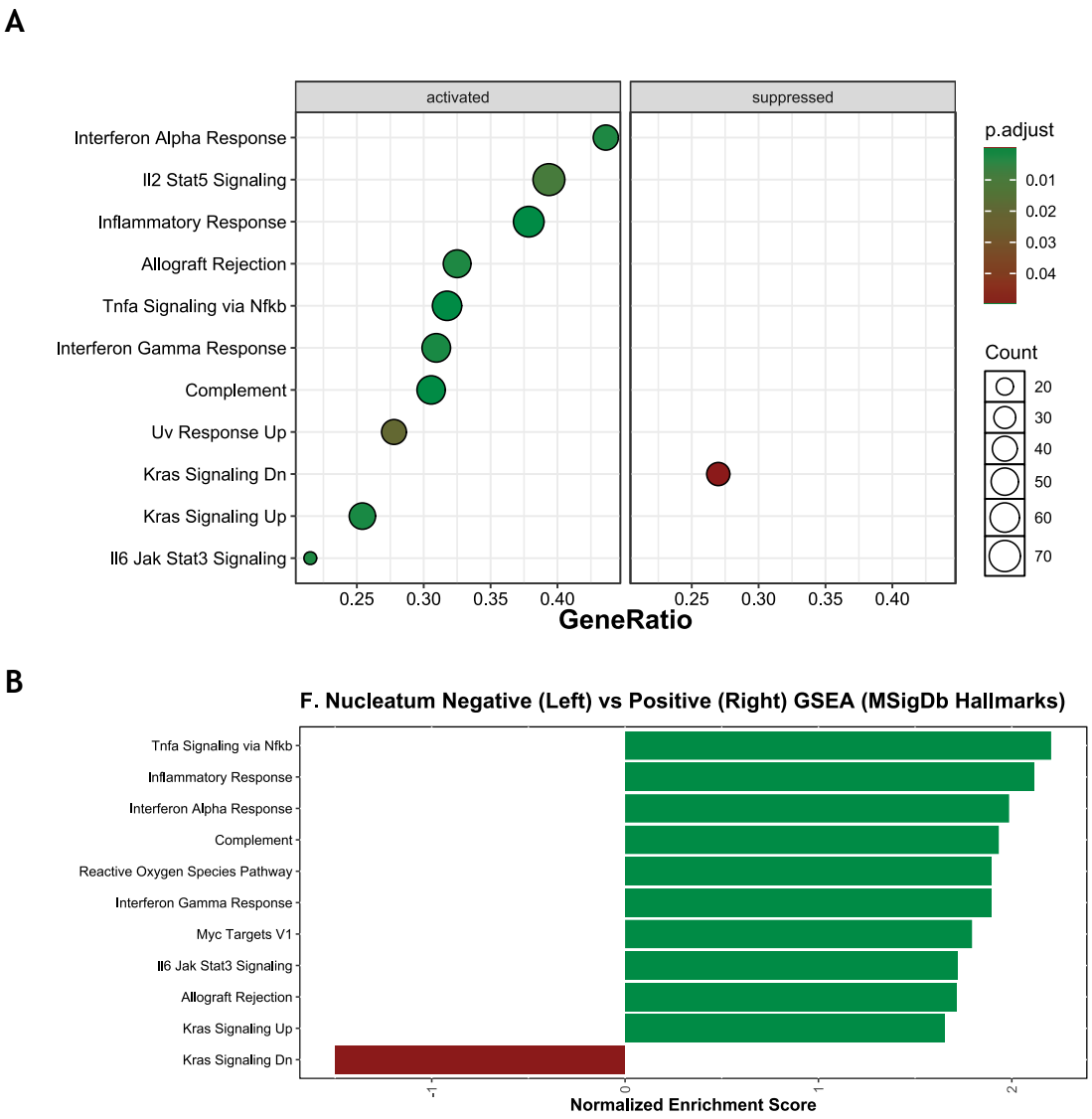


Figure 3.18. MSigDb Hallmarks GSEA Summary.
[A] Dot plot of top 10 activated (left) and supressed (right) MSigDb Hallmarks gene sets ranked by gene ratio. The x-axis shows the calculated gene ratio of input genes corresponding to the gene set in each term. Dot size refers to gene counts while colour refers to pAdj. **[B]** Bar plot of top 10 positively enriched (green) and top 10 negatively enriched (red) MSigDb Hallmarks gene sets ranked by NES. The *F. nucleatum*+ group was compared to the *F. nucleatum*- group. Total n=190, *F. nucleatum*+ group n=20, and *F. nucleatum*- group n=170.

Of the leading-edge genes overlapping between four or more gene sets, three were from the CXC chemokine family, *CXCL9, CXCL10,* and *CXCL11* all three of which were shared between “IL6 JAK STAT3 Signalling”, “Interferon Gamma Response”, and “Inflammatory Response”.

Three overlapping leading-edge genes were common across six of the positively enriched gene sets, *CXCL10, IL6,* and *IRF1*. *CXCL10* was shared between “KRAS Signalling Up”, “IL6 JAK STAT3 Signalling”, “Interferon Gamma Response”, “Interferon Alpha Response”, “Inflammatory Response”, and “TNFA Signalling via NF-KB”. *IL6* was common between “Allograft

Rejection”, “IL6 JAK STAT3 Signalling”, “Interferon Gamma Response”, “Complement”, “Inflammatory Response”, and “TNFA Signalling via NF-KB”. *IRF1* was shared between “IL6 JAK STAT3 Signalling”, “Interferon Gamma Response”, “Complement”, “Interferon Alpha Response”, “Inflammatory Response”, and “TNFA Signalling via NF-KB”. This is summarised in Figure 3.19.

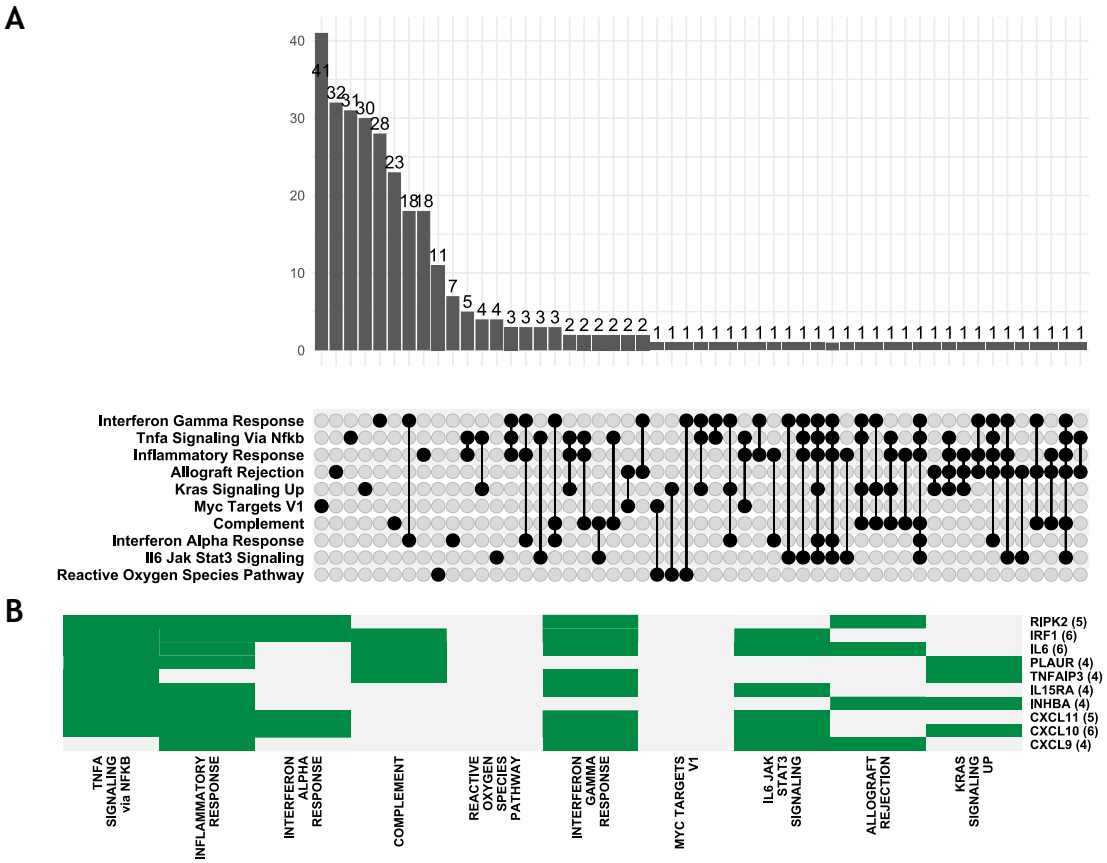


Figure 3.19. Overlapping Leading Edge Genes in Positively Enriched MSigDb Hallmarks Gene Sets.
[A] UpSet plot showing number of overlapping genes as bars across intersecting MSigDb Hallmarks gene sets (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least four MSigDb Hallmarks gene sets.

Since only one gene set was negatively enriched, overlapping leading edge gene analysis was not possible. However, GSEA plots (Figure 3.20) confirm that genes within the positively enriched gene sets are highly ranked, while the genes in the negatively enriched gene set are ranked lower.

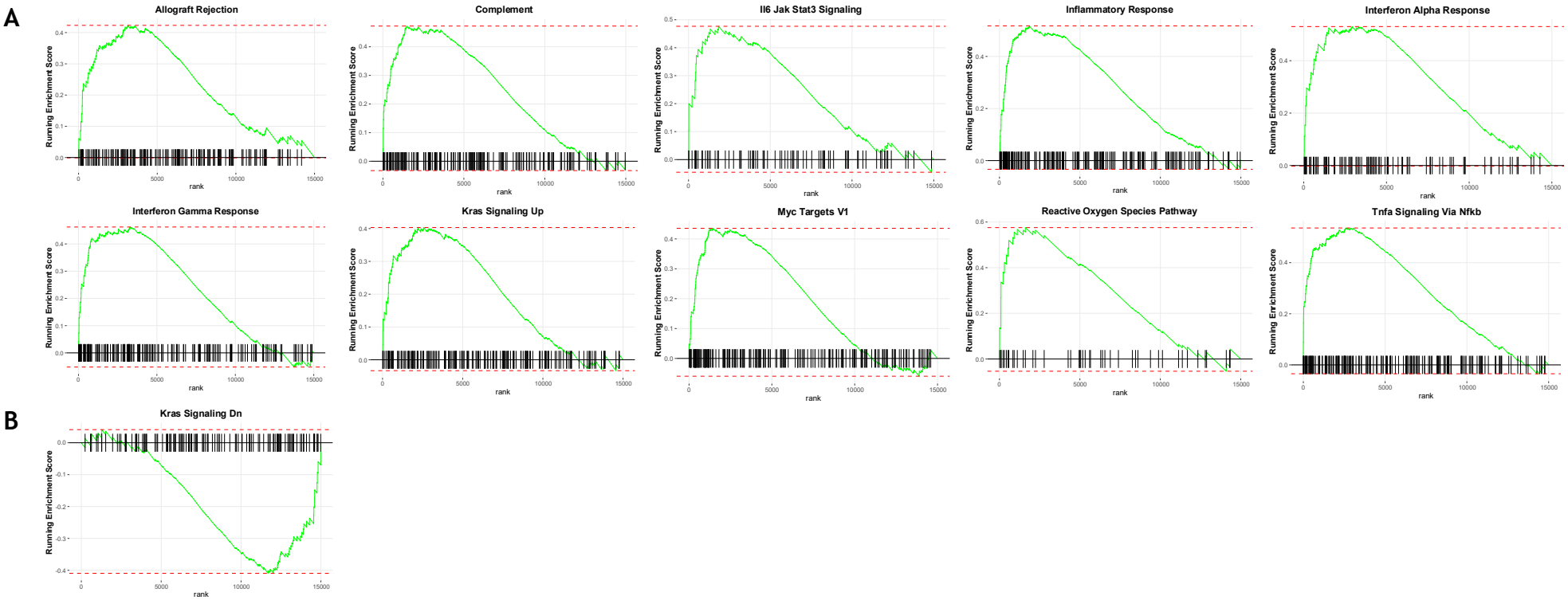


Figure 3.20. GSEA Plots of Top Enriched MSigDb Hallmarks Gene Sets.
[A] Positively enriched MSigDb Hallmarks Gene Sets. **[B]** Negatively enriched MSigDb Hallmarks Gene Sets. The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The horizontal red lines represent the point highest or lowest enrichment scores for genes within the set.

3.5.2.3 Kyoto Encyclopaedia of Genes and Genomes

Finally, KEGG was used to investigate the differences in higher level cellular pathways seen in the *F. nucleatum*+ group. Figure 3.21A shows the top ten activated KEGG pathways within the group. Three activated pathways were categorised as “Human Diseases: Infectious Disease - Viral”: “Epstein-Barr Virus Infection”, “Human Immunodeficiency Virus 1 Infection”, and “Human T-cell Leukaemia Virus 1 Infection”. Two activated pathways were categorised as “Organismal Systems: Immune System”: “Neutrophil Extracellular Trap Formation” and “Chemokine Signalling Pathway”, and two other were categorised as “Cellular Processes: Cell Growth and Death”: “Cell Cycle” and “Cellular Senescence”. One pathway “Prion Disease” was categorised as “Human Diseases: Neurodegenerative Disease”, one pathway “Tuberculosis” was categorised as “Human Diseases: Infectious Disease - Bacterial”, and the final activated pathway “Cytokine-Cytokine Receptor Interaction” was categorised as “Environmental Information Processing: Signalling Molecules and Interaction”.

Of the five suppressed pathways, “Cytoskeleton in Muscle Cells” was categorised as “Cellular Processes: Cell Motility”, while “Focal Adhesion” was categorised as “Cellular Processes: Cellular Community - Eukaryotes”. “Hormone Signalling” was categorised as “Environmental Information Processing: Signalling Molecules and Interaction”, while “PI3K-AKT Signalling Pathway” was categorised as “Environmental Information Processing: Signal Transduction”. Finally, “Axon Guidance” was categorised as “Organismal Systems: Development and Regeneration”.
When ranked by NES (Figure 3.21B), the top ten positively enriched pathways were the same as those shown in the dot plot, with “Prion Disease” dropped and “Measles” gained under “Human Diseases: Infectious Disease - Viral”. The negatively enriched pathways mirror those shown in Figure 3.21A.

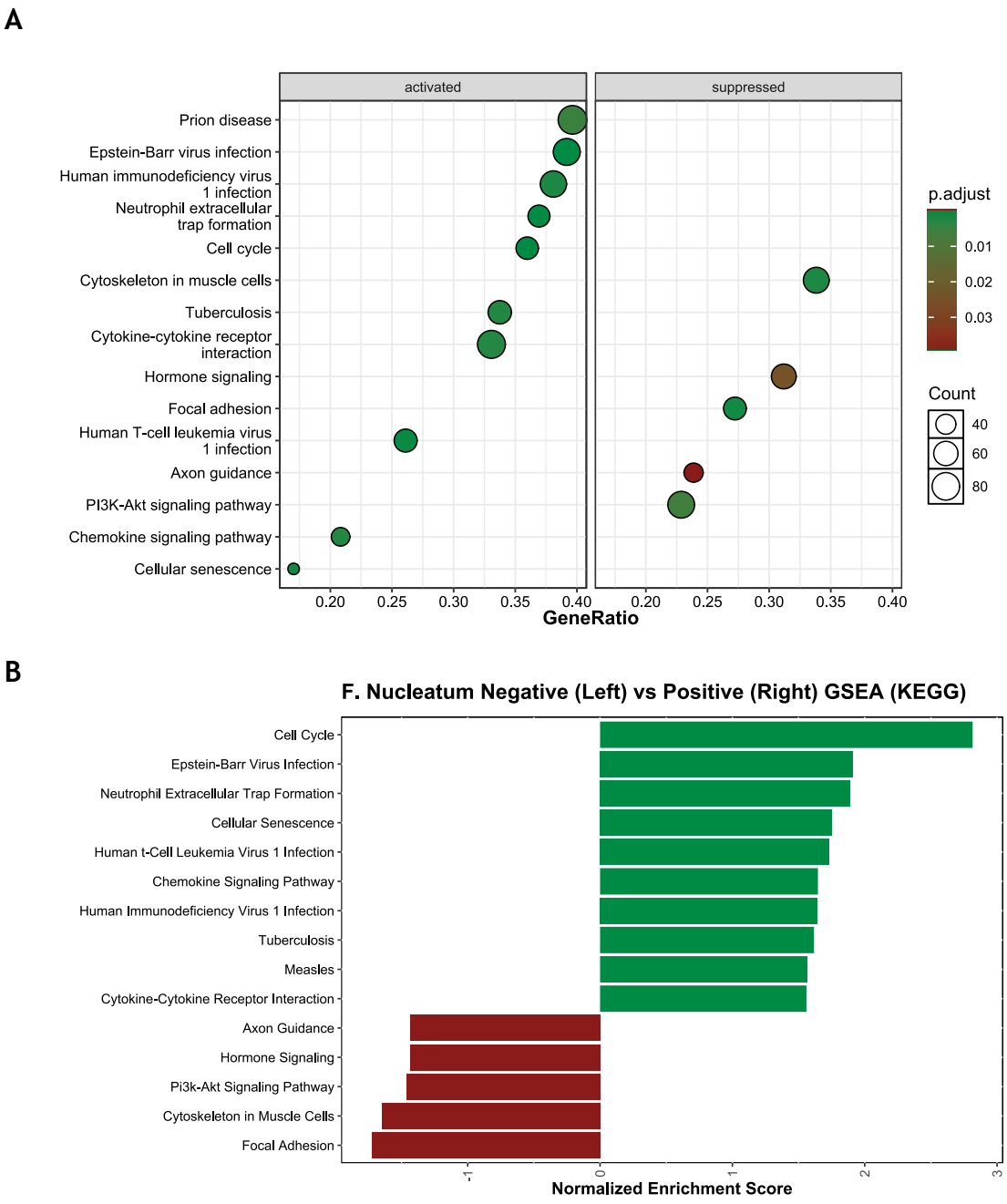


Figure 3.21. KEGG Pathways GSEA Summary.
[A] Dot plot of top 10 activated (left) and supressed (right) KEGG pathways ranked by gene ratio. The x-axis shows the calculated gene ratio of input genes corresponding to the gene set in each term. Dot size refers to gene counts while colour refers to pAdj. **[B]** Bar plot of top 10 positively enriched (green) and top 10 negatively enriched (red) KEGG pathways ranked by NES. The *F. nucleatum+* group was compared to the *F. nucleatum-* group. Total n=190, *F. nucleatum+* group n=20, and *F. nucleatum-* group n=170.

The UpSet plot (Figure 3.22A) shows multiple intersecting genes between the positively enriched pathways. As shown in Figure 3.22B, six leading edge genes were common between five pathways: *CCNE1*, *CCNE2*, *CDK4*, *IFNA21*, *TNF*, *IL6*, and *TLR2*. *CCNE1*, *CCNE2*, and *CDK4* were shared between "Measles", "Human T-cell Leukaemia Virus 1 Infection", "Cellular Senescence", "Epstein-Barr Virus Infection", and "Cell Cycle". While

"Epstein-Barr Virus Infection", "Tuberculosis", "Measles", and "Cytokine-Cytokine Receptor Interaction" shared both *IFNA21* and *IL6, IFNA21* was also a leading-edge gene in "Human Immunodeficiency Virus 1 Infection", while *IL6* was a leading-edge gene in "Human T-cell Leukaemia Virus 1 Infection". *TLR2* was shared between "Epstein-Barr Virus Infection", "Neutrophil Extracellular Trap Formation", "Human Immunodeficiency Virus 1 Infection", "Tuberculosis", and "Measles". *TNF* was common between "Epstein-Barr Virus Infection", "Human T-cell Leukaemia Virus 1 Infection", "Human Immunodeficiency Virus 1 Infection", "Tuberculosis", and "Cytokine-Cytokine Receptor Interaction". Unsurprisingly, *CD3D, CD3E,* and *CD3G* are shared between the same four pathways: "Epstein-Barr Virus Infection", "Human T-cell Leukaemia Virus 1 Infection", "Human Immunodeficiency Virus 1 Infection", and "Measles". *CCNA2, MYC, E2F1*, and *ETF3* are shared between "Cell Cycle", "Epstein-Barr Virus Infection", "Cellular Senescence", and "Human T-cell Leukaemia Virus 1 Infection", while *MAPK12* and *MAPK13* were shared between "Epstein-Barr Virus Infection", "Neutrophil Extracellular Trap Formation", "Human Immunodeficiency Virus 1 Infection", and "Tuberculosis".

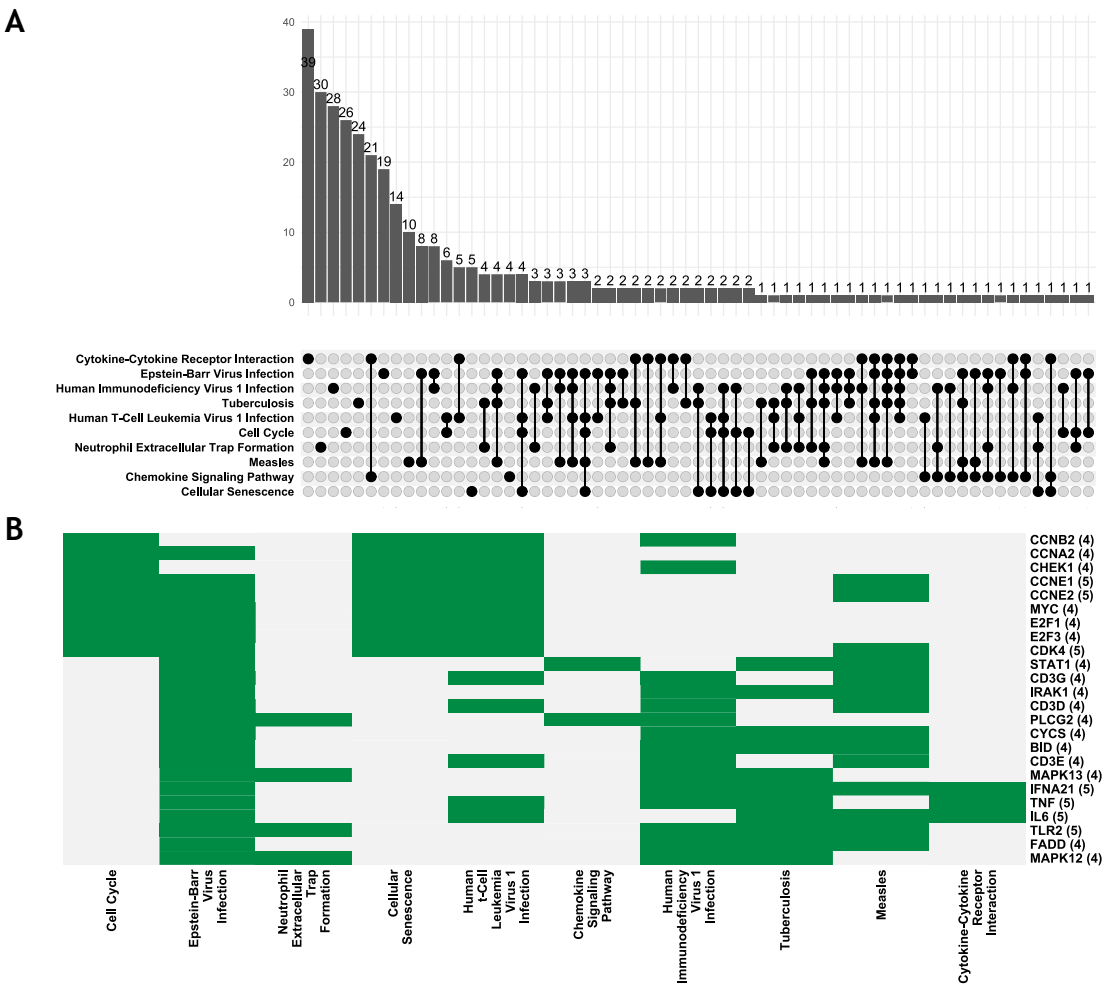


Figure 3.22. Overlapping Leading Edge Genes in Positively Enriched KEGG Pathways. **[A]** UpSet plot showing number of overlapping genes as bars across intersecting KEGG pathways (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least four KEGG pathways.

Of the five negatively enriched pathways, two leading edge genes, *ITGAV* and *PIK3R1*, were common among four of them: “Focal Adhesion”, “Cytoskeleton in Muscle Cells”, “PI3K-AKT Signalling Pathway”, and “Hormone Signalling”. Seven members of the collagen family were leading edge genes shared across three pathways, “Focal Adhesion”, “Cytoskeleton in Muscle Cells”, and “PI3K-AKT Signalling Pathway”, those being *COL1A1, COL1A2, COL4A5, COL4A6, COL6A1, COL6A2,* and *COL6A3*. Four genes, *ITGA3, ITGA7, ITGB5*, and *LAMA2* were common between “Focal Adhesion”, “Cytoskeleton in Muscle Cells”, and “PI3K-AKT Signalling Pathway”. Finally, three members of the thrombospondin family, *THBS2, THBS3*, and *THBS4* were shared between “Focal Adhesion”, “Cytoskeleton in Muscle Cells”, and “PI3K-AKT Signalling Pathway”.

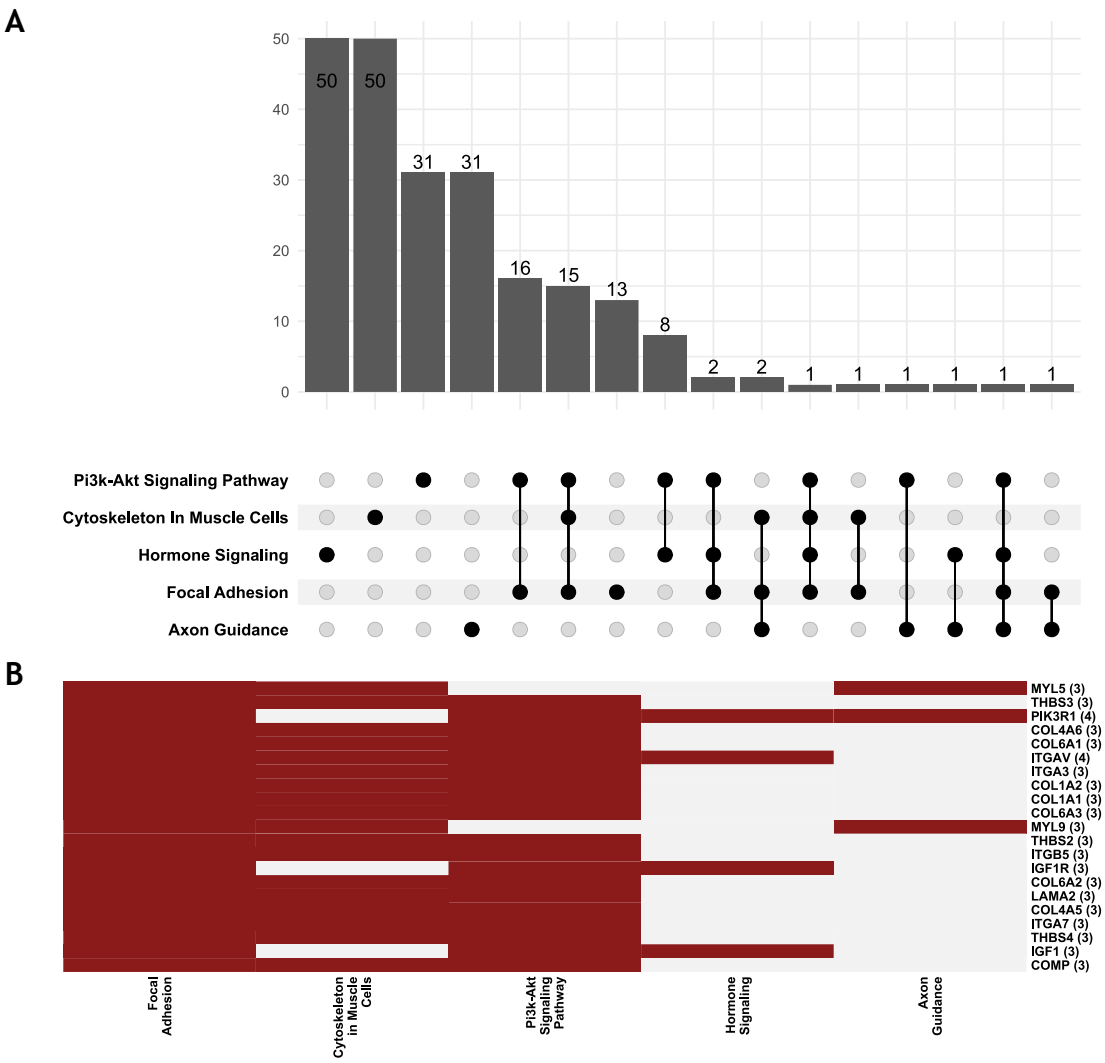


Figure 3.23. Overlapping Leading Edge Genes in Negatively Enriched KEGG Pathways.
[A] UpSet plot showing number of overlapping genes as bars across intersecting KEGG pathways (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least three KEGG pathways.

The GSEA plots in Figure 3.24, demonstrate that most of the genes with the positively enriched pathways are highly ranked, while most of the genes in the negatively enriched pathways are ranked lower.

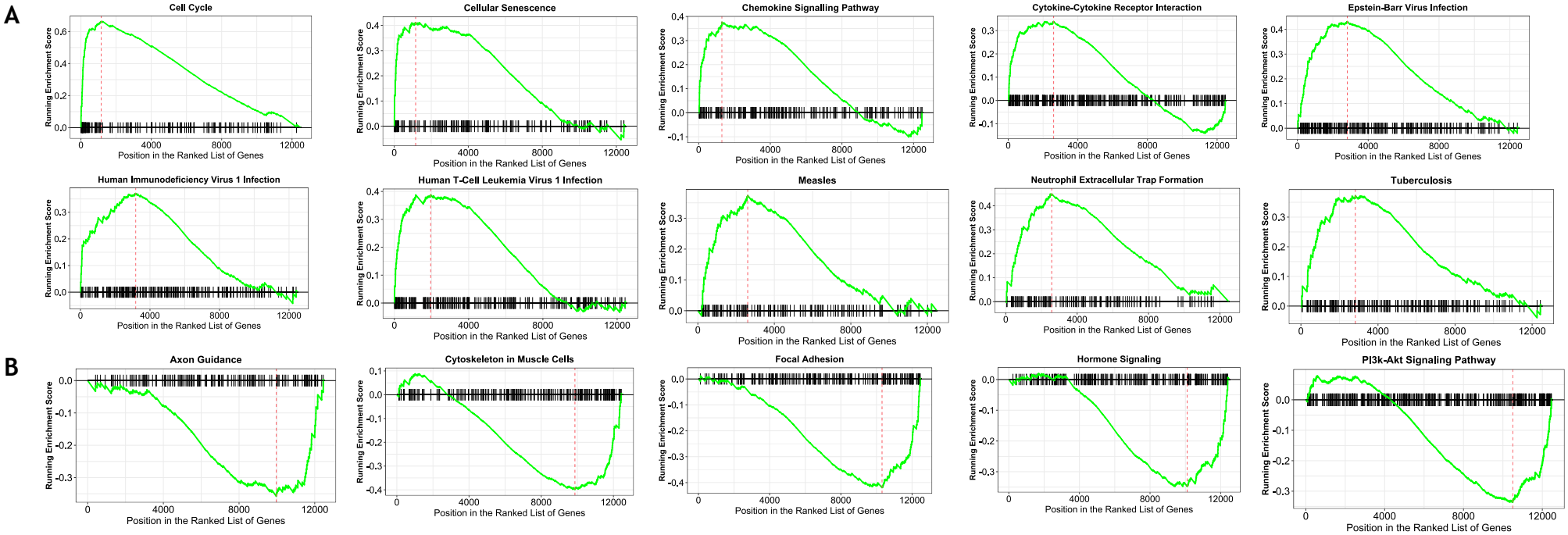


Figure 3.24. GSEA Plots of Top Enriched KEGG Pathways.
[A] Positively enriched KEGG pathways. **[B]** Negatively enriched KEGG pathways. The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The vertical red line represents the point where gene enrichment is most positive or negative.

3.6 Cytokine Stimulation of FAP PDOs

GSEA revealed multiple immune-related pathways to be positively enriched in the *F. nucleatum*+ patients, with a focus on cytokines. GO terms like “neutrophil migration” and “cytokine activity”, MSigDb gene sets like “IL6 JAK STAT3 Signalling”, “IL2 JAK STAT5 Signalling”, “TNFA Signalling via NF-KB”, and KEGG pathways “Cytokine-Cytokine Receptor Interaction”, all drove this work into an immune-centred direction.

In order to investigate if the immune response being enriched in these adenomas is *F. nucleatum*-driven or independent from bacterial influence, PDOs from FAP patients were used to investigate the effects of stimulation by inflammatory cytokines.

Four cytokines: LIGHT, TNF-α, Lymphotoxin, and IL-6 were chosen as they are known to be secreted by a variety of activated immune cells including lymphocytes, macrophages, and neutrophils, and were used as described in 2.2.9.4.1. The optimum concentrations were determined within the lab group (Dr. Kathryn Pennel, Dr. Guang-Yu Lian, and Molly McKenzie) on CRC cell lines. Metabolic activity was used as an indicator of cell viability of PDOs and measured using CellTiter-Blue® 72 hours after treatment.

As shown in Figure 3.25, PDO18235 and PDO19448 grew at an accelerated rate when treated with the cytokines in comparison to the media control, although this was not statistically significant (PDO18235: LIGHT p=0.9992, TNFα p=0.6826, Lymphotoxin p=0.5805, IL-6 p=0.4282; PDO19448: LIGHT p=0.8736, TNFα p=0.5192, Lymphotoxin p=0.5319, IL-6 p=0.4701). For both PDOs, IL-6 had the greatest effect on cell viability, and when visually examined, on organoid size and number. Both TNF-α and Lymphotoxin had similar effects on cell viability and were higher than the vehicle control for both PDOs. LIGHT had the least effect on cell viability, and on visual examination, on organoid size and number.

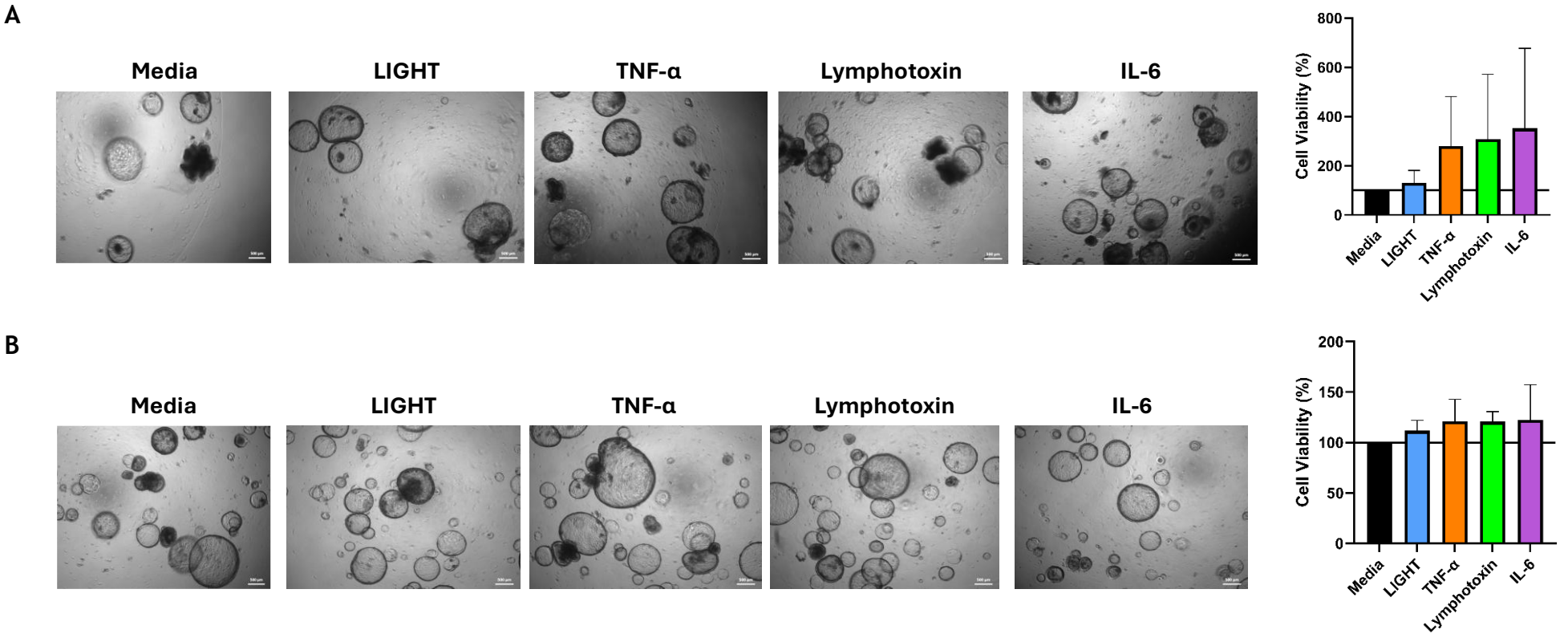


Figure 3.25. FAP Patient PDOs Treated with Cytokines.
[A] Brightfield microscopy images (left) of PDO18235 treated with LIGHT, TNF-α, Lymphotoxin, and IL-6 with vehicle control. Cell viability (right) of PDO18235 after treatment with the cytokines at either concentration normalised to the media control. **[B]** Brightfield microscopy images (left) of PDO19448 treated with LIGHT, TNF-α, Lymphotoxin, and IL-6 with media control. Cell viability (right) of PDO19448 after treatment with the cytokines normalised to the media control. All experiments were performed as biological triplicates from three different passages and as technical triplicates where each well was replicated three times. Images were taken and cell viability was measured 72 hours post-treatment. Scale bar = 500μm. Error bars represent standard error of mean. One-way ANOVA was used to test significance against the media control. Non-significant associations were not graphically shown.

3.7 T Cell Densities on *F. nucleatum* Patient Subset

Next, it was logical to investigate the CD3+ T cells, CD4+ T cells, and CD8+ T cells (lymphocytic immune cells) in the context of these adenomas with known *F. nucleatum* status to understand any effects on the adaptive immune response.

3.7.1 Antibody Validation

The antibody validation for CD3, CD4, and CD8 antibodies was not required as these were diagnostic-grade and do not require research validation.

3.7.2 Total Immune Markers Expression Is Not Associated with the Detection of Metachronous Polyps

3.7.2.1 IHC Staining and Assessment of Immune Markers

Staining of the tissue on INCISE TMA slides for CD3, CD4, and CD8 was carried out as described in 2.2.2.3 and 2.2.2.4. Figure 3.26A, Figure 3.27A, Figure 3.28A and shows representative images of the staining pattern for CD3+ cells, CD4+ cells, and CD8+ cells, respectively. As expected, there were more positive cells in the lamina propria than within the epithelium although there was some epithelial infiltration for CD8. The positive cell classifiers built in QuPath by Marjolein de Koning, Arwen Shirmohammad Damavandi, and Lewis Irvine successfully detected positive cells for each of the stains (Figure 3.26B for CD3, Figure 3.27B for CD4, and Figure 3.28B for CD8). For CD3, CD4, and CD8, no-antibody negative controls (Figure 3.26C, Figure 3.27C, and Figure 3.28C respectively) and matched isotype negative controls (Figure 3.26D, Figure 3.27D, and Figure 3.28D respectively) were included to verify the observed signal was a true positive. A tonsil positive control was included to verify the antibody can detect its target (Figure 3.26E, Figure 3.27E, and Figure 3.28E respectively).

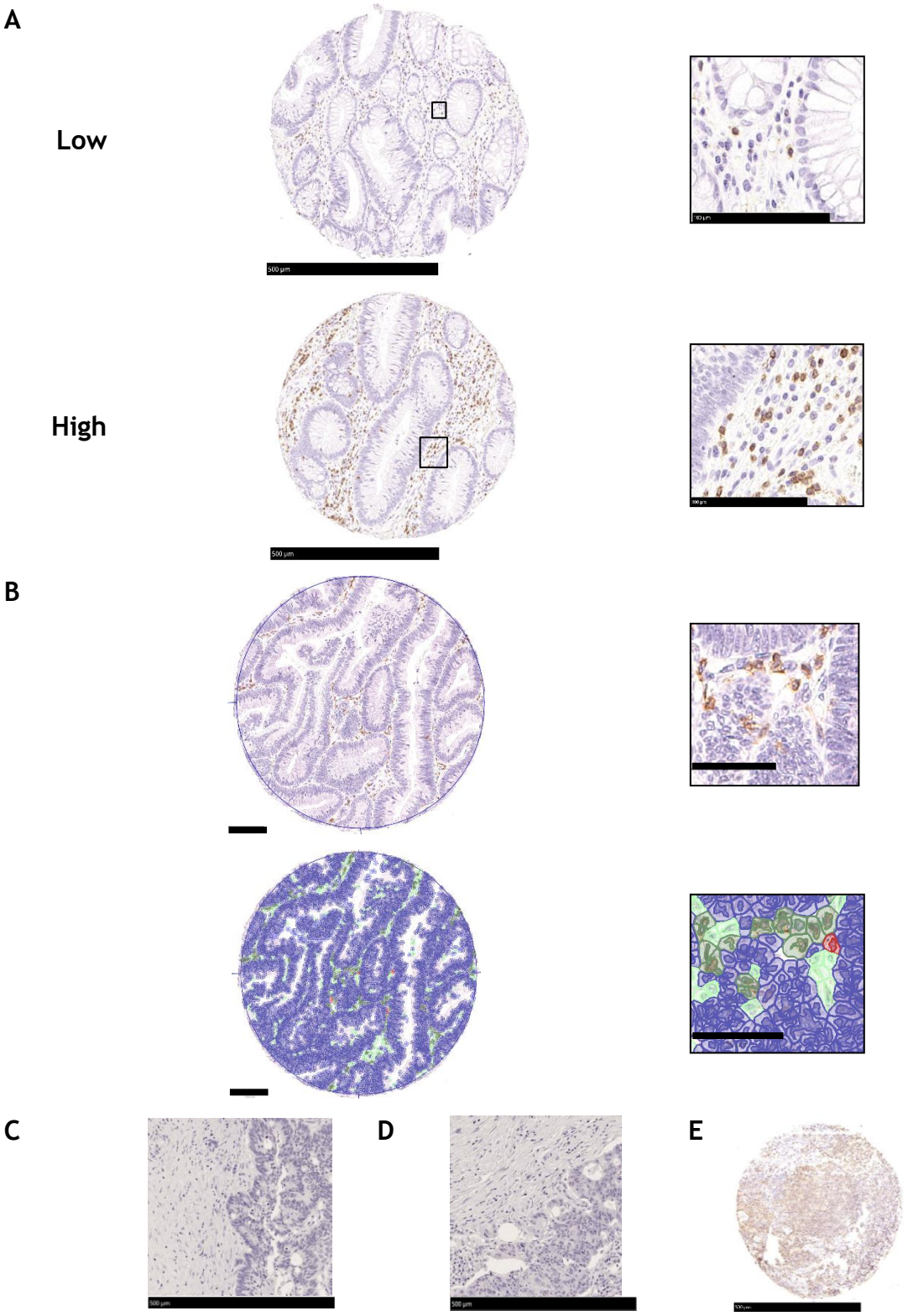


Figure 3.26. IHC of CD3.
[A] Examples of low- and high-density cores for CD3 in polyp tissue. Black squares denote higher magnification areas for clearer visualisation. Scale bars under cores represent 500µm, scale bars under square panels represent 100µm. **[B]** Example of positive cells and positive cell detection classifier on QuPath (Performed by Marjolein de Koning and Arwen Shirmohammad Damavandi). Dark green objects depict positive cells in the lamina propria, while red objects depict positive cells in the epithelium. Light green objects depict weaker stained positive cells in the lamina propria, and blue objects depict negative cells. Scale bars under cores represent 100µm, scale bars under square panels represent 50µm. **[C]** No-antibody negative control of CRC surplus tissue. Scale bars under cores represent 500µm. **[D]** Negative matched iso-type control of CRC surplus tissue. Scale bars under cores represent 500µm. **[E]** Positive control of tonsil surplus tissue. Scale bars under cores represent 500µm.

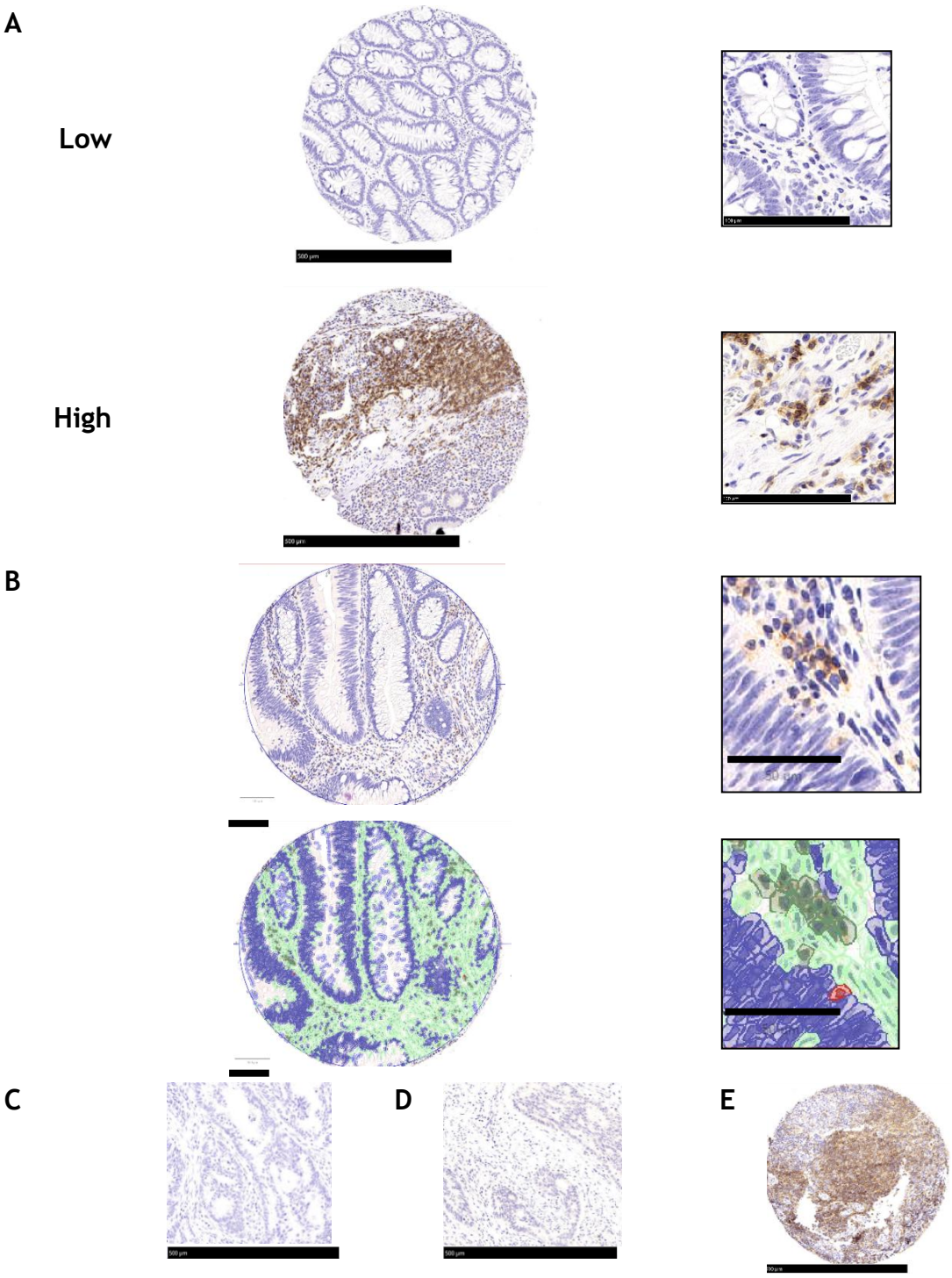


Figure 3.27. IHC of CD4.
[A] Examples of low- and high-density cores for CD4 in polyp tissue. Black squares denote higher magnification areas for clearer visualisation. Scale bars under cores represent 500µm, scale bars under square panels represent 100µm. **[B]** Example of positive cells and positive cell detection classifier on QuPath (Performed by Arwen Shirmohammad Damavandi). Dark green objects depict positive cells in the lamina propria, while red objects depict positive cells in the epithelium. Light green objects depict weaker stained positive cells in the lamina propria, and blue objects depict negative cells. Scale bars under cores represent 100µm, scale bars under square panels represent 50µm. **[C]** No-antibody negative control of CRC surplus tissue. Scale bars under cores represent 500µm. **[D]** Negative matched iso-type control of CRC surplus tissue. Scale bars under cores represent 500µm. **[E]** Positive control of tonsil surplus tissue. Scale bars under cores represent 500µm.

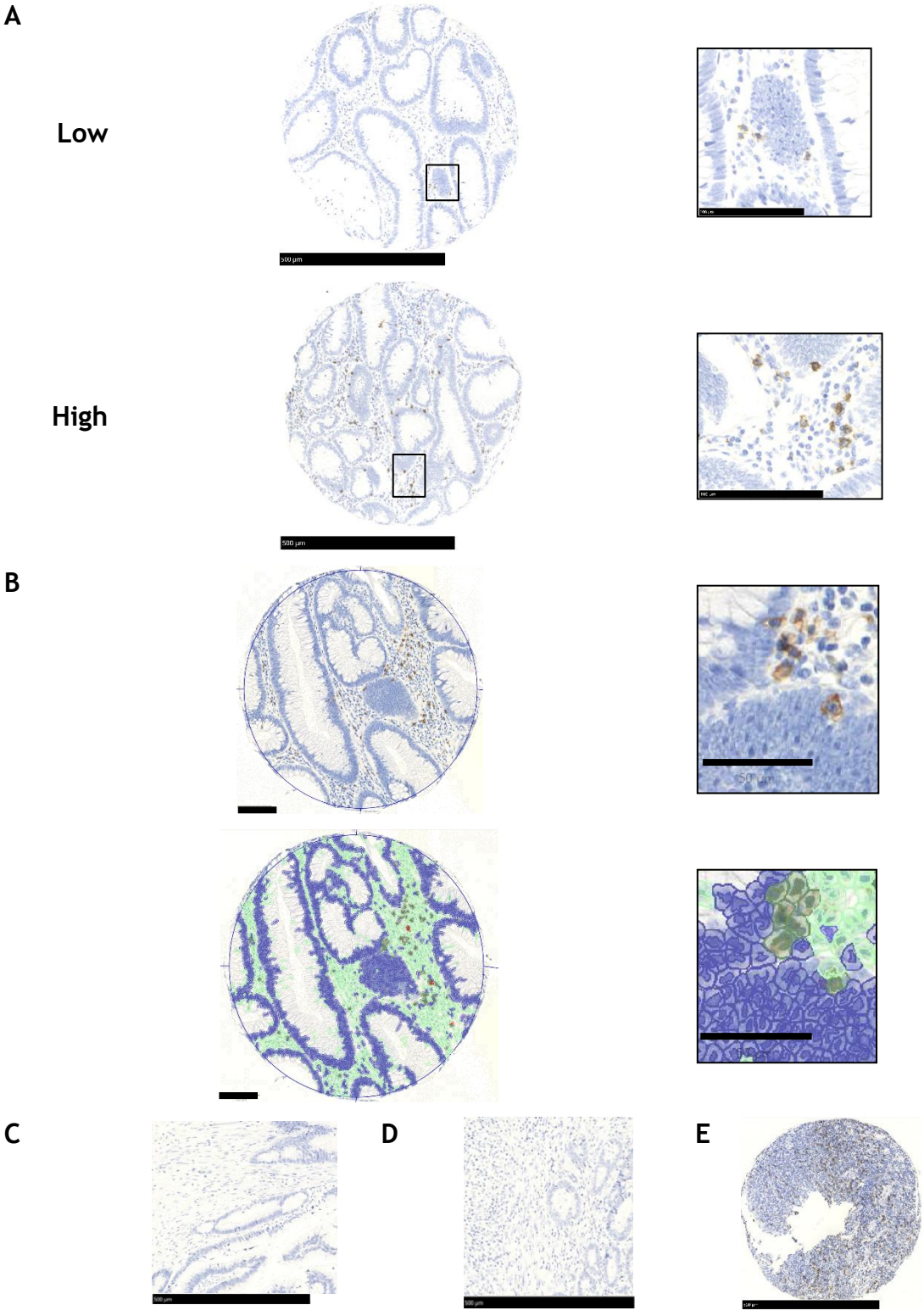


Figure 3.28. IHC of CD8.
[A] Examples of low- and high-density cores for CD8 in polyp tissue. Black squares denote higher magnification areas for clearer visualisation. Scale bars under cores represent 500µm, scale bars under square panels represent 100µm. **[B]** Example of positive cells and positive cell detection classifier on QuPath (Performed by Lewis Irvine). Dark green objects depict positive cells in the lamina propria. Light green objects depict weaker stained positive cells in the lamina propria, and blue objects depict negative cells. Scale bars under cores represent 100µm, scale bars under square panels represent 50µm. **[C]** No-antibody negative control of CRC surplus tissue. Scale bars under cores represent 500µm. **[D]** Negative matched iso-type control of CRC surplus tissue. Scale bars under cores represent 500µm. **[E]** Positive control of tonsil surplus tissue. Scale bars under cores represent 500µm.

3.7.2.2 Quality Control of Raw Data

Scatter plots of raw immune marker counts in the luminal and basal epithelium were generated against the total (combined luminal and basal epithelium) adenoma counts (Figure 3.29), to determine whether it would be representative to use the total counts. As the counts from the histological locations correlated with the total counts, total counts were taken forward for analysis.

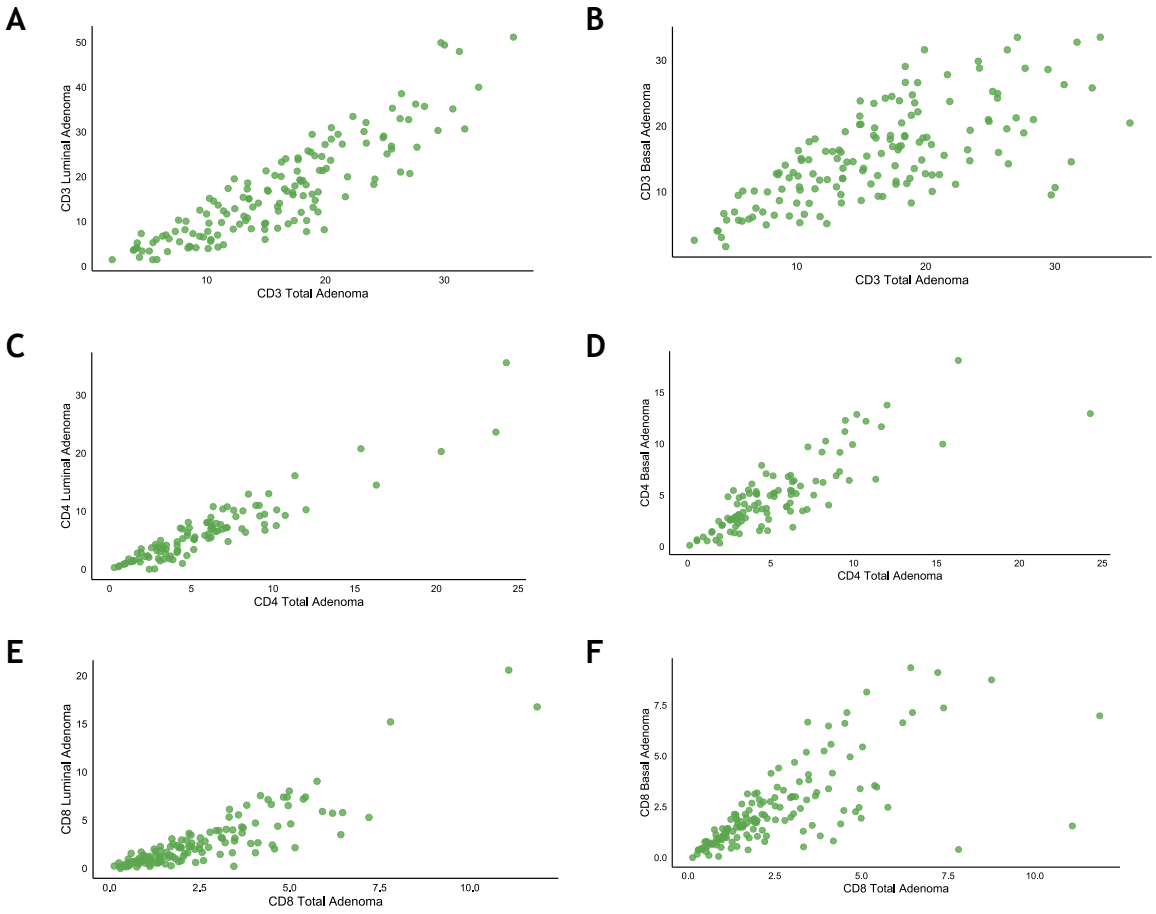


Figure 3.29. **Correlation of Polyp Histological Regions Against Total Polyp Adenomas in the INCISE *F. nucleatum* Subset.**
[A] Scatter plot of CD3+ cell density in the luminal epithelium of the polyp vs CD3+ cell density in the total adenoma of the polyp. **[B]** Scatter plot of CD3+ cell density in the basal epithelium of the polyp vs CD3+ cell density in the total adenoma of the polyp. **[C]** Scatter plot of CD4+ cell density in the luminal epithelium of the polyp vs CD4+ cell density in the total adenoma of the polyp. **[D]** Scatter plot of CD4+ cell density in the basal epithelium of the polyp vs CD4+ cell density in the total adenoma of the polyp. **[E]** Scatter plot of CD8+ cell density in the luminal epithelium of the polyp vs CD8+ cell density in the total adenoma of the polyp. **[F]** Scatter plot of CD8+ cell density in the basal epithelium of the polyp vs CD8+ cell density in the total adenoma of the polyp. CD3 N=152. CD4 N=131. CD8 N=154.

The distribution of the raw counts for CD3+ cells, CD4+ cells, and CD8+ cells is represented in the histograms shown in Figure 3.30. Raw counts for all three immune markers skew left, representing the low counts of each immune cell marker in adenoma tissue.

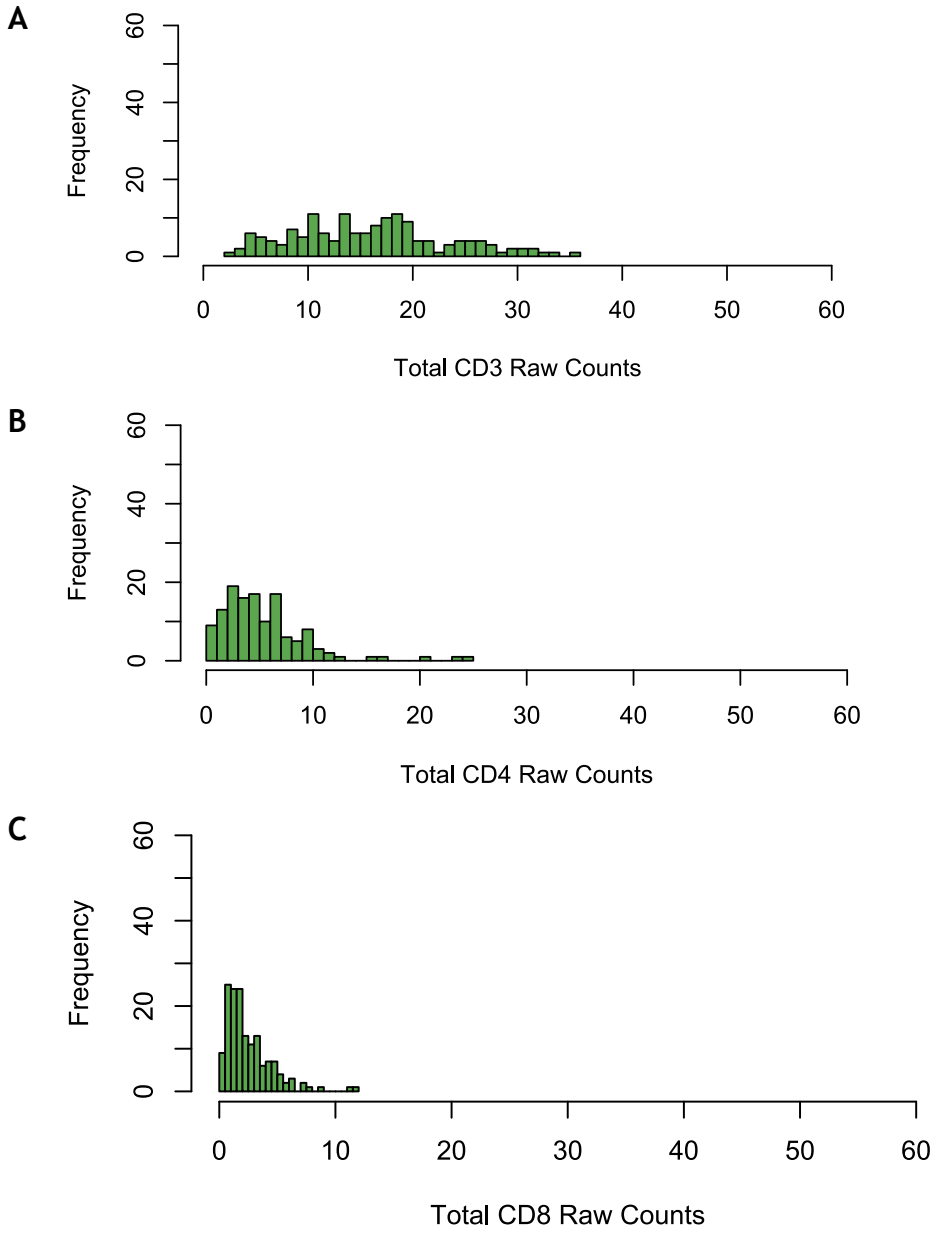


Figure 3.30. Distribution of Raw Immune Markers Counts.
Histograms of the distribution of **[A]** CD3+ cell counts (n=152), **[B]** CD4+ cell counts (n=131), and **[C]** CD8+ cell counts (n=154).

3.7.2.3 Dichotomisation of Raw Immune Counts

Raw counts of positive cells per core were dichotomised as described in 2.2.10. Patients with ≤8.66 CD3+ cells, 8.16 CD4+ cells, and 1.39 CD8+ cells were considered low (Figure 3.31).

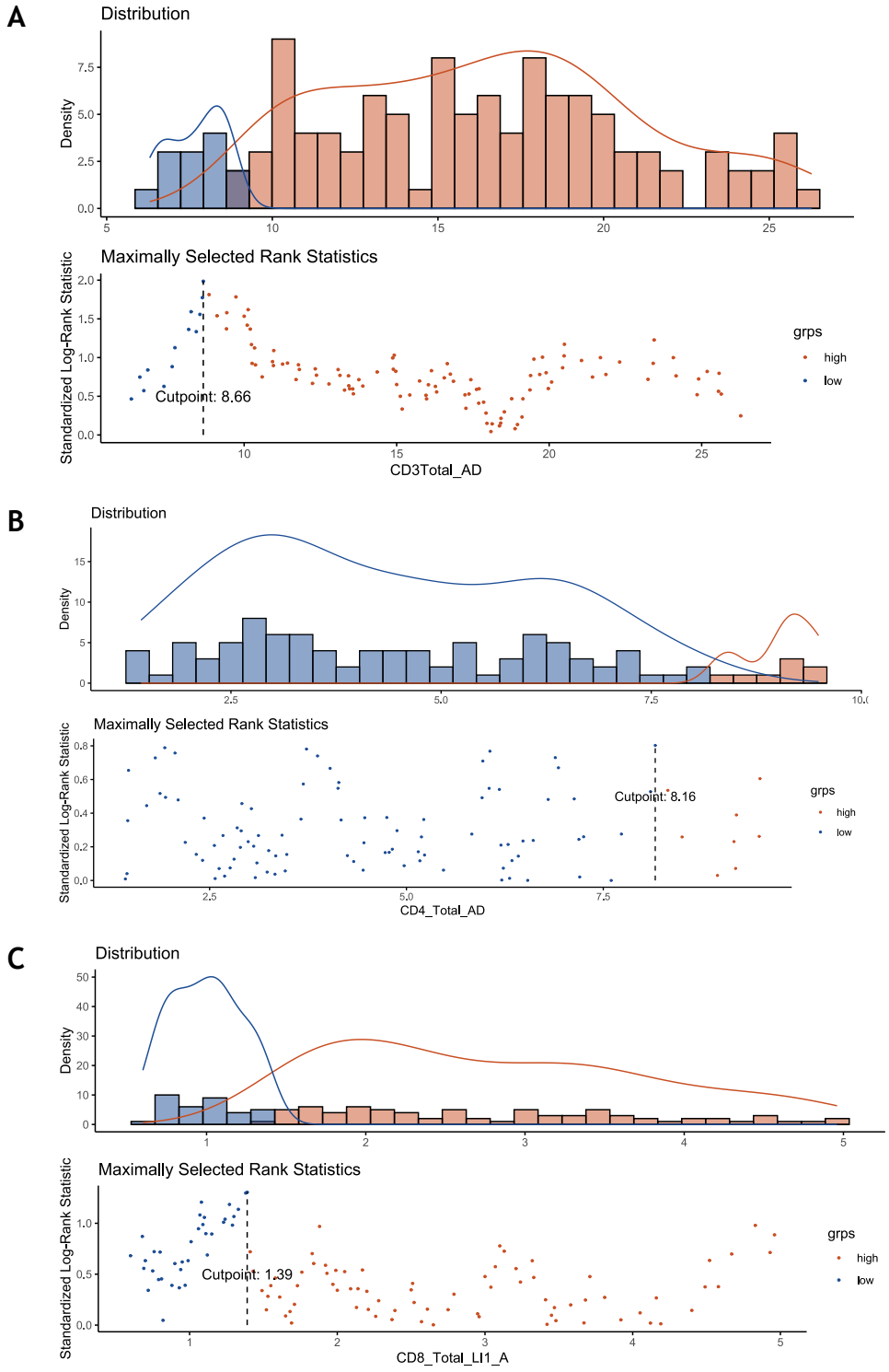


Figure 3.31. Thresholding of Raw Counts.
Threshold generation of **[A]** total CD3+ cell counts (n=152), **[B]** total CD4+ cell counts (n=131), and **[C]** total CD8+ cell counts (n=154), using maximally selected rank statistics.

3.7.2.4 Cell Density and *F. nucleatum* Status

The associations between T cell densities and *F. nucleatum* status were explored. As shown in Figure 3.32, no significant correlations between the *F. nucleatum*+ patients and the densities of the CD3+ T cells, CD4+ T cell, and CD8+ T cells are observed.

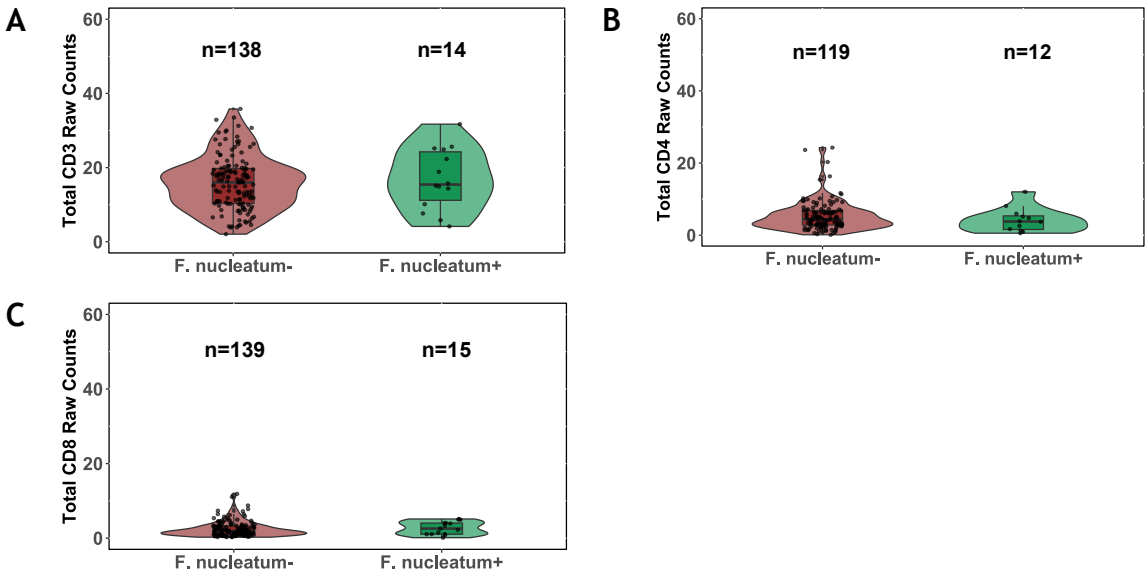


Figure 3.32. Violin Plots of *F. nucleatum* Status Against T Cell Raw Counts.
[A] CD3+ T cell counts. N=152. **[B]** CD4+ T cell counts. N=131. **[C]** CD8+ T cell counts. N=154. All distribution comparisons were not statistically significant (p>0.05).

F. nucleatum status does not appear to be driving the differences in T cell marker densities.

3.7.2.5 T Cell Density and Clinical Characteristics

To reveal any correlations between T cell density and relevant clinical characteristics, violin plots were plotted. As seen in Figure 3.33, there were no statistically significant associations between CD3+ T cell counts and sex, age, polyp site, high grade dysplasia, number of polyps at index colonoscopy, polyp type, or risk according to current BSG2020 Guidelines.

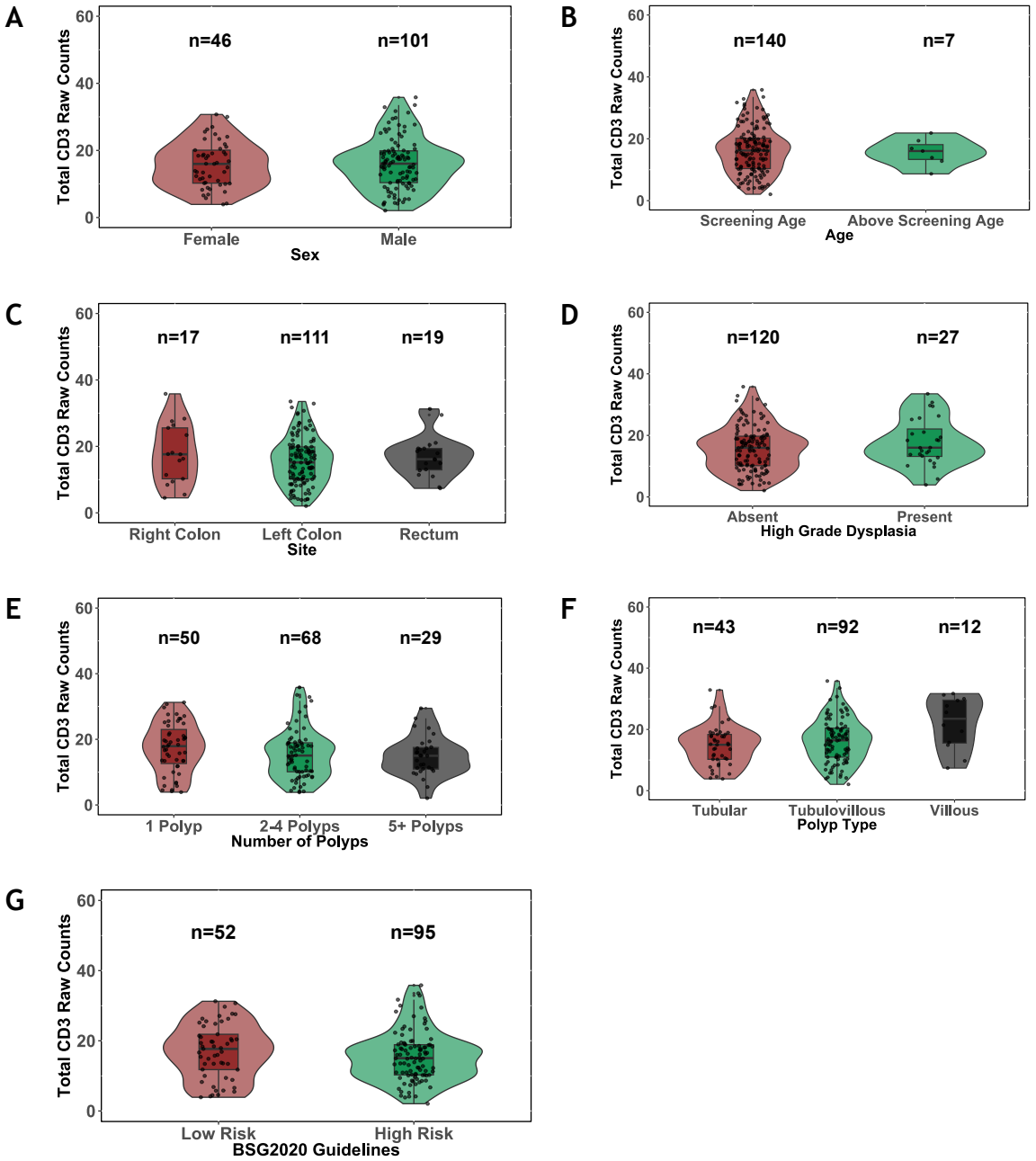


Figure 3.33. Violin Plots of CD3+ T Cell Raw Counts Against Clinical Characteristics. **[A]** Sex, **[B]** age, **[C]** site, **[D]** high grade dysplasia, **[E]** number of polyps, **[F]** polyp type, and **[G]** BSG2020 Guidelines. Only cases with all clinical information were included in this analysis. N=147. All distribution comparisons were not statistically significant (p>0.05).

The same was true for sex, age, polyp site, high grade dysplasia, number of polyps at index colonoscopy, polyp type, or risk according to current BSG2020 Guidelines, when compared to CD4+ T cell counts (Figure 3.34).

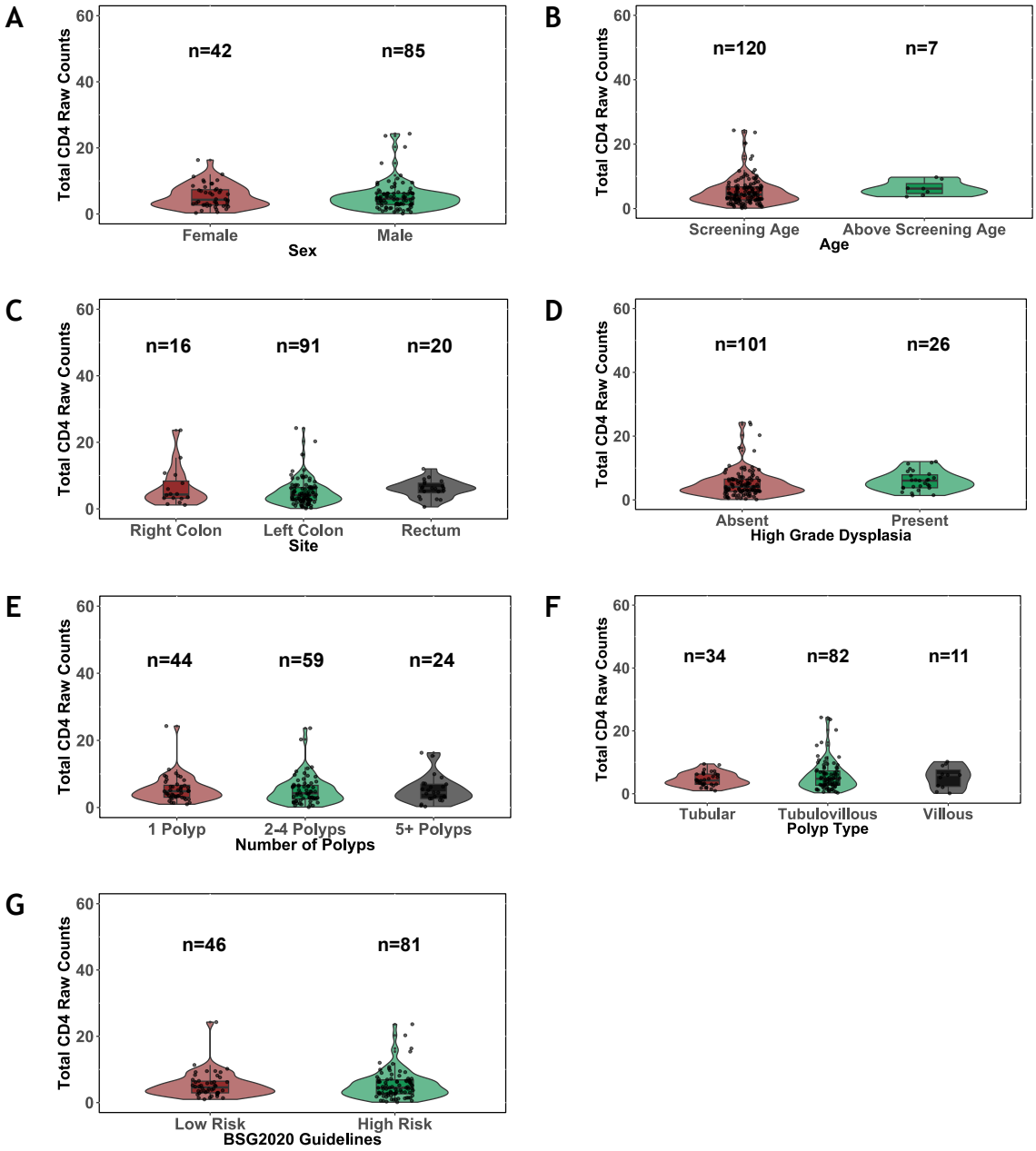


Figure 3.34. Violin Plots of CD4+ T Cell Raw Counts Against Clinical Characteristics. **[A]** Sex, **[B]** age, **[C]** site, **[D]** high grade dysplasia, **[E]** number of polyps, **[F]** polyp type, and **[G]** BSG2020 Guidelines. Only cases with all clinical information were included in this analysis. N=127. All distribution comparisons were not statistically significant (p>0.05).

However, there was a significant correlation between sex and CD8+ T cell counts (p=0.010; Figure 3.35), where male patients had a higher count of CD8+ T cells. Age, polyp site, high grade dysplasia, number of polyps at index colonoscopy, polyp type, or risk according to current BSG2020 Guidelines did not statistically correlate with CD8+ T cell counts.

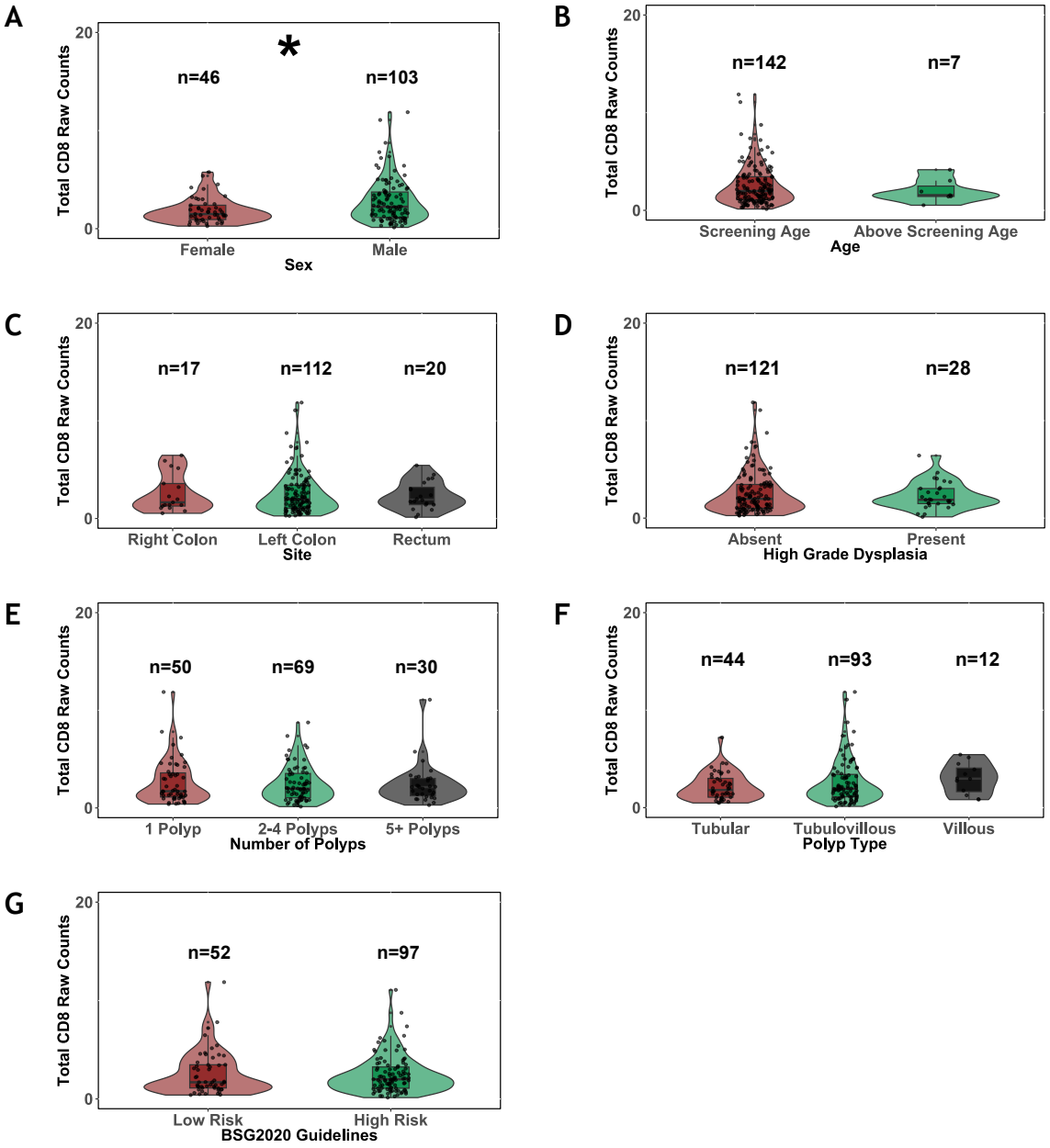


Figure 3.35. Violin Plots of CD8+ T Cell Raw Counts Against Clinical Characteristics. **[A]** Sex (p=0.010), **[B]** age, **[C]** site, **[D]** high grade dysplasia, **[E]** number of polyps, **[F]** polyp type, and **[G]** BSG2020 Guidelines. * denotes a statistically significant association. Only cases with all clinical information were included in this analysis. N=149. All other distribution comparisons were not statistically significant (p>0.05).

3.7.2.6 T Cell Density and Metachronous Polyp or CRC Detection

Upon removal of all invalid cases (n=46 for CD3+ T cells, n=67 for CD4+ T cells, and n=44 for CD8+ T cells), 152 cases, 131 cases, and 154 cases were included for CD3+ T cells, CD4+ T cells, and CD8+ T cells, respectively.

Associations between the density of the T cells and the detection of metachronous polyps or CRC were investigated using log rank survival analysis carried out on the dichotomised groups. As shown in Figure 3.36A, a higher density of total CD3+ cells was not significantly associated with the detection of metachronous polyps or CRC (HR 1.146, 95% CI 0.748-1.757; p=0.061), with a median detection time of 43 months. The patients with a low CD3+ cell density did not reach the median of time to detection of metachronous polyps or CRC.

CD4+ T cell density was also not associated with the detection of metachronous polyps or CRC (Figure 3.36B, HR 1.202, 95% CI 0.758-1.907; p=0.43), with a median detection time of 41 months for patients with low CD4+ T cell density and 43 month for patients with high CD4+ T cell density.

CD8+ cell density was not associated with the detection of metachronous polyps or CRC (Figure 3.36C, HR 1.055, 95% CI 0.428-2.601; p=0.19), with a median detection time of 41 months for patients with low CD8+ T cell density and 54 months for patients with high CD8+ T cell density.

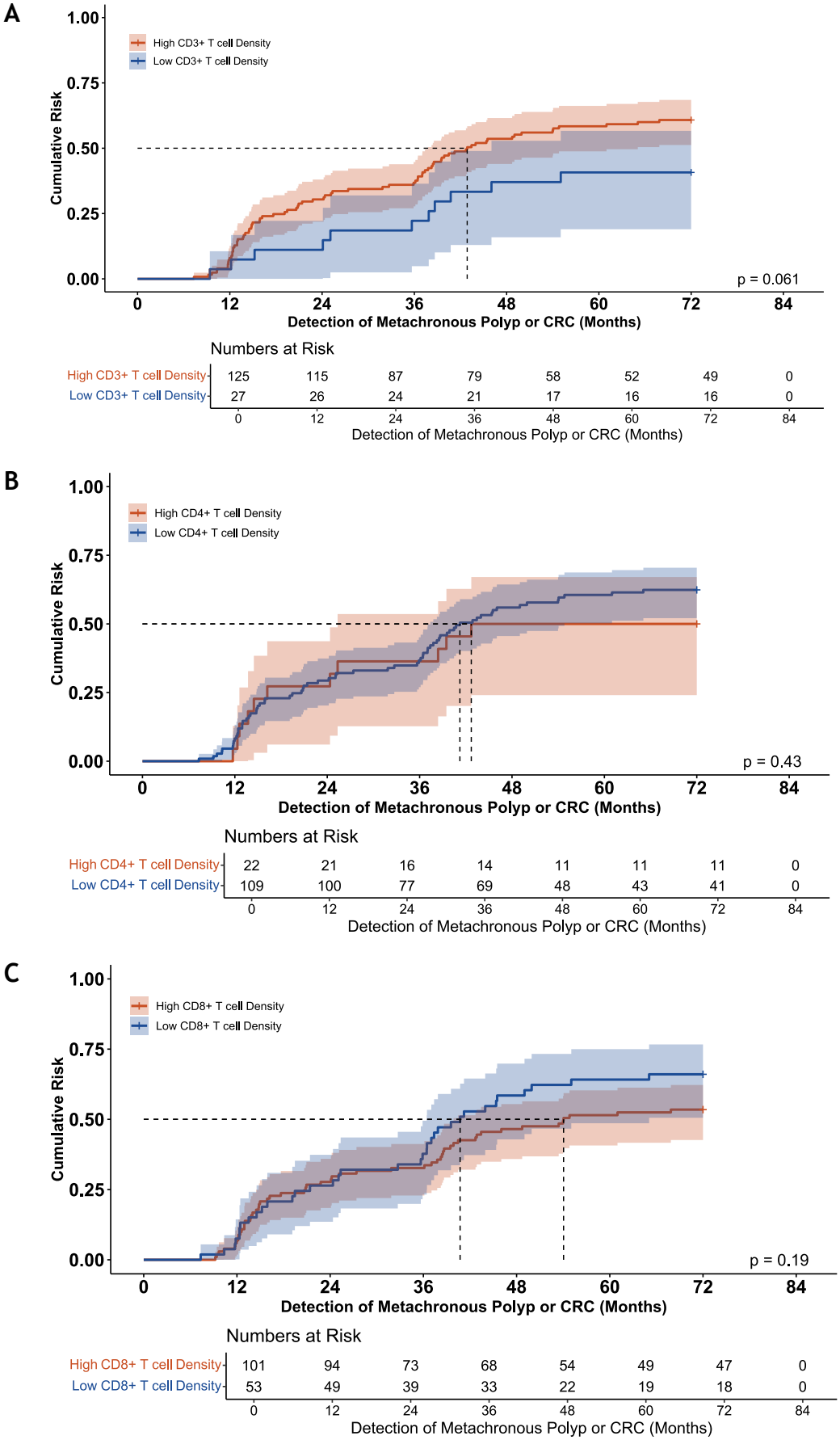


Figure 3.36. T Cell Density in Adenomas.
[A] CD3+ T cell density (n=152), **[B]** CD4+ T cell density. (n =131), and **[C]** CD8+ T cell density (n=154) and the detection of metachronous polyps or CRC. Shading is indicative of 95% CI. Dotted lines are time to median risk.

3.8 Spatial Transcriptomics Analysis CosMx®

Following the investigations of the immune response on these patients, the transcriptomic profile at the single cell level was carried put using Spatial transcriptomics platform NanoString CosMx® SMI platform to identify any *F. nucleatum*-driven spatial changes. The TMA was constructed by GTRF, the experiment was performed by Claire Kennedy Dietrich as part of the Jamieson Spatial Lab service (2.2.13.3), and the analysis was performed with the help of John Cole of the MVLS Shared Research Facilities and Dr. Assya Legrini from the Jamieson Spatial Lab (2.2.13.4).

3.8.1 Tissue Set-up

A TMA of the same patients (n=10) that were included in the RNAscope® experiment was constructed with the RNAscope®-localised *F. nucleatum* at the centroid position of the core (or as close as possible). Two *F. nucleatum*+ cores and one *F. nucleatum*- core were taken from each patient, for a total of 30 cores on the TMA slide (Figure 3.37A). An H&E slide was included to aid in orientating the cores for analysis. Two cores were lost during the experiment, leaving 28 cores available for analysis (Figure 3.37B-C).

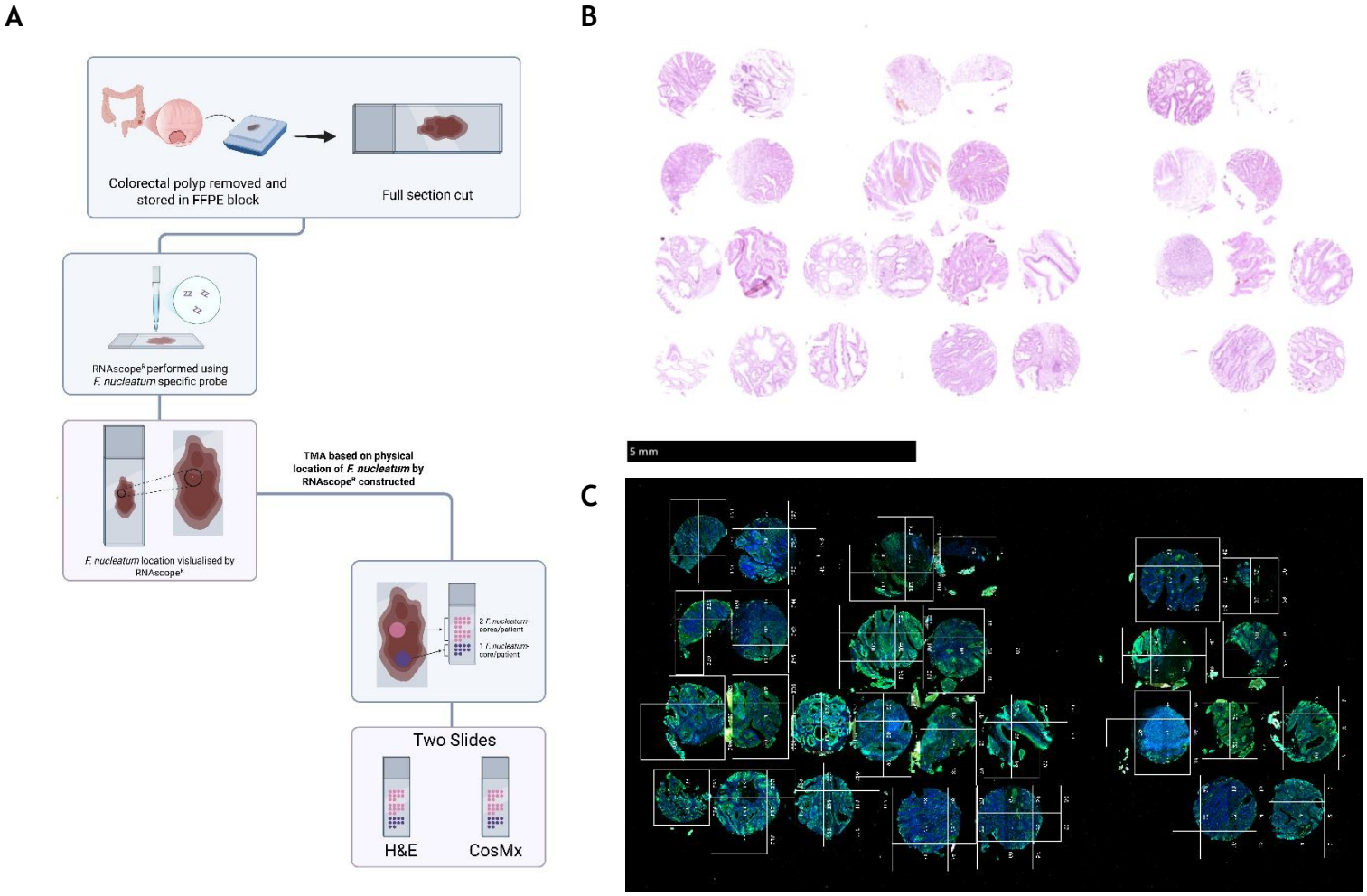


Figure 3.37. Experimental Set-up of INCISE CosMx®.
[A] Schematic of steps involved in TMA construction. **[B]** H&E image of CosMx® TMA to aid in orientation. **[C]** Image of TMA slide following CosMx®. B and C are orientated to match.

3.8.2 Quality Check, Data Normalisation, and Clustering

The violin plots of the “per cell” gene counts against sample and number of genes against samples (Figure 3.38A-B) showed a consistent sample quality, despite the differences in cell counts driven by tissue size variation across samples (Figure 3.38C). SCTransform was used to normalise the data. The elbow plot (Figure 3.38D) suggests that 15-20 dimensions are potentially informative, and 20 dimensions were carried forward to support the exploratory nature of this analysis. An unintegrated UMAP showed a clear per sample batch effect (Figure 3.38E) and, upon integration, a high level of consistency was seen between samples (Figure 3.38F).

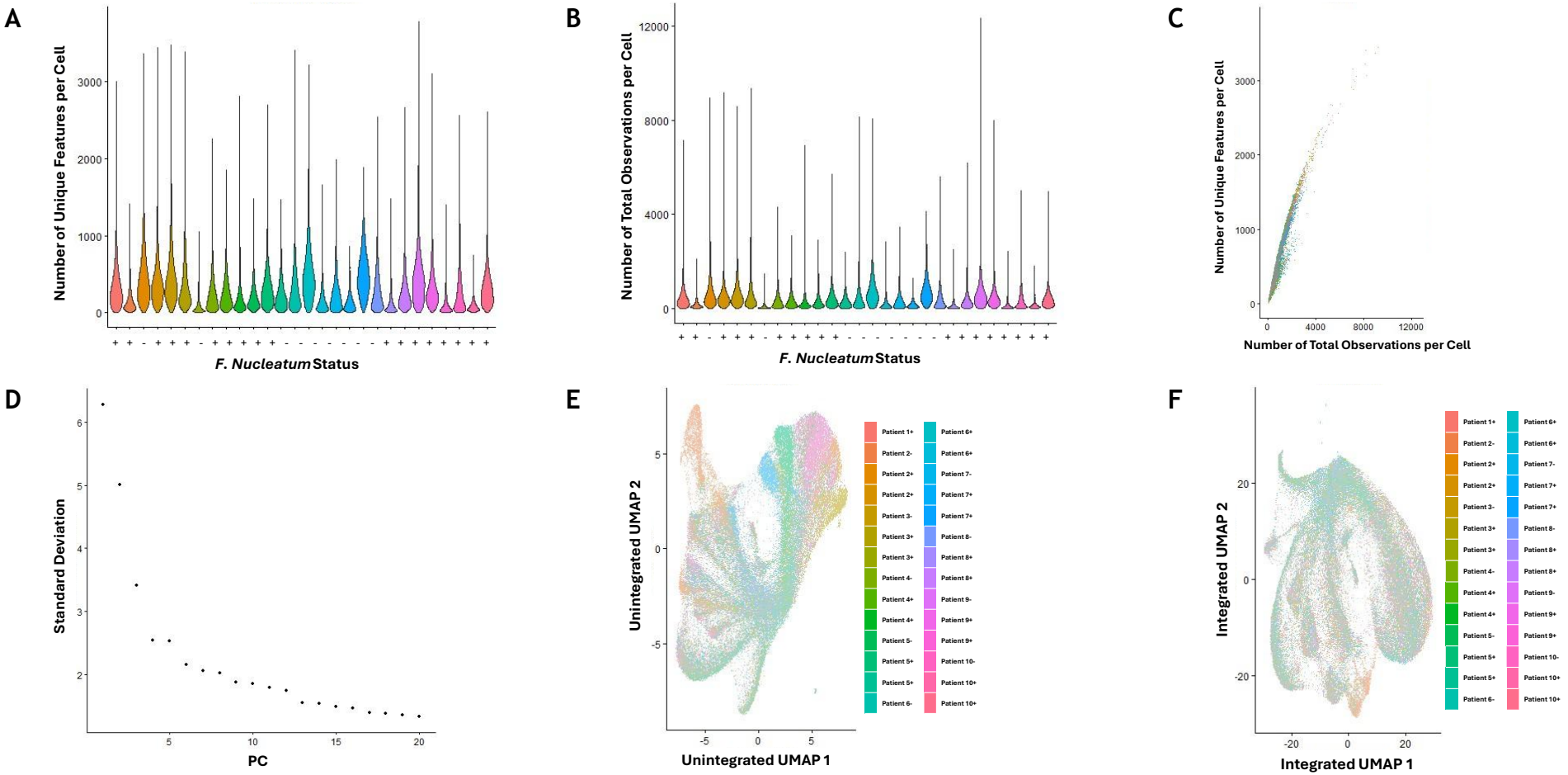


Figure 3.38. Quality Check, Normalisation, and Clustering of CosMx® Dataset.
[A] Violin plot of the number of unique genes (features) detected per cell and sample. **[B]** Violin plots of the total number of expression observations per cell. **[C]** Scatterplot between number of counts and number of features per cell. Cells from the same sample are the same colour. **[D]** PCA scree plot of PC1-PC20. **[E]** UMAP from unintegrated PCA. **[F]** UMAP from integrated PCA. This quality check was performed by John Cole from Shared Research Facilities.

3.8.3 Cell Typing

Cell types between clusters were identified as described in 2.2.13.4. Figure 3.39 shows a heatmap of the clusters by markers and the untyped clusters of cells on the tissue. As there was a clear pattern of histological association between clusters, the most expressed markers of each cluster were used for manual curation of cell types (Table 3.3). Based on these markers, five immune cell types (B cells, macrophage-dendritic-monocytes, plasma B cells, immune T-NK cells, and mast cells), three epithelial cell types (epithelial, intestinal epithelia, and epithelial inflammation cells), two stromal cell types (myo-fibroblasts and stromal, and smooth muscle fibroblasts) were identified.

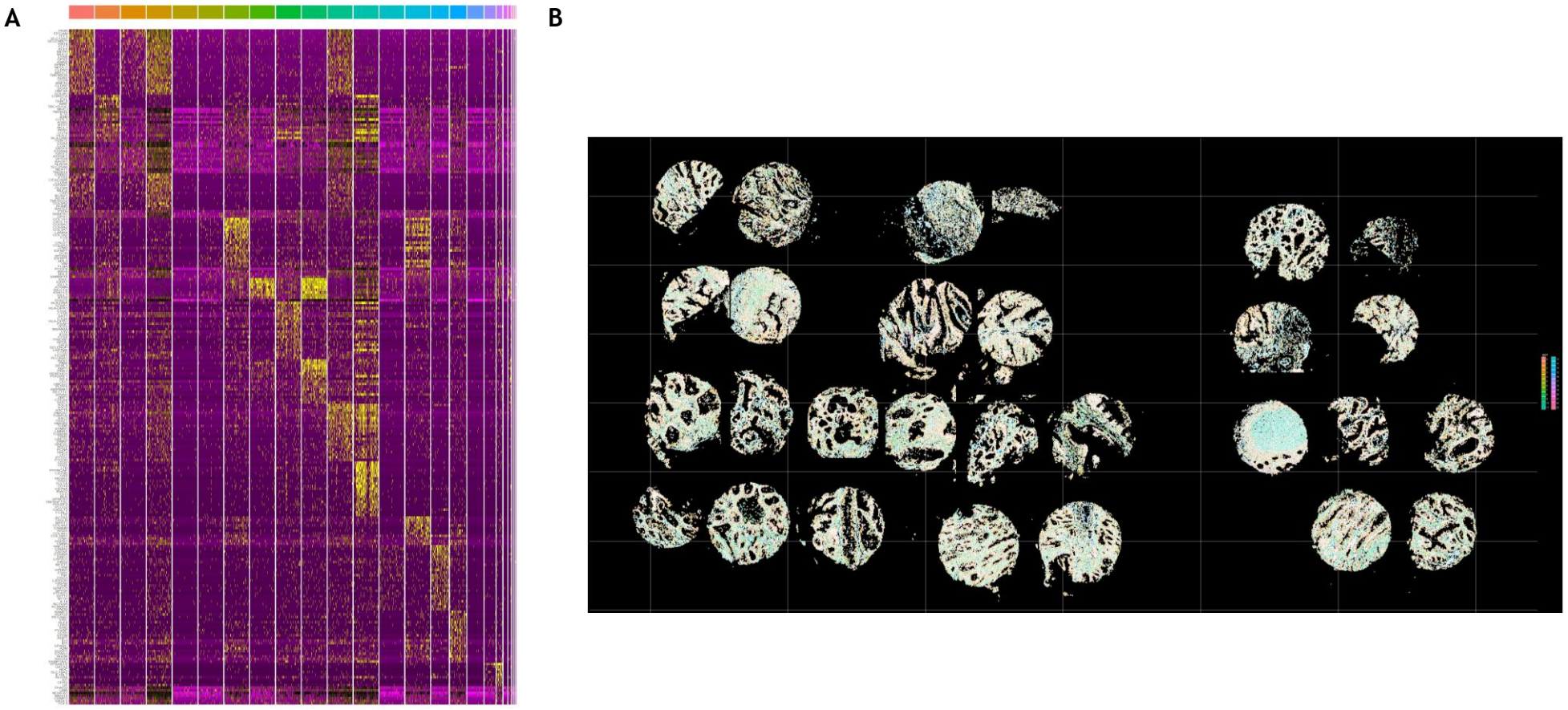


Figure 3.39. Cell Clustering.
[A] Heatmap of markers and untyped cell clusters colour-coded above the heatmap. Yellow denotes highly expressed markers, and purple denotes low expressed markers. **[B]** Image of CosMx® slide with untyped clusters overlaid on the cores.

Table 3.3. Main Markers Used for Cell Typing of Clusters

Cell Type and Colour Code		Markers for Cell Typing				
	B Cells	CLU	MS4A1	CR2	CLU	MS4A1
		CD22	TCL1A	LTF	CD22	TCL1A
		IRAG2	CD37	PAX5	IRAG2	CD37
	Intestinal Epithelial	OLFM4	SLC12A2	PIGR	OLFM4	SLC12A2
		LCN2	MUC4	AXIN2	LCN2	MUC4
		MUC2	AGR2	KLF5	MUC2	AGR2
	Myo-Fibroblasts & Stromal	F3	COL6A3	COL6A1	F3	COL6A3
		CXCL14	LAMA4	COL3A1	CXCL14	LAMA4
		COL6A2	DCN	COL1A2	COL6A2	DCN
	Epithelial	COX2	PIGR	COX1	TFF3	EPCAM
		S100A6	SLC25A6	ALDOA	RACK1	NEAT1
		VPS29	ATP5IF1	SRSF6	SRSF2	TMSB10
	Macrophage-Dendritic-Monocytes	STAB1	C1QC	HLA-DRB	HLA-DRA	HLA-DPA1
		CD74	PSAP	CTSB	SAT1	LYZ
		COTL1	GPX1	TMSB4X	B2M	MHC 1
	Endothelial	RAMP3	HSPG2	PECAM1	ENG	KLF2
		IGFBP7	VIM	ID1	ID3	HSPB1
		RASA4	NPDC1	LMNA	TMSB10	MHC 1
	Plasma-B	IGLC1/2	IGHA1	IGLL5	JCHAIN	IGHG1/2
		IGKC	MZB1	DERL3	IGLL1	PDK1
		PIM2	BTG2	XBP1	HERPUD1	OGT
	Immune-T-NK	TRBC2	EVL	MAF	TBC1D10C	CORO1A
		IL32	COTL1	HLA-F	B2M	MHC 1
		MCL1	TMSB4X	PFN1	A1BG	H2BC5
	Neuron-Glial	DRD1	PTPRZ1	CNTN5	CBLN2	NEUROG3
		DSG3	HIGD1B	EDNRA	LY96	FHIT
		CRHR2	OPHN1	BEST1	TRPC6	L2HGDH
	Smooth muscle fibroblasts	ACTA2	TAGLN	MYL9	COL1A1	TPM2
		COL6A1	PTEN	A1BG	GPX1	
	Epithelial Inflammation	COX2	S100A6	AIBG	ATP5IF1	PTEN
		PIGR	NEAT1	SNRPA1	ALDOA	
	All	SNRPA1				
	Mast	TPSAB1/2	HDC	SLC18A2	GATA2	AIBG
		DUSP2	GPX1			

The 13 distinct cell types were used for further analysis (Figure 3.40).

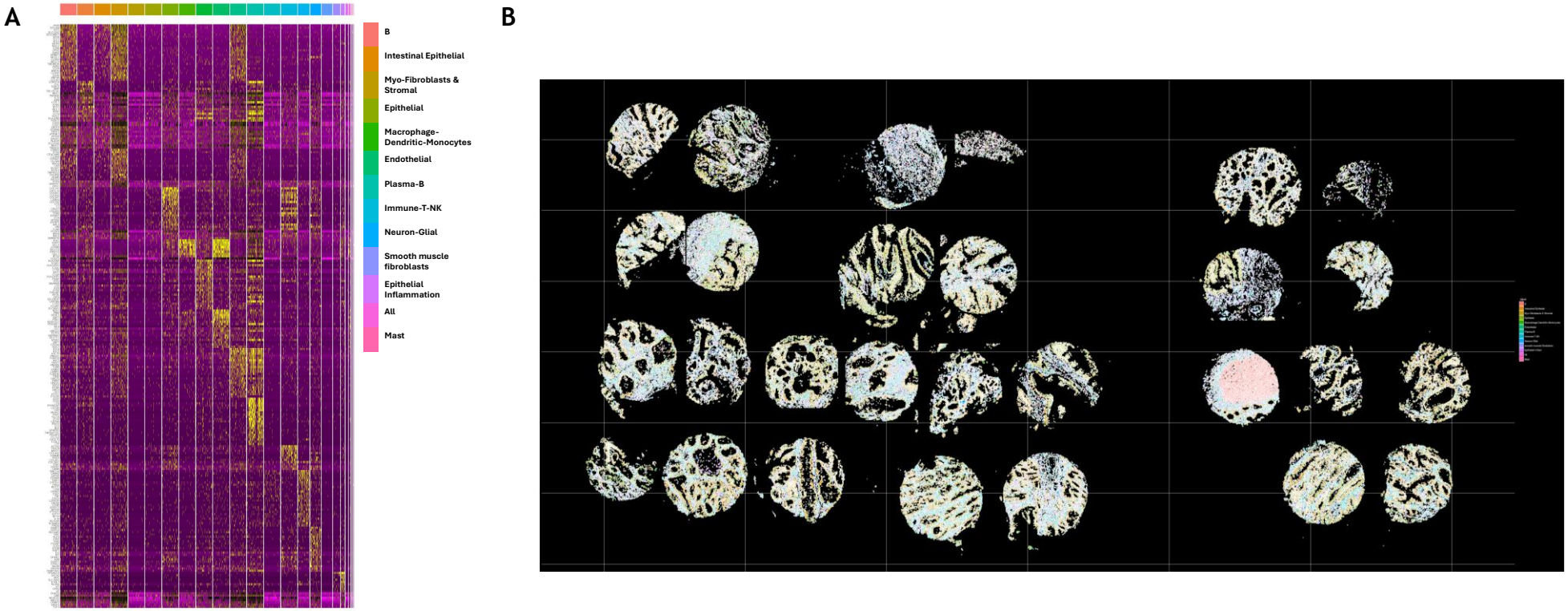


Figure 3.40 Cell Clustering with Cell Typing.
[A] Heatmap of markers and typed cell clusters colour-coded above the heatmap. Yellow denotes highly expressed markers, and purple denotes lowly expressed markers. **[B]** Image of CosMx® slide with typed clusters overlaid on the cores.

3.8.4 Pseudo Bulk Analysis by Spatial Region

To investigate the effects of *F. nucleatum* spatially, each core was divided into circles of increasing diameters starting from the centroid of each core (where the *F. nucleatum* is likely located). As shown in Figure 3.41, all areas within the same circle from all the *F. nucleatum*+ cores were combined, with the same treatment being applied to the next circle up, all the way to the widest, furthest circle from centre. This allowed for the *F. nucleatum*+ zones to be explored regardless of which patient the tissue was from. Although the *F. nucleatum*- cores were not centred around bacteria like the *F. nucleatum*+ cores, it was still important to use the concentric zones in these cores. This approach allowed for a truer comparison as the same zone in *F. nucleatum*+ and *F. nucleatum*- cores would have a similar cell density because the zones were of the same area. If a different approach was used, for instance, if the concentric zones were used only in the *F. nucleatum*+ cores and the *F. nucleatum*- cores were treated as one singular zone, the *F. nucleatum*- cores would have a much larger number of cells than the number of cells in any single zone in the *F. nucleatum*+ cores. Large differences in cell numbers in the *F. nucleatum*+ and *F. nucleatum*- will affect statistical power, increasing the risk of false positives, as well as heavily skewing the results. This analysis was done on a “by-gene” basis, with gene expression explored at increasing distance from the *F. nucleatum*.

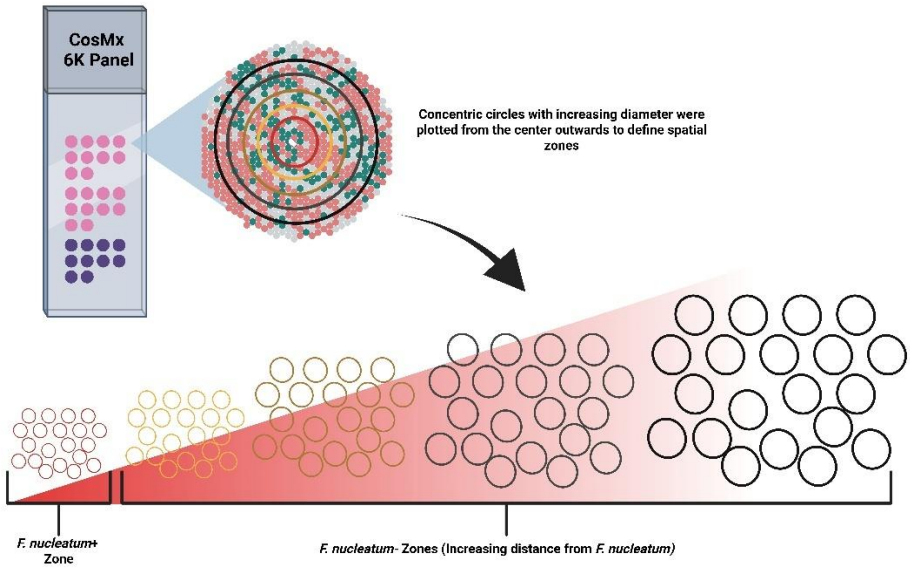


Figure 3.41. Schematic of Pseudo Bulk Analysis by Spatial Region.
All cores were divided into spatial zones based on concentric circles radiating from the centroid of the core. For the *F. nucleatum*+ cores, the centroid of the core is where the bacterium was detected using RNAscope®. The zones were generated in 1 FOV increments. The same zone from the different *F. nucleatum*+ cores was analysed as one, and so were the same zones from the different *F. nucleatum*- cores. The white area in the upper part of the schematic represents the *F. nucleatum* location. The number of zones shown in this figure is not the exact number of zones generated. This figure is only a schematic representation.

Compared to *F. nucleatum*+ cores, 16 genes were highly expressed in *F. nucleatum*- cores (Figure 3.42). When compared to the different zones of the core, *AKT2, ENSA, MBD3, MYEF2, NFYC, NNT, PABN1, RBM17,* and *SAE1* were all expressed higher at the centre of the core of the *F. nucleatum*- cores. The expression of *ARHGEF7, NDUFA11, PPP2R5C, SLC38A1,* and *YBX1* increased with increasing distance from the centroid of the *F. nucleatum*- cores. *IST1* and *STT3B* displayed varying expression levels across the core. It was interesting however not surprising to see the varying transcript levels of genes across the different zones in the *F. nucleatum*- cores. Although these observations are likely not influenced by *F. nucleatum*, the zones have different cell populations, and the cells making up those populations will have different transcript levels, account for the differences seen.

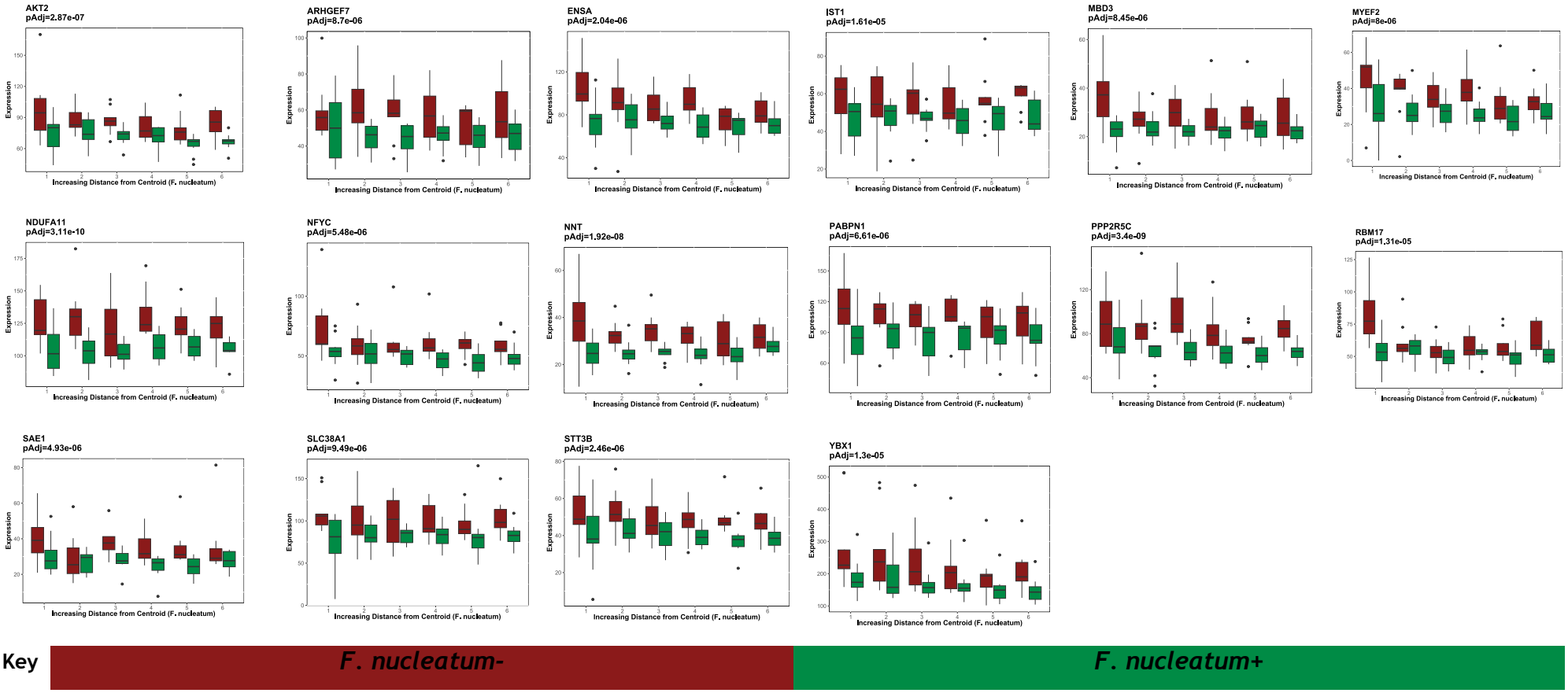


Figure 3.42. Genes Upregulated in the *F. nucleatum*- Cores.
A variable expression pattern of these genes was seen through the different zones of the core, with some genes being expressed more in the innermost zone, and others being expressed more in the outermost zones. INCISE patients n=10, with 2 cores from each patient for the *F. nucleatum*+ cores (n=20 cores) and 1 core for the *F. nucleatum*- cores (n=10 cores).

Thirty-three genes were overexpressed in the *F. nucleatum*+ cores. Of those, nine (Figure 3.43) were more highly expressed in the innermost zone (which contains *F. nucleatum*): *APIP*, *CASP10*, *DUSP2*, *FAP*, *FGF20*, *HK2*, *PHKG2*, *SLC10A6*, and *SLPI*.

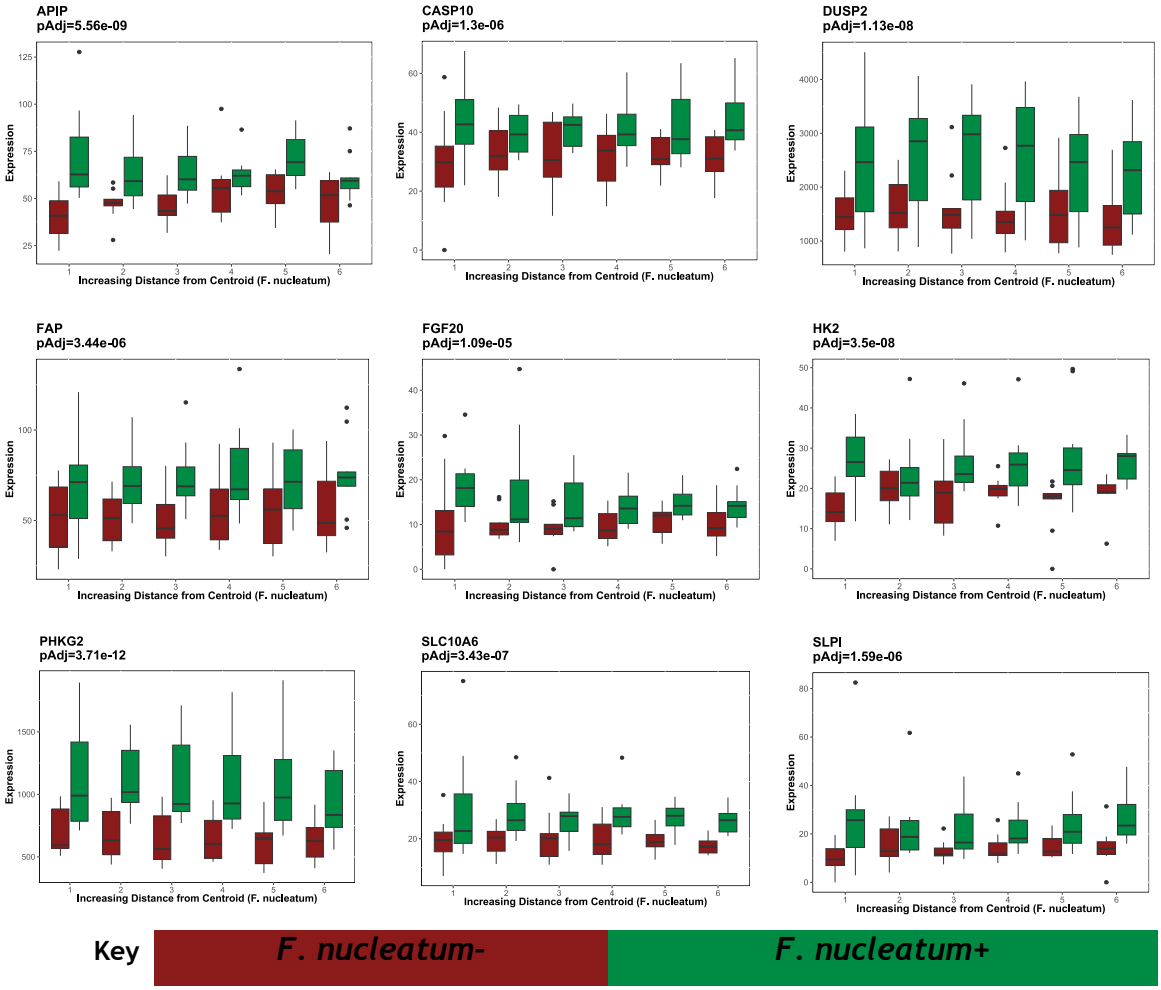


Figure 3.43. Genes Upregulated in the *F. nucleatum*+ Cores.
These genes were expressed higher in the spatial zones close to *F. nucleatum*. INCISE patients n=10, with 2 cores from each patient for the *F. nucleatum*+ cores (n=20 cores) and 1 core for the *F. nucleatum*- cores (n=10 cores).

Twenty-six genes were overexpressed in the outer zones, further away from the location of the *F. nucleatum* (Figure 3.44). *CDKN1A*, *GIPC1*, *IL22RA1*, *KRT6A*, *KRT8*, *KRT19*, *MVP*, *OXR1*, *PON2*, *SFN*, *SNRPA1*, and *TFF1* were all expressed at a higher rate in the outermost zones, while the remaining genes showed more elevated expression towards the middle of the cores.

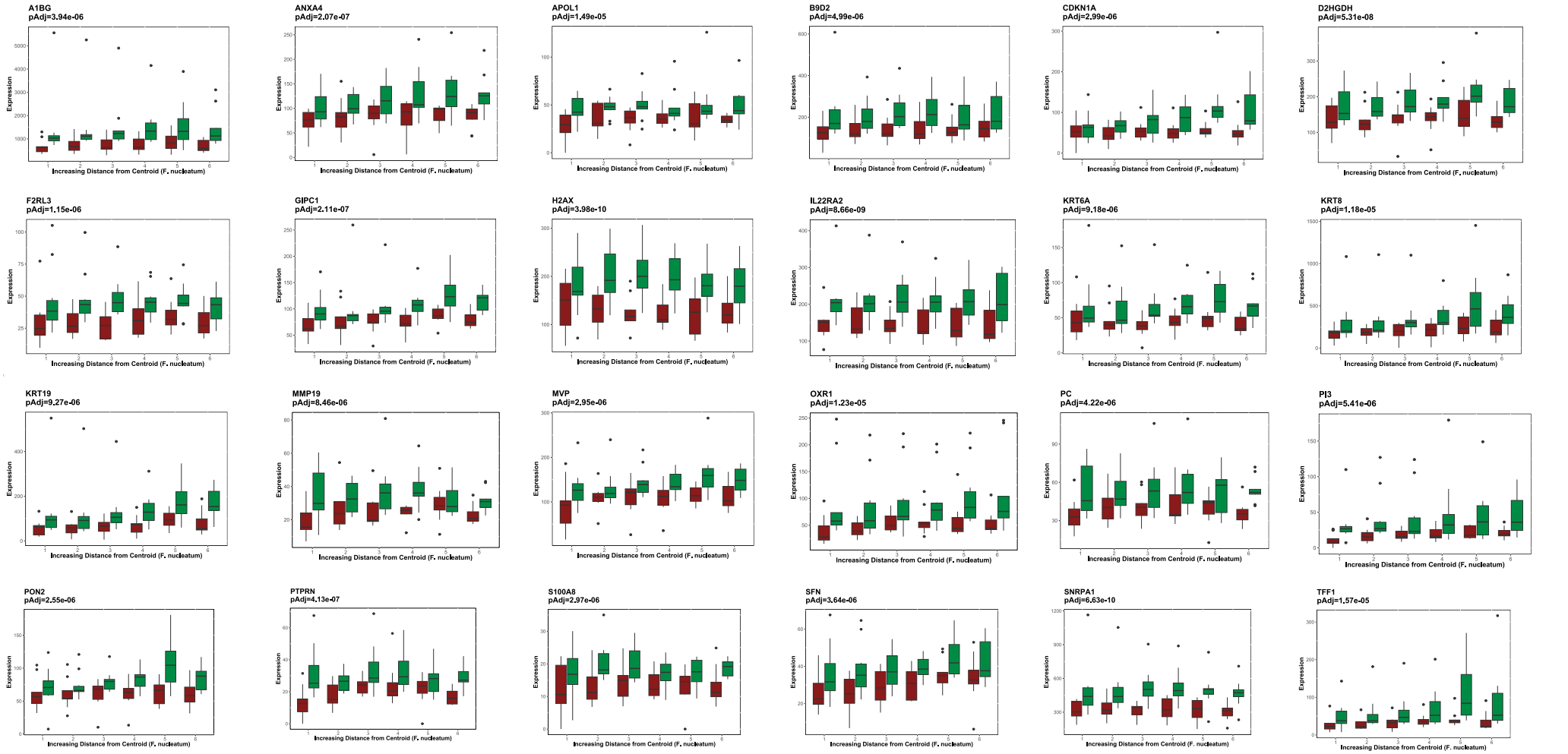


Figure 3.44. Genes Upregulated in the *F. nucleatum*+ Cores.
These genes were expressed higher in the spatial zones away from *F. nucleatum*. INCISE patients n=10, with 2 cores from each patient for the *F. nucleatum*+ cores (n=20 cores) and 1 core for the *F. nucleatum*- cores (n=10 cores).

Figure 3.45 shows a “per-cell” demonstration of the spatial expression of selected genes on the *F. nucleatum*+ and *F. nucleatum*- cores from the same patient (Patient 10). The expression based on spatial zones depicted in the boxplots is similar to the spatial distribution.

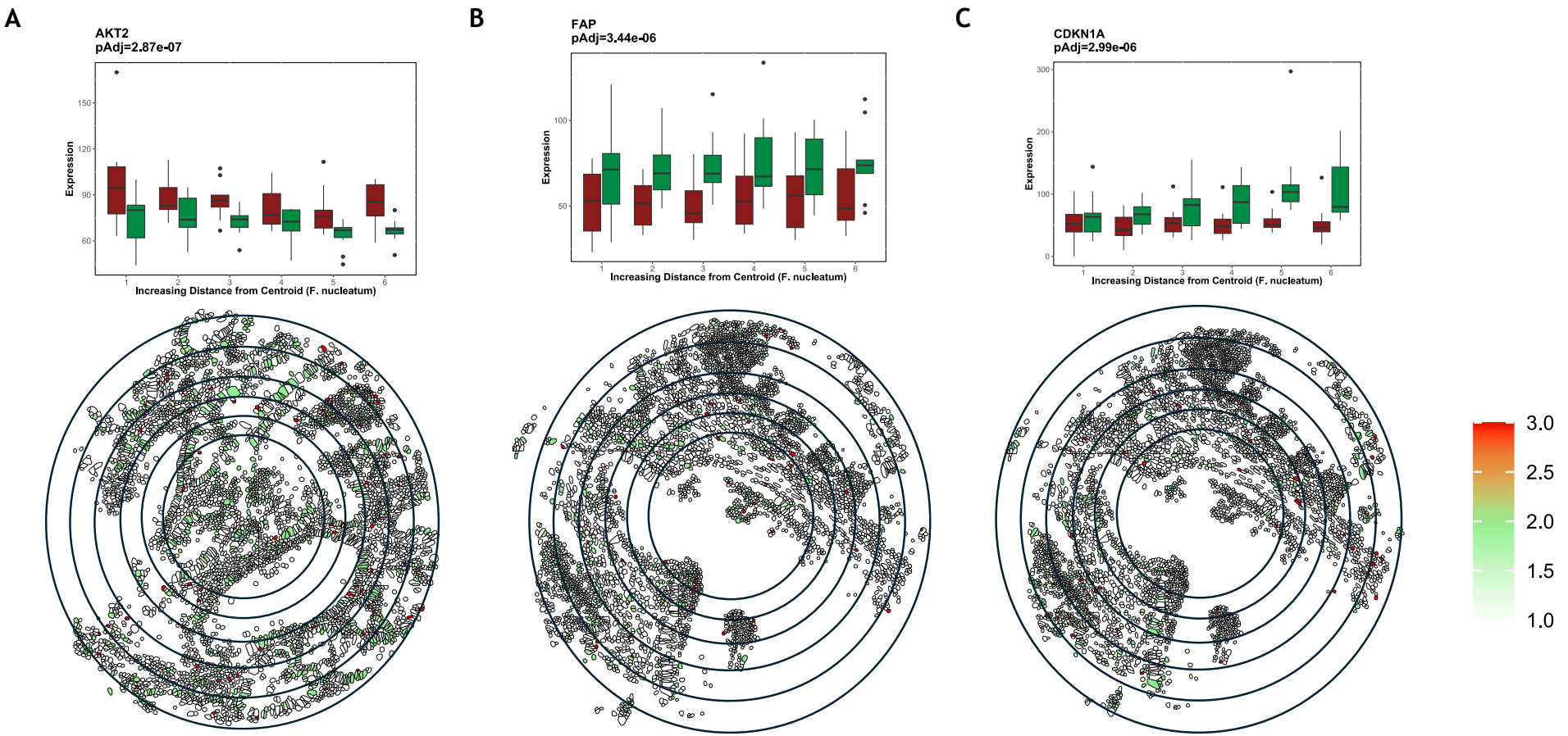


Figure 3.45. "Per-cell" Gene Expression Levels of Selected Genes.
[A] Expression of *AKT2*, which is upregulated in the *F. nucleatum*- core of Patient 10, as shown in Figure 3.42 (upper), corresponds spatially to the expression in the cells in the different spatial zones (lower). **[B]** Expression of *FAP*, which is upregulated in the zones closest to the centroid *F. nucleatum*+ core of Patient 10, as shown in Figure 3.43 (upper), corresponds spatially to the expression in the cells in the different spatial zones (lower). **[C]** Expression of *CDKN1A*, which is upregulated in the zones furthest from the centroid in the *F. nucleatum*+ core of Patient 10, as shown in Figure 3.44 (upper), corresponds spatially to the expression in the cells in the different spatial zones (lower).

3.8.5 Neighbourhood Analysis

Neighbourhood analysis was then performed to identify the members of the most abundant neighbourhoods in each zone. As shown in Figure 3.46, ten neighbourhoods were identified with varying levels of cell types populating each one. Neighbourhood 1 was rich in intestinal epithelial cells, followed by plasma-B cells and general epithelial cells. This was similar to Neighbourhood 2, which was mostly comprised of intestinal epithelial cells, epithelial cells and macrophage-dendritic-monocyte cells. Neighbourhood 3 was mostly populated with plasma-B cells and immune-T-NK cells, interspersed with inflamed epithelial cells. Neighbourhood 4 mostly consisted of epithelial cells (intestinal and general) with macrophage-dendritic-monocyte cells. Neighbourhood 5 was scarcely populated with some neuronal-glial cells, and smooth muscle fibroblasts, while Neighbourhood 6 consisted mainly of intestinal epithelial cells, epithelial cells, and fibroblasts. Neighbourhood 7 consisted of intestinal epithelium and epithelium with plasma-B cells as the second most abundant cell type. Neighbourhood 8 had a high inflammation signature, with B cells, immune-T-NK cells, and inflamed epithelial cells being the three most abundant cell types. Neighbourhood 9 was dominated by epithelial and smooth muscle fibroblasts, while Neighbourhood 10 had an abundance of intestinal epithelial cells, epithelial cells, and macrophage-dendritic-monocytes.

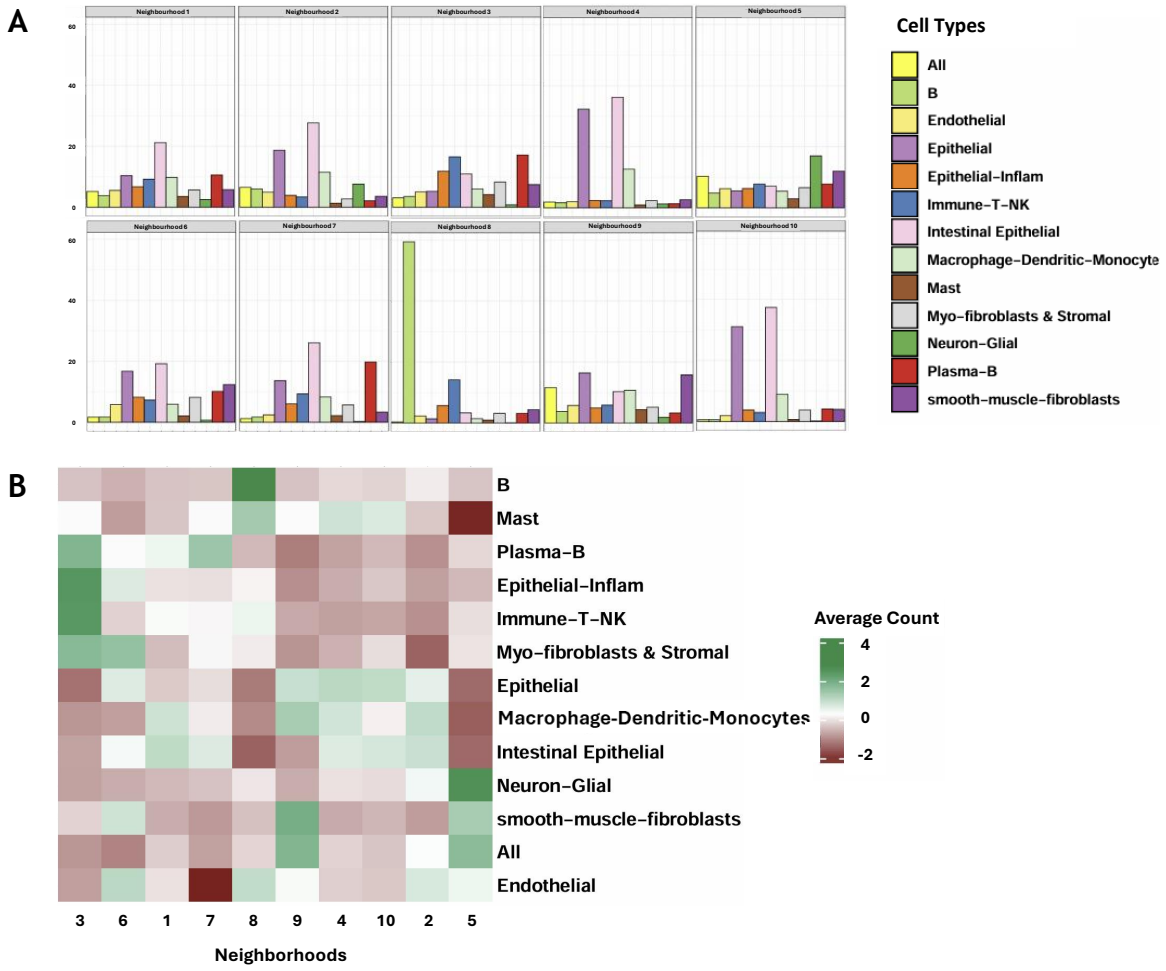


Figure 3.46. Neighbourhood Analysis of Cell Types.
[A] Bar charts of the constituents of each of the 10 identified neighbourhoods. **[B]** Clustered heatmap of the distribution of each cell type within each neighbourhood. NK: Natural Killer.

Once the neighbourhoods were identified, the most common neighbourhood within each spatial zone (as defined in 3.8.1) was investigated. Within the *F. nucleatum*- cores (Figure 3.47A), in the first five zones the immune-rich Neighbourhood 8 was represented most, with the epithelium rich Neighbourhoods 1 and 2 being the most abundant in zones 6 and 7. Finally the epithelial-rich and plasma-B cell rich Neighbourhoods 3 and 7 dominated zone 8.

The *F. nucleatum*+ cores showed a significant shift away from immune-rich environments. With the exception of very low presence in zones 2 and 3, the other zones (especially the innermost zone) showed no presence of the immune-dominant Neighbourhood 8. The zones within these cores were populated by epithelial cell-rich neighbourhoods.

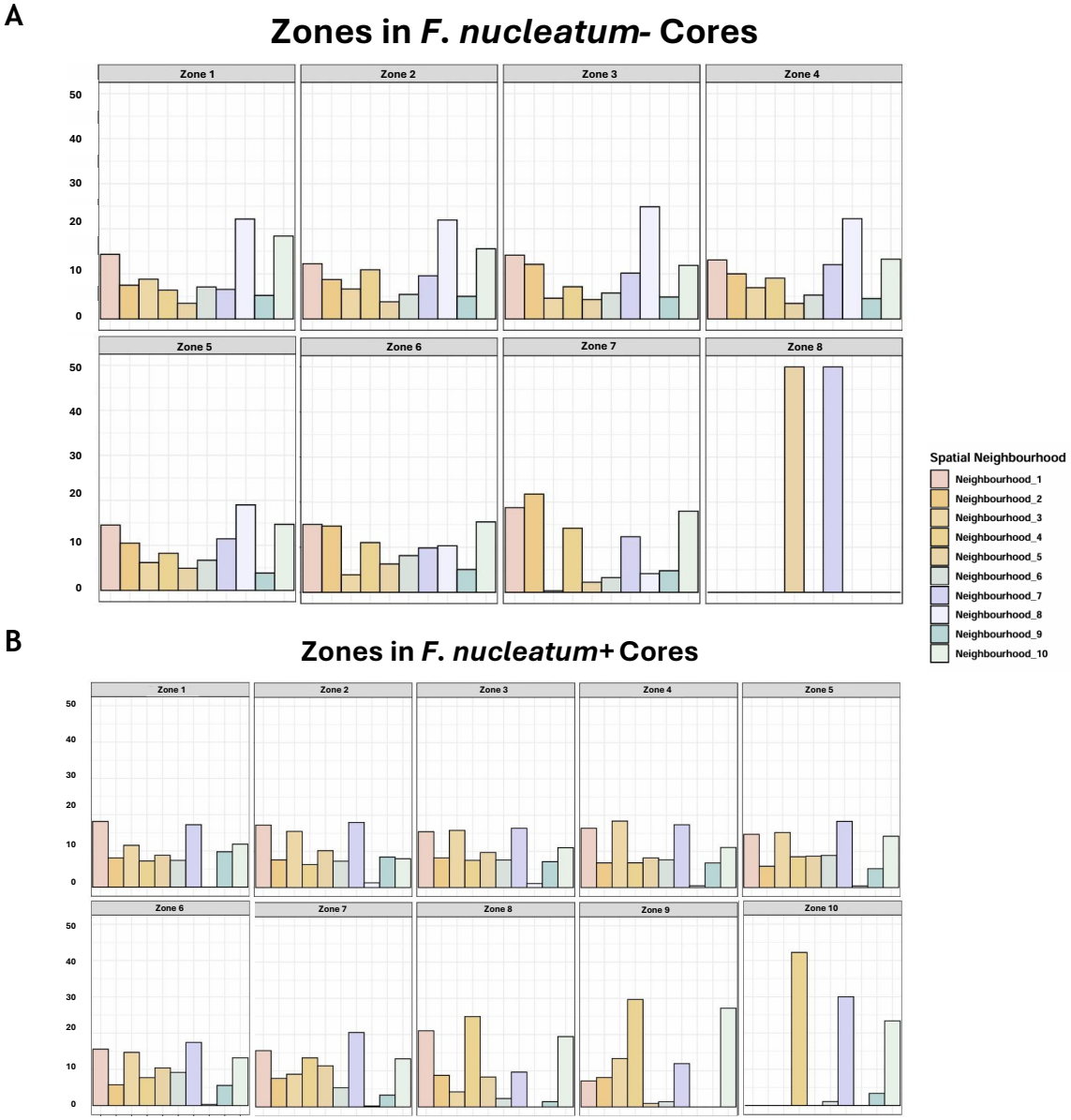


Figure 3.47. Neighbourhood Distribution Across Spatial Zones.
[A] Neighbourhoods distributed within zones in the *F. nucleatum*- cores. **[B]** Neighbourhoods distributed within the zones of the *F. nucleatum*+ cores.

3.8.6 Pathway Analysis

Pathway analysis was performed to investigate whether the presence of *F. nucleatum* has any effect on the regulation of signalling pathways in the areas surrounding *F. nucleatum*. Figure 3.48A shows that within the spatial zones in the *F. nucleatum*- cores, the highest pathway activity was seen in the zone furthest away from the centre. Within that zone, Wnt signalling, VEGF, p53, and NFκB pathways were most upregulated. Within the same zone, there was downregulation of the immune-related JAK-STAT, TGFβ, and TRAIL pathways. This is reflected in Figure 3.48B, where the mean expression for Wnt, VEGF, p53, and NFκB is highest in the 8th spatial zone., while it is lowest for JAK-STAT, TGFβ, and TRAIL. In the *F. nucleatum*+ cores, no pathways were altered (Figure 3.48B). However, stronger regulation trends were seen in the zones further away from the central *F. nucleatum*+ zone. In zones 7-10, the hypoxia pathway was upregulated with no pathways showing an obvious downregulation. This is supported by the mean expression in Figure 3.48C, where the average mean expression of the downregulated pathways is -0.1.

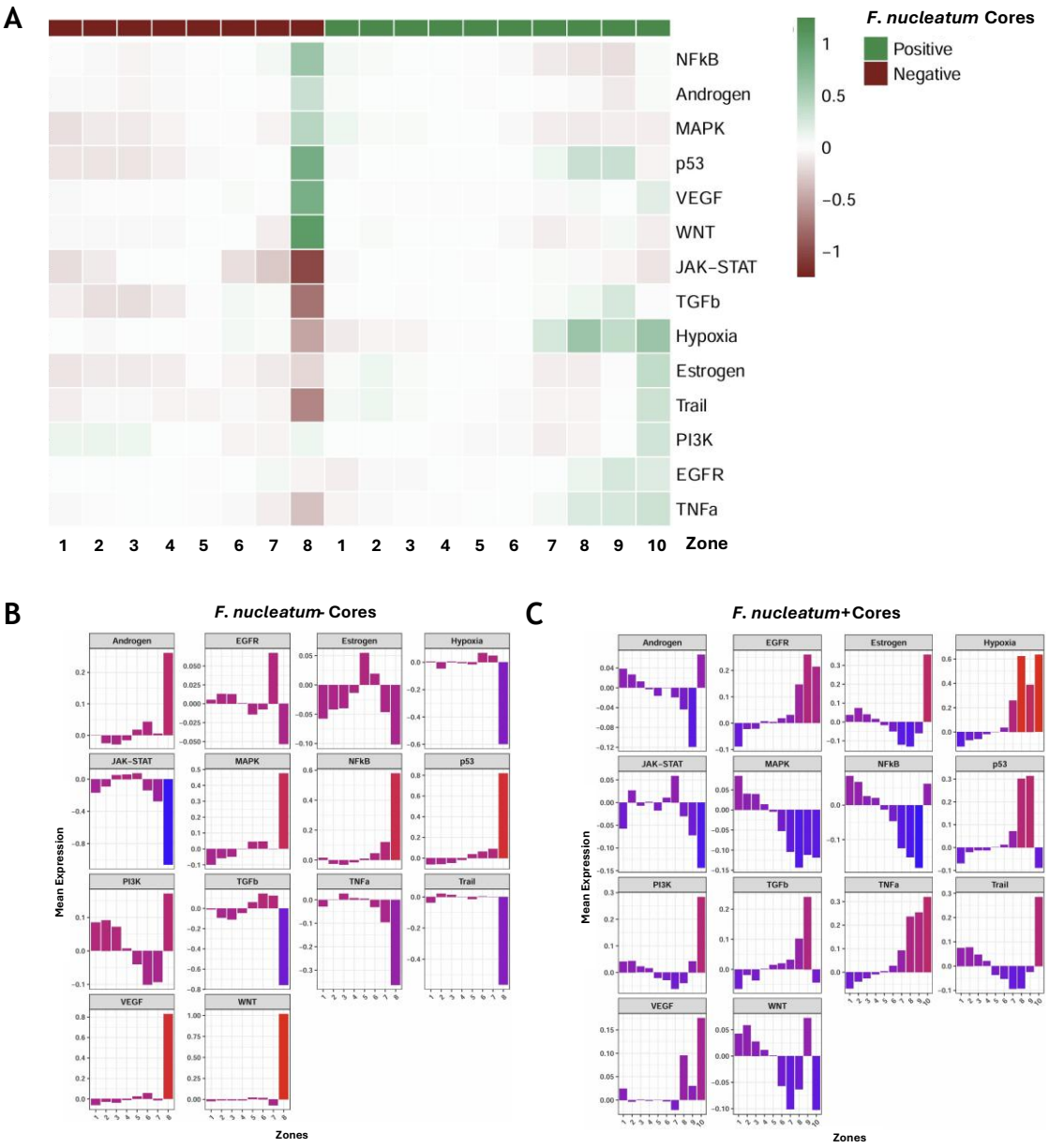


Figure 3.48. Commonly Investigated Cancer Pathway Analysis of the Spatial Zones. **[A]** Heatmap of the pathways of interest and their regulation patterns within the spatial zones of the *F. nucleatum*- and *F. nucleatum*+ cores. **[B]** Bar charts of the mean expression of the pathways of interest within each zone of the *F. nucleatum*- cores. **[C]** Bar charts of the mean expression of the pathways of interest within each zone of the *F. nucleatum*+ cores.

3.9 *F. nucleatum* Is Not Associated with the Detection of Metachronous Polyps or CRC

With the novel bacterial detection technique validated and the genomic and transcriptomic profiles revealing interesting associations between *F. nucleatum* and adenomas, it was time to reveal any associations between *F. nucleatum* status and the detection of metachronous polyps or CRC.

Statistical analysis on the subset of INCISE patients with TempO-Seq™-based *F. nucleatum* status was carried out to investigate whether *F. nucleatum* status was associated with the detection of metachronous polyps or CRC, to reveal any associations between the presence of *F. nucleatum* and clinical factors, and if suitable, to examine the effects its addition to current BSG2020 Guidelines for risk stratification might have.

Survival analysis (2.2.10) revealed no association between *F. nucleatum* status and the detection of metachronous polyps (p=0.51, Figure 3.49). While *F. nucleatum*- patients had a median time to detection of metachronous polyps or CRC of 54 months, *F. nucleatum*+ patients did not reach median cumulative risk. Although patients within this subset who had 5 or more index polyps were at least twice as likely to have metachronous polyps or CRC (p=0.002), no other clinical factors, or *F. nucleatum* status were associated with the detection of metachronous polyps or CRC upon univariate Cox regression analysis (Table 3.4).

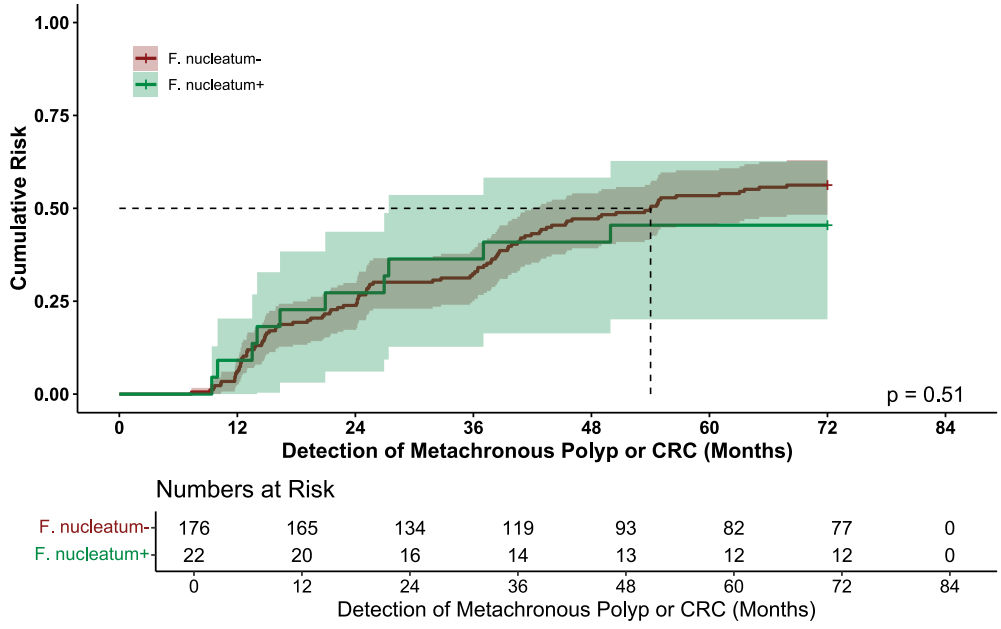


Figure 3.49. *F. nucleatum* Status in Colorectal Polyps.
The presence of *F. nucleatum* was not associated with the detection of metachronous polyps or CRC. Shading is indicative of 95% CI. Dotted lines are time to median risk. N=198.

Table 3.4. *F. nucleatum* Status in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps of CRC

		Univariate		
		HR	95% CI	P
Sex	Female (n=64)	1.0	-	-
	Male (n=134)	0.862	0.700-1.061	0.160
Age	Screening Age (n=188)	1.0	-	-
	Above Screening Age (n=10)	0.775	0.316-1.902	0.578
Site	Right Colon (n=27)	1.0	-	-
	Left Colon (n=145)	0.722	0.426-1.222	0.225
	Rectum (n=25)	0.780	0.426-0.384	0.491
High Grade Dysplasia	Absent (n=161)	1.0	-	-
	Present (n=37)	0.699	0.416-1.173	0.175
Number of Polyps	1 (n=66)	1.0	-	-
	2-4 (n=96)	1.140	0.731-1.777	0.563
	5+ (n=36)	2.240	1.338-3.749	**0.002**
Polyp Type	Adenoma (n=189)	1.0	-	-
	Non-adenoma* (n=9)	0.926	0.377-2.272	0.867
BSG2020 Guidelines	Low Risk (n=73)	1.0	-	-
	High Risk (n=125)	1.427	0.954-2.134	0.084
F. nucleatum Status	Negative (n=176)	1.0	-	-
	Positive (n=22)	0.805	0.420-1.544	0.515

* Non-adenoma refers to all polyps with non-conventional adenoma features. This includes serrated polyps, hyperplastic polyps, etc. Screening Age=50-74 years old. HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted bold. N=198. Cox regression analysis was carried out to obtain hazard ratios.

3.9.1 *F. nucleatum* Status and Clinical Characteristics

To explore associations between *F. nucleatum* status and clinical characteristics, X^2 analysis was performed on the subset of INCISE patients with TempO-Seq™-based *F. nucleatum* status (Table 3.5). Despite small numbers, a statistically significant association between *F. nucleatum*+ patients and polyps with non-adenoma features (such as serrated or hyperplastic polyps) was identified (p=0.003). No associations were observed with other clinical factors, although high-grade dysplasia in *F. nucleatum*+ patients trended towards significance (p=0.094; Table 3.5, Figure 3.50).

Table 3.5. Relationship between *F. nucleatum* Status and Clinical Characteristics in the INCISE Cohort Subset with TempO-Seq™ based *F. nucleatum* Status

		F. nucleatum Status (n=198)				
		Negative		Positive		P
		176	89%	22	11%	
Sex	Female	57	32%	7	32%	0.957
	Male	119	68%	15	68%	
Age	50-74 (Screening Age)	167	95%	21	96%	0.909
	75+ (Above Screening Age)	9	5%	1	4%	
Site	Right Colon	23	13%	4	18%	0.734
	Left Colon	129	74%	16	73%	
	Rectum	23	13%	2	9%	
High Grade Dysplasia	Absent	146	83%	15	68%	0.094
	Present	30	17%	7	32%	
Number of Polyps	1	59	33%	7	32%	0.156
	2-4	82	47%	14	64%	
	5+	35	20%	1	4%	
Polyp Type	Non-adenoma*	6	3%	4	18%	**0.003**
	Adenoma	170	97%	18	82%	
BSG2020 Guidelines	Low Risk	65	37%	8	36%	0.958
	High Risk	111	63%	14	64%	
Metachronous Polyps or CRC Detection	No	77	44%	12	55%	0.337
	Yes	99	56%	10	45%	

* Non-adenoma refers to all polyps with non-conventional adenoma features. This includes serrated polyps, hyperplastic polyps, etc. P-values were generated by Pearson' Correlation test. Site information was not available for 1 patient in the *F. nucleatum*- group, hence information on site for the *F. nucleatum*- group was included for n=175. Chi-squared analysis was carried out to obtain correlations between clinical factors and *F. nucleatum* groups.

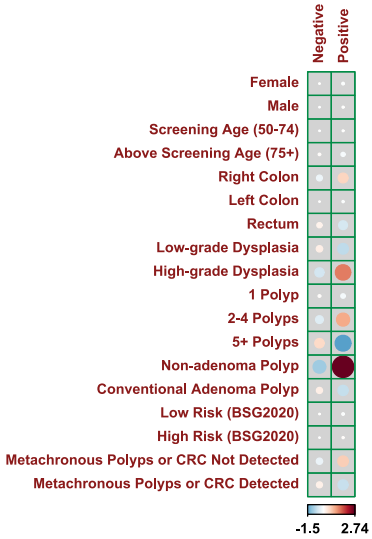


Figure 3.50. Corrplot of Associations Between *F. nucleatum* Status Against Clinical Characteristics.
F. nucleatum- and *F. nucleatum*+ in the subset of INCISE patients with TempO-Seq™ -based *F. nucleatum* status. Dot size suggests a high correlation between variables, while colour intensity indicates negative (blue) or positive (red) associations. N=198.

As *F. nucleatum* status was not associated with the detection of metachronous polyps or CRC, further analysis combining it with current BSG2020 Guidelines was not performed.

3.10 Discussion

Bacterial detection has been at the forefront of medical diagnosis for centuries, and methods to identify bacterial pathogens have evolved from classification by metabolites or cell wall structure, to identification by sequencing techniques (Sidebottom, 2023). The microbiome plays a crucial role in cancer and has been considered one of its enabling characteristics (Hanahan, 2022). This chapter aimed to detect bacteria in FFPE archival polyp tissue and investigate its association with the development and detection of metachronous polyps.

As the focus of this work was the Gram-negative pathogen, *F. nucleatum*, Gram staining was performed on fresh cut archival polyp tissue as a first step to detection. However, no bacteria of any Gram classification were seen. This is not surprising, as recent work on skin tissue has shown that Gram staining on FFPE tissue is not very sensitive (Stagg, Foster-Smith and Tam, 2024).

FFPE is the gold standard in long-term storage of human tissue specimens, and while routinely used for diagnostic H&E and immunohistochemistry staining, RNA extracted from FFPE tissue is notorious for its poor quality and low yield (Evers *et al.*, 2011). RNA-Seq uses extracted RNA to form cDNA libraries that are sequenced into short reads, which can be aligned to reference material, or used as a *de novo* discovery tool (Y. Han *et al.*, 2015). Although this makes RNA-Seq an excellent method for non-targeted transcriptomic studies, challenges have been identified for the use of RNA-Seq. For example, the need for fragmentation of larger RNA molecules introduces bias during library construction. In addition, targeting full genes tends to produce very large amounts of raw data, that require large storage. Deeper sequencing depth translates to greater coverage, leading to more accurate and inclusive results. RNA-Seq continues to be a very expensive approach to transcriptome analysis, despite the recent advances and the reduction in costs (Wang, Gerstein and Snyder, 2009).

TempO-Seq™ works by targeting a specific region in a gene, of every gene, and does so directly from lysates prepared from a variety of starting materials (Yeakley *et al.*, 2017). This not only circumvents issues with library preparation and data storage but also avoids the need for RNA extraction from FFPE tissue, making it a useful non-discovery tool for targeted applications. Previous work demonstrated that TempO-Seq™ is as sensitive in fresh lysates as it is in lysates from fixed tissue and that the output quality was not affected by tissue age or fixation time (Trejo *et al.*, 2019). Turnbull et al., (2020) concluded that for FFPE tissue, TempO-Seq™ was among the best new platforms for expression profile generation.

Results for the detection of *F. nucleatum,* using TempO-Seq™ are presented here, where bespoke probes were considered sensitive and specific and were able to detect their target in FFPE polyp tissue. Quality check experiments demonstrated that the newly designed *F. nucleatum* probes were able to exclusively detect their target in samples where purified *F. nucleatum* RNA was available, suggesting their specificity. Furthermore, they were sensitive to low quantities of their target RNA, as evidenced by the signal observed in samples where *F. nucleatum* RNA was diluted in human RNA (described in 3.2.1.2). This level of sensitivity is crucial, as research on CRC tissue suggests that no more than 13% of samples test positive for *F. nucleatum* (Mima, Cao, *et al.*, 2016; Mima, Nishihara, *et al.*, 2016). Interestingly, this was comparable to the percentage of *F. nucleatum*+ patients within the tested INCISE subset (11%).

This novel detection technique was validated by RNAscope®, where of the 10 available cases that underwent RNAscope®, only three were misclassified. The TempO-Seq™-based *F. nucleatum* status assignments were stringent, with only cases who had a signal from two probes above the identified threshold being considered *F. nucleatum*+. Since RNAscope® is a visualisation technique, it picked up very low bacterial loads. This does not allude to the TempO-Seq™ probes being poor detectors, but more so that the bacterial RNA signal was too low to pass the threshold.

The mutational landscape of the *F. nucleatum*- and *F. nucleatum*+ patients revealed some interesting results. *KRAS* was not among the top 10 most mutated genes in *F. nucleatum*+ samples. This is contrary to previous studies that

reported enrichment of *F. nucleatum* in *KRAS*-mutant CRC (Chen, Zheng and Liu, 2024; Zhu *et al.*, 2024). However, this could point to a major difference between colorectal adenomas and CRC, and although it is the result of a very small sample size, it warrants further investigation and validation in a larger sample group. However, there were mutations in immune-related genes like *HLA* genes which alludes to the possibility of an altered immune response.

Furthermore, host transcriptomic analysis identified DEGs in the *F. nucleatum*+ group. Of the upregulated genes, *CXCL1* and *CXCL2*, both members of the CXC chemokine subfamily, were in the top ten most upregulated genes and they function as neutrophil chemoattractants involved in pro-tumour inflammation (Zhou *et al.*, 2023). *SOD2* was linked to neutrophil infiltration and CRC progression (Su *et al.*, 2021), and *PSMA7* was shown to directly inhibit *NOD2* which functions to recognise bacteria and induce the release of relevant pro-inflammatory cytokines (Yang *et al.*, 2013). *RNF114* has been shown to be a T-cell regulator (Rodriguez *et al.*, 2014), and a promoter of cell proliferation and stemness in CRC cell line experiments (Cheng Liu et al., 2023).

The immune-supportive signature was reflected in GSEA. Of the GO terms, MSigDb Hallmarks gene sets, and KEGG pathways, ten members of the CXC chemokine subfamily, seven members of the CC chemokine subfamily, five interleukin family members, and three CD3 genes, were flagged in the leading-edge analysis. Further insights were gained into the nature of this immune-dominated transcriptomic profile from GO GSEA, as nine microbial terms relating to bacterial response and defence were positively enriched. *LTF*, the gene encoding lactoferrin, was shared across all nine terms. Lactoferrin is an important anti-microbial protein that defends against bacteria by a variety of mechanisms, such as chelating iron to deprive bacteria of its growth-restoring potential (Gruden and Ulrih, 2021).

Another commonly shared leading-edge gene, *SERPINE1*, has been shown to be protective against Gram-negative pneumonia infections (Renckens *et al.*, 2007), supporting the results presented here, where it is shared across anti-bacterial and defence terms. However, *SERPINE1* has been shown to promote CRC through Notch activation (Su *et al.*, 2024a). While *SERPINE1* could be acting in a protective capacity against *F. nucleatum* in adenoma tissue, this could lead to

poor outcomes, should the adenoma progress into malignancy, suggesting the presence of *F. nucleatum* could be a poor prognostic factor in adenomas due to the response it elicits.

Interestingly, *WNT5A* was shared across seven anti-microbial GO terms. Involved in various signalling pathways, such as non-canonical Wnt signalling, *WNT5A* has been implicated in many cancers, including CRC, in a context-dependent manner. For example, the Wnt5a molecule can support an immune-rich environment by binding to ROR1 and activating an Akt-NFκB cascade that leads to the secretion of cytokines like IL-6 (Lopez-Bergami and Barbero, 2020), which was a leading-edge gene in *F. nucleatum*+ patients, across GSEA.

An interesting link was observed between MSigDb Hallmarks and KEGG GSEA. Positively enriched was the “Reactive Oxygen Species (ROS) Pathway” gene set in MSigDb Hallmarks GSEA and the “Neutrophil Extracellular Traps (NETs) Formation” pathway in KEGG. First described by Brinkmann et al. (2004), NETs resemble webs and are protective structures that capture and kill pathogens. NETs form due to many different stimuli, one of which is induction by ROS (Azzouz, Khan and Palaniyar, 2021). *F. nucleatum* has been shown to induce ROS production in CRC cell line experiments (Tang *et al.*, 2016), and NETs were shown to be formed during sepsis through LPS from bacteria binding to TLR4, which leads to the activation of platelets as they associate with neutrophils (Clark *et al.*, 2007). While a *F. nucleatum*/ROS/NETs formation axis has been reported for CRC (Kong *et al.*, 2023), these results suggest an earlier involvement for *F. nucleatum* in ROS-induced NETosis.

The upregulation of replication fork regulator *RTF2* and downregulation of DNA methyltransferase *DMAP1* and lipid kinase *AGK*, suggest a proliferative signature in *F. nucleatum*+ patients. *RTF2* is involved in cell growth and replication stress (Conti *et al.*, 2024), and both *DMAP1* —involved in DNA repair and stability (Negishi *et al.*, 2009)- and *AGK* —involved in homeostasis regulation – have tumour suppressing functions in cancer (Wang et al., 2016; Chu, Hong and Zheng, 2021; Sun et al., 2022). The differential expression of these genes supports the proliferative signature highlighted in GSEA from both MSigDb Hallmarks, and KEGG. As previously described in (1.6.2), *F. nucleatum* can influence cell signalling pathways involved in cell proliferation. This was

supported by the positive enrichment of cell cycle regulatory genes such as *CDK4 and MYC,* cyclin-encoding genes *CCNA2, CCNB2, CCNE1,* and *CCNE2*, as well as transcription factors *E2F1* and *E2F3*, all of which are involved in the decision to enter and commit to the cell cycle, and have been shown to be mutated in CRC (Matthews, Bertoli and de Bruin, 2021).

Negative enrichment also highlighted interesting trends. The negatively enriched GO terms were all related to serotonin. *SLC6A4* (a DEG downregulated in the *F. nucleatum*+ patients), which codes for 5-HTT, a serotonin receptor, was the leading-edge gene across all four terms. This is of interest, as serotonin and the gut microbiota have previously been linked (Waclawiková and El Aidy, 2018). The negative enrichment of *SLC6A4* in GO terms in *F. nucleatum*+ patients shown here reflect results by (Yano *et al.*, 2015), who observed that germ-free mice had higher levels of *SLC6A4*.

CAPN9 was downregulated in the *F. nucleatum*+ patients. A calpain involved in gastric injury defence (Hata *et al.*, 2010), *CAPN9* was a leading-edge gene in the negatively enriched gene set from MSigDb Hallmarks. Furthermore, negatively enriched KEGG pathways relating to cellular structure and localisation offer insights into the relationship between *F. nucleatum* and the eventual migration of cells from the colorectal epithelium to distant sites. Although *F. nucleatum* was found in liver metastasis from CRC primaries (Bullman *et al.*, 2017), and integrins such as *IGTAV* were associated with poor CRC survival (Linhares *et al.*, 2015), the negative enrichment observed here may suggest that the migratory effects reported in *F. nucleatum*+ CRC patients develop with progression to malignancy.

Following up this immune-enriched transcriptomic signature, and the multiple activated cytokine-related pathways, FAP PDOs were used to test the effects of cytokines on adenoma growth. Cytokines have been shown to promote tumorigenesis (Borowczak *et al.*, 2022). The results of cytokine-mediated stimulation of the PDOs led to a marked increase in cell viability, with IL-6 being the best stimulant. Although a systematic review of circulating levels of IL-6 or TNF-α did not find a correlation to risk of adenomas, IL-6 is known to be elevated in CRC patients when compared to those with adenomas (Zhang, Liu and Zhou, 2016). The results shown here suggest that cytokines play a role in

supporting the growth of adenomas. This is an interesting finding, and future work on PDOs co-cultured with immune cells will offer a deeper understanding of any growth-supporting mechanisms.

The adaptive immune response is a crucial component of anti-tumour defences, and previous research indicates that cell densities are valuable predictors of prognosis in CRC. A study by Kasurinen et al. (2022) found that higher densities of CD3+ and CD8+ T cells were associated with disease-specific survival in CRC patients, while another study reported that high densities of CD3+ and CD4+ T cells associated with relapse-free survival and CD4+ and CD8+ T cells associated with disease-specific survival (Kuwahara *et al.*, 2019). The clinical utility of these cell populations has been demonstrated by their use in the Immunoscore®, based on the work of Galon et al. (2006). This raised the question as to whether such cells could be of value in the detection of metachronous polyps or CRC with colorectal adenomas.

While the results shown here do not clearly identify a significant role for immune cells in colorectal polyps with *F. nucleatum* status, it is important to consider this within the context of their function. Interestingly, it has been reported in the literature that *F. nucleatum*+ CRCs not only have lower CD3+ and CD8+ T cell densities (Mima *et al.*, 2015) but also showed a higher expression of the cell exhaustion markers PD-1 and TIM-3 (Kim *et al.*, 2023). While the correlation between *F. nucleatum*+ patients and immune cell density within this cohort lacked statistical significance (probably due to limited numbers), future work could investigate whether the low immune response observed here is in fact due to some *F. nucleatum*-mediated effect on T cell exhaustion. Additionally, the small size of this patient subset is a great limitation, and a repeat of this study on a larger, more comprehensive cohort is required to reach a validated conclusion.

Alternative causes of a weak immune response irrespective of *F. nucleatum* status must also be considered. Intestinal stem cells are known to express LGR5 (Barker *et al.*, 2007), and LGR5+ cells in the gut are targets for immune-mediated clearance (Agudo *et al.*, 2018). Interestingly, Goto et al. (2024) demonstrated that the generation of LGR5- cells occurs in colorectal adenomas through the function of SOX17, which was associated with low MHC-I levels. This

suggests that LGR5- cells can evade immune destruction through decreased antigen-presentation. While investigations of such depth are beyond the scope of this thesis, it may be of interest to investigate the expression of T cell activating antigens (like MHC-I and MHC-II) at the protein level, to assess whether the low immune response seen here is the result of decreased activating antigen levels.

Since *F. nucleatum* has been shown to influence the surrounding environment, an investigation of the spatial landscape in areas close to and further away from *F. nucleatum* was performed. Cell typing revealed two fibroblast subpopulations to be present in these samples, and in the *F. nucleatum*+ cores in the *F. nucleatum*+ zones, *FAP* (also known as fibroblast-activating protein) and *FGF20* (also known as fibroblast growth factor 20) were upregulated. *FAP* is considered a cancer-associated fibroblast (CAF) marker, since its expression has been shown to be arrested during healthy adulthood, except for in the case of cancer, where it has been detected in a variety of stromal cells (Puré and Blomberg, 2018). In CRC patients, high levels of *FAP* were associated with a poor overall survival (Chen *et al.*, 2017). CAFs, known for their role in modulation of the tumour microenvironment, secrete IL-6 (X. Mao *et al.*, 2021) which has been shown to be involved with anti-tumour immune cells such as CD8+ T cells (Orange *et al.*, 2023). A study of CRC patients revealed that levels of *FAP* were higher in CAFs than in normal fibroblasts, and that IL-6 levels produced by CAFs treated with *F. nucleatum* were higher than in normal fibroblasts (Eskandari-Malayeri *et al.*, 2024). The results shown here suggest that the effects of the immune-associated CAFs may be taking place at an earlier step of CRC formation than previously reported.

Neighbourhood analysis in the *F. nucleatum*+ cores revealed that there was little to no immune response in the inner-most zone. This was surprising as the bulk transcriptomics and the spatial transcriptomics cell typing suggested an immune signature present in *F. nucleatum*+ cores. This was supported by pathway analysis showing downregulated pathways such as JAK-STAT. In CRC, activation of the JAK-STAT pathway was shown to be associated with upregulation of the inhibitory immune checkpoint PD-L1 (Mimura *et al.*, 2017), and expression of PD-L1 was positively correlated with *F. nucleatum* abundance in CRC (Gao *et al.*, 2023). It is important to note that that within adenoma tissue, the amount of *F.*

nucleatum was very small and the tissue is benign, possibly leading to the contradicting effects with cancer studies seen here.

The hypoxia pathway was upregulated within the *F. nucleatum*+ cores (albeit in the zones furthest away from the bacteria). Recent work by Udayasuryan et al. (2024) has suggested that hypoxic environments in CRC can enhance *F. nucleatum* infections and that *F. nucleatum* can influence gene expression to generate an anaerobic environment. While hypoxic upregulation was seen in the zones furthest away from the bacteria, it suggests an additional mechanism by which *F. nucleatum* can enrich itself within hyperproliferating colorectal tissue.

A key finding was that both Wnt and NFκB signalling were upregulated in the *F. nucleatum*- cores. As previously discribed in Figure 1.10 and Figure 1.11, *F. nucleatum* mimics Wnt signalling and activates NFκB signalling, resulting in increased proliferation. This suggests that these proliferative pathways may be upregulated within adenoma tissue regardless of external influences such as *F. nucleatum*. Additionally, there is spatial evidence supporting the proliferative signature as seen through GSEA. The upregulation of *CDKN1A* (which encodes p21) increased with increasing distance from the *F. nucleatum*+ zone. p21 is a cell cycle regulator that controls G1/S phase passage (Bueno-Fortes *et al.*, 2022). This, along with the upregulation of p53 in the spatial zones furthest away from the *F. nucleatum* in the *F. nucleatum*+ cores, suggests that the cell cycle is less regulated in cells proximal to *F. nucleatum*.

However, the most interesting finding revealed by spatial transcriptomic analysis involves *F. nucleatum*-induced ROS generation, which is a known cause of DNA damage (Galasso *et al.*, 2025). *OXR1* is a gene known for its crucial role as a protector against ROS (Yang *et al.*, 2015), with its loss leading to developmental defects in neurons (Lin *et al.*, 2023). Spatially, *OXR1* was upregulated in the zones further away from the *F. nucleatum*, suggesting a reduction in *OXR1*-mediated protection in areas likely to experience ROS-mediated damage. Interestingly, p21 is also involved in cell cycle arrest as a response to DNA damage, as is *H2AX*, a histone protein whose phosphorylated form (γH2AX) is a known marker for DNA damage (Dickey *et al.*, 2009). *F. nucleatum*-induced DNA damage has been reported through multiple mechanisms (Guo *et al.*, 2020; Xu *et al.*, 2024), and the distance seen here between the location of *F. nucleatum* and

the genes involved in protection from ROS and DNA damage repair could provide an explanation as to why *F. nucleatum*-induced DNA damage is not always repaired, and how *F. nucleatum* primes its surroundings into a malignancy-supporting environment.

No associations between the presence of *F. nucleatum* and the detection of metachronous polyps, or most clinical characteristics like site, were observed. However, *F. nucleatum*+ patients tended to have serrated and hyperplastic polyps. While most CRC cases develop from traditional adenomas, through the adenoma-carcinoma pathway, serrated polyps confer a poor outcome as not only are they more likely to be missed during endoscopy (Zhao *et al.*, 2019), but they are thought to give rise to both CRC with poor survival outcomes, and to interval cancers (Kang *et al.*, 2021). The results shown here corroborate previous research that identified higher levels of *F. nucleatum* in serrated lesions than in traditional adenomas (Yu *et al.*, 2016).

This chapter provided evidence of the possibility of using novel techniques to detect and visualise *F. nucleatum* in archival tissue of colorectal adenomas, and while the findings presented here must be validated in a larger cohort, TempO-Seq™ and RNAscope® offered specific and sensitive detection and visualisation of *F. nucleatum*. Despite the small sample size and large size bias between the *F. nucleatum*+ and *F. nucleatum*- patients making it difficult to draw concrete conclusions, the transcriptomic profile of *F. nucleatum*+ patients suggests a role for *F. nucleatum* in preparing and supporting a proliferative, inflamed environment in the adenoma stage. This immune environment was shown to influence adenoma growth through cytokine stimulations of FAP PDOs. However, the adaptive immune response of patients with *F. nucleatum* status known is not of clinical relevance, and this was supported by the lack of lymphocytic immune cell populations in the spatial transcriptomic analysis of the areas closest to the *F. nucleatum*, indicating that there are some transcriptomic changes being made by the bacterial presence that were not yet reflected on the whole tissue. Ultimately, no associations were found between the presence of *F. nucleatum* and the detection of metachronous polyps, however a correlation was found to non-conventional adenoma polyps, corroborating results in the literature.

Within this patient cohort, *F. nucleatum* did not present a viable option for the prediction of metachronous polyps or CRC. Hence, the next chapter will involve a larger patient subset and will delve into the mutational profile of colorectal adenomas, in the hopes of identifying new targets of investigation.

Chapter 4 Exploration of Mutational Landscape of Colorectal Polyps

4.1 Introduction

As a common cancer, the development of CRC has been the subject of investigation for many years, with its origins and development at the forefront of research. It is widely accepted that specific mutations in oncogenes and tumour suppressor genes are early events in colorectal carcinogenesis. Fearon and Vogelstein (1990) suggested a genetic alterations model for tumour development that involved the accumulation of alterations on chromosomes 5q, 12p, 18q, and 17p. As discussed in 1.3.2, *APC* is regarded as an essential tumour suppressor gene, controlling important cellular functions and pathways such as Wnt signalling (Raji, Sasikumar and Jacob, 2018) and mutations in *APC* that lead to the loss of its regulatory functions have been implicated in both hereditary CRC syndromes such as FAP (Olkinuora *et al.*, 2021) and sporadic CRC (Aghabozorgi *et al.*, 2019). Similarly, *KRAS*, frequently undergoes genetic alterations in glycine 12, glycine 13, or glutamine 61, all of which can lead to increased *KRAS* activation in CRC (Zhu *et al.*, 2021). While *APC* and *KRAS* mutations have been identified as leading events in adenoma development, other genetic mutations in genes like *SMAD4* and *TP53* (Muzny *et al.*, 2012) have been reported in the transition of adenomas into the carcinomas (Figure 1.6).

Since there was no link between *F. nucleatum* and the detection of metachronous polyps or CRC, the focus on this work shifted to investigating intrinsic factors within the polyps. Although the evidence of a genetically altered genome in CRC is well-established, little is known about the value of mutations in the prediction of metachronous polyps or CRC. This chapter presents a mutational analysis of an INCISE patient subset of 722 patients, with stratifications based on adenoma site, BSG2020 risk, and outcome, and will examine the value of any mutations as targets for further investigation.

4.2 Mutational Landscape of INCISE Adenomas

DNA extracted from adenomas from 722 INCISE patients was prepared and sequenced as described in 2.2.7, and analysed as described in 2.2.11. 97% (n=700) of sequenced patients harboured mutations. As shown in Figure 4.1A, most mutations were "missense" mutations. Of the top ten most mutated genes, *APC* and *KRAS* were the top two, with 87% and 32% of patients having mutations in those genes, respectively (Figure 4.1B). Multi-hit mutations were most abundant for *APC*, while "missense" were more common for *KRAS*. Human leukocyte antigen genes *HLA-A* and *HLA-B* were the third and eighth most mutated genes, with mutations in 24% and 16% of samples, respectively, most being "missense". Mismatch repair gene *MSH3* was mutated in 20% of the samples, making it the fourth most mutated gene, with most of the mutations being "in-frame deletions". In fifth and sixth place and mutated in 18% of samples, were transcription factor *ZFHX3* and histone methyltransferase *KMT2C*, respectively, both with mostly "missense" mutations. *SOX9*, a stemness marker and cell signalling regulator, was the seventh most mutated gene and was altered in 17% of patients, with most of the alterations being "frameshift insertions". *TP53* was ninth, and was mutated in 16% of patients, while transcription factor *MGA* was the tenth most mutated gene in 14% of patients. Both had mostly "missense" mutations.

To investigate any co-occurring or mutually exclusive genetic mutations, somatic interactions were plotted (statistical significance at pAdj<0.05 determined using pairwise Fisher's Exact test; Figure 4.1C, Table 4.1). As *APC* and *KRAS* are the first two genes to acquire mutations in the Vogelstein adenoma-carcinoma pathway, it was not surprising that they co-mutated at a rate of 48% (pAdj=0.019). *HLA-A* and *HLA-B* co-mutated at a rate of 21% (pAdj=0.009) and *KMT2C* and *MGA* co-mutated at a rate of 17% (pAdj=0.018). While *HLA-B* co-mutated with *KRAS* at a rate of 23% (pAdj=0.002), only *SOX9* co-mutated with both *APC* (23%, pAdj=0.004) and *KRAS* (22%, pAdj=0.019), suggesting a link between *SOX9* mutations and the earliest stages of carcinogenesis. No genes were mutated in a mutually exclusive, statistically significant manner.

Oncogenic pathway analysis of all sequenced patients (Figure 4.1D) revealed that the Wnt pathway, in which genes such as *APC* and *SOX9* are involved, was

affected in 92% of patients. This was followed by the RTK-RAS pathway in which KRAS is a crucial member, which was affected in 72% of patients, and the NOTCH pathway, which was affected in 47% of patients.

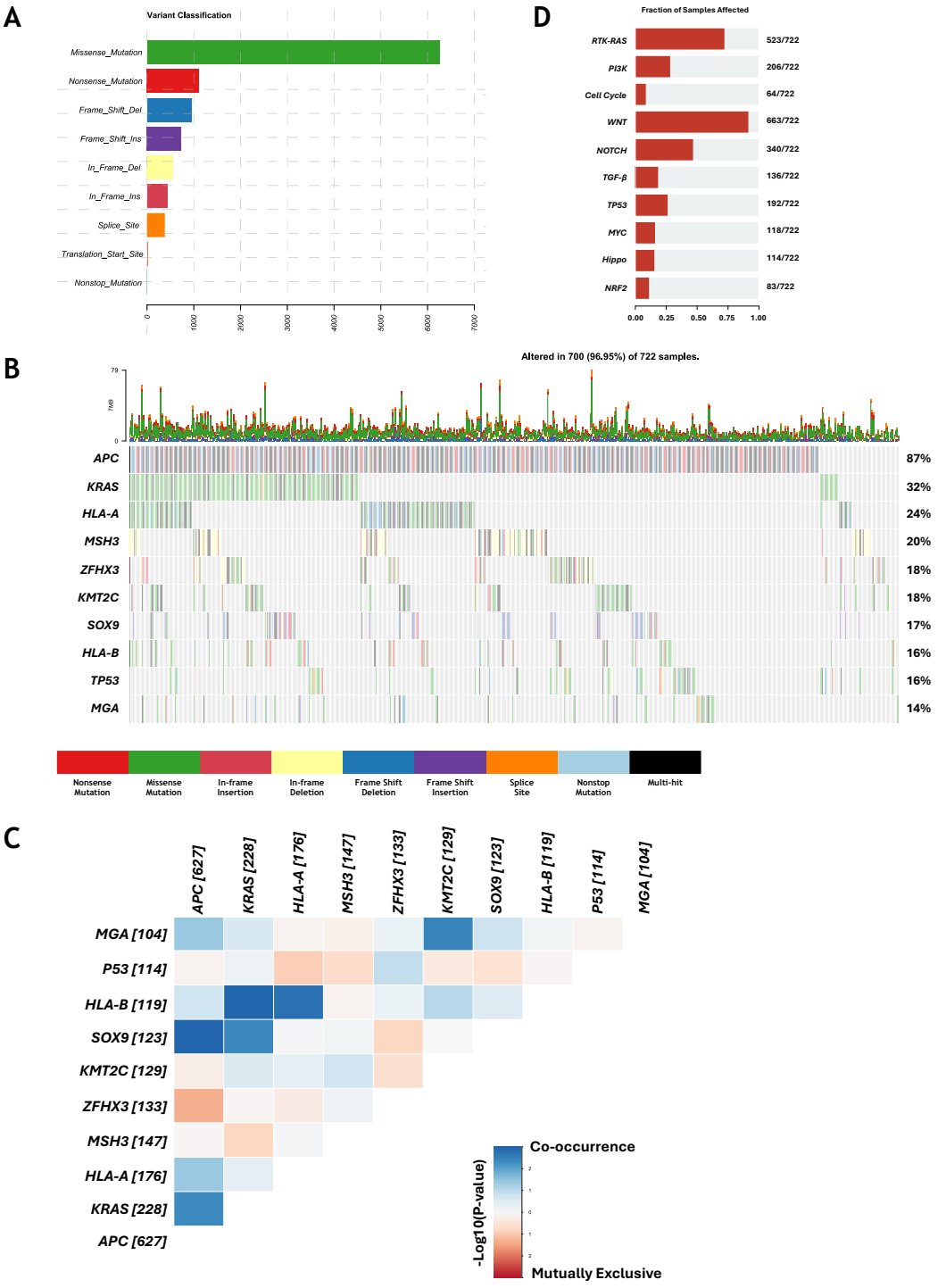


Figure 4.1. Mutational Analysis of INCISE Patient Polyps.
[A] Variant classification of entire sequenced cohort. **[B]** Oncoplot depicting the top 10 most mutated genes ranked by the percentage of patients in which the gene is affected. "Multi-hit" denotes genes with multiple mutations in a sample. **[C]** Somatic interactions depicting co-occurring or mutually exclusive gene pairs as statistically true events. **[D]** Fraction of samples affected for each oncogenic pathway. N=722.

Table 4.1. Top 10 Gene Pair Mutations for Adenomas from the INCISE Cohort

Rank	Gene Pair	pAdj	Significant	Event	Event Ratio
1	*HLA-B/KRAS*	0.0028	Yes	Co-occurrence	55/237
2	*APC/SOX9*	0.0044	Yes	Co-occurrence	118/514
3	*HLA-A/HLA-B*	0.0098	Yes	Co-occurrence	43/209
4	*KMT2C/MGA*	0.0185	Yes	Co-occurrence	30/173
5	*KRAS/SOX9*	0.0194	Yes	Co-occurrence	53/245
6	*APC/KRAS*	0.0196	Yes	Co-occurrence	210/435
7	*APC/HLA-A*	0.1556	No	Co-occurrence	161/481
8	*APC/MGA*	0.1556	No	Co-occurrence	97/537
9	*APC/ZFHX3*	0.1631	No	Mutually Exclusive	108/544
10	*HLA-B/KMT2C*	0.2948	No	Co-occurrence	28/192

Gene pairs are ranked by most significant pAdj.

As the effects of some mutations are governed by their location, lollipop plots were generated to visualise mutation hotspots on the top 10 most mutated genes of the sequenced cohort (Figure 4.2). While *APC, MSH3, ZFHX3, KMT2C,* and *MGA* all harboured mutations across the length of the gene, most *MSH3* mutations appeared to occur at alanine 89. Similarly, most *ZFHX3* mutations occurred at glycine 3512, and 11 samples had *MGA* mutations at arginine 2529. *KRAS* carried most of its mutational burden towards the N-terminus, with 76 samples having mutations at glycine 12. Most *HLA-A* and *HLA-B* - mutations are found in the first half of the gene, where 53 samples had mutations at arginine 41 and 31 samples in glutamine 49, respectively. While *SOX9* appears to harbour mutations across the gene, the mutational burden is slightly higher in the distal two-thirds, towards the C-terminus. Finally, *TP53* mutations appear to aggregate at the centre of the gene, with 13 samples having mutations in arginine 175.

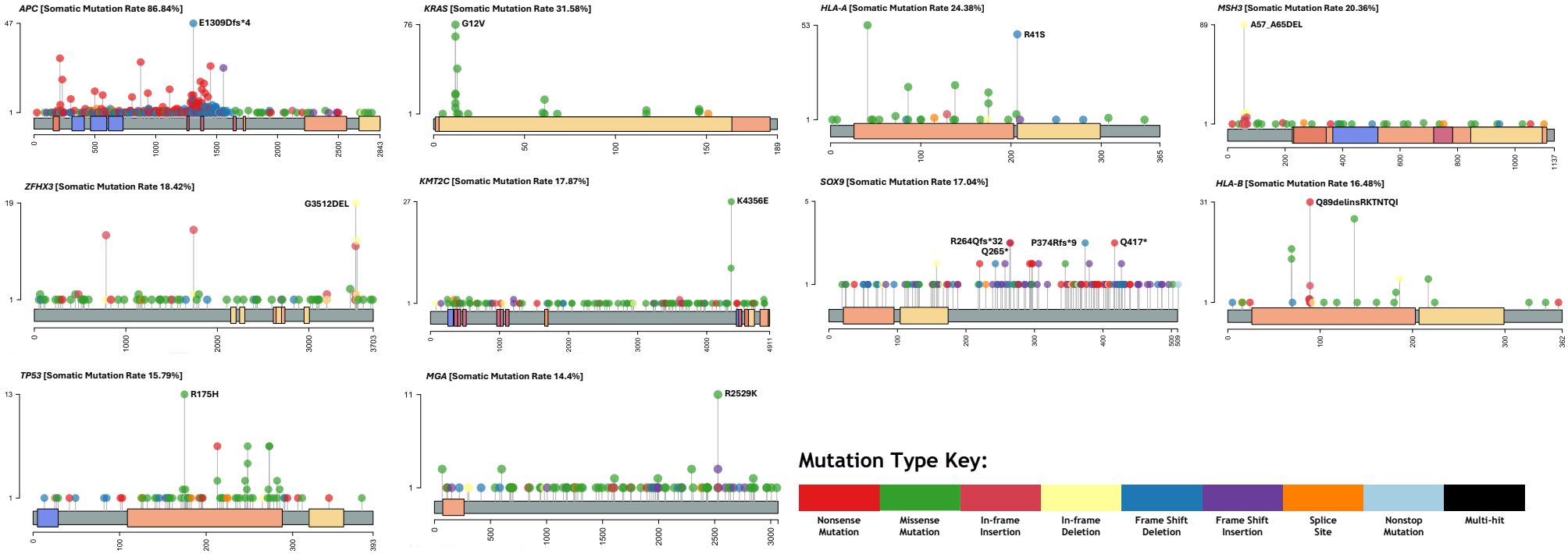


Figure 4.2. Lollipop Plots Displaying the Location of Mutational Hits for the Top 10 Most Mutated Genes.
The somatic mutation rate denotes the percentage of patients within the cohort with mutations in that gene. The y-axis of each lollipop plot denotes how many samples had a mutation at a specific location. The longer the lollipop stick, the more samples had a specific mutation at a specific location. The x-axis denotes the amino acids making up the gene. The coloured boxes on the gene denote functionally relevant locations. The colours of the lollipops denote the mutation type and are explained by the mutation type key. The mutations occurring in most patients mutated for that gene are included. Human Genome Variation Society nomenclature was followed.

To reveal any differences in the mutational landscape of polyps that might influence their growth profile, oncoplots and somatic interactions were investigated in subsets stratified by index polyp *F. nucleatum* status, site, risk according to current BSG2020 Guidelines, and outcome.

4.2.1 Adenomas by Site

4.2.1.1 Right Colon

As shown in Figure 4.3A, 97% (n=122) of right colonic adenomas were mutated. *APC* (86%) and *KRAS* (40%) were the top two most mutated genes with "multi-hit" and "missense" events leading the alterations, respectively. Three members of the KMT2 family of histone methyltransferases were in the top ten most mutated genes in right colonic adenomas: *KMT2C* (22%), *KMT2D* (16%), and *KMT2A* (14%), were the third, seventh, and tenth most mutated genes respectively, all with mostly "missense" mutations. *MSH3* (21%) was the fourth most mutated gene with mainly "in-frame deletions". *HLA-A* (21%) and *HLA-B* (14%) were the fifth and ninth most mutated genes respectively, with mostly "missense" alterations. "Frame shift insertions" were the most common alteration in the sixth most mutated gene *SOX9* (17%), and gene suppressor *BCOR* (15%) was the seventh most mutated genes with mainly "missense" mutations. Seven of these genes (*APC, KRAS, KMT2C, MSH3, HLA-A, SOX9,* and *HLA-B*) were in the top ten genes in the whole cohort, however *TP53* was not in the top ten mutations found in right colonic adenomas.

Somatic interactions revealed that only *APC* and *KRAS* co-mutations were statistically significant (pAdj=0.01; Figure 4.3D, Table 4.2), occurring at an event ratio of 82%. Similar to the full cohort, no two genes were mutating in a mutually exclusive, statistically significant manner.

4.2.1.2 Left Colon

The oncoplot in Figure 4.3B shows that 97% (n=494) of left colonic adenomas were mutated, and *APC* (86%) and *KRAS* (27%) were the top two most mutated genes. Similar to right colonic adenomas, left colonic adenomas had mutations in *HLA-A* (26%, third), *MSH3* (21%, fourth), *SOX9* (18%, sixth), *HLA-B* (17%, seventh), and *KMT2C* (17%, eighth). *HLA-A* and *HLA-B* both had mainly "missense"

mutations, while “in-frame deletions” and “frame shift insertions” were the most common mutation type in *MSH3* and *SOX9,* respectively. Both *ZFHX3* (19%, fifth) and *TP53* (17%, ninth) mutations were present in left colonic adenomas but not right colonic adenomas, with “missense” alterations being most abundant. WNT signalling transcription factor *TCF7L2* (16%) was the tenth most mutated gene, with mainly “missense” mutations. Interestingly, *APC* and *KRAS* were not significantly co-mutated in left colonic adenomas (pAdj=0.497). However, *SOX9* was co-mutated with *APC* (pAdj=0.006) at an event ratio of 24%, and with *KRAS* (pAdj=0.001) at an event ratio of 26%. *KRAS* was also co-mutated with *TCF7L2* (pAdj=0.009) at an event ratio of 23%, and with *HLA-B* (pAdj=0.005) at an event ratio of 24%. Again, no gene pair mutations were mutually exclusive (Figure 4.3E, Table 4.2).

4.2.1.3 Rectum

Rectal adenomas were highly mutated, at a rate of 99% (n=81; Figure 4.3C). *APC* (90%) and *KRAS* (49%) were not only the top two most mutated genes but also were mutated at a higher rate in the rectum than in the colon. *TP53* (24%, third), *ZFHX3* (23%, fourth), *HLA-A* (20%, seventh), *HLA-B* (18%, eighth), and *KMT2C* (18%, ninth) were mutated in the rectum, as they were in the left colon, all with a mainly “missense” mutations. *MGA* (22%, fifth) was mainly mutated by “missense” events , as it was in the full cohort, suggesting a driving role for the rectal polyps in the results seen in the full cohort. Oncoprotein degrader *FBXW7* (21%, sixth) was mutated in rectal adenomas only, mainly by “missense” events. Unlike mutations in the colon, somatic interactions revealed no statistically significant gene pair mutation events in the rectum (Figure 4.3F, Table 4.2).

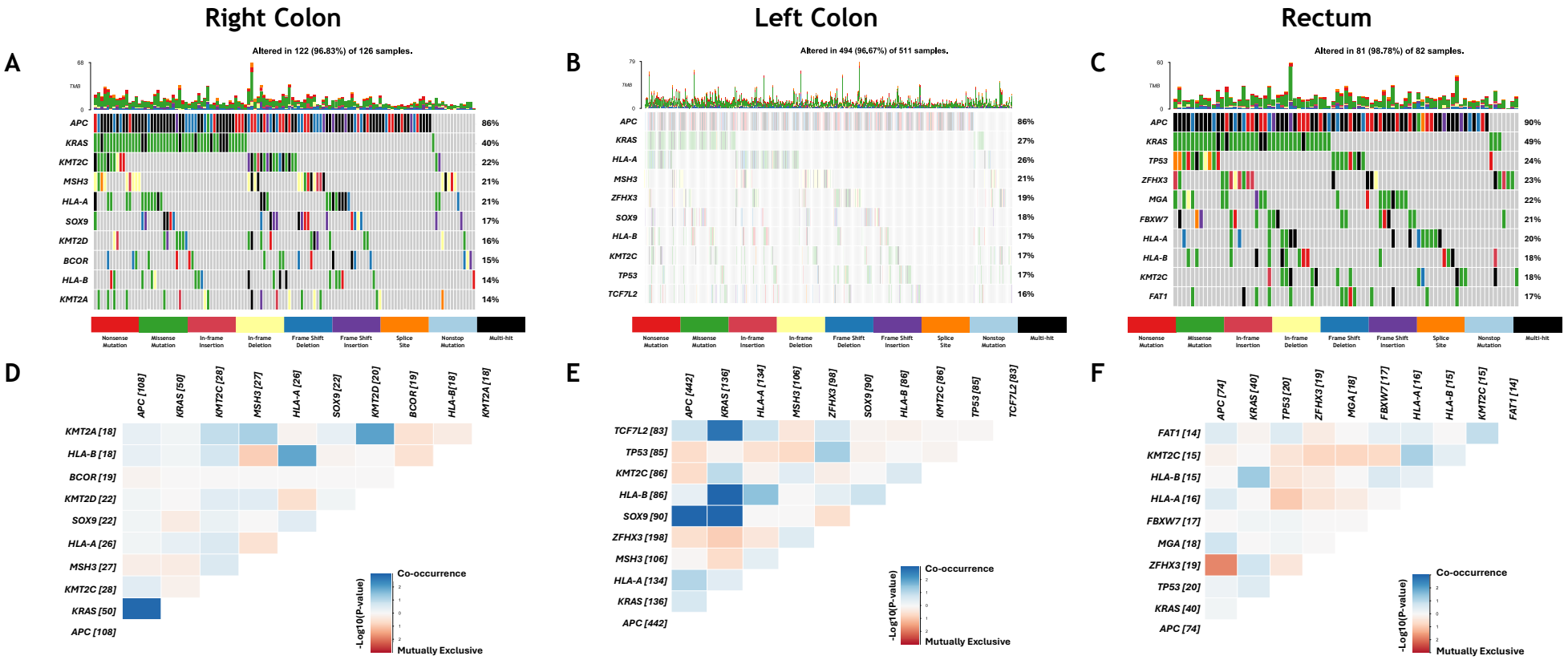


Figure 4.3. Comparison of the Mutational Landscape of Adenomas by Location.
Oncoplots depicting the top 10 most mutated genes in **[A]** right colonic adenomas, **[B]** left colonic adenomas, and **[C]** rectal adenomas. Genes are ranked by the percentage of patients in which the gene is affected. “Multi-hit” denotes genes with multiple mutations in a sample. Somatic interactions depicting co-occurring or mutually exclusive gene pairs as statistically true events in **[D]** right colonic adenomas, **[E]** left colonic adenomas, and **[F]** rectal adenomas.

Table 4.2. Top 5 Gene Pair Mutations for Adenomas Stratified by Location

Rank	Gene Pair	pAdj	Significant	Event	Event Ratio
Right Colonic Adenomas					
1	*APC/KRAS*	0.0102	Yes	Co-occurrence	49/60
2	*KMT2A/KMT2D*	0.0668	No	Co-occurrence	7/24
3	*HLA-A/HLA-B*	0.0762	No	Co-occurrence	8/28
4	*KMT2A/MSH3*	0.3604	No	Co-occurrence	7/31
5	*HLA-B/MSH3*	0.5493	No	Mutually Exclusive	1/43
Left Colonic Adenomas					
1	*KRAS/SOX9*	0.0016	Yes	Co-occurrence	39/148
2	*HLA-B/KRAS*	0.0054	Yes	Co-occurrence	36/150
3	*APC/SOX9*	0.0063	Yes	Co-occurrence	87/358
4	*KRAS/TCF7L2*	0.0098	Yes	Co-occurrence	34/151
5	*HLA-A/HLA-B*	0.1545	No	Co-occurrence	31/158
Rectal Adenomas					
1	*APC/ZFHX3*	0.1211	No	Mutually Exclusive	14/65
2	*HLA-B/KRAS*	0.3337	No	Co-occurrence	11/33
3	*HLA-A/KMT2C*	0.4023	No	Co-occurrence	6/19
4	*HLA-A/TP53*	0.5628	No	Mutually Exclusive	1/34
5	*FAT1/KMT2C*	0.6031	No	Co-occurrence	5/19

Gene pairs are ranked by most significant pAdj.

4.2.2 Adenomas by Risk According to BSG2020 Guidelines

4.2.2.1 Low Risk

The mutational profile of patients stratified as low risk according to current BSG2020 Guidelines was investigated. 96% (n=244) had mutations, with *APC* (87%) and *KRAS* (32%) the top two most mutated genes with “multi-hit” and “missense” mutations, respectively (Figure 4.4A). *HLA-A* (28%, third) had mostly “missense” mutations, while *MSH3* (24%, fourth) was mostly mutated by “in-frame deletions”. Altered in 17% of low-risk patients, *SOX9* (fifth) and *TP53* (sixth) were affected by “frameshift insertions” and “missense” mutations, respectively. *FBXW7* (seventh), *MGA* (eighth), and *TCF7L2* (ninth) were all mutated in 15% of low-risk patients. While “missense” mutations were most common for *FBXW7* and *MGA, TCF7L2* was more likely to be altered by “frameshift insertions”. Finally, *KMT2C* (15%, tenth) had mostly “missense” events. Somatic interactions revealed that while one gene pair *FBXW7/KRAS* was significantly co-mutating (pAdj=0.043) at an event ratio of 26%, no other gene pairs were mutated in a co-occurring or mutually exclusive manner (Figure 4.4C, Table 4.3).

4.2.2.2 High Risk

Alterations were identified in 97% (n=453) of high-risk patients according to current BSG2020 Guidelines (Figure 4.4B). As reflected in previous stratifications and in the whole sequenced cohort, *APC* (87%) and *KRAS* (31%) were the top two most mutated genes with “multi-hit” and “missense” mutations, respectively. *HLA-A* (23%, third) and *HLA-B* (18%, seventh) were mostly altered through “missense” mutations. *ZFHX3* (21%, fourth), and *KMT2C* (19%, fifth) both had mostly “missense” mutations, while *MSH3* (19%, sixth) was mostly mutated by “in-frame deletions”. Altered by “frameshift insertions” in 17% of high-risk patients, *SOX9* was the seventh most mutated genes in this subset. *TP53* (15%, ninth) was altered by “missense” mutations, while *ARID1A* (14%, tenth) was mostly altered by “frameshift deletions”. While no mutually exclusive gene mutation pairings were statistically significant, four pairs of genes were significantly co-mutating (Figure 4.4D, Table 4.3). *SOX9* co-mutated with both the top two most mutated genes, *APC* (pAdj=0.004) at an event ratio of 24%, and *KRAS* (pAdj=0.031) at an event ratio of 23%. *HLA-B* also co-mutated with *KRAS* (pAdj=0.019) at an event ratio of 25%, as well as with *HLA-A* (pAdj=0.032) at an event ratio of 22%.

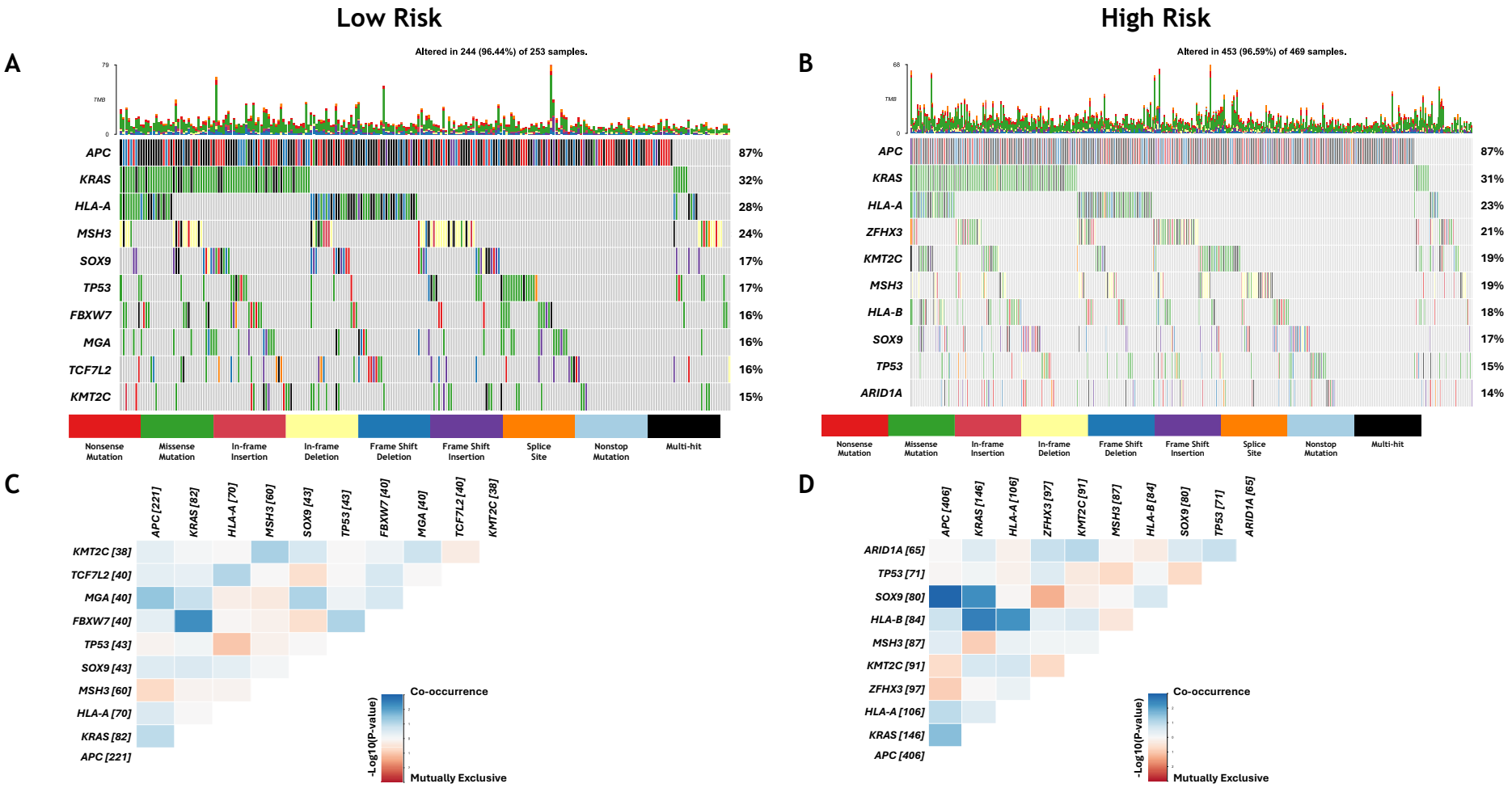


Figure 4.4. Comparison of the Mutational Landscape of Adenomas by BSG2020 Risk Guidelines.
Oncoplots depicting the top 10 most mutated genes in **[A]** low risk patients according to current BSG2020 Guidelines, **[B]** high risk patients according to current BSG2020 Guidelines. Genes are ranked by the percentage of patients in which the gene is affected. Somatic interactions depicting co-occurring or mutually exclusive gene pairs as statistically true events in **[C]** low risk patients according to current BSG2020 Guidelines, **[D]** high risk patients according to current BSG2020 Guidelines.

Table 4.3. Top Five Gene Pair Mutations for Adenomas Stratified by BSG2020 Risk Guidelines

Rank	Gene Pair	pAdj	Significant	Event	Event Ratio
Low Risk					
1	*FBXW7/KRAS*	0.0434	Yes	Co-occurrence	21/80
2	*APC/MGA*	0.2621	No	Co-occurrence	39/183
3	*KMT2C/MSH3*	0.3315	No	Co-occurrence	14/70
4	*MGA/SOX9*	0.3315	No	Co-occurrence	11/61
5	*FBXW7/TP53*	0.3315	No	Co-occurrence	11/61
High Risk					
1	*APC/SOX9*	0.0046	Yes	Co-occurrence	78/330
2	*HLA-B/KRAS*	0.0195	Yes	Co-occurrence	38/154
3	*KRAS/SOX9*	0.0318	Yes	Co-occurrence	36/154
4	*HLA-A/HLA-B*	0.0330	Yes	Co-occurrence	29/132
5	*APC/KRAS*	0.1404	No	Co-occurrence	134/284

Gene pairs are ranked by most significant pAdj.

4.2.3 Adenomas by Outcome

4.2.3.1 No Metachronous Polyps or CRC Detected

When stratified by outcome, 98% (n=295) of patients with no metachronous polyps or CRC had genetic mutations (Figure 4.5A). *APC* (88%) and *KRAS* (31%) were the top two most mutated genes with “multi-hit” and “missense” mutations, respectively. Mutated in 26% of patients within this subset, *HLA-A* was the third most mutated gene with mainly “missense” mutations. *MSH3* (22%) was fourth and was mostly altered by “in-frame deletions”. *SOX9* (fifth) was mutated by “frameshift insertions” in 18% of patients with no metachronous polyps or CRC detected. 17% of patients had mutations in *ZFHX3* (sixth) and *TP53* (seventh), and these were mostly “missense” mutations. *HLA-B* (eighth) and *KMT2C* (ninth) were altered in 16% of patients mostly through “missense” mutations. *TCF7L2* (tenth) was mutated in 15% of patients, with most alterations being due to “missense” mutations. *TCF7L2* was co-mutated with both *KRAS* (pAdj=0.006) and *ZFHX3* (pAdj=0.028) at an event ratio of 27% and 23%, respectively. No genes were altered in a mutually exclusive manner (Figure 4.5C, Table 4.4).

4.2.3.2 Metachronous Polyps or CRC Detected

Of the patients with metachronous polyps or CRC, 96% (n=410) had mutations (Figure 4.5B). As reflected in the full sequenced cohort, *APC* (86%) and *KRAS* (32%) were mostly mutated through “multi-hit” and “missense” events, respectively, as the top two most mutated genes. Similarly to the patients with no metachronous polyps detected, *HLA-A* (23%) was the third most mutated gene with mainly “missense” mutations. *ZFHX3* (fourth), *KMT2C* (fifth), and *MSH3* (sixth) were mutated in 19% of the patients, with “missense” mutations being most common in *ZFHX3* and *KMT2C*, and *MSH3* being mostly affected by “in-frame deletions”. *HLA-B* (17%, seventh) mostly had “missense” mutations, while *SOX9* (16%, eighth) was mutated by “frameshift insertions”. Altered in 15% of patients within this subset, *TP53* (ninth) and *MGA* (tenth) were both affected by mostly “missense” events. *KRAS* was co-mutated with *HLA-B* (pAdj=0.029, event ratio 24%) and *APC* (pAdj=0.048, event ratio 51%). Interestingly, the same somatic interaction analysis revealed that *SOX9* and *APC* co-mutations were on the cusp of statistical significance (pAdj=0.05) with an event ratio of 22%. As with all previous analysis, no gene pairing was mutated in a mutually exclusive manner (Figure 4.5D, Table 4.4).

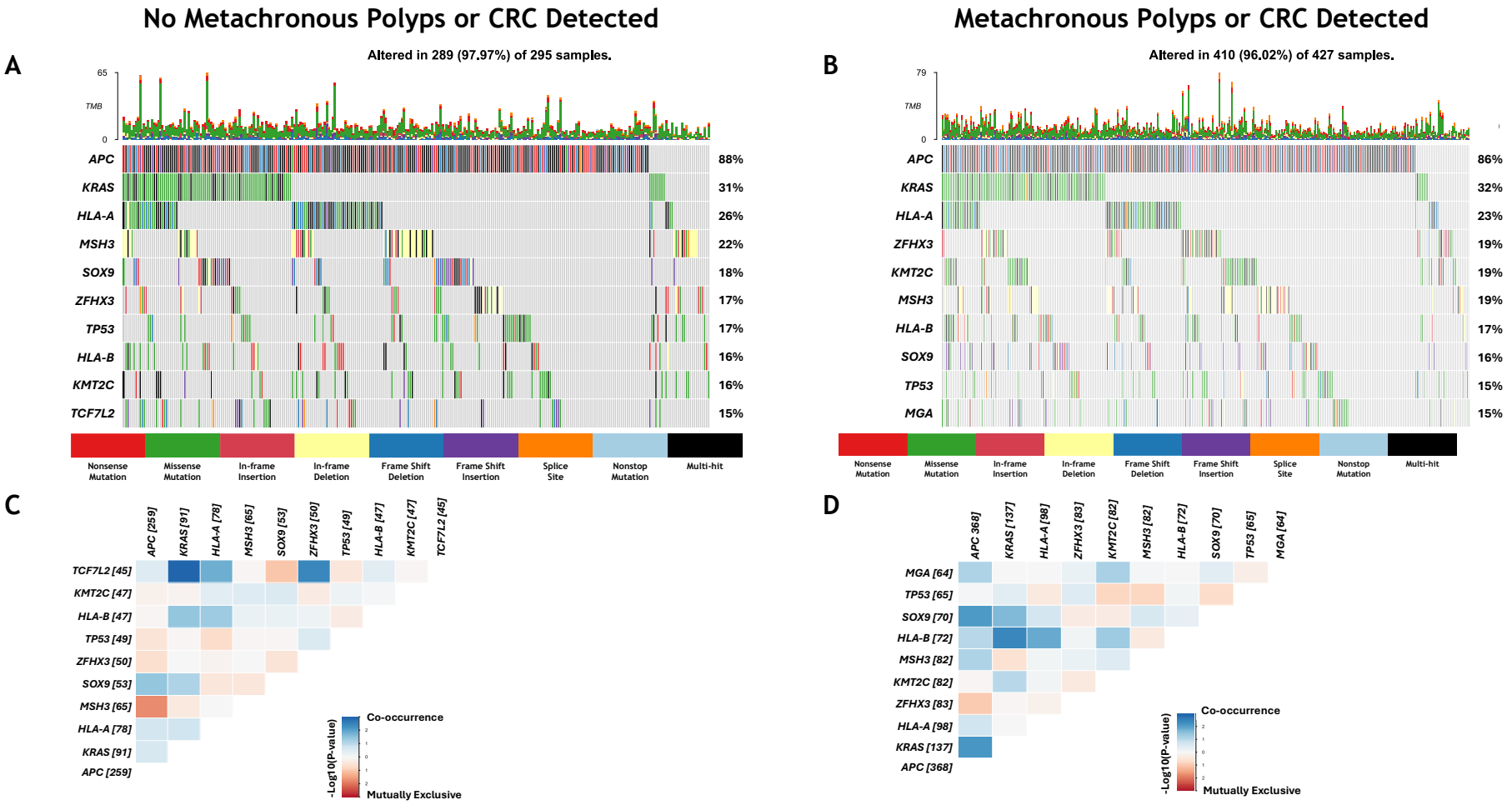


Figure 4.5. Comparison of the Mutational Landscape of Adenomas by Metachronous Polyp or CRC Outcome.
Oncoplots depicting the top 10 most mutated genes in **[A]** patients with no metachronous polyps or CRC detected, and **[B]** patients with metachronous polyps or CRC detected. Genes are ranked by the percentage of patients in which the gene is affected. “Multi-hit” denotes genes with multiple mutations in a sample. Somatic interactions depicting co-occurring or mutually exclusive gene pairs as statistically true events in **[C]** patients with no metachronous polyps or CRC detected, **[D]** patients with metachronous polyps or CRC detected.

Table 4.4. Top Five Gene Pair Mutations for Adenomas Stratified by Metachronous Polyp or CRC Outcome

Rank	Gene Pair	pAdj	Significant	Event	Event Ratio
No Metachronous Polyp or CRC Detected					
1	*KRAS/TCF7L2*	0.0061	Yes	Co-occurrence	24/88
2	*TCF7L2/ZFHX3*	0.0283	Yes	Co-occurrence	15/65
3	*HLA-A/TCF7L2*	0.0915	No	Co-occurrence	19/85
4	*APC/MSH3*	0.0915	No	Mutually Exclusive	51/222
5	*HLA-B/KRAS*	0.1711	No	Co-occurrence	21/96
Metachronous Polyp or CRC Detected					
1	*HLA-B/KRAS*	0.02929	Yes	Co-occurrence	34/141
2	*APC/KRAS*	0.0481	Yes	Co-occurrence	127/251
3	*APC/SOX9*	0.0500	No	Co-occurrence	67/304
4	*HLA-A/HLA-B*	0.0744	No	Co-occurrence	25/120
5	*KRAS/SOX9*	0.1221	No	Co-occurrence	31/145

Gene pairs are ranked by most significant pAdj.

4.3 Discussion

Mutations play an important role in carcinogenesis and these results present a highly altered genetic landscape of index polyps that reflects previous work. Some of the most mutated genes flagged in this analysis are as expected. As previously mentioned, *APC* mutations have long been considered initiating events in the switch from healthy cells to dysplastic growth. First identified as the cause of FAP by Groden et al. (1991), *APC* is the most mutated gene in CRC, and its mutations are correlated with poor outcomes (Li, Zhang and Xu, 2023; Peng *et al.*, 2023). The results presented here show a high mutation rate for *APC* across all samples, supporting previous works implicating it in CRC development.

KRAS has been shown to be mutated in 30-50% of CRCs (Prior, Hood and Hartley, 2020; Pellatt, Bhamidipati and Subbiah, 2024). Its oncogenic effects and direct involvement with the MAPK pathway led to it being the target of approved anti-cancer drugs for lung cancer (Jänne *et al.*, 2022; Dy *et al.*, 2023). A review on studies reporting *KRAS* oncogenic mutations in CRC cited mutations affecting glycine at positions 12 and 13 as being a common occurrence (Meng *et al.*, 2021). The results presented here not only show *KRAS* to be highly mutated in adenomas but also demonstrate that the oncogenic *KRAS* mutations reported in the literature, are present early in the adenoma-carcinoma sequence. The

presence of *APC* and *KRAS* in the first and second positions of the top mutated genes not only in the full cohort, but also in subsets stratified by site, number, risk, and outcome, demonstrates their importance and involvement in colorectal adenomas.

Commonly referred to as “the guardian of the genome”, *TP53* is known for its crucial role in maintaining genomic stability and preventing genetic mutations by controlling DNA repair mechanisms and cell cycle pathways, among others (Sullivan *et al.*, 2017). The tumour-suppressor functions of *TP53* were identified 10 years after its discovery (Lane and Benchimol, 1990), with mention of mutations altering this function. These mutations affect at least 30% of CRC patients (Campbell *et al.*, 2020), and most of them are missense mutations that lead to abnormal protein structures that cannot perform their original functions (H. Wang *et al.*, 2023). *TP53* is known to be involved in more advanced CRC rather than as an initiating event, thus, it is not surprising that *TP53* mutations have been reported in only 16% of adenomas (Naccarati *et al.*, 2012). Not only do the results of this study reflect the literature, the most common *TP53* mutation in this cohort, R175H, is also one of the most common in cancer and it affects the structural integrity of the p53 protein by interfering with its zinc-binding ability (H. Wang *et al.*, 2023). *TP53* was not among the top 10 most mutated genes in right-sided colon cancer. This is not surprising as previous work suggests it is more commonly mutated in the left colon (Loree *et al.*, 2017).

In addition to the expected genes, some interesting patterns emerged. *HLA-A* and *HLA-B* are main proteins in antigens of the MHC and are expressed by tumour cells as the recognised target of CD8+ T cells, playing a major role in the adaptive anti-tumour immune response (Romero and Coulie, 2014). Alterations to MHC and the loss of its function have been established as contributors to the immune-evasive nature of CRC (Aptsiauri, Ruiz-Cabello and Garrido, 2018). The results presented here support these findings in colorectal adenomas, where genetic alterations are present in most samples, however, not in those who are stratified as low risk according to current guidelines. This alludes to the importance of an intact immune response to risk stratification.

Hypermutated CRCs are often characterised by poor prognostic factors like high-MSI, MMR-deficiency, and CIMP phenotypes (Carethers and Jung, 2015; Campbell

et al., 2017). *MSH3*, part of the mismatch repair system, is involved in recognition of DNA mismatches and initiation of their repair (Li, 2007). *MSH3* mutations have been reported in 40% of hypermutated sporadic CRCs (Muzny *et al.*, 2012; Carethers and Jung, 2015), and germline mutation in *MSH3* have been implicated in familial CRC syndromes (Adam *et al.*, 2016; Perne *et al.*, 2021). The presence of *MSH3* in the top 10 most mutated genes in various patient subsets (excluding those with rectal polyps) not only suggests a role in the loss of mismatch repair in adenoma development but also offers an insight to the possible different biological processes underlying colonic tumours and rectal tumours.

A stemness marker involved in skeletal development, *SOX9* is a member of the SOX family (Jo *et al.*, 2014). *SOX9* has been shown to be involved in cell proliferation by regulating Wnt signalling by phosphorylating nuclear β-Catenin, preventing it from binding to the main Wnt transcription factor TCF7 (Akiyama *et al.*, 2004; Topol *et al.*, 2009). Akiyama et al. (2004) reported that *SOX9* mutations that led to the loss of the C-terminus resulted in SOX9 protein not sufficiently binding to and decreasing β-Catenin levels. Conversely, Topol et al. (2009) reported that the C-terminus was not required for the physical interaction between SOX9 and β-Catenin. However, Abdel-Samad et al. (2011), identified miniSOX9, a SOX9-variant missing its C-terminus that was overexpressed in colon cancer cell lines. Truncated SOX9 was later associated with CRC, where truncations were mostly the result of frameshift mutations (Muzny *et al.*, 2012; Javier *et al.*, 2016). These findings from cancer studies align with the results shown here, where *SOX9* frameshift mutations are common in colonic adenomas. Interestingly, in left colonic adenomas, mutations in *TCF7L2* were present. This is a Wnt transcription factor that transcribes Wnt target genes upon binding to β-Catenin (Liu et al., 2022). This suggests a role for Wnt signalling pathway effectors in adenoma formation.

Other commonly mutated genes were *ZFHX3* (a zinc-finger-heavy transcription factor), *KMT2C* (a histone methyltransferase), and *MGA* (a transcription factor). *ZFHX3* mutations were not only common in CRC, but they were correlated with better overall survival in a pan-cancer cohort (Principe, 2022), while *KMT2C* mutations were found to be associated with a higher immune response in CRC, suggesting a possible role as a therapeutic biomarker (Ni *et al.*, 2024). *MGA*

mutations were reported in lung cancer models (Llabata *et al.*, 2019) and as a driver gene in CRC (Cornish *et al.*, 2024). Together this supports the idea that the mutational landscape of adenomas is both unique but still reflective of the signatures observed in cancer.

As the two most mutated genes, *APC* and *KRAS* were expected to co-mutate in the full cohort, however, this was only observed in right colonic polyps and metachronous polyps. Hypermutated CRCs have been reported to have *APC*/*KRAS* co-mutations (Nunes *et al.*, 2024). However, some somatic interactions signatures appeared consistently throughout this analysis. *SOX9* was co-mutated with both *APC* and *KRAS* in left colonic polyps, and in polyps from patients stratified as high-risk according to current guidelines. This reflects results reported in CRC, where Javier et al. (2016) found a significant correlation between *KRAS* mutations and *APC* mutations in *SOX9*-mutant CRC. *HLA-B* was co-mutated with *KRAS* in left colonic polyps and polyps from patients stratified as high-risk according to current guidelines. *HLA* expression has been reported to be reduced in *KRAS* mutated CRC (Fu *et al.*, 2020), and as *KRAS* mutations lead to neoantigens that are presented by *HLA* molecules (Zhu *et al.*, 2025), co-mutations between the two genes may be expected and could affect the anti-tumour immune response.

This analysis focussed solely on the mutational landscape of adenomas, and some of the stratified subsets presented here are underpowered. The expansion of both the sample pool and a comparison between adenomas and serrated lesions are warranted and could lead to more robust results. Regardless, the mutational patterns observed here lead to the follow-up question of whether the proteins made by the mutated genes in these index adenomas influence the development of metachronous polyps or CRC. Of special interest are the genes involved in the Wnt signalling pathway. Not only was Wnt signalling the most mutated pathway in this cohort, but also its effector *SOX9* was co-mutated with *APC* and *KRAS* in the high-risk patients. Hence, Wnt signalling presented itself as a viable route of investigation in the upcoming chapters within this work.

Chapter 5 Examination of Wnt Signalling in Colorectal Adenomas

5.1 Introduction

Wnt signalling is a major biological process that controls various functions from cell proliferation and cell fate determination during embryonic development to adult homeostasis maintenance (MacDonald, Tamai and He, 2009; Liu et al., 2022). It is mainly achieved through the Wnt/ß-Catenin pathway. This pathway depends on the translocation of ß-Catenin from the cytoplasm into the nucleus and its subsequent interaction with the transcription factor T cell factor/lymphoid enhancer-binding factor (TCF/LEF). In the absence of a Wnt molecule, a ß-Catenin destruction complex comprised of Axin, APC, GSK3, and CKIα (Figure 5.1) regulates the levels of cytosolic ß-Catenin by ubiquitination, thus priming it for proteasomal digestion. However, when the pathway is activated, the cell surface proteins LRP and FZD recruit the ß-Catenin destruction complex to the cell membrane, leading to its destabilisation and the loss of its ability to degrade cytosolic ß-Catenin. ß-Catenin then translocates to the nucleus, where it binds to TCF/LEF and begins transcription of genes such as *MYC*, that are involved in a host of cellular processes, including cell proliferation and differentiation.

ß-Catenin is not only important for cell proliferation, but it also has a role in cell-cell adhesion. In 1989, Ozawa, Barbault, and Kemler discovered α-Catenin, ß-Catenin, and γ-Catenin in their capacity as structural linkers. At the cell surface, ß-Catenin is bound to the intracellular portion of E-Cadherin, as a link in a chain between the latter and α-Catenin and the actin cytoskeleton. However, E-Cadherin-bound ß-Catenin can be released into the cytosol should an extracellular molecule, such as *F. nucleatum,* bind to E-Cadherin, as discussed previously in Figure 1.10. When the ß-Catenin destruction complex is available, the levels of ß-Catenin are controlled. However, where mutations are present e.g. in the *APC* gene, the destruction complex becomes nonfunctional, allowing ß-Catenin to accumulate in the cytosol and eventually translocate to the nucleus (Zhang *et al.*, 2024).

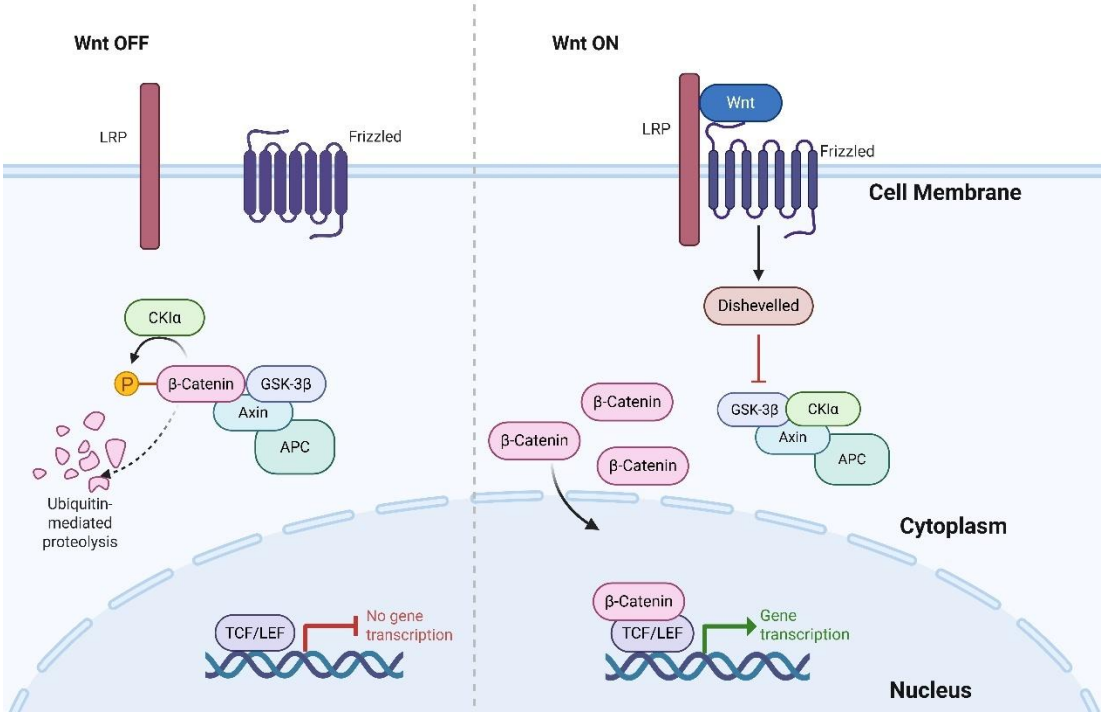


Figure 5.1. Schematic Representation of Wnt Signalling.
[OFF] When there is no Wnt molecule present, cell surface LRP and Frizzled remain unbound. The β-Catenin destruction complex (APC, Axin, GSK-3β, CD1α) is active in the cytosol, ubiquitinating β-Catenin, leading to its degradation, ultimately preventing β-Catenin-mediated gene transcription. **[ON]** Upon Wnt binding to LRP and Frizzled, the β-Catenin destruction complex translocates to the membrane, where phosphorylation of Axin binds it to the internal tail of LRP, decreasing its stability. Dishevelled is phosphorylated and activated, which in turn inhibits GSK. This allows β-Catenin to accumulate in the cytosol. It then localises to the nucleus, where it binds to TCF/LEF transcription factor, and begins the transcription of target genes involved in cell proliferation.

While the effects of APC on colorectal adenoma development are well-established, further investigation into components and effectors of Wnt signalling was warranted. This chapter will investigate E-Cadherin and ß-Catenin at the protein level for roles in regulating downstream Wnt signalling effects using IHC on INCISE patient tissue, and *in vitro* using Wnt stimulated PDOs.

5.2 Protein Expression of Selected Wnt Signalling Pathway Proteins and the Detection of Metachronous Polyps or CRC

5.2.1 Antibody Specificity and Sensitivity

The specificity and sensitivity of the antibodies used for tissue staining were confirmed using the methods described in 2.2.1. DepMap was used to identify cell lines expressing E-Cadherin, and ß-Catenin varying levels (Figure 5.2).

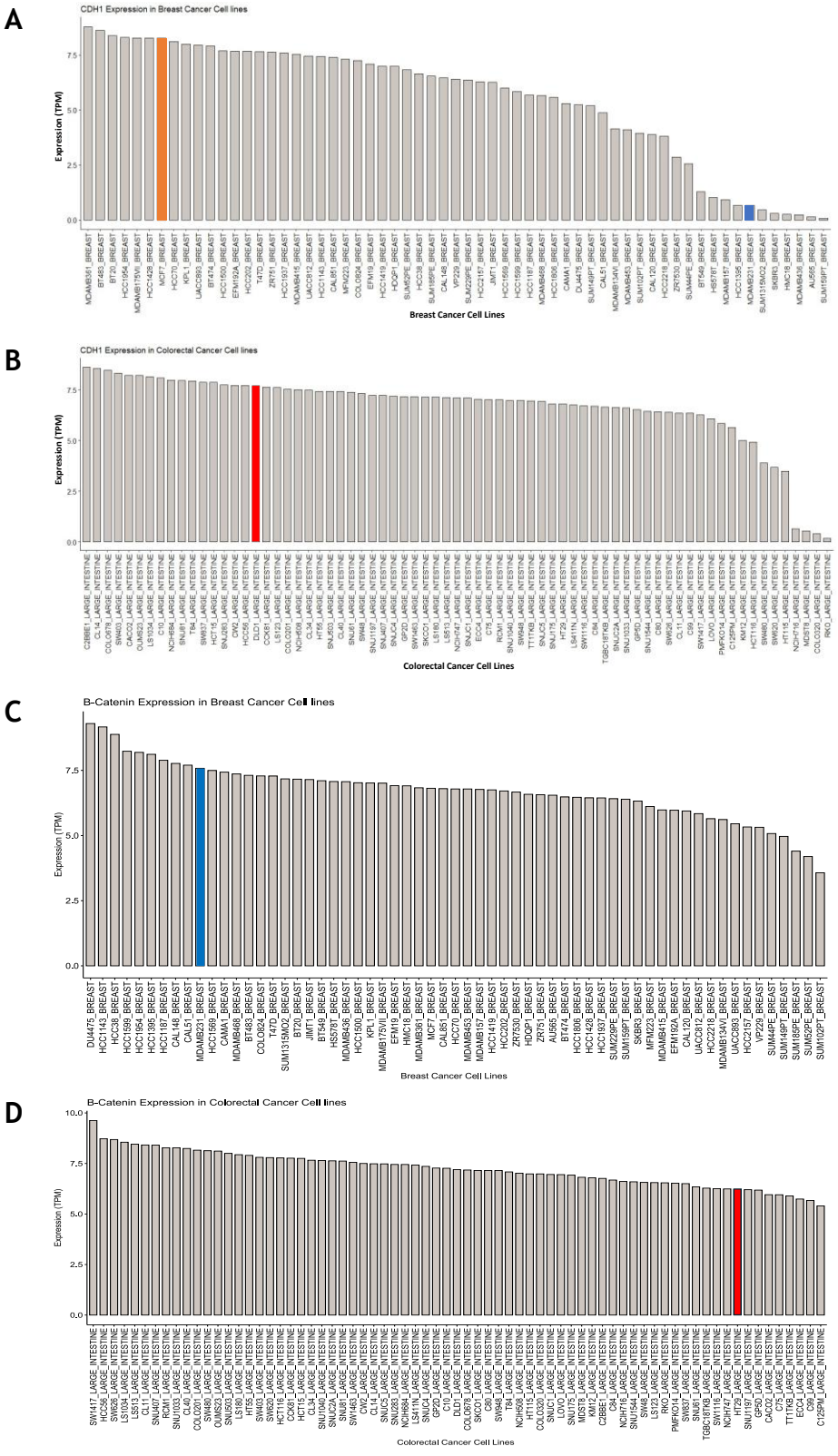


Figure 5.2. DepMap Cell Line Selector for Cell Lines Expressing Target Proteins. **[A]** E-Cadherin expression in breast cancer cell lines. A low-expressing cell line, MDA-MB-231 (blue), and a medium-expressing cell line, MCF-7 (orange), were selected for antibody validation. **[B]** E-Cadherin expression in colorectal cancer cell lines. A high-expressing cell line, DLD-1 (red), was selected for antibody validation. **[C]** β-Catenin expression in breast cancer cell lines. A low-expressing cell line, MDA-MB-231 (blue), was selected for antibody validation. **[D]** β-Catenin expression in colorectal cancer cell lines. A high-expressing cell line, HT-29 (red), was selected for antibody validation.

Lysates from the selected cell lines were generated as described in 2.2.1.3. Western blotting was carried out as described in 2.2.1.4. As shown in Figure 5.3A-B, E-Cadherin was more weakly expressed in both MDA-MB-231 and MCF-7 cells than in DLD-1 cells. This validated the specificity of the antibody to its target. Similarly, β-Catenin expression was weaker in MDA-MB-231 than HT29 (Figure 5.3C).

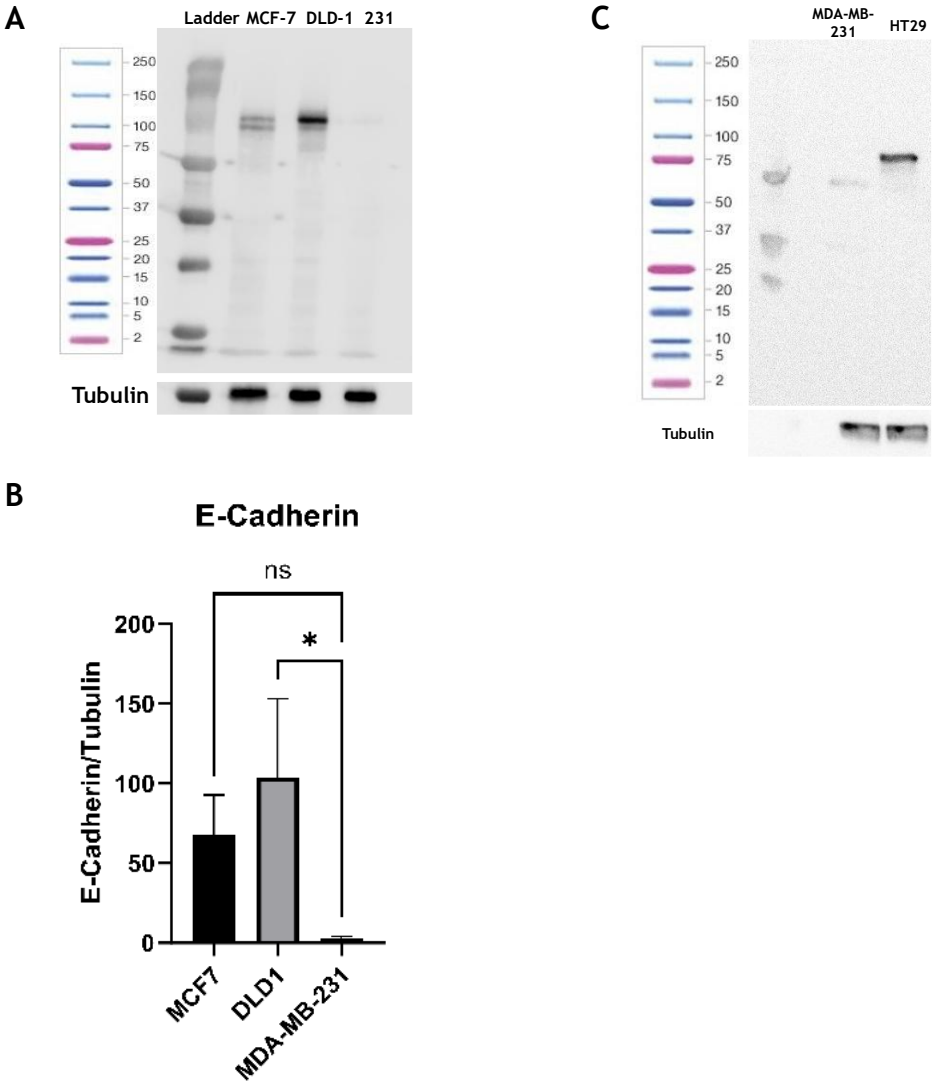


Figure 5.3. Antibody Specificity Test.
Western blotting was used to validate the specificity of the selected antibodies using lysates from cell lines identified using DepMap. **[A]** Western blot of E-Cadherin expression in lysates from MCF-7 (medium expression), DLD-1 (high expression), and MDA-MB-231 (low expression) at 80-135kDa. Tubulin was used as a loading control. **[B]** Expression of E-Cadherin against the loading control. N=3. Error bars represent standard deviation. ns denotes $p>0.05$ on a One-way ANOVA. **[C]** Western blot of β-Catenin expression in lysates from MDA-MB-231 (low expression) and HT-29 (high expression) at ~80kDa. Tubulin was used as a loading control. Antibody validation of β-Catenin was carried out by Dr. Amna Matly (Matly, 2024). β-Catenin/Tubulin bar chart was unfortunately not available for β-Catenin.

To further test the specificity of the antibodies, cell pellets were stained as described in 2.2.1.5. The difference in expression between identified low- and high-expressing cell lines is presented in Figure 5.4.

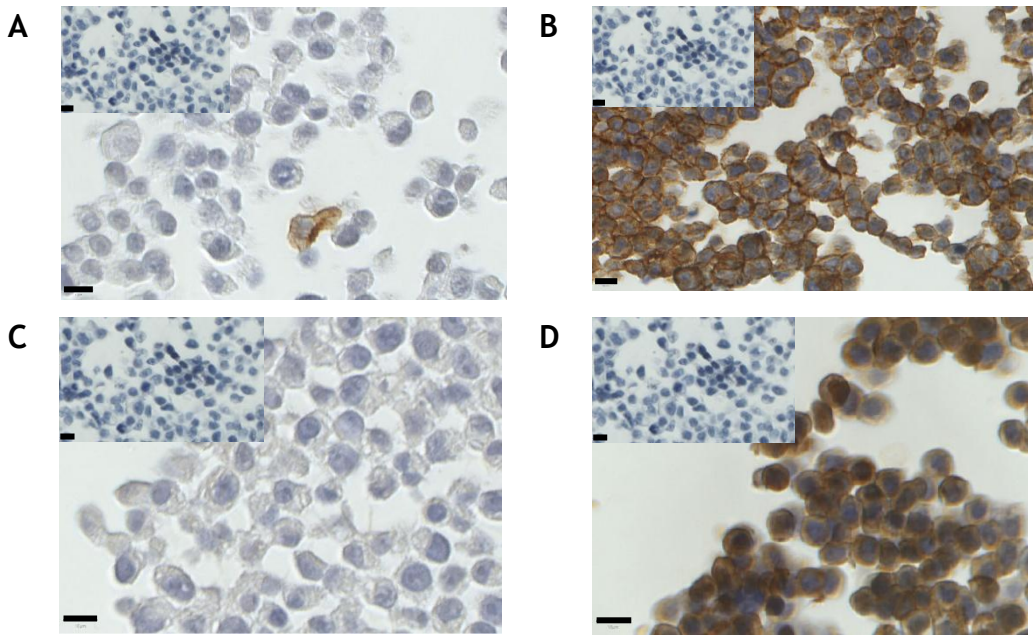


Figure 5.4. Antibody Specificity Tests.
IHC on cell pellets of cell lines identified through DepMap. No-antibody controls were included and are shown at the top left corner of each image. All staining was carried out under the conditions described in Table 2.5. **[A]** MCF-7 cell pellets (low expression) stained for E-Cadherin. **[B]** DLD-1 cell pellets (high expression) stained for E-Cadherin. **[C]** MDA-MB-231 cell pellets (low expression) stained for β-Catenin. **[D]** HT29 cell pellets (high expression) stained for β-Catenin. (Performed by Dr. Amna Matly). Black scale bars = 10μm.

5.2.2 E-Cadherin Protein Expression Is Not Associated with the Detection of Metachronous Polyps or CRC

5.2.2.1 IHC and Staining Assessment of E-Cadherin

E-Cadherin staining was performed as described in 2.2.2.1. Expression was observed in the membrane and cytoplasm of the epithelium of adenomas (Figure 5.5A). Staining assessment was carried out digitally using Visiopharm by Dr. Christopher Bigley to generate H-Scores (Figure 5.5B). The negative no-antibody control verifies that the signal observed is not due to the interaction between the tissue and any of the reagents used in the staining process. The positive tissue control further confirms the antibody can detect its true target (Figure 5.5C).

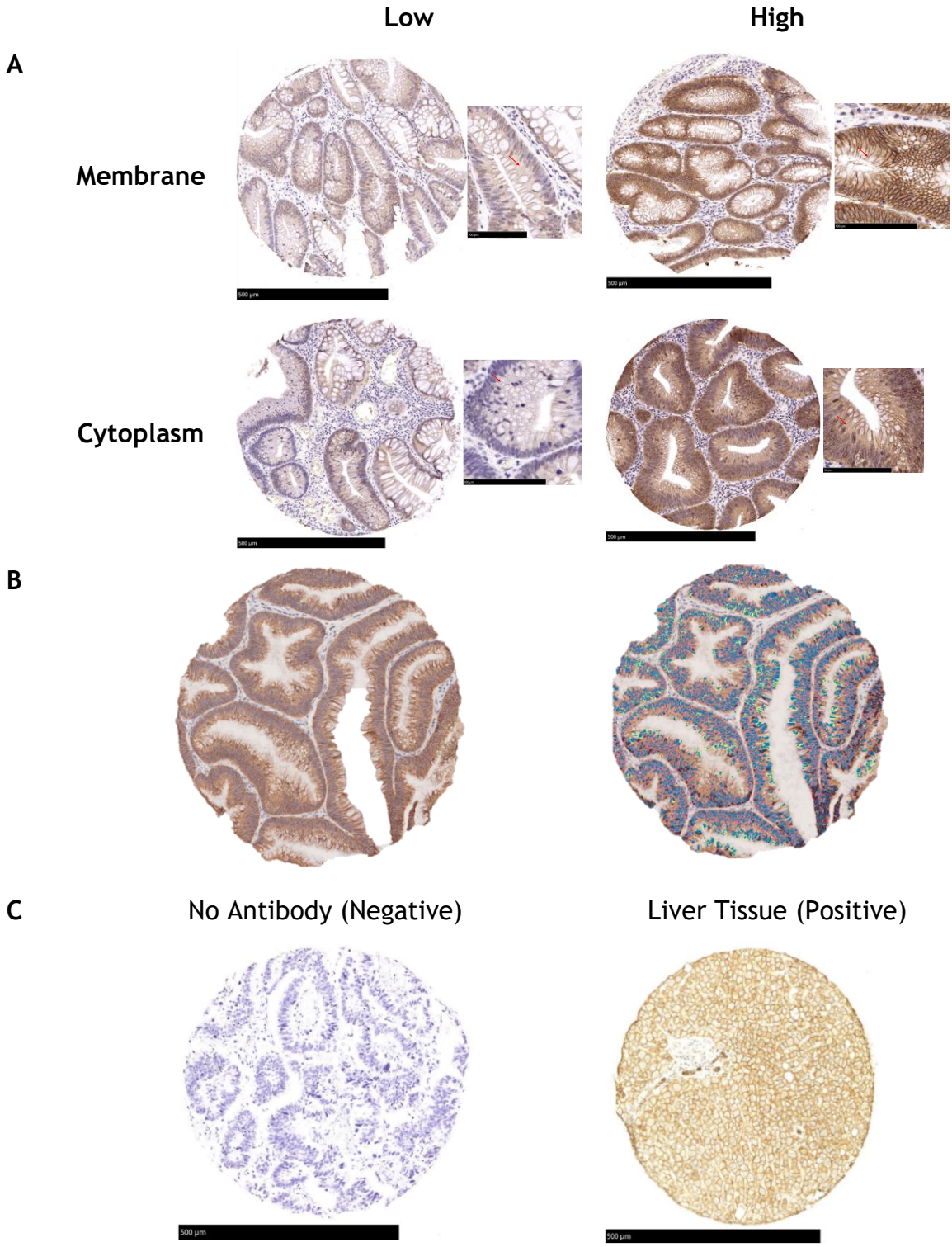


Figure 5.5. IHC and Staining Assessment of E-Cadherin.
[A] Examples of low and high expression of membranous and cytoplasmic E-Cadherin in polyp tissue, with zoomed-in images (right) detailing the membranous and cytoplasmic staining patterns clarified with red arrows. **[B]** Example of nuclear detection and stain scoring using Visiopharm (Performed by Dr. Christopher Bigley). **[C]** Two controls were included in the staining. The negative no-antibody control used surplus CRC tissue, while the positive control used surplus liver cancer tissue. Scale bars of full cores = 500µm, scale bars of zoomed-in images = 100µm.

5.2.2.2 Quality Control of Raw Data

As described in 2.1.1.3, the INCISE cohort was split into a training and a test dataset. The INCISE TMA comprised of cores taken from both the luminal and

basal regions of the adenomas to better capture tissue heterogeneity (Figure 2.3). Scatter plots (Figure 5.6) of raw H-scores of membranous and cytoplasmic E-Cadherin expression in the luminal and basal epithelium were generated against the total (combined luminal and basal epithelium) adenoma scores, to investigate whether the datasets reflect each other and whether there was value in investigating each histological region separately. This not only showed that E-Cadherin protein expression is not specific to histological region but also allowed for the use of a total adenoma H-score comprised of the average of each patient's luminal and basal core scores. Additionally, the test dataset is matched to the characteristics of the training dataset.

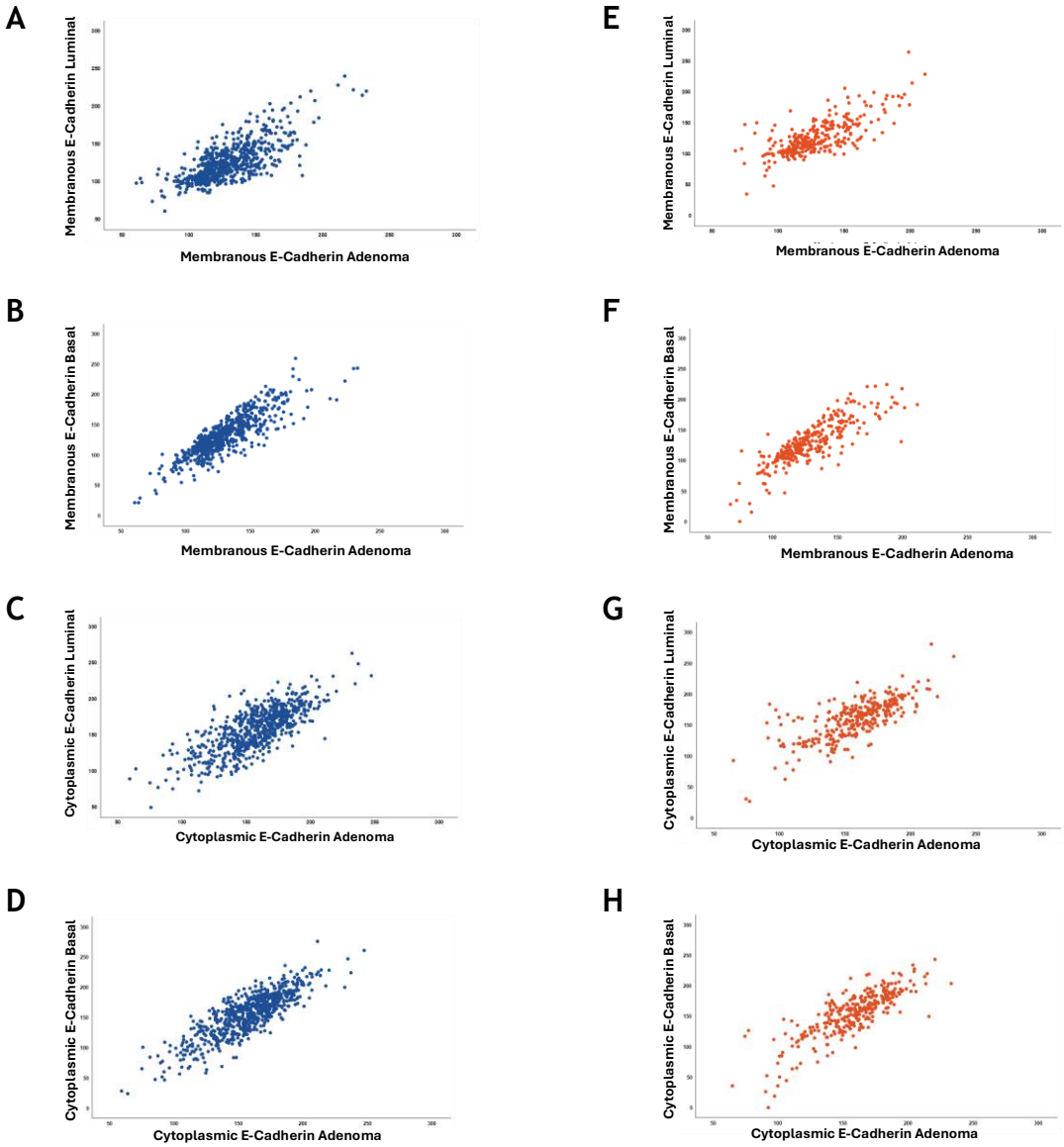


Figure 5.6. Correlation of Polyp Histological Regions Against Total Polyp Adenomas in the INCISE Training and Test Datasets.
Scatter plot of membranous E-Cadherin expression in the luminal epithelium of the polyp vs membranous E-Cadherin expression in the total adenoma of the polyp in the **[A]** training and **[E]** test datasets. Scatter plot of membranous E-Cadherin expression in the basal epithelium of the polyp vs membranous E-Cadherin expression in the total adenoma of the polyp in the **[B]** training and **[F]** test datasets. Scatter plot of cytoplasmic E-Cadherin expression in the luminal epithelium of the polyp vs cytoplasmic E-Cadherin expression in the total adenoma of the polyp in the **[C]** training and **[G]** test datasets. Scatter plot of cytoplasmic E-Cadherin expression in the basal epithelium of the polyp vs cytoplasmic E-Cadherin expression in the total adenoma of the polyp in the **[D]** training and **[H]** test datasets. Both the x-axis and y-axis range from 0 to 300 in increments of 50, to align with the range of scores generated by the weighted H-Score method.

Figure 5.7A-B demonstrates the distribution of membranous and cytoplasmic E-Cadherin expression in the whole cohort. While membranous E-Cadherin is slightly skewed to the left, cytoplasmic E-Cadherin is more normally distributed. These trends are reflected in both the training and test datasets (Figure 5.7C-F).

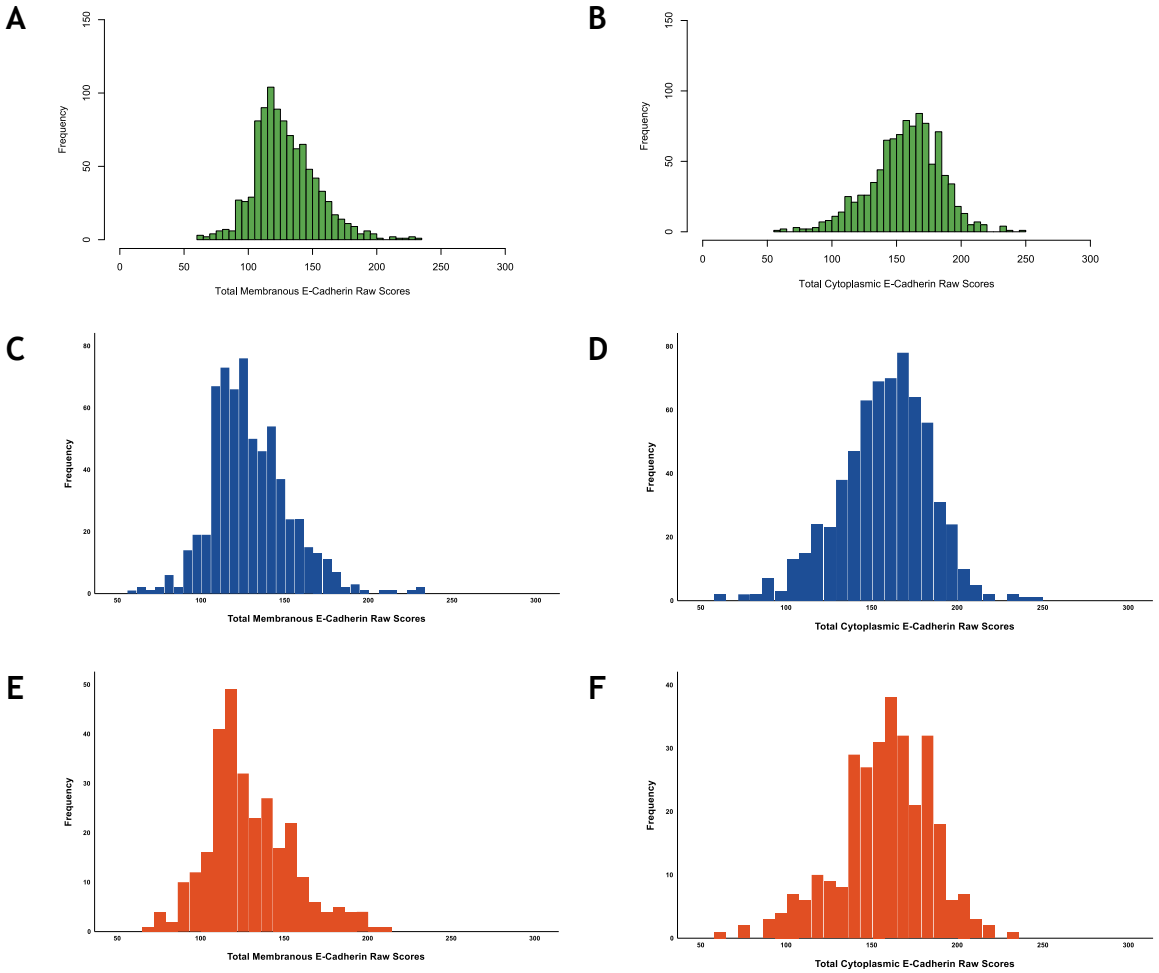


Figure 5.7. Distribution of Raw E-Cadherin H-scores.
Histogram of the distribution of **[A]** membranous H-scores and **[B]** cytoplasmic H-scores of E-Cadherin, in the whole cohort. Histograms of total membranous E-Cadherin scores in the **[C]** training and **[E]** test datasets. Histograms of total cytoplasmic E-Cadherin scores in the **[D]** training and **[F]** test datasets.

5.2.2.3 Dichotomisation of E-Cadherin H-scores

H-scores were dichotomised as described in 2.2.10. As shown in Figure 5.8, the generated thresholds are difficult to visually determine by a human observer, and so comparable thresholds were chosen (125 for membranous expression; 170 for cytoplasmic expression). This was done by selecting a threshold within the same maximally selected statistical peak as the original threshold, or one within a comparable peak.

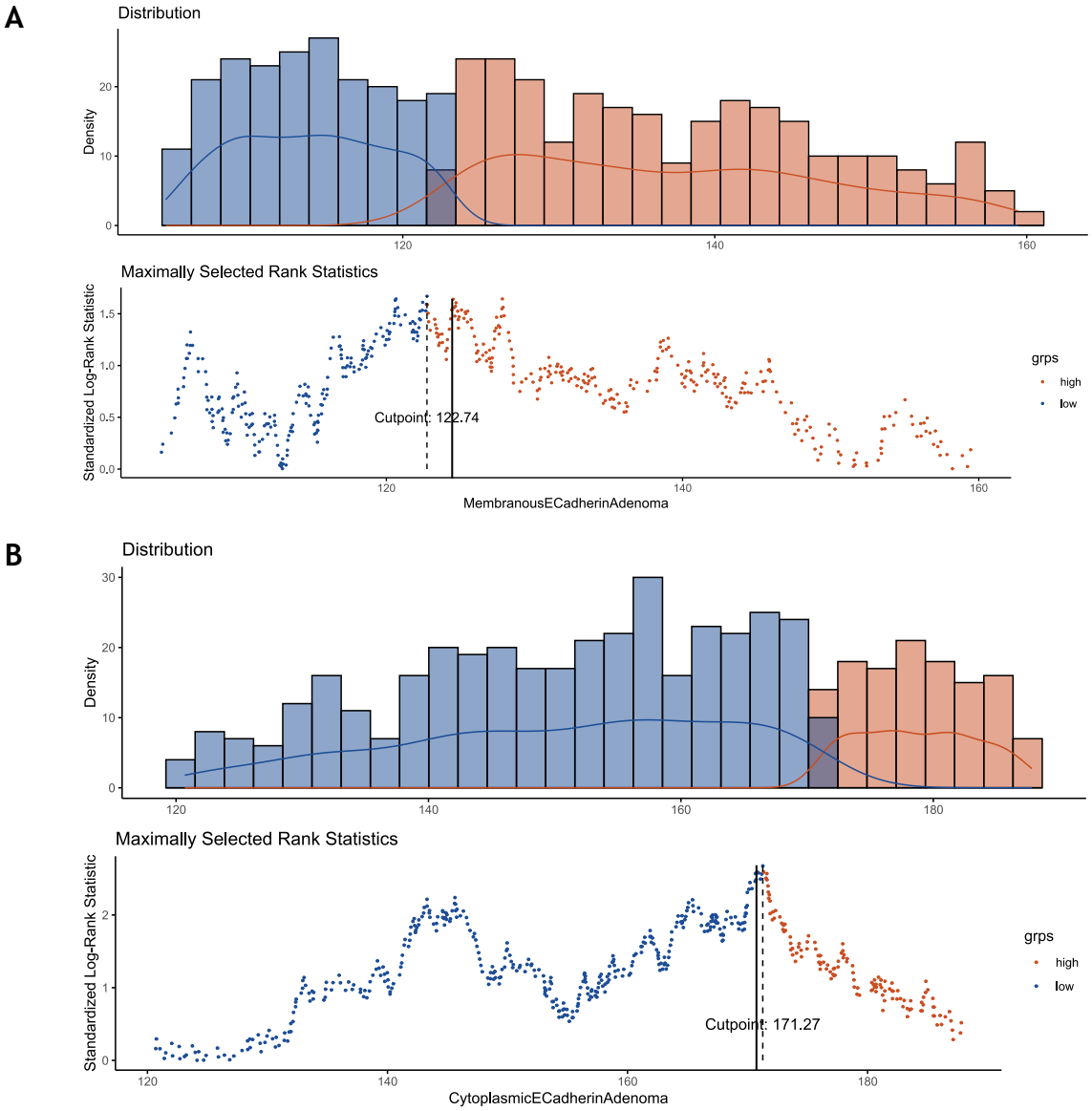


Figure 5.8. Thresholding of H-scores.
Threshold generation of **[A]** membranous and **[B]** cytoplasmic E-Cadherin using maximally selected rank statistics on the training dataset. Distribution by density and standardized log rank maximally selected statistics by E-Cadherin Histo-score distribution were both grouped in intervals of 20. The block lines indicate the chosen membrane threshold of 125 and the cytoplasmic threshold of 170. N=640.

The distribution of scores considered as low expression and high expression of membranous E-Cadherin in the training and test dataset was comparable (Figure 5.9A,E and Figure 5.9B,F respectively). This was true for cytoplasmic E-Cadherin expression (Figure 5.9C,G and Figure 5.9D,H).

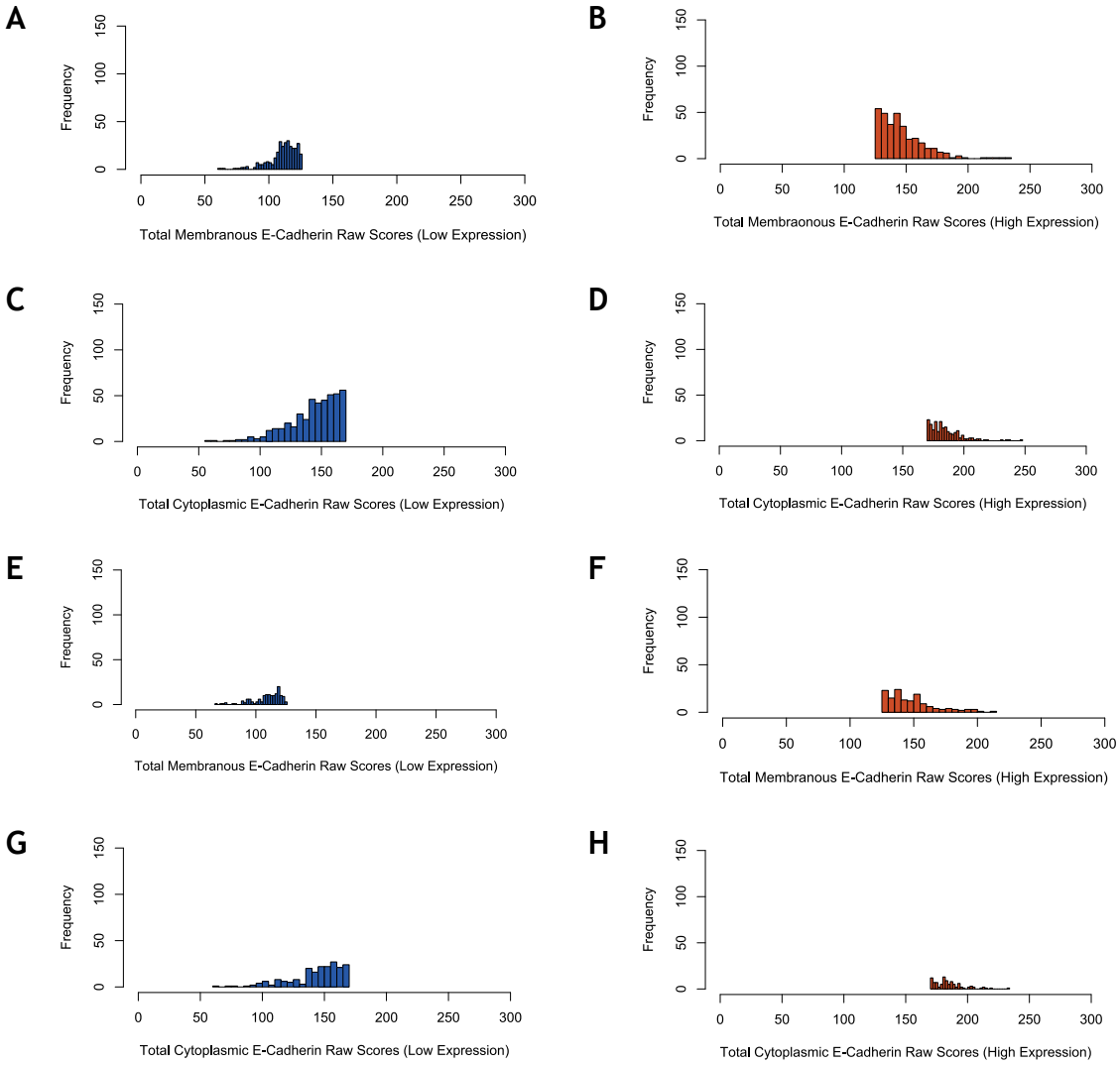


Figure 5.9. Histograms of Distribution of H-scores Across Low and High Expression Groups.
[A] Low and **[B]** high membranous E-Cadherin expression in the training dataset. **[C]** Low and **[D]** high cytoplasmic E-Cadherin expression in the training dataset. **[E]** Low and **[F]** high membranous E-Cadherin expression in the test dataset. **[G]** Low and **[H]** high cytoplasmic E-Cadherin expression in the test dataset.

5.2.2.4 Expression of E-Cadherin and Metachronous Polyp or CRC Detection

After implementing the exclusion criteria described in 2.1.1.3.1 and 2.1.1.3.2 and removing all invalid cases (n=178 in the training dataset and n=66 in the test dataset), 640 cases were included in the training dataset, and 292 cases were included in the test dataset.

Log rank survival analysis was carried out on the dichotomised groups as described in 2.2.10 to explore any association between the expression of E-Cadherin and the detection of metachronous polyps or CRC. As shown in Figure 5.10, membranous expression of E-Cadherin was not associated with

metachronous polyp or CRC detection in either the training (HR 0.852, 95% CI 0.688-1.055; p=0.14) or the test (HR 1.045, 95% CI 0.771-1.417; p=0.77) datasets. There was no consistent trend between the two datasets, with low membranous E-Cadherin expressing patients showing a slightly shorter median time to detection of metachronous polyps (56 months) than those with high membranous E-Cadherin expression, who never reach median risk, in the training dataset. Inversely, high expressors of membranous E-Cadherin in the test dataset have a shorter median time to detection of metachronous polyps (49 months) than their low expressing counterparts (59 months).

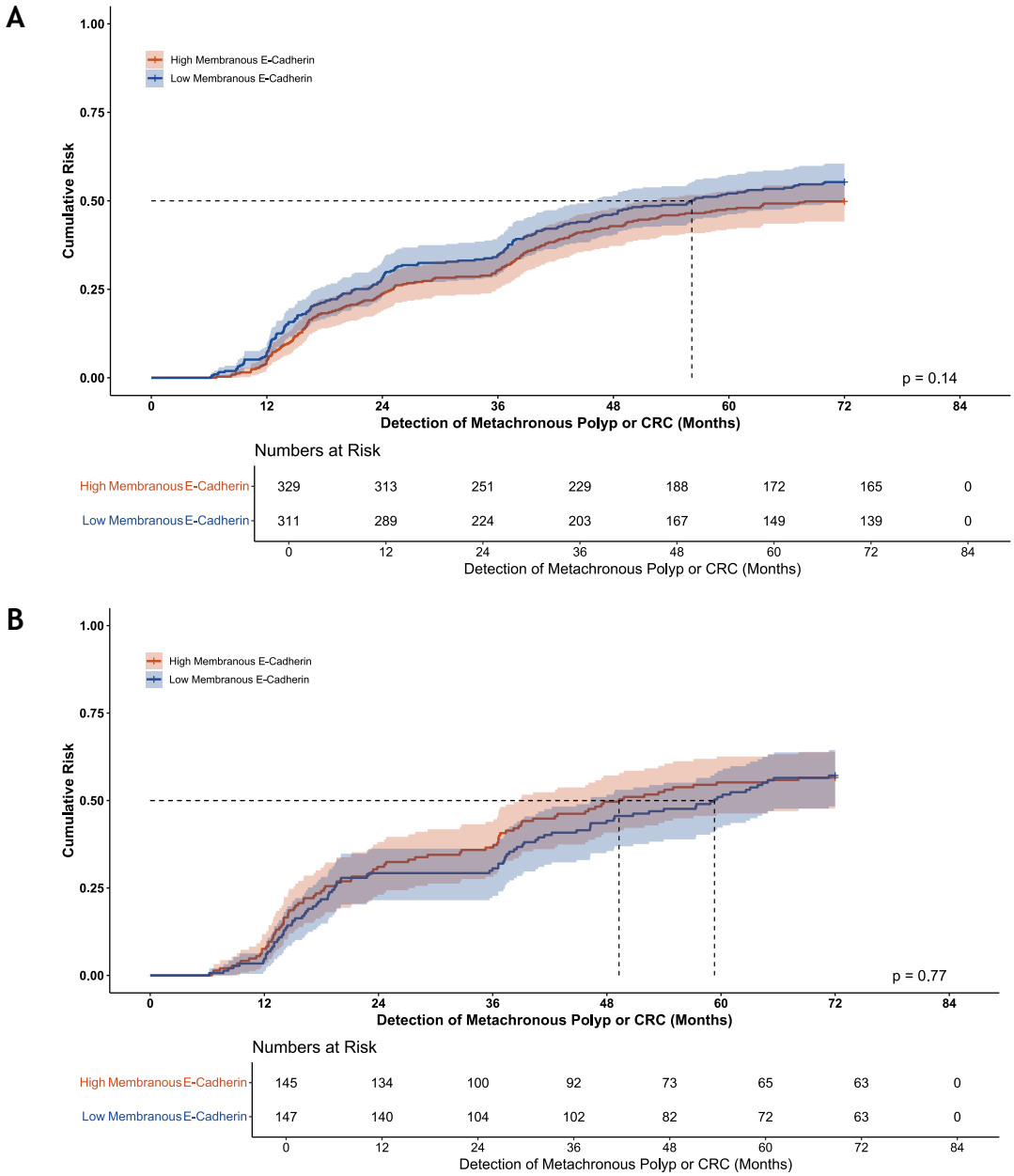


Figure 5.10. Membranous Expression of E-Cadherin in Adenomas.
Expression of E-Cadherin in the epithelia of adenomas in the **[A]** training (n=640) and **[B]** test (n=292) datasets. Shading is indicative of 95% CI. Dotted lines are time to median risk.

Low cytoplasmic expression of E-Cadherin was significantly associated with the detection of metachronous polyps or CRC (Figure 5.11), with a median time to detection of 56 months in the training cohort (HR 0.768, 95% CI 0.603-0.964; p=0.017). While this did not validate in the test dataset (HR 0.821, 95% CI 0.590-1.143; p=0.2), a similar trend was observed with patients expressing low cytoplasmic E-Cadherin having a shorter time to metachronous polyp or CRC detection (49 months) than those with high expression of cytoplasmic E-Cadherin (60 months).

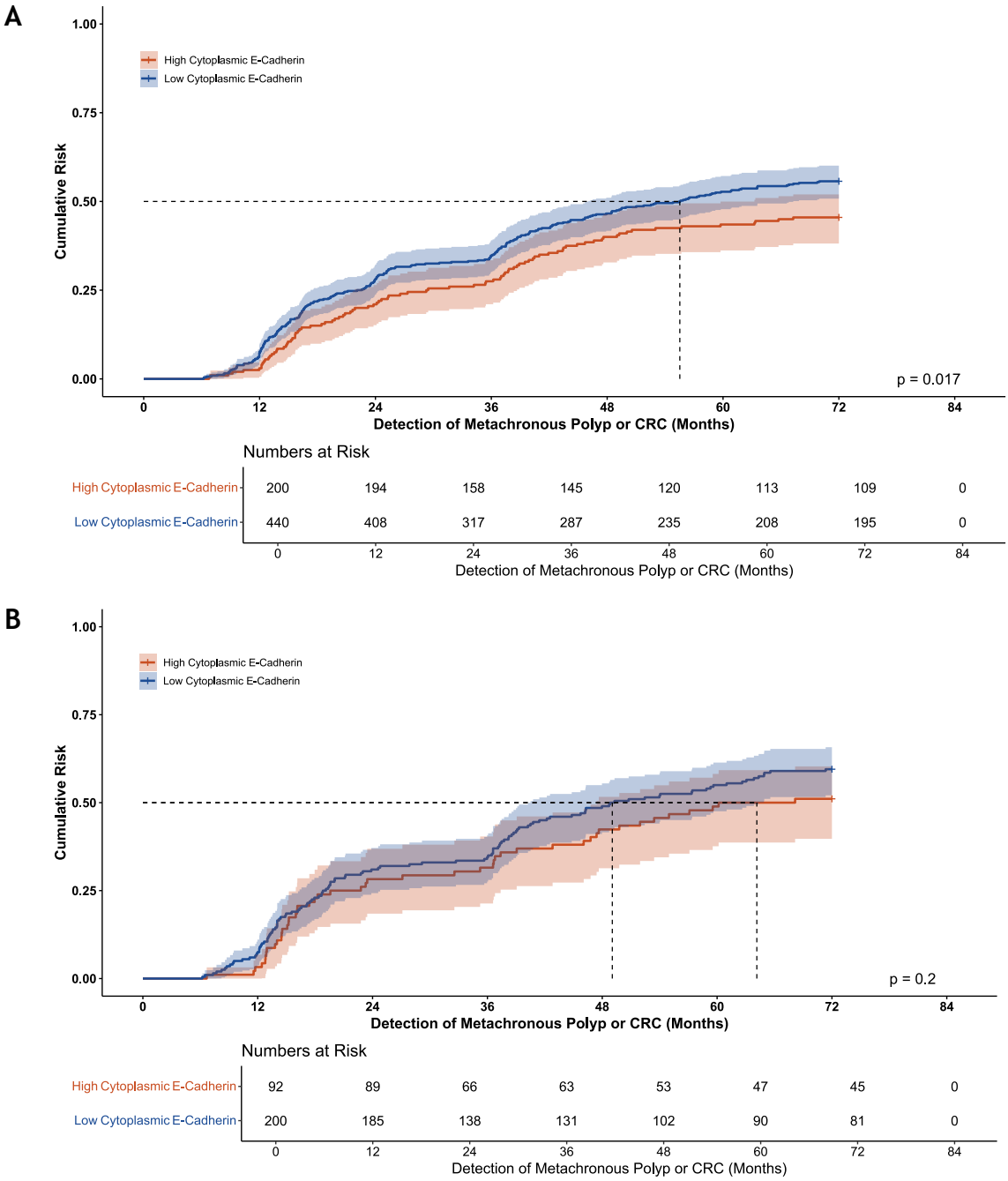


Figure 5.11. Cytoplasmic Expression of E-Cadherin in Adenomas.
Expression of E-Cadherin in the epithelia of adenomas in the **[A]** training (n=640) and **[B]** test (n=292) datasets. Shading is indicative of 95% CI. Dotted lines are time to median risk.

5.2.2.5 Expression of E-Cadherin, Clinical Characteristics, and the Detection of Metachronous Polyps or CRC

Univariate Cox regression was performed to determine associations between E-Cadherin expression in the membrane and the cytoplasm and relevant clinical factors. In the training dataset, (Table 5.1), male sex ($p<0.0001$), left colonic polyps ($p=0.002$), multiple index polyps ($p<0.0001$), and high BSG2020 risk ($p<0.0001$) all significantly associated with the detection of metachronous polyps or CRC. Neither high membranous E-Cadherin expression ($p=0.142$) nor high cytoplasmic E-Cadherin expression ($p=0.23$) was statistically associated with the detection of metachronous polyps or CRC.

In the test dataset (Table 5.1), left colonic polyps ($p=0.003$), multiple index polyps ($p=0.013$ for 2-4 polyps, $p<0.0001$ for 5+ polyps), and high BSG2020 risk ($p=0.002$) all significantly associated with the detection of metachronous polyps or CRC. However, E-Cadherin expression in the membrane or cytoplasm was not.

Table 5.1. Univariate Cox Regression Analysis on E-Cadherin Expression in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE TMA Cohort

		Training Dataset (n=640)				Test Dataset (n=292)			
		N	HR	95% CI	P	N	HR	95% CI	P
Sex	Female	185	1.0	-	-	77	1.0	-	-
	Male	455	1.540	1.228-1.930	**<0.001**	215	1.072	0.778-1.478	0.670
Age	Screening Age	615	1.0	-	-	279	1.0	-	-
	Above Screening Age	25	0.863	0.515-1.445	0.575	13	1.453	0.811-2.603	0.209
Site	Right Colon	75	1.0	-	-	52	1.0	-	-
	Left Colon	467	0.640	0.484-0.847	**0.002**	210	0.596	0.424-0.837	**0.003**
	Rectum	98	0.862	0.607-1.222	0.404	30	0.676	0.402-1.137	0.140
High Grade Dysplasia	Absent	532	1.0	-	-	249	1.0	-	-
	Present	108	1.155	0.897-1.487	0.264	43	1.114	0.778-1.597	0.555
Number of Polyps	1	221	1.0	-	-	85	1.0	-	-
	2-4	340	1.513	1.210-1.891	**<0.0001**	156	1.547	1.095-2.185	**0.013**
	5+	79	2.522	1.859-3.422	**<0.0001**	51	2.584	1.711-3.902	**<0.0001**
Polyp Type	Tubular	270	1.0	-	-	115	1.0	-	-
	Tubulovillous	319	0.980	0.803-1.196	0.841	165	0.846	0.637-1.123	0.248
	Villous	50	0.979	0.658-1.457	0.918	12	1.741	0.929-3.263	0.083
BSG2020 Guidelines	Low Risk	223	1.0	-	-	87	1.0	-	-
	High Risk	417	1.659	1.338-2.056	**<0.001**	205	1.687	1.218-2.338	**0.002**
Membranous E-Cadherin Expression in Adenoma	Low Expression	311	1.0	-	-	147	1.0	-	-
	High Expression	329	0.852	0.688-1.055	0.142	145	1.045	0.771-1.417	0.774
Cytoplasmic E-Cadherin Expression in Adenoma	Low Expression	440	1.0	-	-	200	1.0	-	-
	High Expression	200	0.762	0.603-0.964	0.23	92	0.821	0.590-1.143	0.242

Screening Age=50-74 years old. HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. Cox regression analysis was carried out to obtain hazard ratios.

As E-Cadherin expression in either compartment was not validated, investigation of the association of E-Cadherin expression in colorectal adenomas was not taken further.

5.2.3 β-Catenin Protein Expression Is Not Associated with the Detection of Metachronous Polyps or CRC

5.2.3.1 IHC and Staining Assessment of β-Catenin

ß-Catenin staining was carried out as described in 2.2.2.1. It was expressed in the membrane, cytoplasm, and nucleus of the epithelium of adenomas (Figure 5.12A). Scoring of expression levels of membranous ß-Catenin was performed using Visiopharm by Dr. Christopher Bigley. Cytoplasmic and nuclear scoring of ß-Catenin was performed manually by Sara Al-Badran as described in 2.2.3.1. The negative isotype-matched control ensured that any staining seen was a true signal and not due to non-specific interactions between the tissue. Similarly, the inclusion of a negative no-antibody control ensured that the signal observed is not due to the interaction between the tissue and any of the reagents used in the staining process. A positive tissue control further confirmed the antibody can detect its true target (Figure 5.12B).

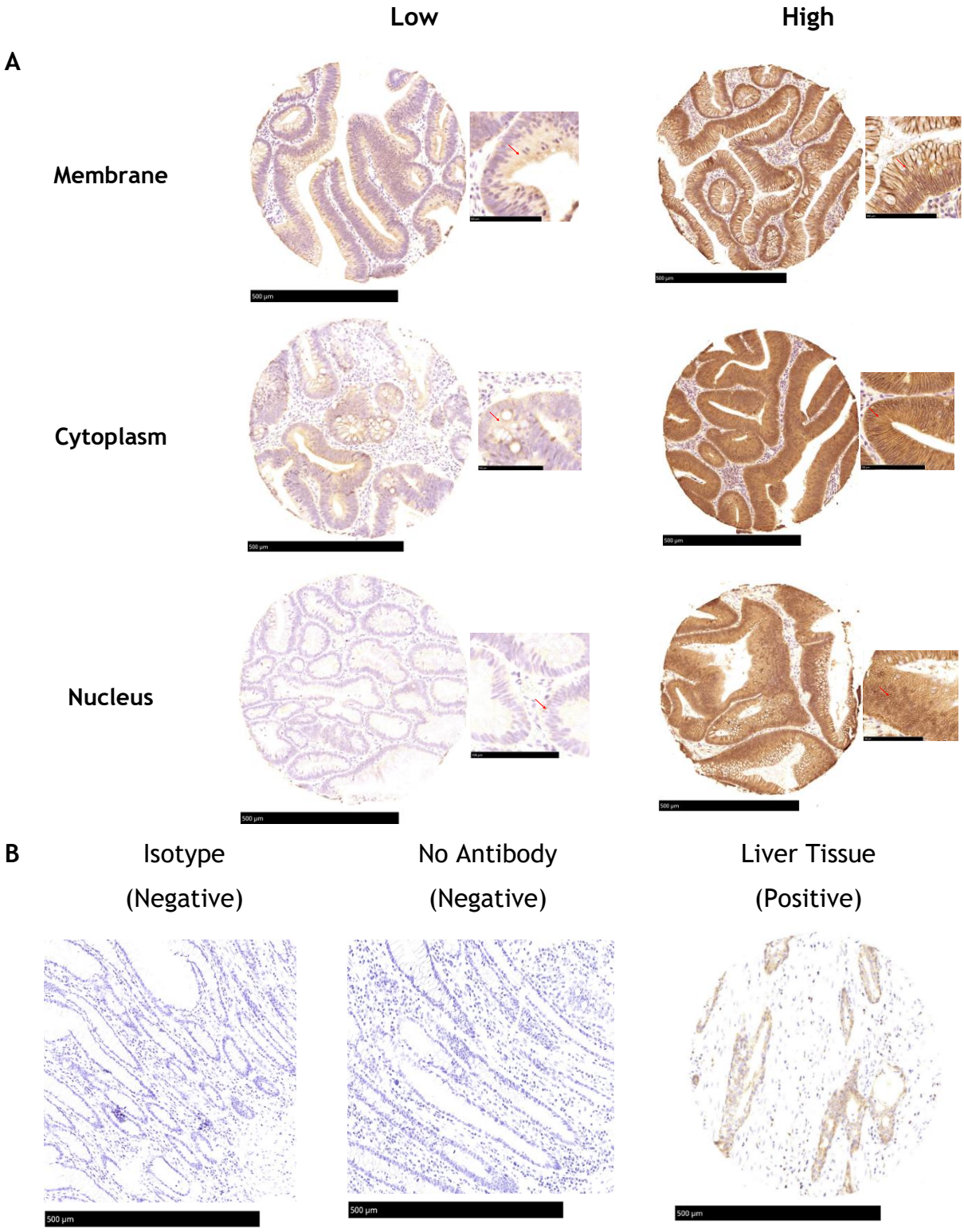


Figure 5.12. IHC and Staining Assessment of β-Catenin.
[A] Examples of low and high expression of membranous, cytoplasmic, and nuclear β-Catenin in polyp tissue, with zoomed-in images (right) detailing the membranous and cytoplasmic staining patterns clarified with red arrows. **[B]** Three controls were included in the staining. The negative matched-isotype control and negative no-antibody control used surplus CRC tissue, while the positive control used surplus liver cancer tissue. Scale bars of full cores = 500µm, scale bars of zoomed-in images = 100µm.

5.2.3.2 Quality Control of Raw Data

Scatter plots (Figure 5.13) of raw H-scores for membranous, cytoplasmic, and nuclear ß-Catenin expression in the luminal and basal epithelium were generated against the total (combined luminal and basal epithelium) adenoma scores in the training and test datasets, for correlation tests. ß-Catenin protein expression was not specific to geographical region, hence, a total adenoma H-score comprised of the average of each patient's luminal and basal core scores was used for further analysis.

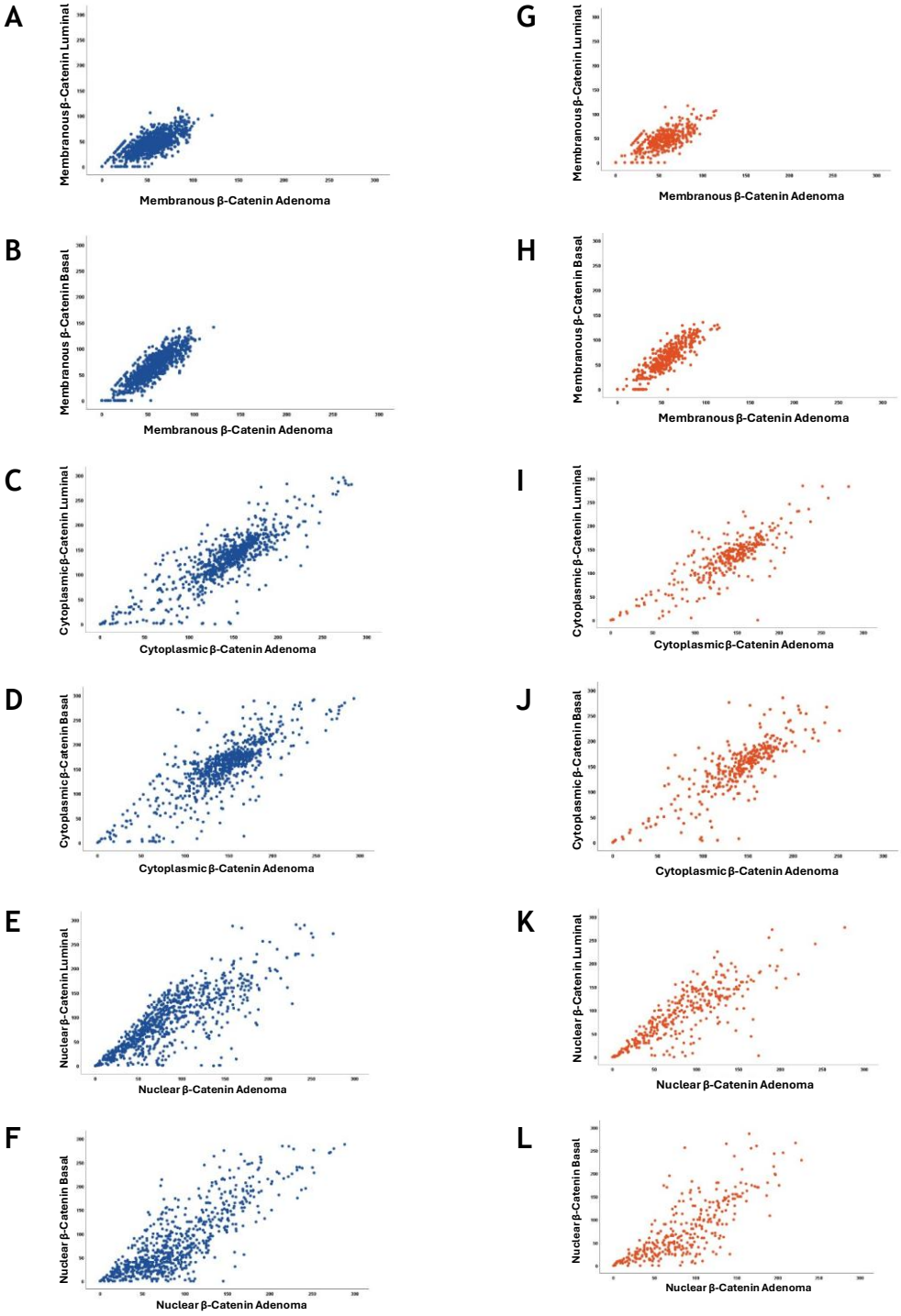


Figure 5.13. Correlation of Polyp Histological Regions Against Total Polyp Adenomas in the INCISE Training and Test Datasets.
Scatter plot of membranous β-Catenin expression in the luminal epithelium vs membranous β-Catenin expression in the total adenoma of the polyp in the **[A]** training and **[G]** test datasets. Scatter plot of membranous β-Catenin expression in the basal epithelium vs membranous β-Catenin expression in the total adenoma of the polyp in the **[B]** training and **[H]** test datasets. Scatter plot of cytoplasmic β-Catenin expression in the luminal epithelium vs cytoplasmic β-Catenin expression in the total adenoma of the polyp in the **[C]** training and **[I]** test datasets. Scatter plot of cytoplasmic β-Catenin expression in the basal epithelium vs cytoplasmic β-Catenin expression in the total adenoma of the polyp in the **[D]** training and **[J]** test datasets. Scatter plot of nuclear β-Catenin expression in the luminal epithelium vs nuclear β-Catenin expression in the total adenoma of the polyp in the **[E]** training and **[K]** test datasets. Scatter plot of nuclear β-Catenin expression in the basal epithelium vs nuclear β-Catenin expression in the total adenoma of the polyp in the **[F]** training and **[L]** test datasets. Both the x-axis and y-axis range from 0 to 300 in increments of 50, to align with the range of scores generated by the weighted H-Score method.

As shown in Figure 5.14A, the distribution of membranous ß-Catenin in the whole cohort skews sharply to the left, representing an abundance of weak membranous staining. This was observed in both the training and test datasets (Figure 5.14D,G). Cytoplasmic ß-Catenin expression was more normally distributed in the whole cohort (Figure 5.14B), which is reflected in the training and test datasets (Figure 5.14E,H). The distribution of nuclear ß-Catenin scores was skewed to the right (Figure 5.14C) albeit not as sharply as membranous ß-Catenin. This is reflected in the training and test datasets (Figure 5.14F,I).

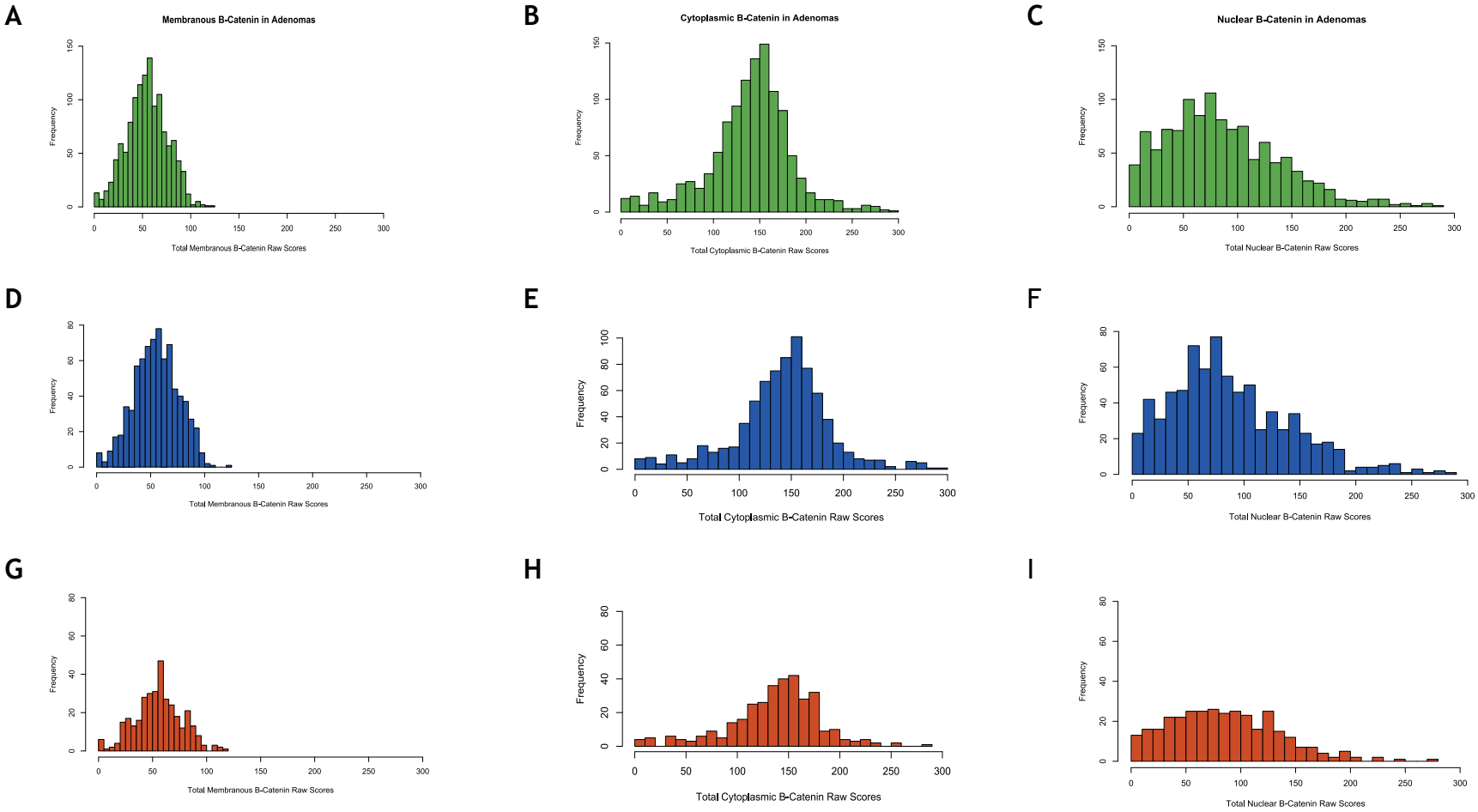


Figure 5.14. Distribution of Raw β-Catenin H-scores.
Histograms of the distribution of **[A]** membranous H-scores of β-Catenin, **[B]** cytoplasmic H-scores of β-Catenin, and **[C]** nuclear H-scores of β-Catenin, in the whole cohort. Histograms of the distribution of membranous β-Catenin score in the **[D]** training and **[G]** test dataset. Histograms of the distribution of cytoplasmic β-Catenin scores in the **[E]** training and **[H]** test dataset. Histograms of the distribution of nuclear β-Catenin scores in the **[F]** training and **[I]** test datasets

5.2.3.3 Dichotomisation of β-Catenin H-scores

Thresholds were generated as in 5.2.2.3. As for E-Cadherin, comparable thresholds to those generated were chosen (80 for membranous expression; 100 for cytoplasmic expression; 150 for nuclear expression) that could be determined by a human observer (Figure 5.15).

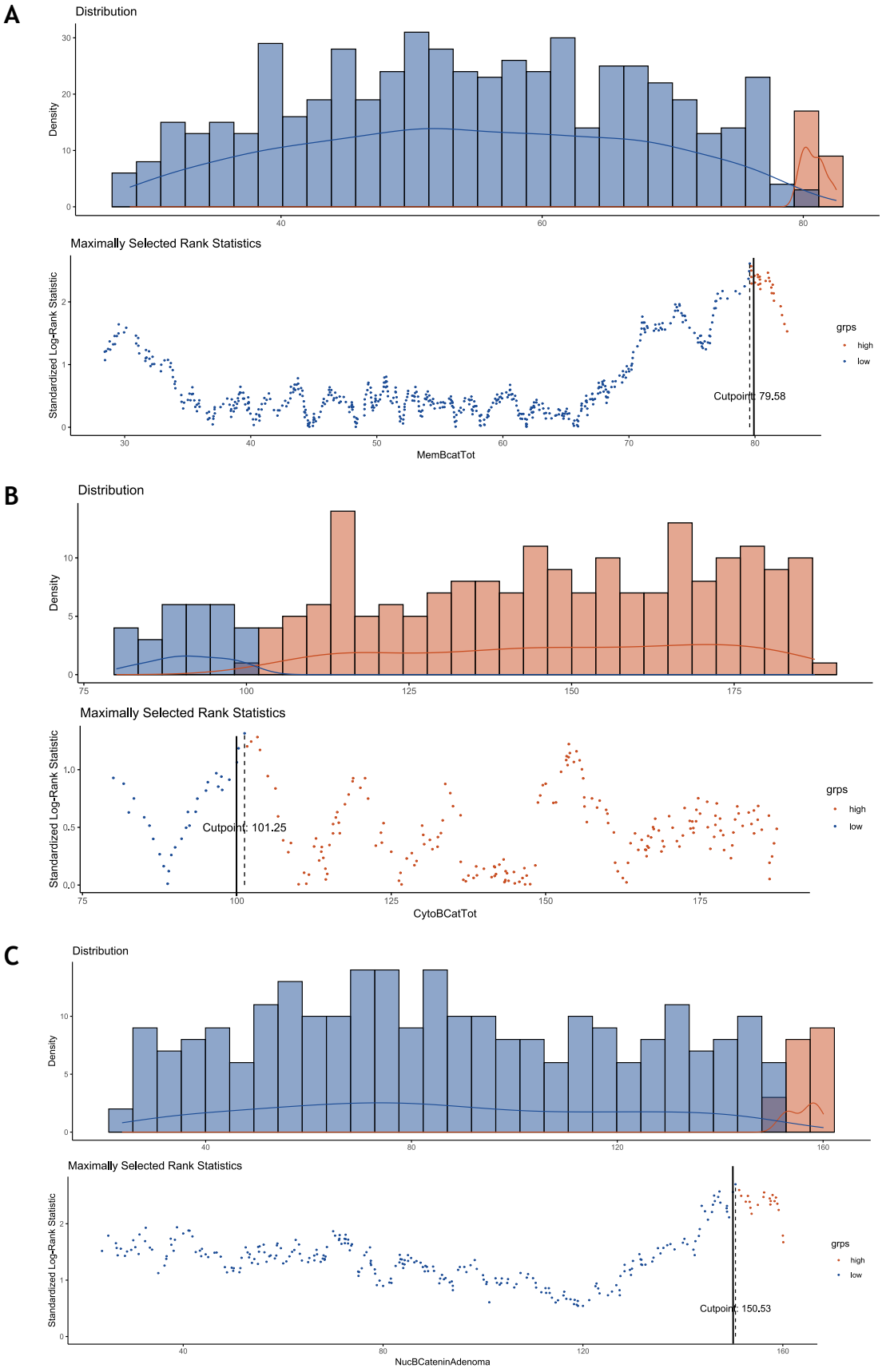


Figure 5.15. Thresholding of H-scores.
Threshold generation of **[A]** membranous, **[B]** cytoplasmic, **[C]** nuclear β-Catenin using maximally selected rank statistics on the training dataset. Distribution by density and standardized log rank maximally selected statistic by β-Catenin Histo-score distribution were both grouped in intervals of 20. The block lines indicate the chosen membranous threshold of 80, the chosen cytoplasmic threshold of 100, and the nuclear threshold of 150. N=767.

The distribution of scores considered as low expression and high expression of membranous ß-Catenin in the training and test dataset was comparable (Figure 5.16A, G and Figure 5.16B, H respectively). This was also true for both cytoplasmic (Figure 5.16C, I and Figure 5.16D ,J), and nuclear ß-Catenin expression (Figure 5.16E, K and Figure 5.16F, L).

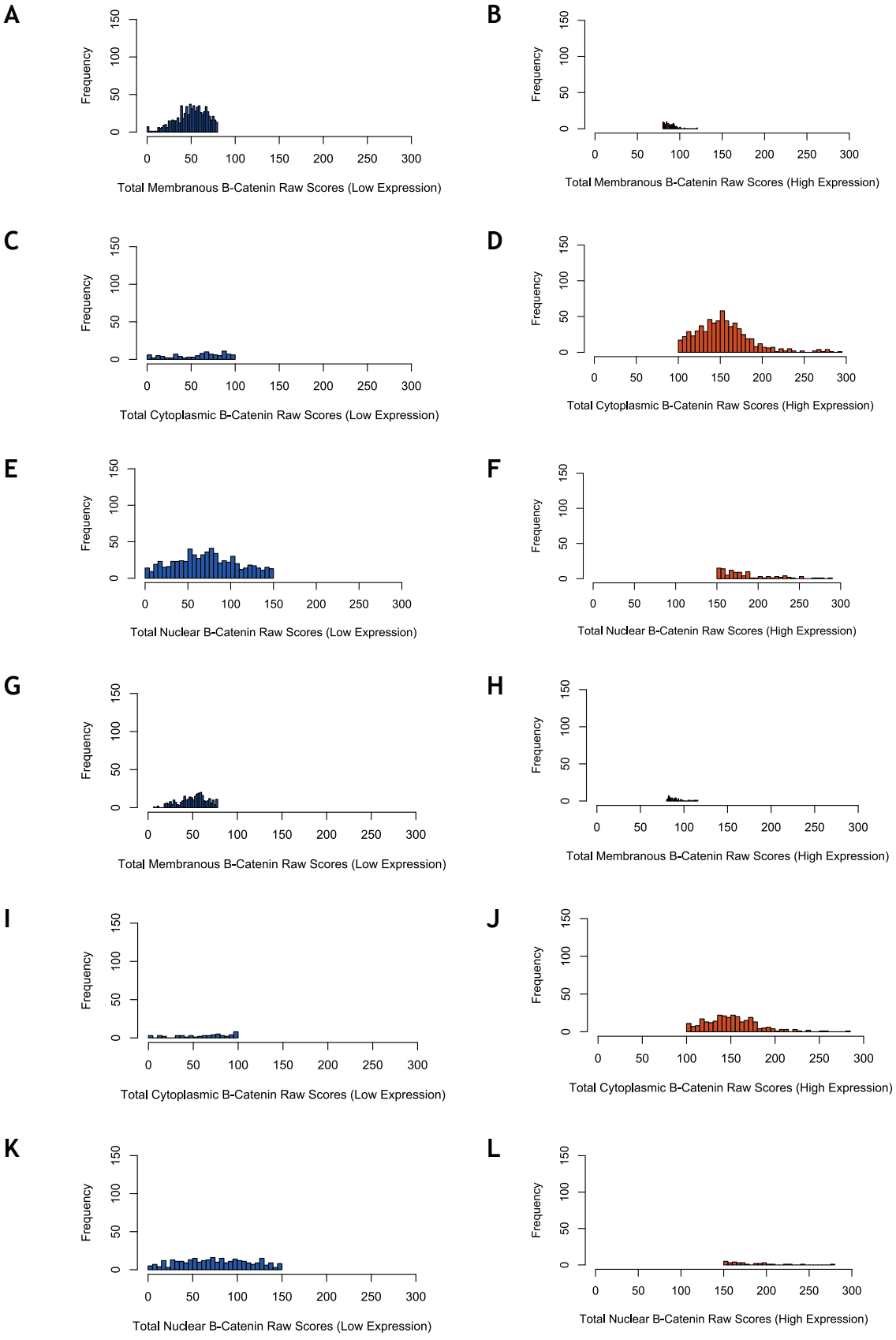


Figure 5.16. Histograms of Distribution of H-scores Across Low and High Expression Groups.
[A] Low and **[B]** high membranous β-Catenin expression in the training dataset. **[C]** Low and **[D]** high cytoplasmic β-Catenin expression in the training dataset. **[E]** Low and **[F]** high nuclear β-Catenin expression in the training dataset. **[G]** Low and **[H]** high membranous β-Catenin expression in the test dataset. **[I]** Low and **[J]** high cytoplasmic β-Catenin expression in the test dataset. **[K]** Low and **[L]** high nuclear β-Catenin expression in the test dataset.

5.2.3.4 Expression of β-Catenin and Metachronous Polyp or CRC Detection

Kaplan-Meier plots were constructed, and log rank survival analysis was carried out after the exclusion criteria detailed in 2.1.1.3.1 and 2.1.1.3.2 were implemented and after removing all invalid cases (n=51 for the training dataset and n=22 for the test dataset), leaving a total of 767 cases in the training dataset and 336 cases in the test dataset.

As shown in Figure 5.17, there was a significant association between high membranous ß-Catenin expression and time to detection of metachronous polyps or CRC in the training dataset (HR 1.409, 95% CI 1.077-1.842; p=0.015). The median time for metachronous polyp detection in patients with high membranous ß-Catenin was 42 months, while it was 61 months for those with low membranous ß-Catenin expression. This observation was not validated in the test dataset (HR 1.176, 95% CI 0.796-1.737; p=0.42), although the median time to polyp or CRC detection was comparable to that of the training dataset (high expression: 39 months, low expression: 57 months).

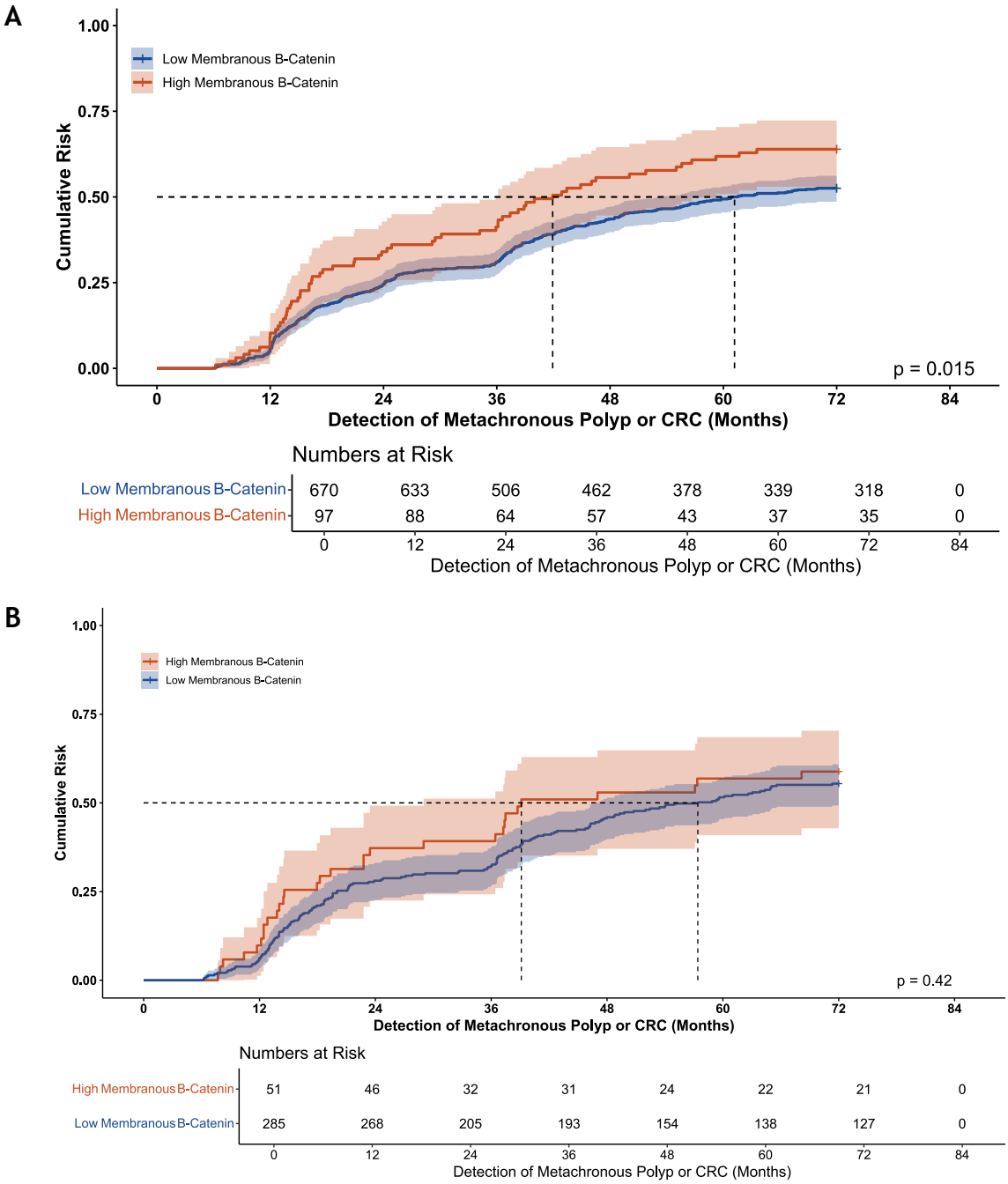


Figure 5.17. Membranous Expression of β-Catenin in Adenomas.
Expression of β-Catenin in the epithelia of adenomas in the **[A]** training and **[B]** test datasets. Shading is indicative of 95% CI. Dotted lines are time to median risk.

β-Catenin expression in the cytoplasm was not associated with the detection of metachronous polyps or CRC (Figure 5.18) in either the training (HR 0.865, 95% CI 0.664-1.127; p=0.283) or test datasets (HR 0.832, 95% CI 0.547-1.207; p=0.33). However, low expression of cytoplasmic β-Catenin had a shorter time to detection of metachronous polyp or CRC in the training dataset (46 months, vs 60 months for high expression) and in the test dataset (41 months, vs 57 months for high expression).

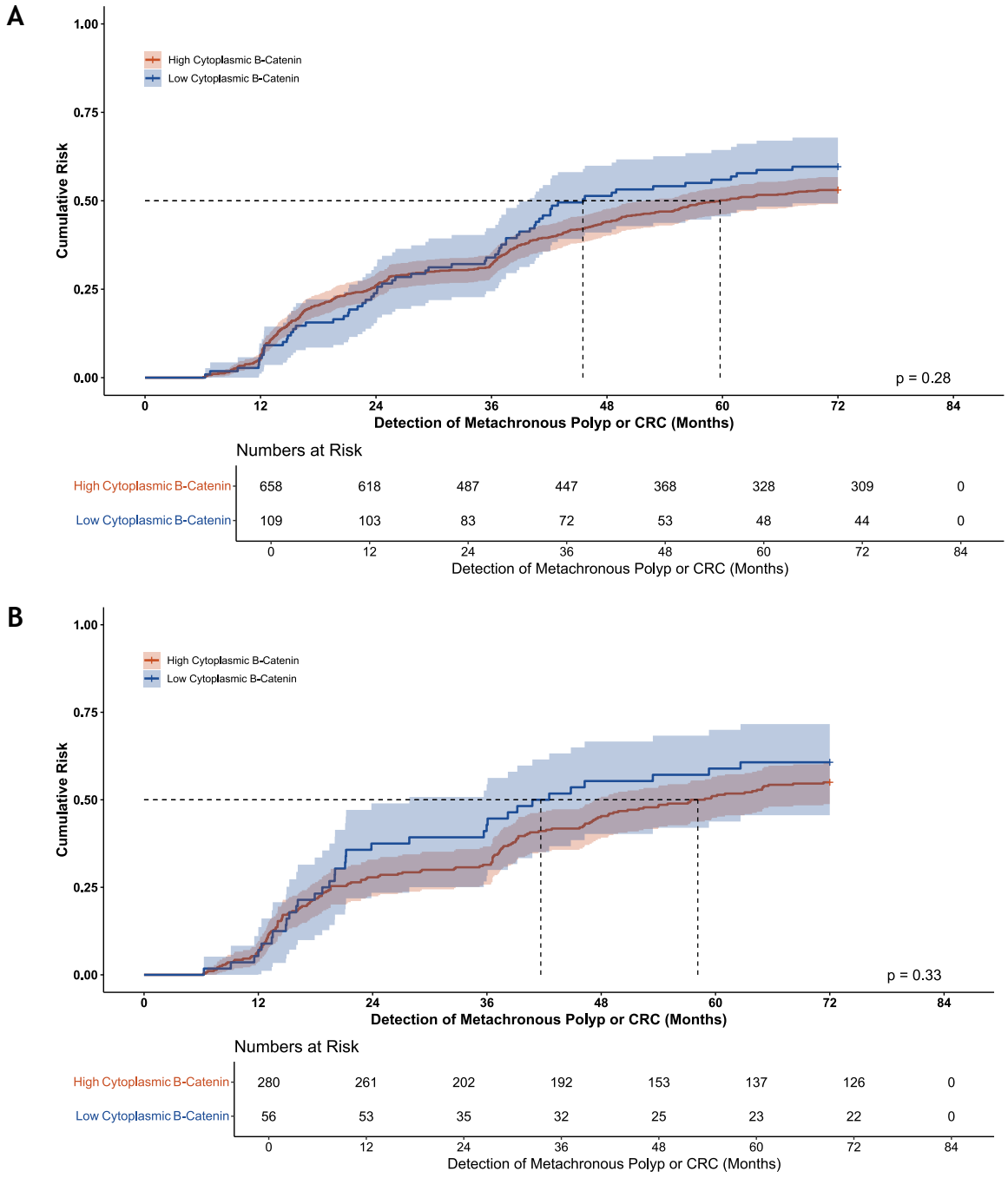


Figure 5.18. Cytoplasmic Expression of β-Catenin in Adenomas.
Expression of β-Catenin in the epithelia of adenomas in the **[A]** training and **[B]** test datasets. Shading is indicative of 95% CI. Dotted lines are time to median risk.

Finally, while low nuclear expression of ß-Catenin (Figure 5.19) in the training dataset was associated with detection of metachronous polyps or CRC (HR 0.671, 95% CI 0.488-0.923; p=0.014), this was not validated in the test dataset (HR 0.889, 95% CI 0.540-1.463; p=0.64). Patients with low expression of nuclear ß-Catenin had a median time to detection of metachronous polyps or CRC of 55 months while their high-expressing counterparts did not reach median risk, in the training dataset. In the test dataset, the median time to detection of

metachronous polyps or CRC in the patients with low expression of nuclear β-Catenin was comparable to that of the training dataset (53 months). In the test dataset, patients with high expression of nuclear β-Catenin had a median time to detection of metachronous polyps or CRC of 61 months.

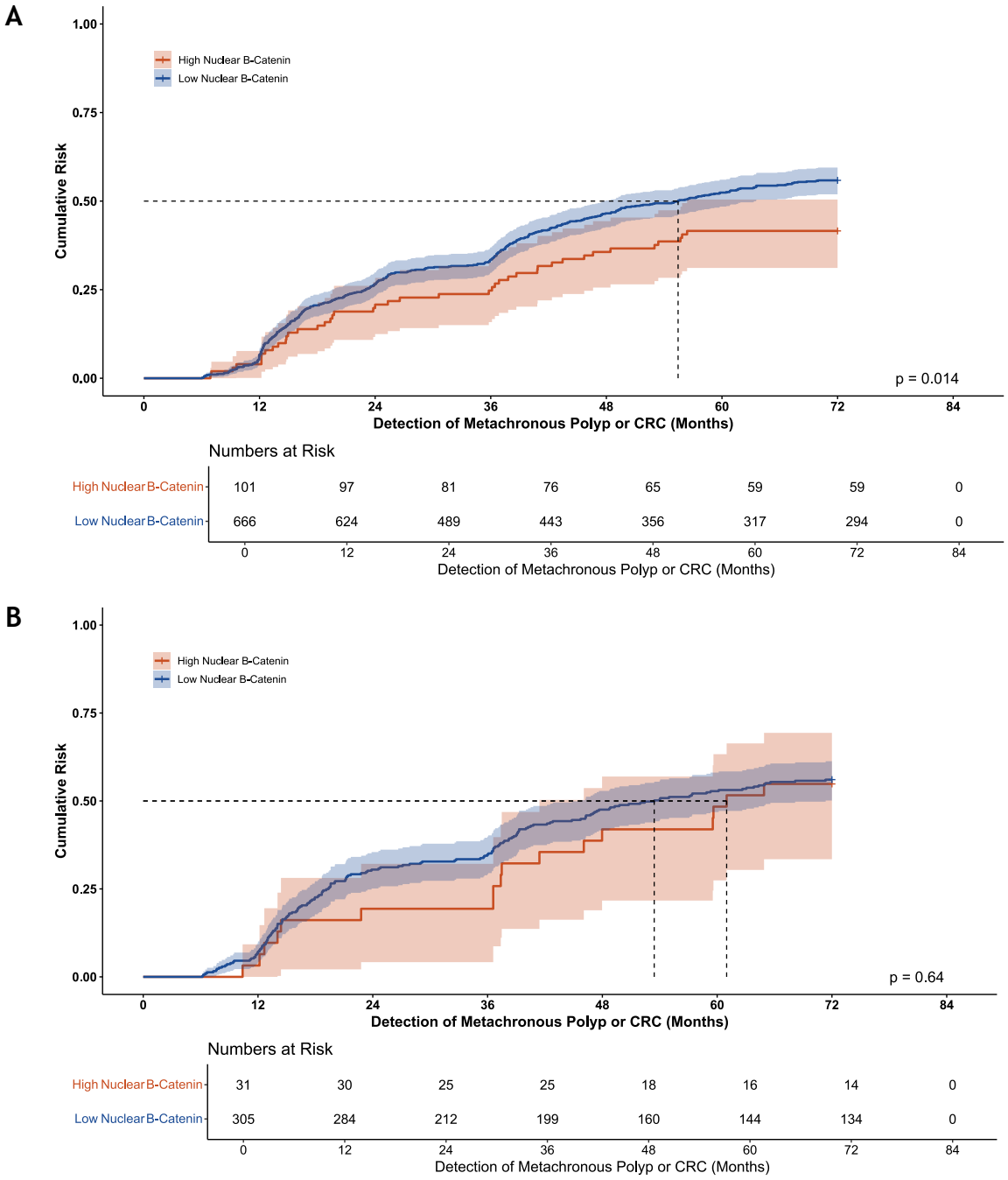


Figure 5.19. Nuclear Expression of β-Catenin in Adenomas.
Expression of β-Catenin in the epithelia of adenomas in the **[A]** training and **[B]** datasets. Shading is indicative of 95% CI. Dotted lines are time to median risk.

5.2.3.5 Expression of β-Catenin, Clinical Characteristics, and the Detection of Metachronous Polyps or CRC

Univariate Cox regression was performed to determine associations between ß-Catenin expression in the membrane, the cytoplasm, and the nucleus and relevant clinical factors. Table 5.2 shows that male sex ($p<0.0001$), left colonic polyps ($p=0.002$), multiple index polyps ($p<0.0001$), and high BSG2020 risk ($p<0.0001$) all significantly associated with detection of metachronous polyps or CRC. High membranous ß-Catenin expression ($p=0.012$) and high nuclear ß-Catenin expression ($p=0.014$) were also statistically associated with the same clinical outcome.

Multivariate Cox regression determined that male sex ($p=0.002$), left colonic polyps ($p=0.013$), multiple index polyps ($p=0.001$ for 2-4 polyps and $p<0.0001$ for 5+ polyps), high membranous ß-Catenin expression ($p=0.020$), and high nuclear ß-Catenin expression ($p=0.016$) were all independent of other factors in model 1. Male sex ($p=0.001$), left colon ($p=0.009$), high BSG2020 risk ($p<0.0001$), high membranous ß-Catenin expression ($p=0.028$), and high nuclear ß-Catenin expression ($p=0.032$) were independent predictors in model 2.

Table 5.2. β-Catenin Expression in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE Training Dataset

		Univariate			Multivariate					
					Model 1: BSG2020 Guidelines Components			Model 2: BSG2020 Guidelines		
		HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Sex	Female (n=222)	1.0	-	-	1.0	-	-	1.0	-	-
	Male (n=545)	1.540	1.228-1.930	**<0.0001**	1.444	1.149-1.813	**0.002**	1.463	1.165-1.837	**0.001**
Age	Screening Age (n=734)	1.0	-	-	-	-	-	-	-	-
	Above Screening Age (n=33)	0.863	0.515-1.445	0.575	-	-	-	-	-	-
Site	Right Colon (n=91)	1.0	-	-	1.0	-	-	1.0	-	-
	Left Colon (n=565)	0.640	0.484-0.847	**0.002**	0.699	0.528-0.926	**0.013**	0.689	0.520-0.912	**0.009**
	Rectum (n=111)	0.862	0.607-1.222	0.404	0.992	0.697-1.411	0.963	0.985	0.692-1.401	0.932
High Grade Dysplasia	Absent (n=640)	1.0	-	-	-	-	-	Omitted From Model		
	Present (n=127)	1.155	0.897-1.487	0.264	-	-	-			
Number of Polyps	1 (n=262)	1.0	-	-	1.0	-	-	Omitted From Model		
	2-4 (n=413)	1.513	1.210-1.891	**<0.0001**	1.444	1.154-1.808	**0.001**			
	5+ (n=92)	2.522	1.859-3.422	**<0.0001**	2.2442	1.791-3.328	**<0.0001**			
Polyp Type	Tubular (n=333)	1.0	-	-	-	-	-	-	-	-
	Tubulovillous (n=379)	0.980	0.803-1.196	0.841	-	-	-	-	-	-
	Villous (n=55)	0.979	0.658-1.457	0.918	-	-	-	-	-	-
BSG2020 Guidelines	Low Risk (n=264)	1.0	-	-	Omitted From Model			1.0	-	-
	High Risk (n=503)	1.659	1.338-2.056	**<0.0001**				1.583	1.276-1.965	**<0.001**
Membranous B-Catenin Expression in Adenoma	Low Expression (n=670)	1.0	-	-	1.0	-	-	1.0	-	-
	High Expression (n=97)	1.409	1.077-1.842	**0.012**	1.379	1.051-1.810	**0.020**	1.354	1.033-1.776	**0.028**
Cytoplasmic B-Catenin Expression in Adenoma	Low Expression (n=109)	1.0	-	-	1.0	-	-	1.0	-	-
	High Expression (n=658)	0.865	0.664-1.127	0.283	-	-	-	-	-	-
Nuclear B-Catenin Expression in Adenoma	Low Expression (n=666)	1.0	-	-	1.0	-	-	1.0	-	-
	High Expression (n=101)	0.671	0.488-0.923	**0.014**	0.673	0.488-0.929	**0.016**	0.703	0.510-0.969	**0.032**

Screening Age=50-74 years old. HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. N=767. Cox regression analysis was carried out to obtain hazard ratios.

While left colonic polyps (p=0.003), multiple index polyps (p=0.013 for 2-4 polyps, p<0.0001 for 5+ polyps), and high BSG2020 risk (p=0.002) all significantly

associated with the detection of metachronous polyps or CRC in the test dataset (Table 5.3), high β-Catenin expression in neither the membrane nor nucleus did not.

Table 5.3. β-Catenin Expression in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE Test Dataset

		Univariate		
		HR	95% CI	P
Sex	Female (n=89)	1.0	-	-
	Male (n=247)	1.072	0.778-1.478	0.670
Age	Screening Age (n=322)	1.0	-	-
	Above Screening Age (n=14)	1.453	0.811-2.603	0.209
Site	Right Colon (n=53)	1.0	-	-
	Left Colon (n=248)	0.596	0.424-0.837	**0.003**
	Rectum (n=35)	0.676	0.402-1.137	0.140
High Grade Dysplasia	Absent (n=284)	1.0	-	-
	Present (n=52)	1.114	0.778-1.597	0.555
Number of Polyps	1 (n=100)	1.0	-	-
	2-4 (n=180)	1.547	1.095-2.185	**0.013**
	5+ (n=56)	2.584	1.711-3.902	**<0.0001**
Polyp Type	Tubular (n=130)	1.0	-	-
	Tubulovillous (n=191)	0.846	0.637-1.123	0.248
	Villous (n=15)	1.741	0.929-3.263	0.083
BSG2020 Guidelines	Low Risk (n=103)	1.0	-	-
	High Risk (n=233)	1.687	1.218-2.338	**0.002**
Membranous B-Catenin Expression in Adenoma	Low Expression (n=285)	1.0	-	-
	High Expression (n=51)	1.176	0.796-1.737	0.415
Cytoplasmic B-Catenin Expression in Adenoma	Low Expression (n=56)	1.0	-	-
	High Expression (n=280)	0.832	0.574-1.207	0.333
Nuclear B-Catenin Expression in Adenoma	Low Expression (n=305)	1.0	-	-
	High Expression (n=31)	0.889	0.540-1.463	0.643

Screening Age=50-74 years old. HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. N=336. Cox regression analysis was carried out to obtain hazard ratios.

As the association between the expression of β-Catenin and the detection of metachronous polyps or CRC did not validate in any of the cellular compartments, no further investigation was carried out.

5.5 Effects of Wnt Signalling Stimulation on Colorectal Adenomas

It has been previously discussed in this thesis that Wnt signalling is an important pathway for cell proliferation in CRC and as Wnt signalling-related genes such as *APC* and *TCF7L2* are frequently mutated in INCISE adenomas (4.2). Although the

protein expression of E-Cadherin and ß-Catenin was not predictive of metachronous polyps or CRC, the basic function of the Wnt signalling pathway made it possible to hypothesise that it might influence the growth of adenomas. Thus, PDOs from FAP patients were used to investigate the effects of Wnt-signalling stimulation on adenoma growth.

Four Wnt signalling stimulants were selected for their abilities to either mimic Wnt molecules or inhibit negative regulation of Wnt signalling. These were CHIR-99201 (Ring *et al.*, 2003), BML-284 (Liu *et al.*, 2005), AZD2858 (Marsell *et al.*, 2012), and IM-12 (Schmöle *et al.*, 2010), and their mechanisms of action are presented in Figure 5.20.

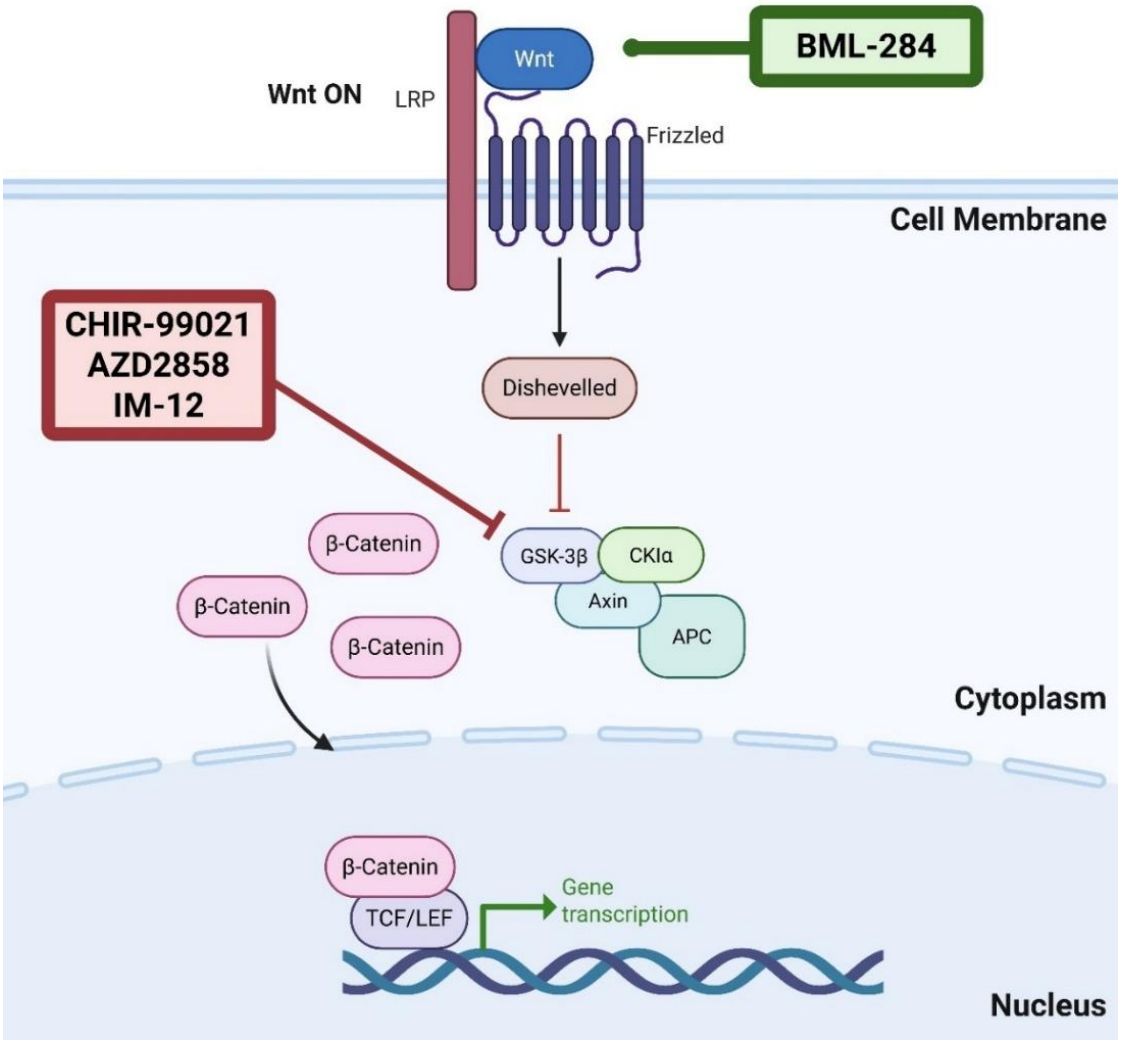


Figure 5.20. Mechanism of Action of Wnt Signalling Stimulants.
BML-284 is a Wnt molecule mimic and induces β-Catenin-dependent and TCF4-dependent gene transcription. CHIR-99021 is a GSK-3α/GSK-3β inhibitor that leads to Wnt activation and induces autophagy. AZD2858 is a GSK-3 inhibitor that activates Wnt signalling. IM-12 is a GSK-3β inhibitor that enhances canonical Wnt signalling. GSK3 is part of the β-Catenin destruction complex, and its inhibition disrupts the complex's function, allowing β-Catenin levels to increase.

5.5.1 Optimisation of Concentrations

To identify the most effective stimulant and its optimal concentration, all four stimulants were tested at a range of concentrations as described in 2.2.9.2.2 using three different CRC cell lines: HT29, HCT116, and DLD-1. Cell viability was used as a reflection of metabolic activity via Wnt stimulation 72 hours after exposure to the compounds. As shown in Figure 5.21A-B, neither HT29 nor HCT116 cell lines responded to stimulation as evidenced by the cell viability in treated cells decreasing and never surpassing the vehicle control. DLD-1 cells were more responsive to Wnt stimulation in terms of viability as determined by WST-1 assay (Figure 5.21C). Of the four tested stimulants, CHIR-99201 and AZD2858 were the most effective, although not statistically significant when compared to the vehicle control (CHIR-99201: 1μM p=0.2407, 3μM p=0.0944, 10μM=0.9656; AZD2858: 0.1μM p=0.9282, 0.5μM p=0.9988, 1μM p=0.7952, 3μM p=0.9863). The optimal concentrations for CHIR-99201-stimulated cell viability were 1μM and 3μM, with a decrease in viable cells at 10μM. AZD2858-stimulated cell viability was highest at 1μM followed by 0.1μM with 0.5μM resulting in the least stimulation of cell metabolic activity. Based on these observations, CHIR-99201 at 3μM and 1μM, and AZD2858 at 1μM and 0.1μM were taken forward for further investigation.

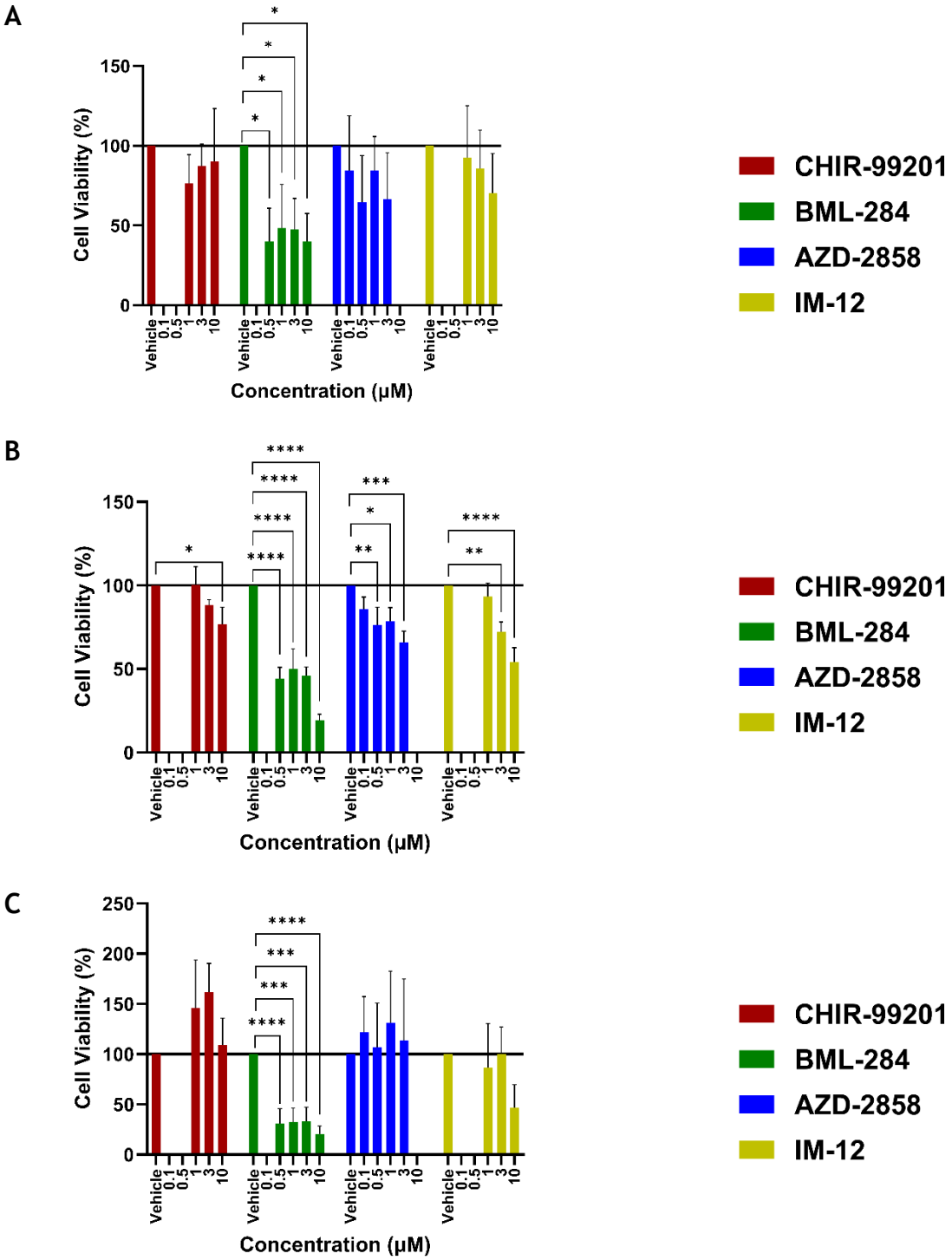


Figure 5.21. Optimisation of Wnt Signalling Stimulators on CRC Cell Lines Tested by Cell Viability.
[A] HT29 CRC cell lines stimulated by increasing concentrations of four different Wnt signalling stimulants. **[B]** HCT116 CRC cell lines stimulated by increasing concentrations of four different Wnt signalling stimulants. **[C]** DLD-1 CRC cell lines stimulated by increasing concentrations of four different Wnt signalling stimulants. Cell viability was measured 72 hours post-treatment and normalised to the vehicle control (1% DMSO). Three concentrations were tested for CHIR-99201 and IM-12. Four concentrations were tested for BML-284 and AZD-2858. All experiments were conducted in biological triplicates from 3 different passages and experimental triplicates. Error bars represent standard error of the mean. One-way ANOVA was used to test significance against the vehicle control.* $p<0.05$, ** $p<0.01$, ** $p<0.001$, and **** $p<0.0001$. Non-significant associations were not graphically shown.

To determine if the change in cell viability/metabolic activity by treating with CHIR-99201 and AZD2858 was via their reported effect to disrupt the function of the β-Catenin destruction complex by inhibiting GSK, the effects of these stimulants on the expression of selected downstream proteins of the Wnt signalling pathway were investigated.

Figure 5.22 displays quantification of the protein levels of the Wnt signalling target gene Myc, Wnt signalling effector and stemness marker and downstream target of Wnt signalling SOX9, and Wnt signalling pathway member, β-Catenin, in stimulated DLD-1 cell lysates. When stimulated by CHIR-99201, Myc levels increased when compared to the vehicle control, and of the two concentrations, 3μM led to a higher increase than 1μM (3μM p=0.519; 1μM p=0.8939), mirroring the cell viability results (Figure 5.21C). Stimulation by AZD2858 also led to overexpression of Myc, with both concentrations resulting in comparable levels of overexpression (Figure 5.22A; 1μM p=0.2476; 0.1μM p=0.1072).

SOX9 levels in stimulated DLD-1 cells increased when stimulated with both CHIR-99201 and AZD2858. Interestingly, it was the lower of the two concentrations for both CHIR-99201 (1μM p=0.9383) and AZD2858 (0.1μM, p=0.2283) that led to the higher overexpression of SOX9 (Figure 5.22B).

Finally, β-Catenin levels did not increase when stimulated with CHIR-99201 (3μM p=0.9997; 1μM p=0.9974). However, treatment with AZD2858 at a concentration of 1μM led to a decrease in β-Catenin expression levels when compared to the vehicle control (p=0.5844), while β-Catenin levels were observed to increase when stimulated with the lower concentration of AZD2858 (0.1μM p=0.999; Figure 5.22C). Although none of the observed changes were statistically significant, the consistent increase in target protein expression was observed when Wnt was stimulated.

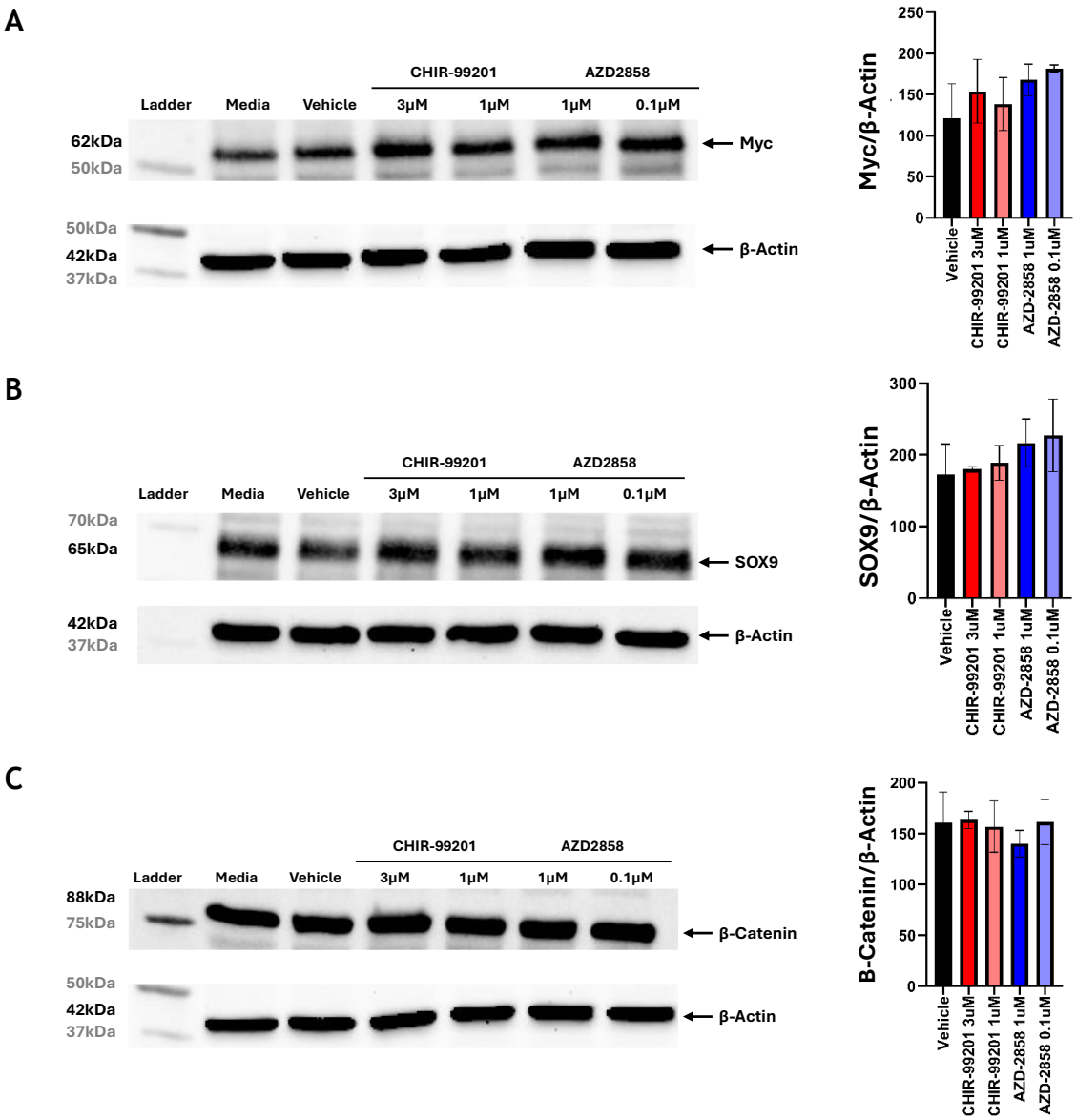


Figure 5.22. Protein Quantification of Wnt Signalling Proteins and Targets Downstream of Stimulatory Effects.
[A] Western blot (left) of Myc (Molecular weight (MW): 62kDa) protein expression in DLD-1 cells treated with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 with vehicle control. Relative expression (right) of Myc protein in cells treated with either stimulator at either condition compared to the vehicle. **[B]** Western blot (left) of SOX9 (MW: 65kDa) protein expression in DLD-1 cells treated with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 with vehicle control. Relative expression (right) of SOX9 protein in cells treated with either stimulator at either condition compared to the vehicle. **[C]** Western blot (left) of β-Catenin (MW: 88kDa) protein expression in DLD-1 cells treated with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 with vehicle control. Relative expression (right) of β-Catenin protein in cells treated with either stimulator at either condition compared to the vehicle. Cells were lysed 72 hours post-treatment. All experiments were performed as biological triplicates from 3 different passages. Vehicle control was 1% DMSO. β-Actin was used as a loading control. Molecular weights given in grey denote the ladder band, while molecular weights given in black denote the weight of the experimental bands. Error bars represent standard error of the mean. One-way ANOVA was used to test significance against the vehicle control. Non-significant associations were not graphically shown.

5.5.2 Stimulation of PDOs

The effects of CHIR-99201 and AZD2858 on Wnt signalling were next investigated in PDOs from FAP patients acquired and maintained as described in 2.2.8. Treatment with CHIR-99201 and AZD2858 was carried out as described in 2.2.9.4.2. Metabolic activity was used as an indicator of cell viability of PDOs and measured using CellTiter-Blue® 72 hours after treatment rather than WST-1 (as was the case for 2D cell lines) to reduce penetration issues and matrix interference.

Two FAP PDOs were used, PDO18235 and PDO19448. As shown in Figure 5.23, in both PDOS, stimulation with CHIR-99201 at 3μM or 1μM led to an increase in cell viability when compared to the vehicle control, with the two concentrations giving comparable results to each other when compared to the vehicle control although this was not statistically significant (PDO18235: 3μM p=0.3578, 1μM p=0.4827; PDO19448: 3μM p=0.9137, 1μM p=0.9979). When examined visually, the stimulated PDOs were larger and more numerous than the vehicle control.

When the PDOs were stimulated with AZD2858 at 1μM, cell viability decreased (PDO18235 p=0.9946; PDO19448 p=0.891) and on visual examination, the organoids were smaller in number and in size, when compared to the vehicle control. However, treatment with the lower concentration of AZD2858, 0.1μM, led to an increase in cell viability (PDO18235 p=0.4064; PDO19448 p=0.9999), organoid size and number, as examined visually when compared to the higher concentration (1μM). These effects were more pronounced in PDO18235 than PDO19448. When stimulated with 0.1μM of AZD2858, the cell viability of PDO18235 increased to a level comparable to that seen with CHIR-99201. This was reflected on visual examination, where the organoids stimulated with 0.1μM of AZD2858 were larger and more numerous.

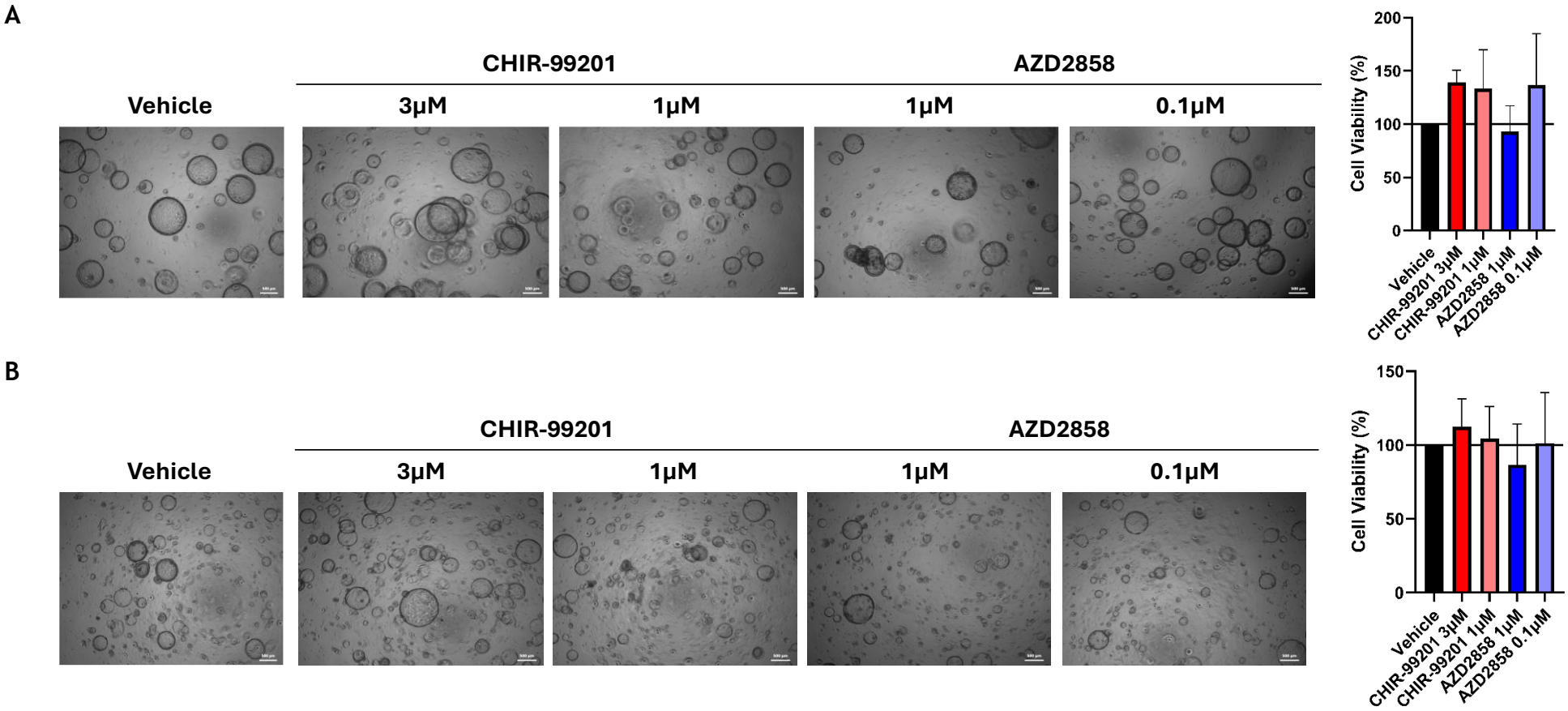


Figure 5.23. FAP Patient PDOs Treated with Wnt Signalling Stimulants.
[A] Brightfield microscopy images (left) of PDO18235 treated with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 with vehicle control. Cell viability (right) of PDO18235 after treatment with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 normalised to the vehicle control. **[B]** Brightfield microscopy images (left) of PDO19448 treated with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 with vehicle control. Cell viability (right) of PDO19448 after treatment with 3μM and 1μM of CHIR-99201 and 1μM and 0.1μM of AZD2858 normalised to the vehicle control. Vehicle control was 1% DMSO. All experiments were performed as biological triplicates from 3 different passages and technical triplicates. Images were taken and cell viability was measured 72 hours post-treatment. Scale bar = 500μm. Error bars represent standard error of mean. One-way ANOVA was used to test significance against the vehicle control. Non-significant associations were not graphically shown.

5.6 Discussion

Since current BSG2020 Guidelines depend on important yet insufficient polyp properties such as number, size, and dysplasia (Rutter *et al.*, 2020), this chapter examined whether Wnt-associated proteins were viable additions to these guidelines for the detection of metachronous polyps or CRC.

While associations between the expression of E-Cadherin or ß-Catenin and the detection of metachronous polyps or CRC were not validated in the test dataset, interesting patterns still emerged. E-Cadherin is a cell surface molecule crucial for cell adhesion, and a reduction in its expression has been associated not only with CRC, but especially more aggressive and invasive CRC behaviour (Kim *et al.*, 2016). While E-Cadherin expression was not significantly associated with the detection of metachronous polyps in the training or test datasets in the membrane or cytoplasm, Figure 5.10A and Figure 5.11 demonstrate that low expression of membranous and cytoplasmic E-Cadherin (respectively) could be linked to metachronous polyps, as is the case with CRC.

However, the findings presented here regarding ß-Catenin expression are variable. As shown in Figure 5.1, ß-Catenin is important for the transcription of downstream Wnt targets genes such as *MYC* which result in pro-cancer outcomes like cell-proliferation (Shah *et al.*, 2015). Indeed, studies on the levels of ß-Catenin within specific cellular compartments show that loss of membranous ß-Catenin and increased levels of nuclear ß-Catenin are associated with reduced CRC-specific survival (Gao *et al.*, 2014; Matly *et al.*, 2021). While this is contrary to the results presented in Figure 5.17A and Figure 5.19A, where high membranous and low nuclear ß-Catenin are both associated with shorter time to detection of metachronous polyps or CRC, it is important to note that not only were these findings not validated in the test dataset (Figure 5.17B and Figure 5.19B), but the numbers in the high expression groups were significantly smaller than their low expression counterparts.

Regardless, a dysregulation of Wnt signalling in colorectal adenomas, as previously reported in the literature (Schneikert and Behrens, 2007; Novellasdemunt, Antas and Li, 2015; H. Zhao *et al.*, 2022), is well-established in CRC. The mechanistic studies presented in this chapter demonstrate the role of

Wnt stimulation on the progression of colorectal adenomas, suggesting a role in supporting adenoma progression.

Three of the four chosen Wnt stimulants were inhibitors of glycogen synthase kinase 3 (GSK3), which phosphorylates ß-Catenin leading to its degradation (Figure 5.20). While previous research has shown that the use of GSK3-inhibitors such as CHIR-99201 on CRC cell lines (e.g. HT29 and HCT116), leads to Wnt stimulation (Jia *et al.*, 2023), this disagreed with the results presented here. This could be due to the mutational status of the cell lines. A characterisation study of 34 CRC cell lines by Berg et al. (2017), identified *APC* and *CTNNB1* (ß-Catenin) mutations in HT29 cells. HCT116 is known to have wild-type *APC* (Zhang *et al.*, 2016), and mutated ß-Catenin (Y. Li *et al.*, 2020). However, Berg et al. (2017) reported both cell lines as having copy number gain alterations in *MYC*, suggesting that such Wnt target genes are already overexpressed and somewhat saturated, implying that further Wnt activation would have little effect, or even in cytotoxicity. The same study identified DLD-1 as having alteration in *APC* but not in ß-Catenin or in *MYC*, suggesting that in DLD-1 cells, Wnt signalling is less activated, and stimulation is more likely to yield pronounced effects.

Another possible reason for the results observed, is the differentiation/proliferation paradigm. It is commonly accepted that differentiation leads to decreased proliferation and that proliferating cells do not prioritise differentiation (Ruijtenberg and van den Heuvel, 2016). The "colon-like" status of HT29 (i.e., leaning towards differentiation) and DLD-1 being considered an undifferentiated cell line (Berg *et al.*, 2017), could further explain the difference in growth observed between DLD-1 and HT29 cells.

Levels of some proteins involved in Wnt signalling were measured after treatment with the two best performing GSK3 inhibitors. While more downstream targets such as Myc and SOX9 had increased levels with treatment, the levels of ß-Catenin did not noticeably change. It is possible that inhibition of the ß-Catenin destruction complex protein GSK3 by CHIR-99201 and AZD2858 could lead to increased ß-Catenin levels, and the efficacy of these two compounds is supported by overexpression of Myc and SOX9.

However, the ß-Catenin results shown in this chapter disagree with this, and experimental limitations must be considered. These experiments utilised whole cell lysates; however, it is nuclear, activated ß-Catenin that confers the effects of Wnt signalling pathway activation (Figure 5.20), and while the total levels of ß-Catenin might seem unchanged, ß-Catenin might have been translocated to the nucleus. Previous research suggests that GSK3 inhibition both increases ß-Catenin retention in the nucleus, and directs ß-Catenin transport in a one-way, nucleus-bound direction (Jamieson, Sharma and Henderson, 2011). This could explain why downstream targets were overexpressed, and it would be of interest to measure the levels of nuclear ß-Catenin with and without stimulation, rather than depending on the total protein levels, by subcellular fractionation before western blotting. Finally, BML-284 (Wnt agonist) consistently decreased proliferation. In addition to the previously discussed possible reasons, it is possible that the cells were exposed to BML-284 (and the other compounds) for too long, leading to cytotoxicity (Liu *et al.*, 2005; Lim *et al.*, 2008).

There is a lack of 2D cell lines established from pre-malignant adenomas, with the only commercially available options being PDOs embedded in 3D cultures. However, PDOs present a more advanced and biologically relevant model of disease than 2D cell lines. They are more recapitulative of tissue development and better reflect the physiological complexity of *in vivo* disease (Kim, Koo and Knoblich, 2020). FAP PDOs were used to study the effects of Wnt stimulation on cells not of a cancer origin. Wnt stimulation led not only to an increase in cell viability but also to an increase in the number and size of the organoids. However, there are some differences between the results observed in 2D cell lines and organoids. For example, the higher concentration of AZD2858 (1µM) seemed to have a lesser effect in organoids. This is likely due to differences in culture conditions between 2D and 3D models that may affect drug-induced effects on proliferation and metabolism (Kapałczyńska *et al.*, 2016). Alternatively, the effects could be due to the involvement of GSK3 in other pathways such as the NFκB pathway, the inhibition of which could lead to reduced proliferation and induction of apoptosis (Brüning-richardson *et al.*, 2021) However, these result support the hypothesis that Wnt signalling plays a crucial role in the development and progression of adenomas. It is also important to consider that a more marked effect was not seen on the FAP PDOs

because, by their nature, they have an *APC* mutation that leads to a constitutive activation of Wnt signalling (Samadder *et al.*, 2019). This could mean that any further stimulation of the pathway will not result in a significant increase. Thus, it would be beneficial to repeat this experiment on PDOs from non-FAP adenomas.

Although these results did not paint a portrait of active involvement of Wnt in colorectal adenoma development, another Wnt-associated target was flagged in mutational analysis in 4.2. The stemness marker and Wnt-regulator *SOX9* was co-mutating with *APC* and *KRAS* in high-risk patients. This interesting observation led to SOX9 becoming the next target for investigation as a possible predictor of metachronous polyp or CRC.

Chapter 6 Evaluation of Wnt-associated Protein SOX9 on Colorectal Adenomas

6.1 Introduction

As elaborated in 5.1, Wnt signalling is an important and well-studied pathway in CRC development. It heavily depends on the levels of ß-Catenin which are mainly regulated by the ß-Catenin destruction complex (Figure 6.1). When the ß-Catenin destruction complex is available, the levels of ß-Catenin are controlled. However, where mutations are present e.g. in the *APC* gene, the destruction complex becomes nonfunctional, allowing ß-Catenin to accumulate in the cytosol and eventually translocate to the nucleus (Zhang *et al.*, 2024). In such cases, redundant regulatory functions become crucial. One such case is that of SRY-box Transcription Factor 9 (SOX9).

SOX9 was discovered in the 1990s (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990) and was identified as the cause of the skeletal malformation condition campomelic dysplasia as well as sex-reversal in humans (Wagner *et al.*, 1994). Interestingly, later work by Blache et al. (2004) identified SOX9 as a downstream target of Wnt signalling and as a possible Wnt regulator through a feedback mechanism. Mechanistic work has shown a physical interaction between ß-Catenin and SOX9 in the nucleus (Prévostel *et al.*, 2016). While there is contention as to whether SOX9 regulates Wnt/ß-Catenin signalling through its transcriptional activity of Wnt repressors (Sinha *et al.*, 2021), or independently of transcription through the translocation of nuclear ß-Catenin back to the cytosol (Prévostel *et al.*, 2016), its involvement in regulating ß-Catenin-mediated gene transcription is clear.

This, combined with its co-mutation trend with both *APC* and *KRAS* as observed in Chapter 4, led to it being selected for further investigation in relation to the detection of metachronous polyps or CRC. This chapter will explore the expression of the SOX9 protein in INCISE polyp tissue, and if associations with the detection of metachronous polyps or CRC are revealed, the SOX9 expression will be tested for correlation to other clinical factors. Furthermore, SOX9 will be added to the current BSG2020 risk stratification guidelines to understand whether they can be improved through such methods. Finally, the transcriptomic landscape of patients with varying degrees of SOX9 expression will be explored.

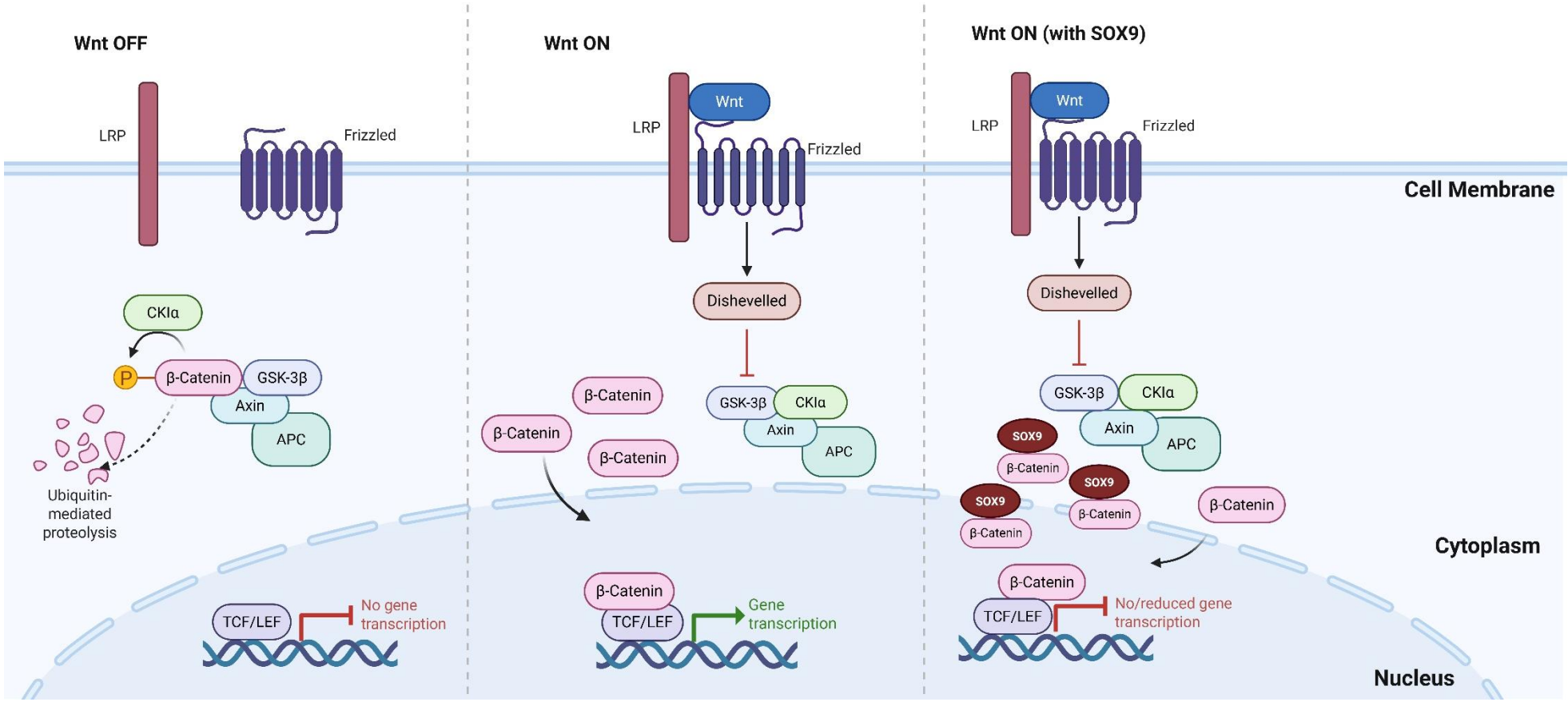


Figure 6.1. Schematic Representation of the Effect of SOX9 on Wnt Signalling.
[OFF] When there is no Wnt molecule present, cell surface LRP and Frizzled remain unbound. The β-Catenin destruction complex (APC, Axin, GSK-3β, CD1α) is active in the cytosol, ubiquitinating β-Catenin, leading to its degradation, ultimately preventing β-Catenin-mediated gene transcription. **[ON]** Upon Wnt binding to LRP and Frizzled, the β-Catenin destruction complex translocates to the membrane, where phosphorylation of Axin binds it to the internal tail of LRP, decreasing its stability. Dishevelled is phosphorylated and activated, which in turn inhibits GSK. This allows β-Catenin to accumulate in the cytosol. It then localises to the nucleus, where it binds to TCF/LEF transcription factor, and begins the transcription of target genes involved in cell proliferation. **[ON with SOX9]** The presence of SOX9 can regulate β-Catenin-mediated transcription, for instance, by physically binding nuclear β-Catenin returning it to the cytosol. Figure made with BioRender.

6.2 Protein Expression of SOX9 and the Detection of Metachronous Polyps or CRC

6.2.1 Antibody Specificity and Sensitivity

The specificity and sensitivity of the antibody used for tissue staining was confirmed using the methods described in 2.2.1. DepMap was used to identify cell lines expressing SOX9 at varying levels (Figure 6.2).

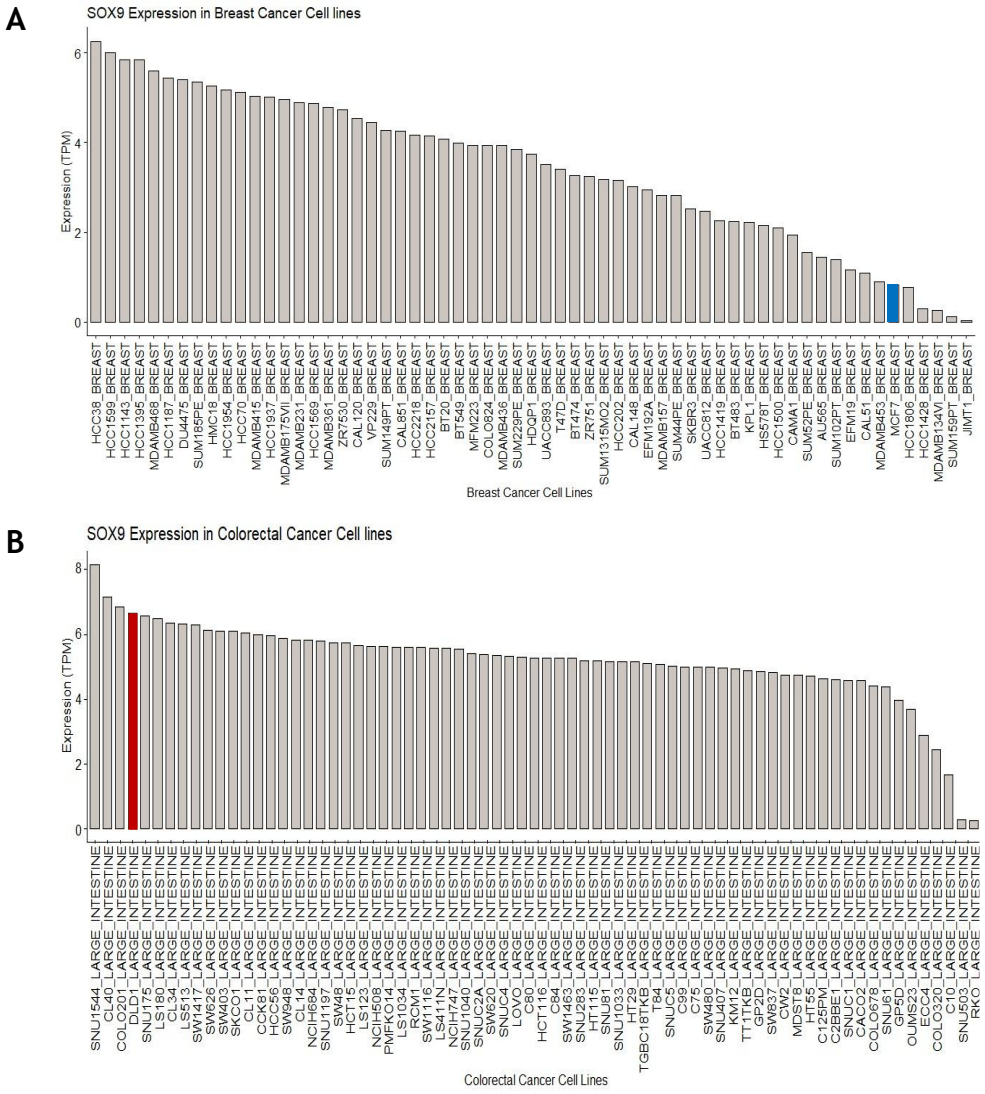


Figure 6.2. DepMap Cell Line Selector for Cell Lines Expressing SOX9.
[A] SOX9 expression in breast cancer cell lines. A low-expressing cell line, MCF-7 (blue), was selected for antibody validation. **[B]** SOX9 expression in colorectal cancer cell lines. A high-expressing cell line, DLD-1 (red), was selected for antibody validation.

Lysates from the selected cell lines were generated as described in 2.2.1.3. Western blotting was carried out as described in 2.2.1.4. As shown in Figure 6.3, SOX9, was expressed more weakly in MCF-7 cells than in DLD-1 cells, validating the specificity of the antibody to its target.

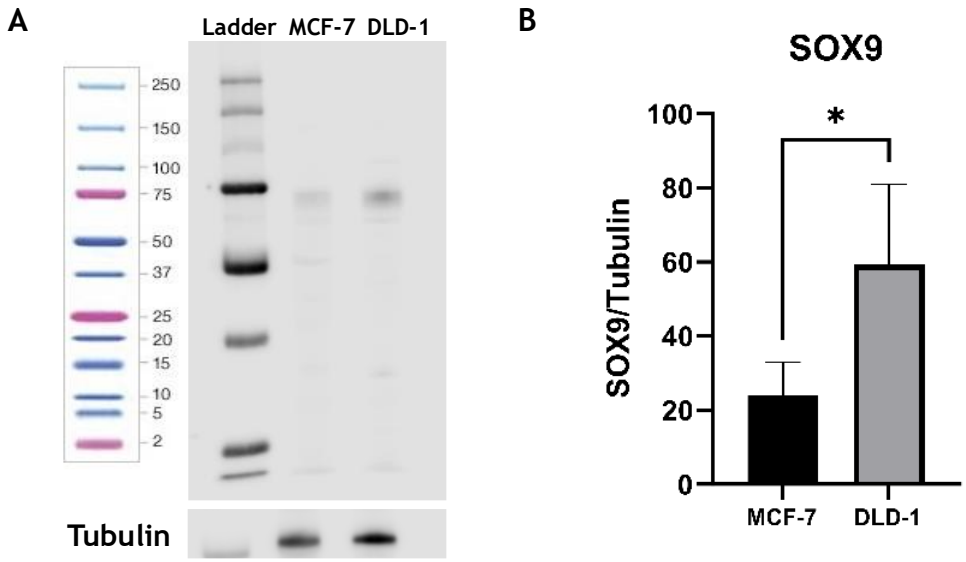


Figure 6.3. Antibody Specificity Test.
Western blotting was used to validate the specificity of the selected antibodies using lysates from cell lines identified using DepMap. **[A]** Western blot of SOX9 expression in lysates from MCF-7 (low expression) and DLD-1 (high expression) at 56-65kDa. Tubulin was used as a loading control. **[B]** Relative expression of SOX9 against the loading control. N=3. Error bars represent the standard deviation. ns denotes $p>0.05$, * denotes $p<0.05$ on a paired t-test.

To further test the specificity of the antibodies, cell pellets were stained as described in 2.2.1.5. The difference in expression between identified low- and high-expressing cell lines is presented in Figure 6.4.

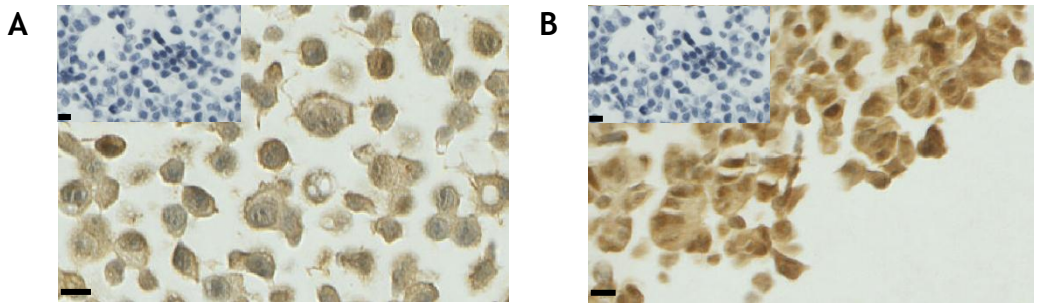


Figure 6.4. Antibody Specificity Tests.
IHC on cell pellets of cell lines identified through DepMap. **[A]** MCF-7 cell pellets (low expression) stained for SOX9. **[B]** DLD-1 cell pellets (high expression) stained for SOX9. No-antibody controls were included and are shown at the top left corner of each image. All staining was carried out under the conditions described in Table 2.5. Black cell bars = 10µm.

6.2.2 SOX9 Protein Expression Is Associated with the Detection of Metachronous Polyps or CRC

6.2.2.1 IHC and Staining Assessment of SOX9

Staining for SOX9 was carried out as described in 2.2.2.4. Examples of tissue staining are shown in Figure 6.5A. SOX9 was expressed in the cytoplasm and nucleus of polyp epithelial cells. Scoring of the staining was carried out digitally using QuPath as described in 2.2.3.2. As shown in Figure 6.5B, the QuPath classifier was able to accurately classify tissue as epithelium (shades of red) or as lamina propria (shades of green), as well as classifying pixels into appropriate intensity thresholds. The negative isotype-matched control ensured that any staining seen was a true signal and not due to non-specific interactions within the tissue. Similarly, the inclusion of a negative no-antibody control ensured that the signal observed was not due to the interaction between the tissue and any of the reagents used in the staining process. Finally, a positive tissue control further confirmed the antibody can detect its true target (Figure 6.5C).

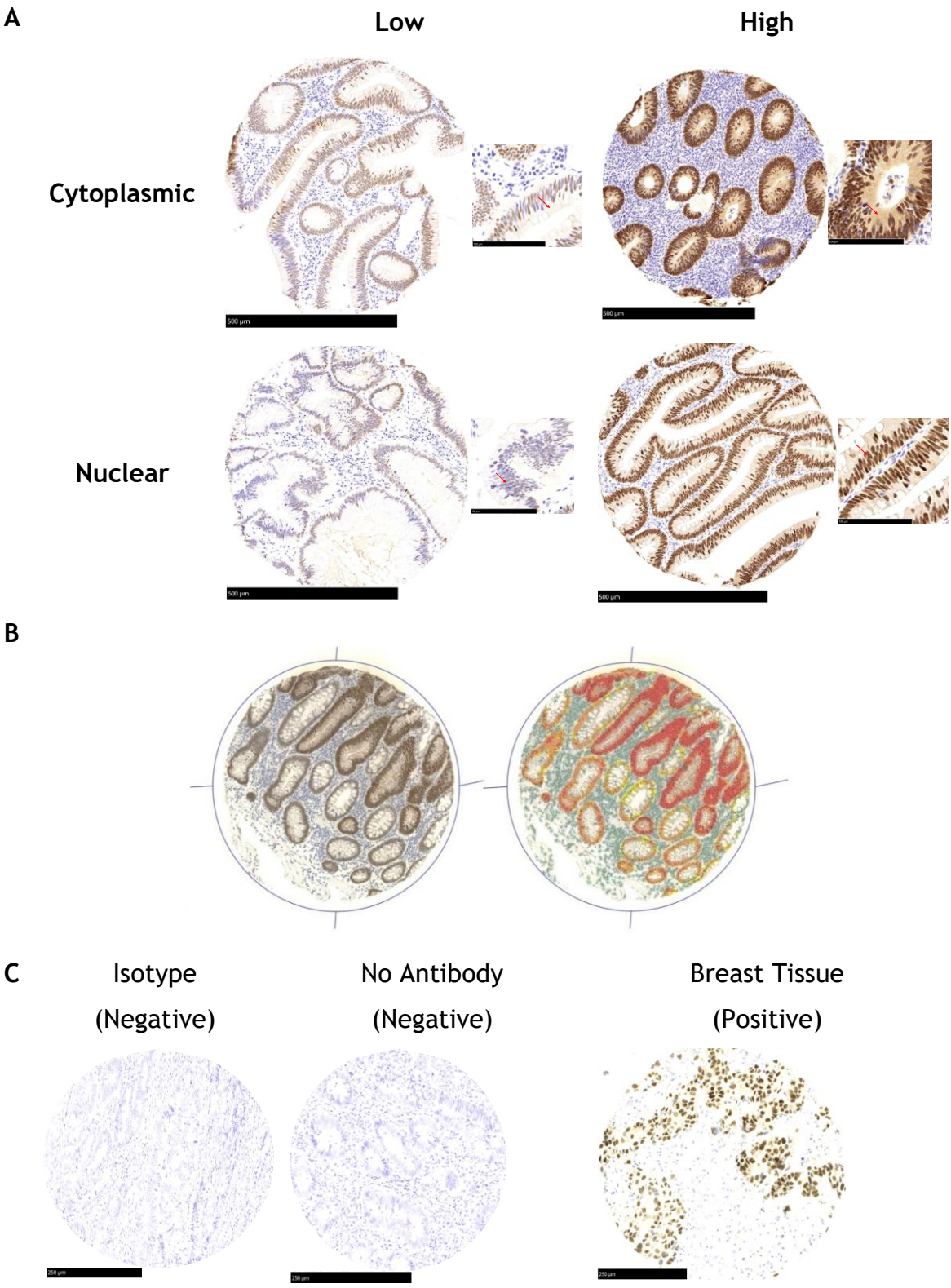


Figure 6.5. IHC and Staining Assessment of SOX9.
[A] Examples of low and high expression of cytoplasmic and nuclear SOX9 in polyp tissue, with zoomed-in images (right) detailing the cytoplasmic and nuclear staining patterns clarified with red arrows. Scale bars of full cores = 500μm, scale bars of zoomed-in images = 100μm.
[B] Example of QuPath classifier used to assess the staining of SOX9. Dearrayed, unclassified core (left), and dearrayed, classified same core (right). Red denoted strongly stained epithelial cells, orange denoted moderately stained epithelial cells, and yellow denoted weakly stained epithelial cells, while green denoted cells in the lamina propria. **[C]** Three controls were included in the staining. The negative isotype and negative no-antibody controls used surplus CRC tissue, while the positive control used surplus breast cancer tissue. Scale bars = 250μm.

6.2.2.2 Quality Control of Raw Data

As shown in Figure 6.6, there was a positive correlation between luminal and total H-scores, as well as between basal and total H-scores, allowing for the use of the total H-scores in place of those of each histological region. Additionally, the test dataset mirrors the patterns observed in the training dataset.

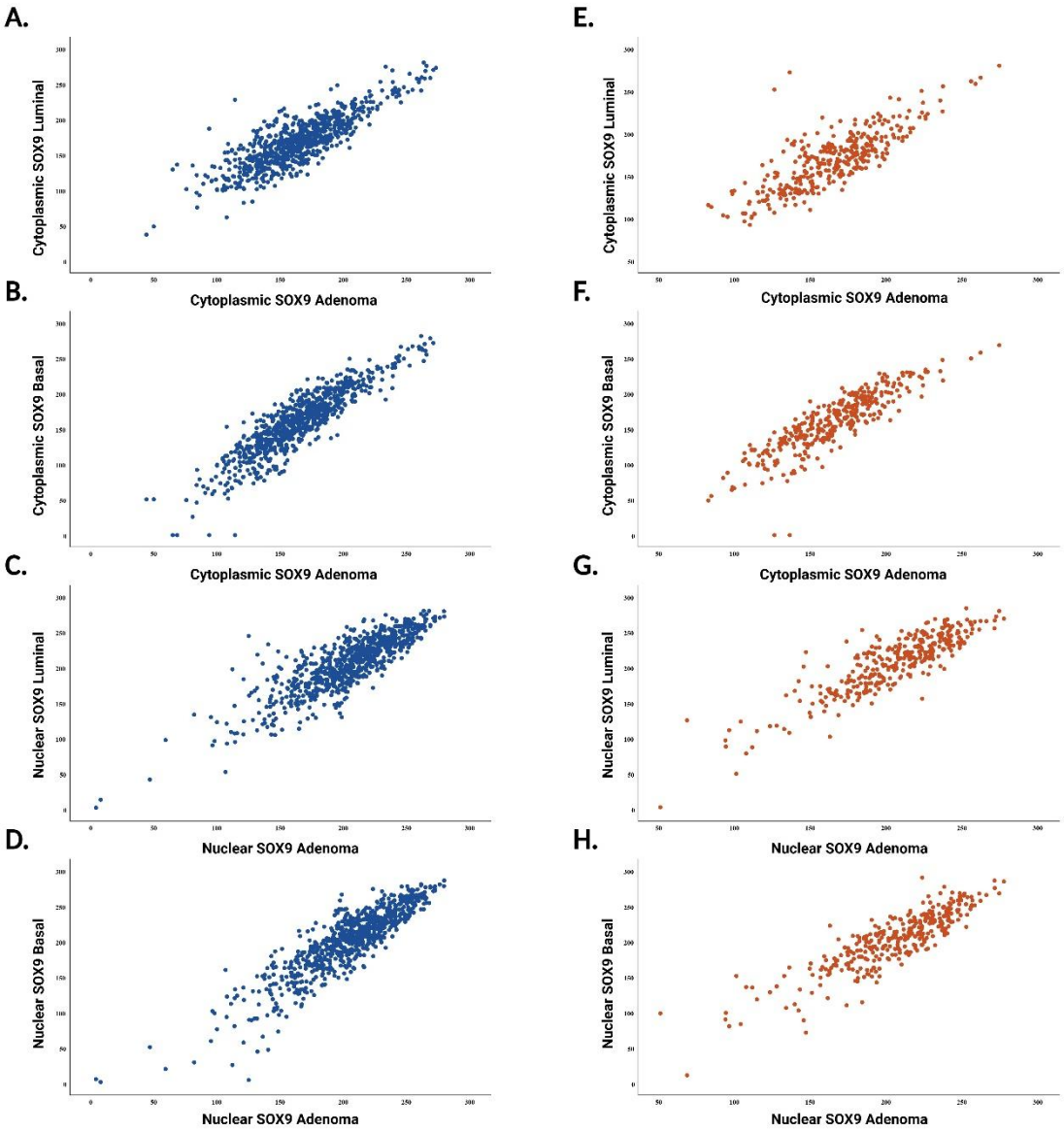


Figure 6.6. Correlation of Polyp Histological Regions Against Total Polyp Adenomas in the INCISE Training and Test Datasets.
Scatter plot of nuclear SOX9 expression in the luminal epithelium of the polyp vs cytoplasmic SOX9 expression in the total adenoma of the polyp in the **[A]** training and **[E]** test datasets. Scatter plot of cytoplasmic SOX9 expression in the basal epithelium of the polyp vs cytoplasmic SOX9 expression in the total adenoma of the polyp in the **[B]** training and **[F]** test datasets. Scatter plot of nuclear SOX9 expression in the luminal epithelium of the polyp vs cytoplasmic SOX9 expression in the total adenoma of the polyp in the **[C]** training and **[G]** test datasets. Scatter plot of nuclear SOX9 expression in the basal epithelium of the polyp vs nuclear SOX9 expression in the total adenoma of the polyp in the **[D]** training and **[H]** test datasets. Both the x-axis and y-axis range from 0 to 300 in increments of 50, to align with the range of scores generated by the weighted H-Score method.

Histograms revealed that while cytoplasmic SOX9 H-scores appeared to be normally distributed in the whole cohort (Figure 6.7A), nuclear SOX9 H-scores were slightly skewed to the right (Figure 6.7B). This is mirrored in the training and test datasets for both cytoplasmic and nuclear SOX9 (Figure 6.7C,E and Figure 6.7D,F respectively).

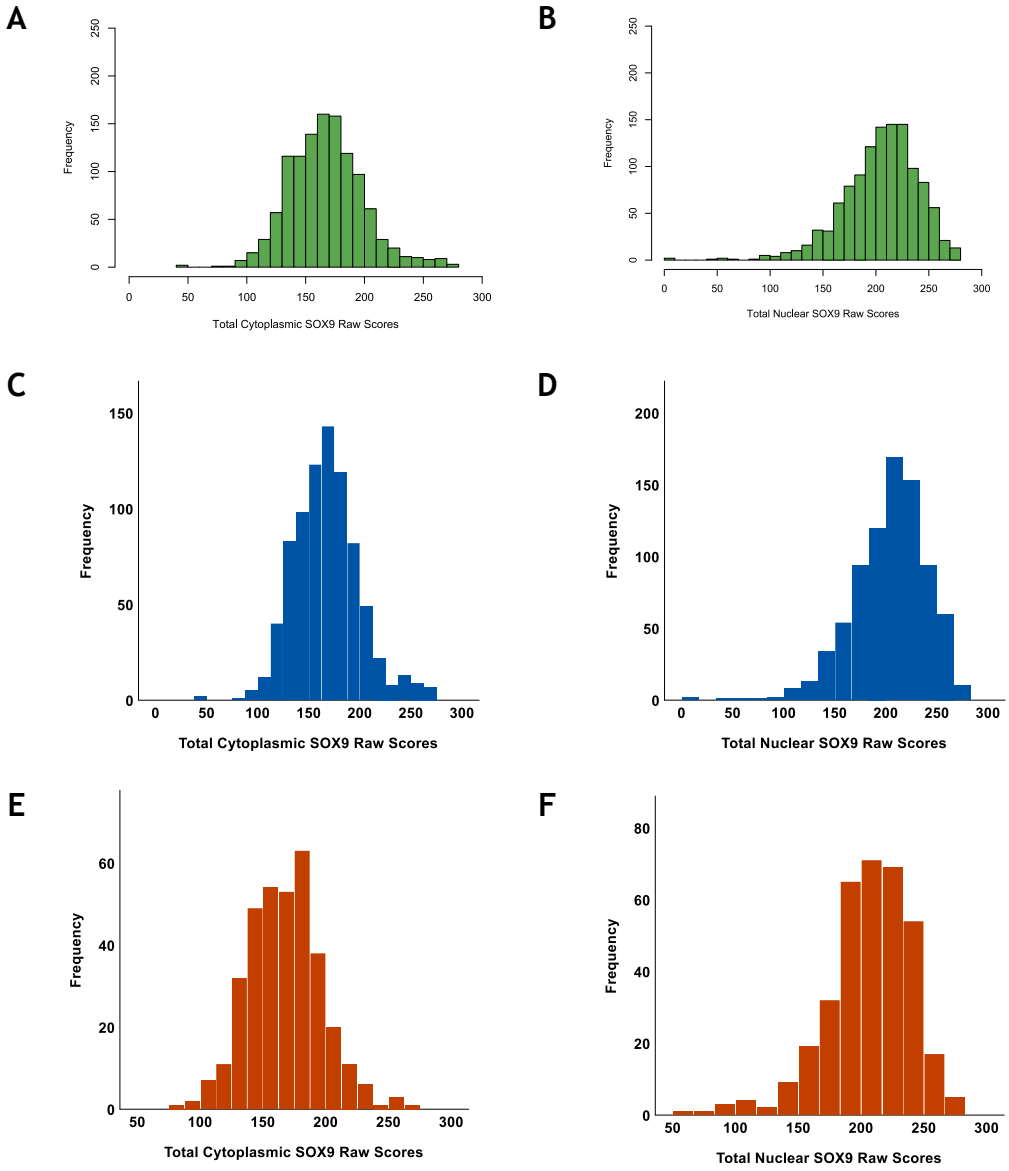


Figure 6.7. Quality Control of Raw SOX9 H-scores.
Histograms of the distribution of **[A]** cytoplasmic and **[B]** nuclear H-scores of SOX9, in the whole cohort. Histograms of the distribution of **[C]** cytoplasmic and **[D]** nuclear H-scores of SOX9 in the training dataset. Histograms of the distribution of **[E]** cytoplasmic and **[F]** nuclear H-scores of SOX9 in the test dataset.

6.2.2.3 Dichotomisation of SOX9 H-scores

Thresholds were generated as in 2.2.10. As for E-Cadherin and ß-Catenin, comparable thresholds to those generated were chosen (215 for nuclear expression; 200 for cytoplasmic expression) that could be determined by a human observer (Figure 6.8).

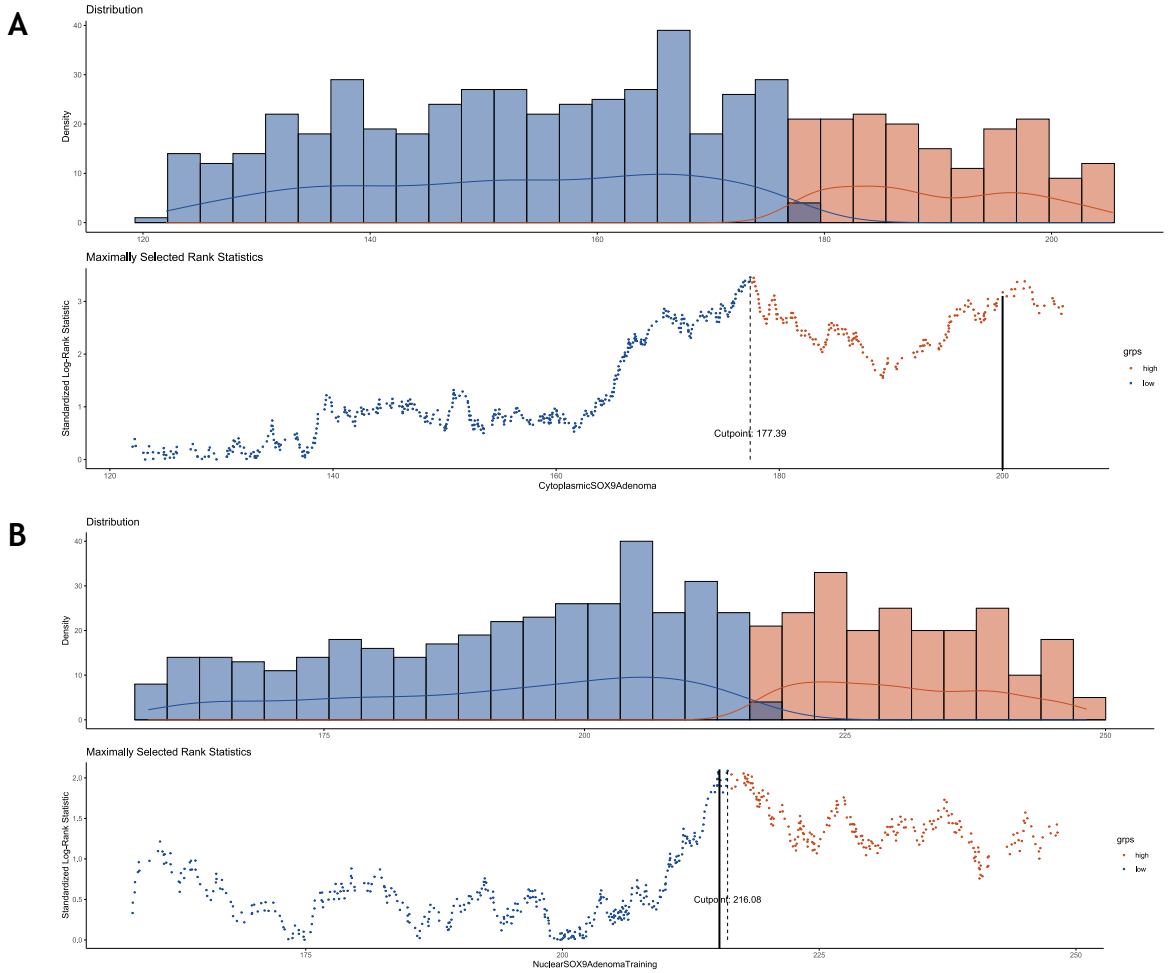


Figure 6.8. Thresholding of H-scores.
Threshold generation of **[A]** cytoplasmic and **[B]** nuclear SOX9 using maximally selected rank statistics on the training dataset. Distribution by density and standardized log rank maximally selected statistic by SOX9 Histo-score distribution were both grouped in intervals of 20. The solid black lines indicate the chosen nuclear threshold of 215 and the cytoplasmic threshold of 200. N=768.

Figure 6.9 shows the score distribution between low and high SOX9 expression in total adenoma epithelium. Low cytoplasmic SOX9 skewed slightly to the right in the training and test datasets (Figure 6.9A, E), while low nuclear SOX9 had a more pronounced skew towards higher scores (Figure 6.9C, G). High cytoplasmic and high nuclear SOX9 both sharply skewed to the right in both training (Figure 6.9B,D) and test datasets (Figure 6.9F, H).

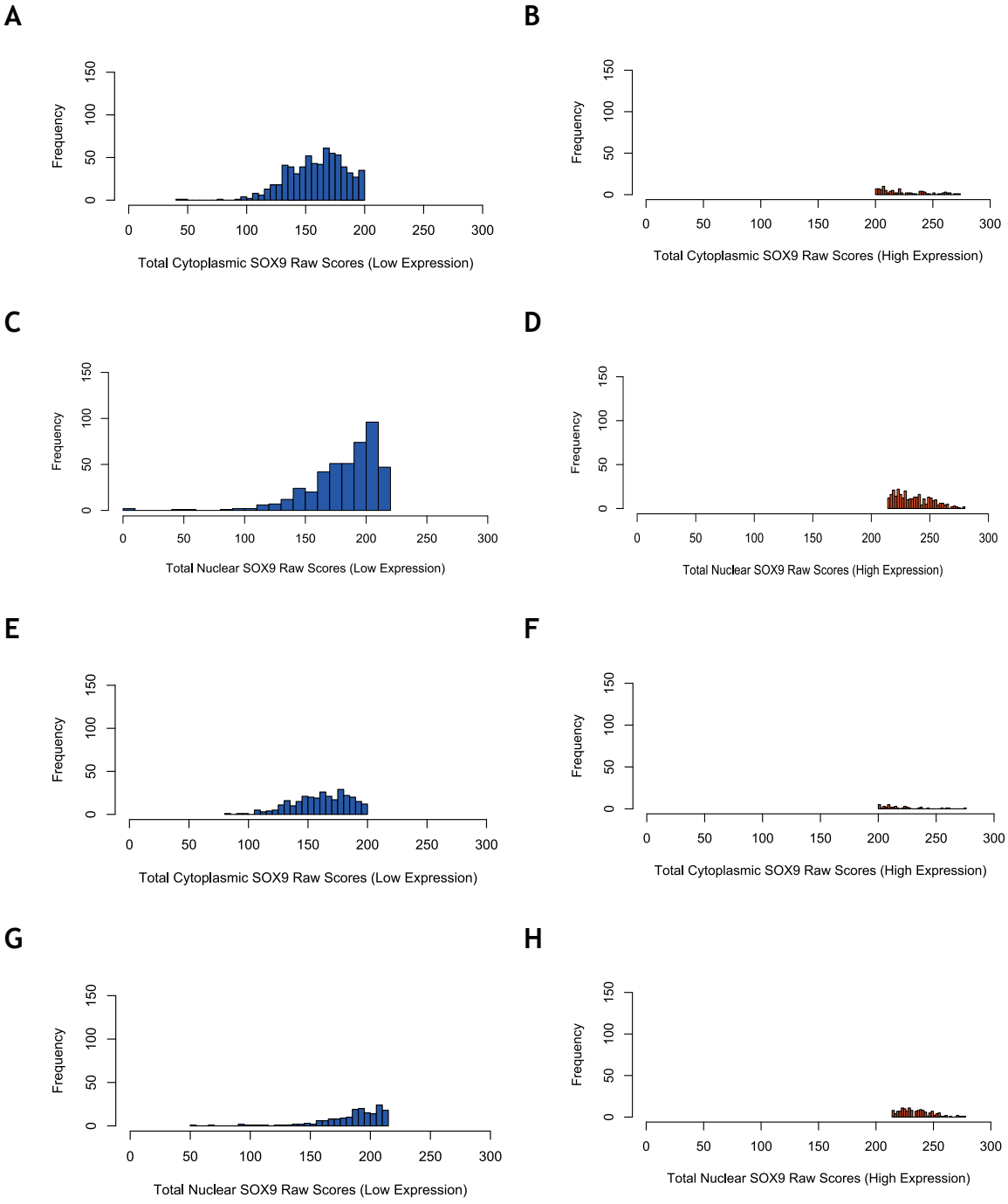


Figure 6.9. Histograms of Distribution of H-scores Across Low and High Expression Groups.
[A] Low and **[B]** high cytoplasmic SOX9 expression in the training dataset. **[C]** Low and **[D]** high nuclear SOX9 expression in the training dataset. **[E]** Low and **[F]** high cytoplasmic SOX9 expression in the test dataset. **[G]** Low and **[H]** high nuclear SOX9 expression in the test dataset.

6.2.2.4 Expression of SOX9 and Metachronous Polyp Detection

After implementing the exclusion criteria described previously in 2.1.1.3.1 and 2.1.1.3.2, and after removing all invalid cases (n=50 in the training dataset and n=22 in the test dataset), 768 cases were included in the training dataset, and 336 cases were included in the test dataset.

Survival analysis was used to investigate the relationship between SOX9 expression and the detection of metachronous polyps or CRC as described in 2.2.10. There was a significant association between cytoplasmic SOX9 expression and the detection of metachronous polyps or CRC (Figure 6.10A; HR 1.547, 95% CI 1.191-2.008; $p<0.001$) in the training dataset. Patients with high expression of SOX9 were at higher risk of detection of metachronous polyps or CRC. The median time to metachronous polyp or CRC detection was 37 months for high cytoplasmic SOX9 expressors, and 59 months for low cytoplasmic SOX9 expressors. This was reflected in the test dataset (Fig.5.29B; HR 1.654, 95% CI 1.113-2.457; $p=0.012$), where the median time to metachronous polyp or CRC detection was 38 months for high cytoplasmic SOX9 expressors, and 59 months for low cytoplasmic SOX9 expressors.

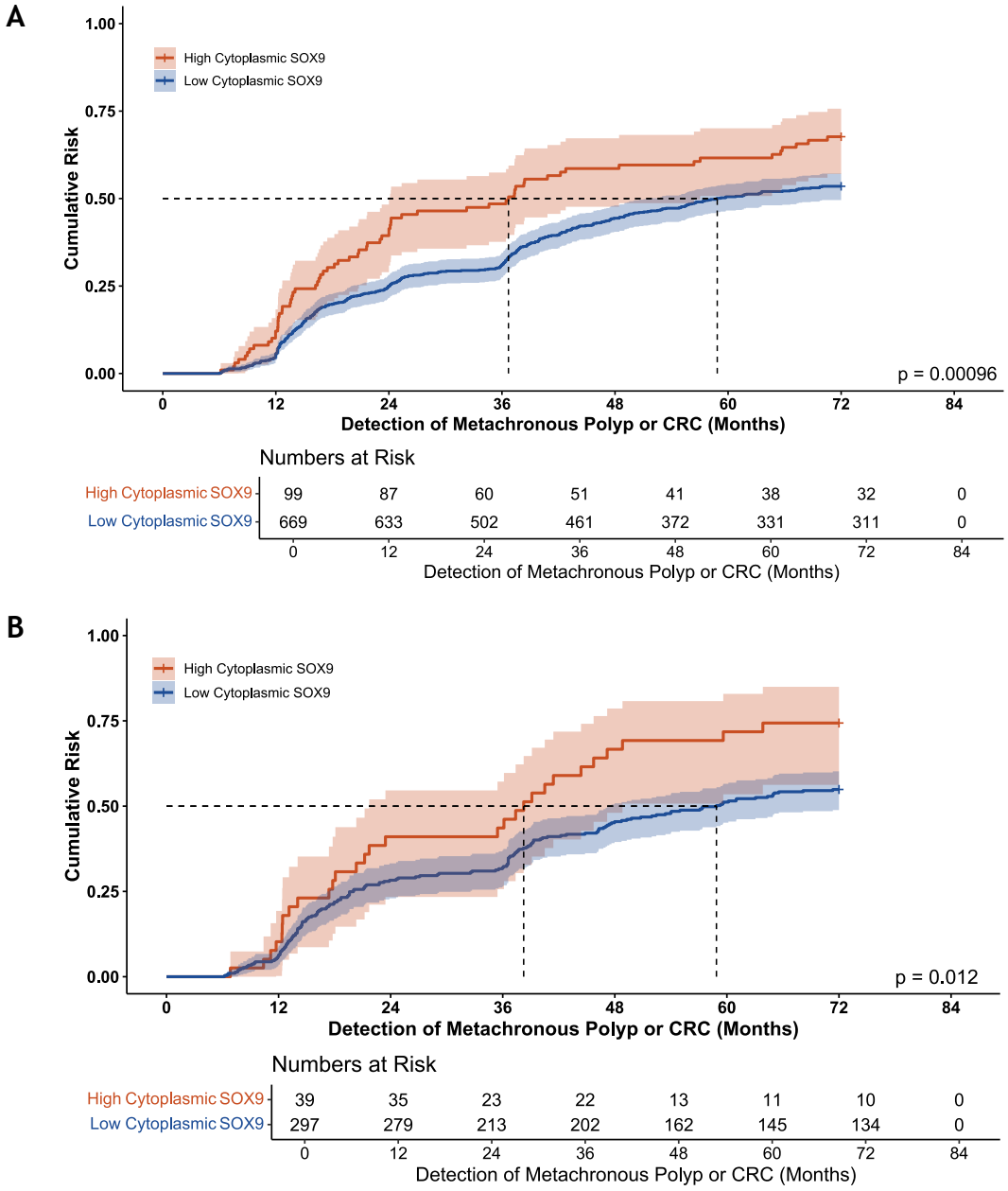


Figure 6.10. Cytoplasmic Expression of SOX9 in Adenomas.
Expression of SOX9 in the epithelia of adenomas in the **[A]** training and **[B]** test datasets. Shading is indicative of 95% CI. Dotted lines are time to median risk.

Survival analysis for nuclear SOX9 expression (Figure 6.11) revealed that nuclear SOX9 expression did not associate with detection of metachronous polyps or CRC in the training dataset (HR 1.197, 95% CI 0.988-1.450; p=0.066). The median time to detection of polyps or CRC for patients with high nuclear SOX9 was 47 months, compared to 60 months for those with low nuclear SOX9. Similarly in the test dataset, there was no significant association between nuclear SOX9 expression and detection of metachronous polyps or CRC (HR 1.266, 95% CI 0.923-1.627; p=0.16). The median time to detection of polyps or CRC for

patients with high nuclear SOX9 expression was 44 months, compared to 60 months for those with low nuclear SOX9 expression.

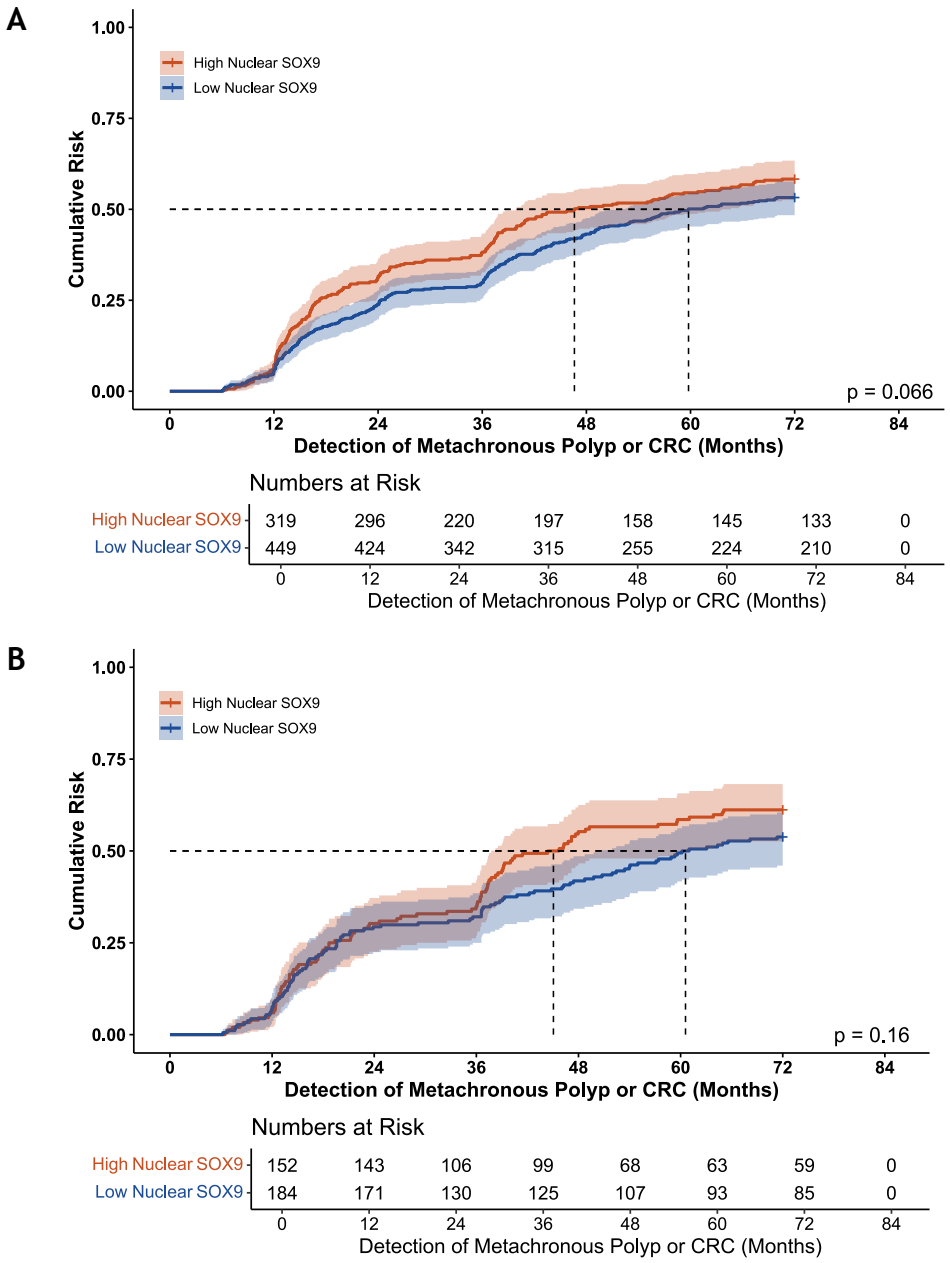


Figure 6.11. Nuclear Expression of SOX9 in Adenomas.
Expression of SOX9 in the epithelia of adenomas in the **[A]** training and **[B]** test datasets. **[B]** Expression of SOX9 in the epithelia of adenomas in the test dataset. Shading is indicative of 95% CI. Dotted lines are time to median risk.

As nuclear expression of SOX9 was not significantly associated with the detection of metachronous polyps or CRC, only cytoplasmic SOX9 expression was taken forward for further analysis.

6.2.2.5 Expression of SOX9 as an Independent Factor in the Prediction of Metachronous Polyps or CRC

Univariate Cox regression was performed to determine associations between cytoplasmic SOX9 expression and relevant clinical factors. As shown in Table 6.1, male sex ($p<0.0001$), left colonic polyps ($p<0.0001$), multiple index polyps ($p<0.0001$), and high BSG2020 risk ($p<0.0001$) all significantly associated with detection of metachronous polyps or CRC. Additionally, high SOX9 expression alone ($p=0.001$) was also statistically associated with the same clinical factors. Figure 6.12 shows equal distribution of SOX9 H-scores against clinical factors. When these factors were entered into multivariate Cox regression to determine which factors were independent predictors of metachronous polyps or CRC, two models were generated, based on whether individual factors were combined into a single factor. Male sex ($p=0.001$), left colon ($p=0.002$), multiple index polyps ($p<0.0001$), and high SOX9 expression ($p=0.015$) were all independent of other factors in model 1. Male sex ($p=0.001$), left colon ($p=0.001$), high BSG2020 risk ($p<0.0001$), and high SOX9 expression ($p=0.005$) were independent predictors in model 2.

Table 6.1. Cytoplasmic SOX9 Expression in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE Training Dataset

		Univariate			Multivariate					
					Model 1: BSG2020 Guidelines Components			Model 2: BSG2020 Guidelines		
		HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Sex	Female (n=218)	1.0	-	-	1.0	-	-	1.0	-	-
	Male (n=550)	1.506	1.214-1.868	**<0.001**	1.461	1.166-1.830	**0.001**	1.477	1.179-1.850	**0.001**
Age	Screening Age (n=735)	1.0	-	-	-	-	-	-	-	-
	Above Screening Age (n=33)	0.844	0.513-1.391	0.506	-	-	-	-	-	-
Site	Right Colon (n=98)	1.0	-	-	1.0	-	-	1.0	-	-
	Left Colon (n=554)	0.615	0.473-0.800	**<0.001**	0.652	0.498-0.853	**0.002**	0.646	0.494-0.845	**0.001**
	Rectum (n=115)	0.844	0.608-1.1.72	0.311	0.923	0.660-1.291	0.640	0.924	0.661-1.292	0.644
High Grade Dysplasia	Absent (n=643)	1.0	-	-	-	-	-	Omitted From Model		
	Present (n=125)	1.126	0.881-1.439	0.342	-	-	-			
Number of Polyps	1 (n=260)	1.0	-	-	1.0	-	-	Omitted From Model		
	2-4 (n=415)	1.546	1.249-1.915	**<0.001**	1.581	1.264-1.976	**<0.0001**			
	5+ (n=93)	2.549	1.905-3.410	**<0.001**	2.376	1.749-3.229	**<0.0001**			
Polyp Type	Tubular (n=342)	1.0	-	-	-	-	-	-	-	-
	Tubulovillous (n=372)	0.935	0.773-1.131	0.492	-	-	-	-	-	-
	Villous (n=54)	0.936	0.638-1.372	0.735	-	-	-	-	-	-
BSG2020 Guidelines	Low Risk (n=262)	1.0	-	-	Omitted From Model			1.0	-	-
	High Risk (n=506)	1.694	1.379-2.081	**<0.001**				1.708	1.378-2.118	**<0.0001**
Cytoplasmic SOX9 Adenoma	Low Expression (n=669)	1.0	-	-	1.0	-	-	1.0	-	-
	High Expression (n=99)	1.547	1.191-2.008	**0.001**	1.398	1.068-1.828	**0.015**	1.465	1.123-1.911	**0.005**

Screening Age=50-74 years old. HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. N=768. Site information was not available for 1 patient, hence information on site was included for n=767. Cox regression analysis was carried out to obtain hazard ratios.

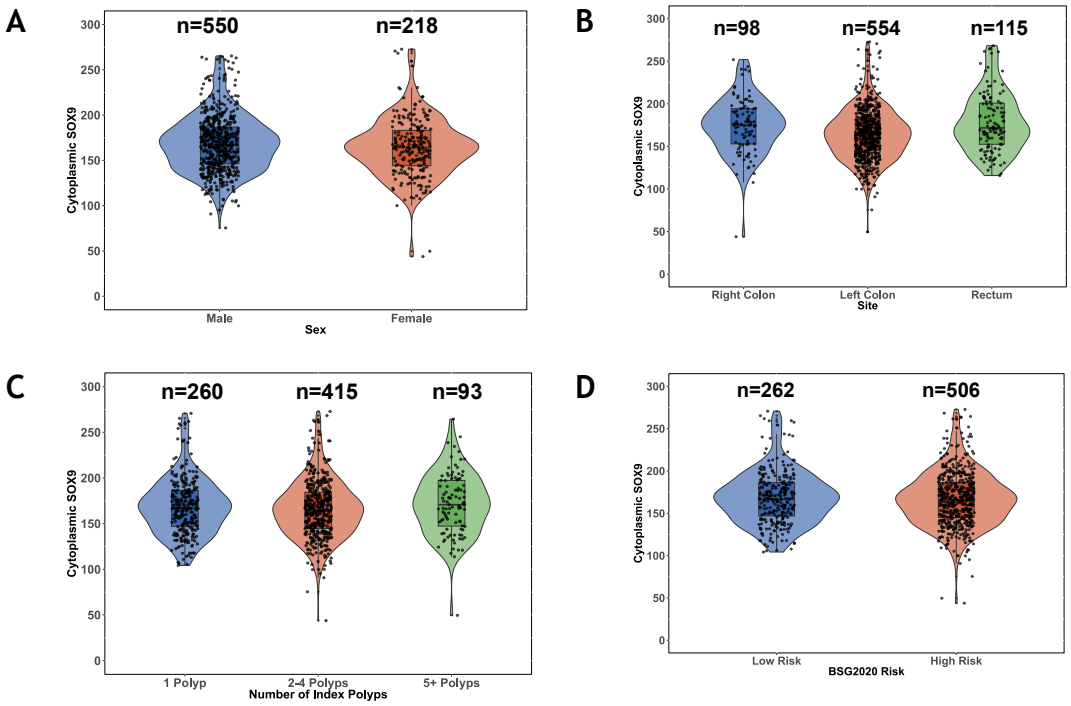


Figure 6.12. Violin Plots of the Distribution of Cytoplasmic H-scores Against Clinical Characteristics Associated with the Detection of Metachronous Polyps in the Training Dataset.
[A] Sex. **[B]** Polyp site. **[C]** Number of index polyps. **[D]** Risk by BSG2020 guidelines. N=768. Site information was not available for 1 patient, hence information on site was included for n=767.

In the test dataset (Table 6.2), left colonic polyps (p=0.003), multiple index polyps (p=0.013 for 2-4 polyps, p<0.0001 for 5+ polyps), and high BSG2020 risk (p=0.002) all significantly associated with the detection of metachronous polyps or CRC. Additionally, high SOX9 expression alone (p=0.013), and in combination with BSG2020 criteria (p=0.009 for the intermediate-risk group, and p=0.001 for the high-risk group) were also statistically associated with the detection of metachronous polyps or CRC. Figure 6.13 shows equal distribution of SOX9 H-scores against clinical factors. Left colon (p=0.001), rectal polyp (p=0.031), 5+ index polyps (p<0.0001), and high SOX9 expression (p=0.032) were all independent of other factors in multivariate model 1. Left colon (p=0.005), high BSG2020 risk (p=0.011), and high SOX9 expression (p=0.019) were independent predictors in model 2.

Table 6.2. Cytoplasmic SOX9 Expression in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE Test Dataset

		Univariate			Multivariate					
					Model 1: BSG2020 Guidelines Components			Model 2: BSG2020 Guidelines		
		HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Sex	Female (n=84)	1.0	-	-	-	-	-	-	-	-
	Male (n=252)	1.072	0.778-1.478	0.670	-	-	-	-	-	-
Age	Screening Age (n=321)	1.0	-	-	-	-	-	-	-	-
	Above Screening Age (n=15)	1.453	0.811-2.603	0.209	-	-	-	-	-	-
Site	Right Colon (n=54)	1.0	-	-	1.0	-	-	1.0	-	-
	Left Colon (n=246)	0.596	0.424-0.837	**0.003**	0.556	0.387-0.798	**0.001**	0.595	0.416-0.851	**0.005**
	Rectum (n=35)	0.676	0.402-1.137	0.140	0.544	0.313-0.945	**0.031**	0.641	0.373-1.100	0.106
High Grade Dysplasia	Absent (n=281)	1.0	-	-	-	-	-	Omitted From Model		
	Present (n=55)	1.114	0.778-1.597	0.555	-	-	-			
Number of Polyps	1 (n=97)	1.0	-	-	1.0	-	-	Omitted From Model		
	2-4 (n=185)	1.547	1.095-2.185	**0.013**	1.388	0.979-1.967	0.066			
	5+ (n=54)	2.584	1.711-3.902	**<0.0001**	2.574	1.667-3.974	**<0.0001**			
Polyp Type	Tubular (n=135)	1.0	-	-	-	-	-	-	-	-
	Tubulovillous (n=187)	0.846	0.637-1.123	0.248	-	-	-	-	-	-
	Villous (n=14)	1.741	0.929-3.263	0.083	-	-	-	-	-	-
BSG2020 Guidelines	Low Risk (n=100)	1.0	-	-	Omitted From Model			1.0	-	-
	High Risk (n=236)	1.687	1.218-2.338	**0.002**				1.531	1.101-2.130	**0.011**
Cytoplasmic SOX9 Adenoma	Low Expression (n=297)	1.0	-	-	1.0	-	-	1.0	-	-
	High Expression (n=39)	1.654	1.113-2.457	**0.013**	1.548	1.039-2.305	**0.032**	1.606	1.080-2.390	**0.019**

Screening Age=50-74 years old. HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. N=336. Site information was not available for 1 patient, hence information on site was included for n=335. Cox regression analysis was carried out to obtain hazard ratios.

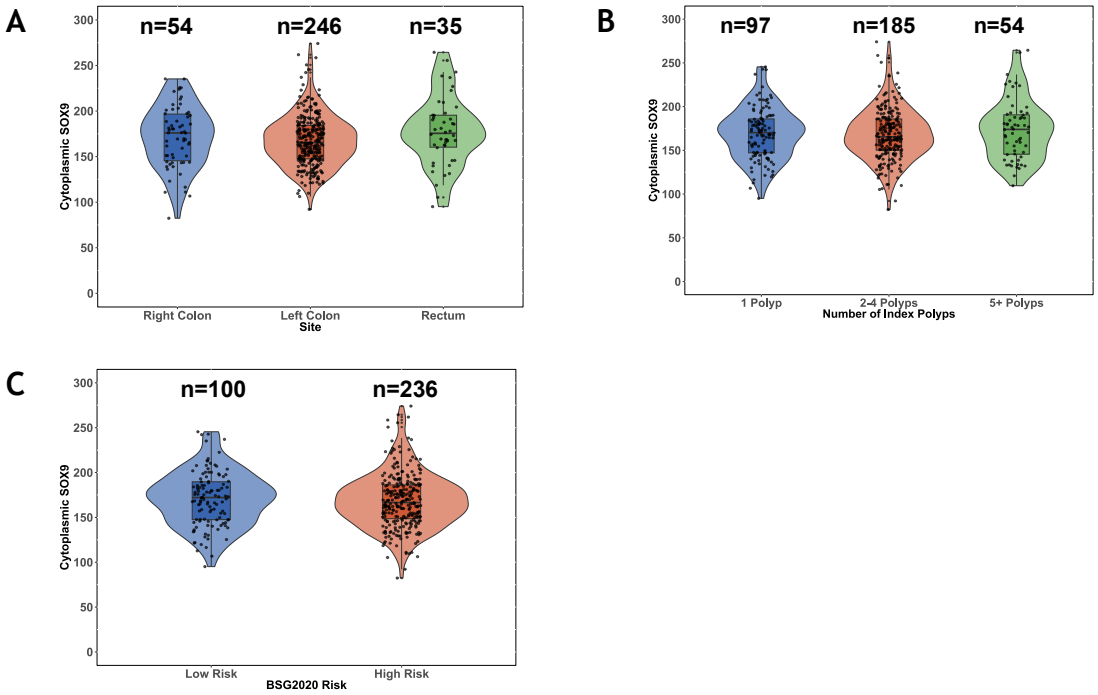


Figure 6.13. Violin Plots of the Distribution of Cytoplasmic H-scores Against Clinical Characteristics Associated with the Detection of Metachronous Polyps in the Test Dataset.
[A] Polyp site. **[B]** Number of index polyps. **[C]** Risk by BSG2020 guidelines. N=336. Site information was not available for 1 patient, hence information on site was included for n=335.

6.2.2.6 Expression of SOX9 and Clinical Characteristics

As high expression of cytoplasmic SOX9 was associated with shorter time to detection of metachronous polyps or CRC, X^2 analysis was carried out to identify any correlations with clinical characteristics. As shown in Table 6.3 and Figure 6.14A, high cytoplasmic SOX9 expression was positively associated with rectal polyps ($p<0.001$) and the detection of metachronous polyps or CRC ($p=0.008$) in the training dataset. These trends were reflected in the test dataset where high cytoplasmic SOX9 correlated with polyps in patients scoped above the screening age ($p=0.007$), polyps with high-grade dysplasia ($p=0.034$), and the detection of metachronous polyps or CRC ($p=0.021$; Table 6.3 Figure 6.14B).

Table 6.3. Relationship between Cytoplasmic SOX9 Expression and Clinical Characteristics in the Training and Test Datasets.

		Training N=768					Test N=336				
		Low Expression		High Expression		P	Low Expression		High Expression		P
		669	87%	99	13%		297	88%	39	12%	
Sex	Female	193	29%	25	25%	0.459	76	26%	8	21%	0.491
	Male	476	71%	74	75%		221	74%	31	79%	
Age	50-74 (Screening Age)	639	96%	96	97%	0.506	287	97%	34	87%	**0.007**
	75+ (Above Screening Age)	30	4%	3	3%		10	3%	5	13%	
Site	Right Colon	79	12%	19	19%	**<0.001**	45	15%	9	23%	0.201
	Left Colon	505	76%	49	50%		222	75%	24	62%	
	Rectum	84	12%	31	31%		29	10%	6	15%	
High Grade Dysplasia	Absent	565	85%	78	79%	0.154	253	85%	28	72%	**0.034**
	Present	104	15%	21	21%		44	15%	11	28%	
Number of Polyps	1	225	34%	35	35%	0.322	88	30%	9	23%	0.693
	2-4	367	55%	48	49%		162	54%	23	59%	
	5+	77	11%	16	16%		47	16%	7	18%	
Adenoma Type	Tubular	301	45%	41	41%	0.104	120	40%	15	39%	0.935
	Tubulovillous	326	49%	46	47%		165	56%	22	56%	
	Villous	42	6%	12	12%		12	4%	2	5%	
BSG2020 Guidelines	Low Risk	227	34%	35	35%	0.781	89	30%	11	28%	0.821
	High Risk	442	66%	64	65%		208	70%	28	72%	
Metachronous Polyps or CRC Detection	No	311	47%	32	32%	**0.008**	134	45%	10	26%	**0.021**
	Yes	358	53%	67	68%		163	55%	29	74%	

Chi-squared analysis to obtain correlations. P-values were generated by Pearson's Correlation test. Site information was not available for 1 patient in either dataset, hence information on site was included for n=767 and n=335 patients in the training and test dataset, respectively.

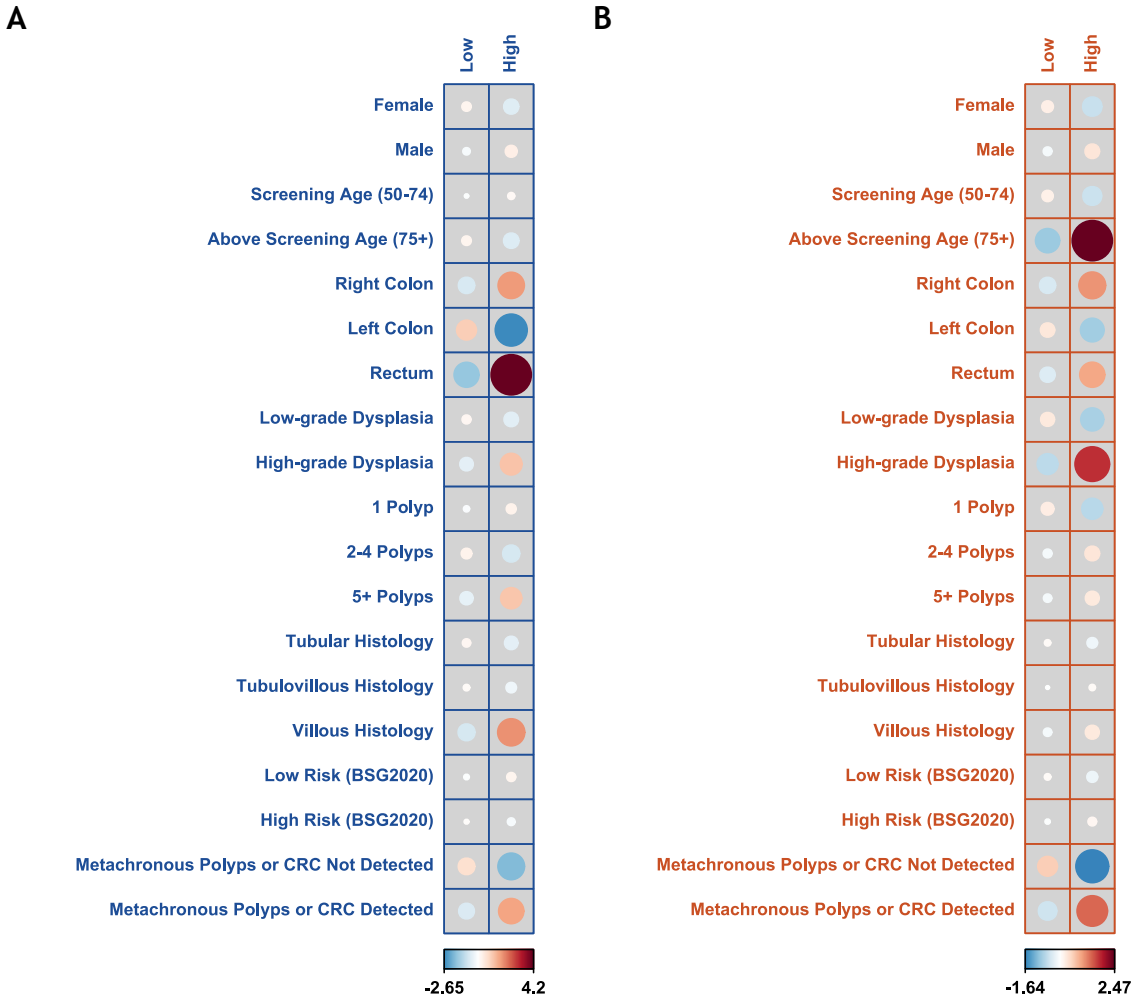


Figure 6.14. Corrplots of Associations Between SOX9 Expression Against Clinical Characteristics.
Low and high expression of SOX9 in the **[A]** training (n=768) and **[B]** test (n=336) dataset. Dot size suggests a high correlation between variables, while and colour intensity indicates negative (blue) or positive (red) associations.

6.2.3 Addition of SOX9 to BSG2020 Guidelines

To improve current patient risk stratification guidelines, the expression of cytoplasmic SOX9 was added to the BSG2020 criteria. This resulted in the following four groups:

1- Both Low, where SOX9 expression is low and BSG2020 risk was low

2- High SOX9 and Low BSG2020

3- Low SOX9 and High BSG2020

4- Both High, where SOX9 expression is high and BSG2020 risk was high

6.2.3.1 SOX9 in Combination with BSG2020 Guidelines and Metachronous Polyps or CRC

As shown in Figure 6.15, there was a distinct overlap between the two “one low one high” groups (green and black; training dataset p=0.967, test dataset p=0.552).

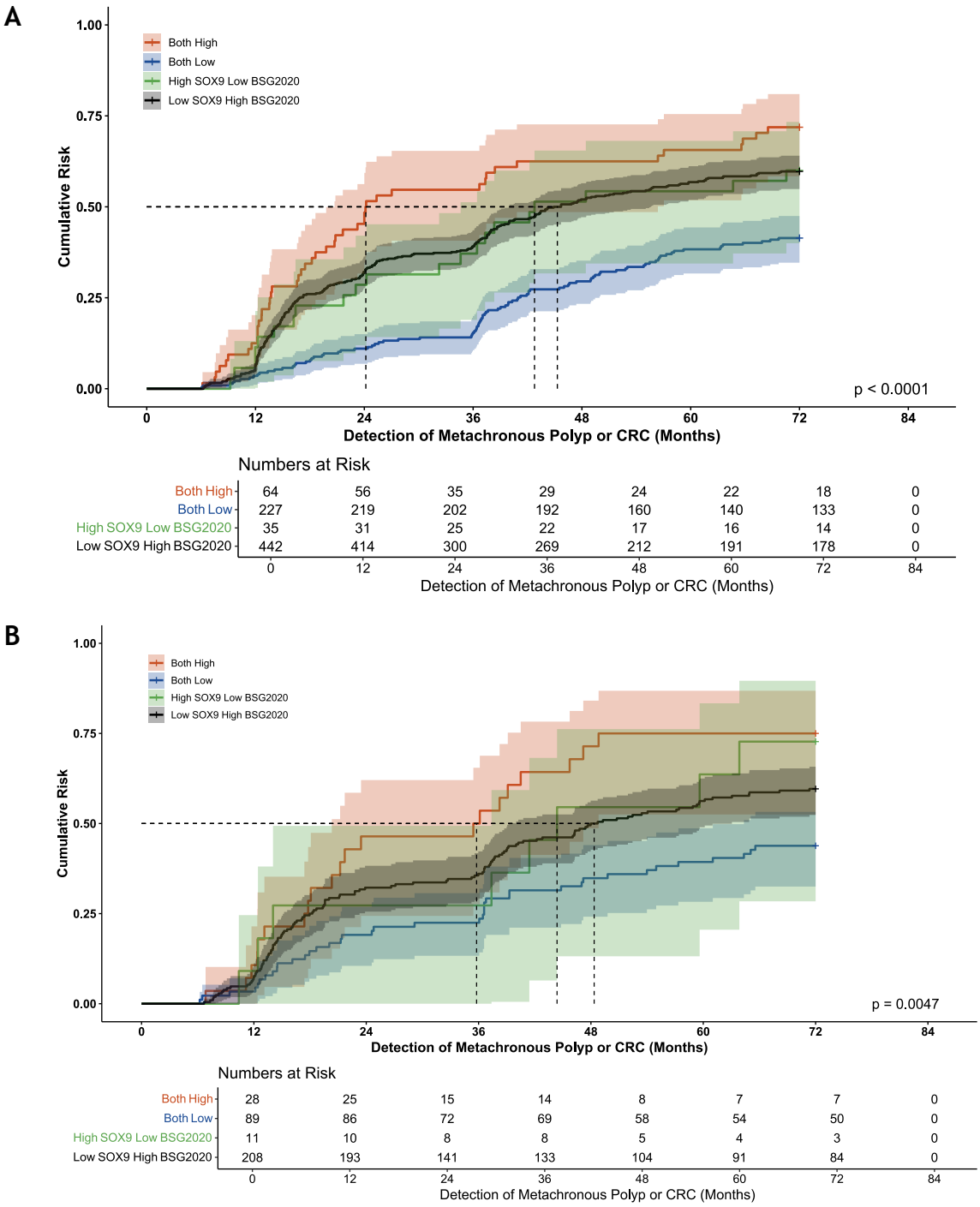


Figure 6.15. Combination of SOX9 with BSG2020 Criteria.
[A] Training dataset (n=768). Both Low-High SOX9 Low BSG2020, p=0.009. Both Low-Low SOX9 High BSG2020, p<0.0001. Both Low-Both High, p<0.0001. High SOX9 Low BSG2020-Low SOX9 High BSG2020, p=0.967. High SOX9 Low BSG2020-Both High, p=0.145. Low SOX9 High BSG2020-Both High, p=0.019. **[B]** Test dataset (n=336). Both Low-High SOX9 Low BSG2020, p=0.070. Both Low-Low SOX9 High BSG2020, p=0.011. Both Low-Both High, p=0.001. High SOX9 Low BSG2020-Low SOX9 High BSG2020, p=0.552. High SOX9 Low BSG2020-Both High, p=0.551. Low SOX9 High BSG2020-Both High, p=0.066. Shading is indicative of 95% CI. Dotted lines are time to median risk.

These two groups were combined into one, resulting in three distinct risk groups:

1- Low Risk, where SOX9 expression is low and BSG2020 risk is low

2- Intermediate Risk, where either SOX9 expression or BSG2020 risk is low

3- High Risk, where SOX9 expression is high and BSG2020 risk is high

As shown in Figure 6.16, these novel risk groups significantly correlated with the detection of metachronous polyps or CRC ($p<0.0001$). The high-risk group had a median time to metachronous polyp or CRC detection of 24 months, the intermediate-risk group 45 months, and the low-risk group did not reach median risk. This was reflected in the test dataset ($p=0.0002$), where the high-risk group had a median time to metachronous polyp or CRC detection of 35 months, the intermediate-risk group 48 months, and the low risk-group again did not reach median risk.

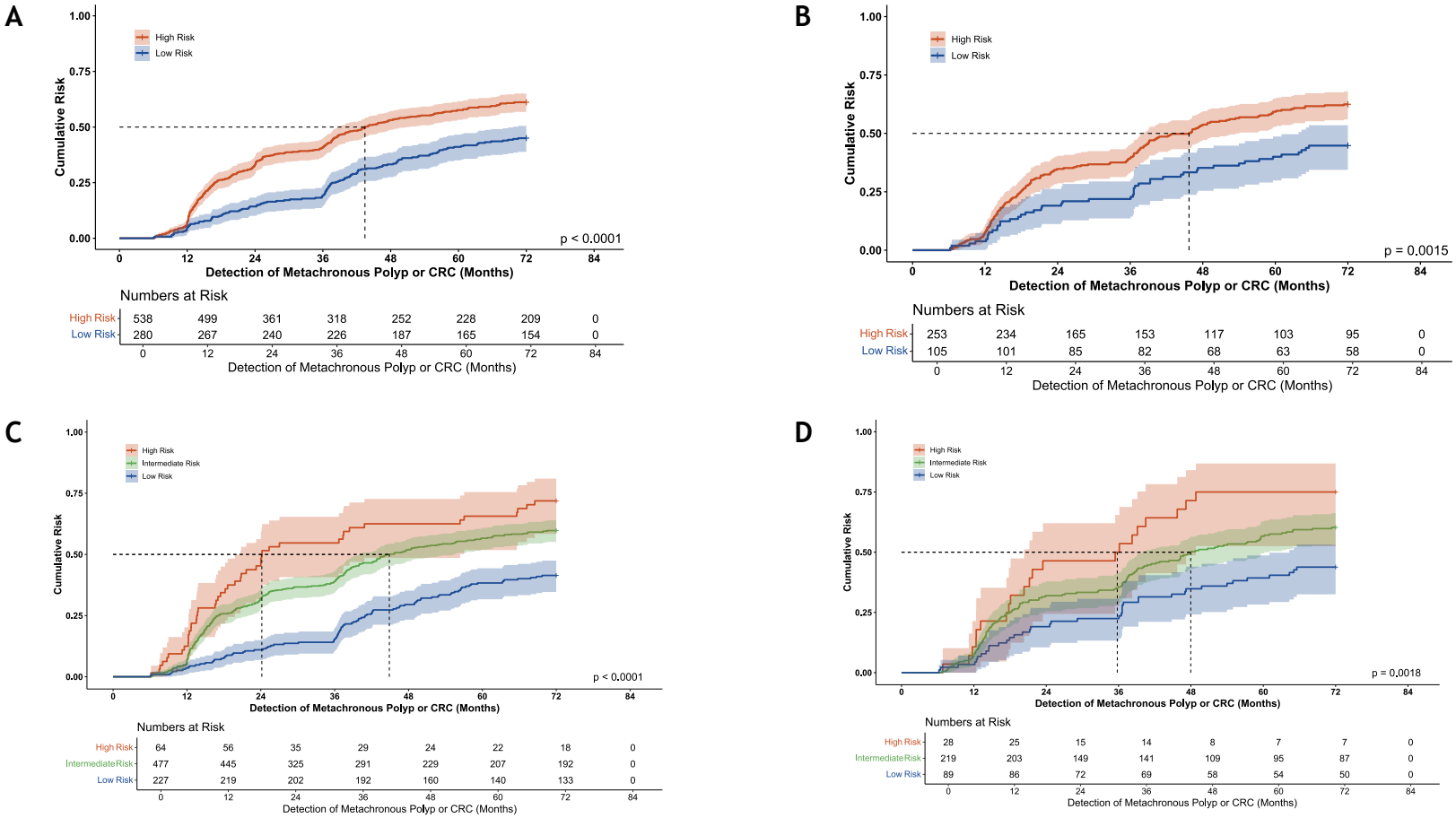


Figure 6.16. Final Risk Groups When SOX9 Expression is Combined with BSG2020 Criteria.
[A] Current BSG2020 Guidelines applied to the INCISE cohort, training dataset (n=818). **[B]** Current BSG2020 Guidelines applied to the INCISE cohort, test dataset (n=358). **[C]** SOX9 in combination with BSG2020 Guidelines, training dataset (n=768). Low Risk-Intermediate Risk, p<0.0001. Low Risk-High Risk, p<0.0001. Intermediate Risk-High Risk, p=0.018. **[D]** SOX9 in combination with BSG2020 Guidelines, test dataset (n=336). Low Risk-Intermediate Risk, p=0.008. Low Risk-High Risk, p=0.001. Intermediate Risk-High Risk, p=0.070. Shading is indicative of 95% CI. Dotted lines are time to median risk.

Univariate Cox regression was performed to determine associations between the three risk groups and relevant clinical factors in the training dataset (Table 6.4). SOX9 combined with BSG2020 criteria (p<0.0001) significantly associated with the detection of metachronous polyps or CRC. A third multivariate model was generated, and in addition to male sex (p=0.001), and left colonic polyps (p=0.002) being independent predictors of the detection of metachronous polyps or CRC, both intermediate and high risk (p<0.0001 for both) were also independent predictors.

Table 6.4. Cytoplasmic SOX9 Expression Combined with BSG2020 Guidelines in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE Training Dataset

		Univariate			Multivariate		
					Model 3: Combined Cytoplasmic SOX9 and BSG2020 Guidelines		
		HR	95% CI	P	HR	95% CI	P
Sex	Female (n=146)	1.0	-	-	1.0	-	-
	Male (n=450)	1.506	1.214-1.868	**<0.001**	1.477	1.179-1.850	**0.001**
Site	Right Colon (n=98)	1.0	-	-	1.0	-	-
	Left Colon (n=554)	0.615	0.473-0.800	**<0.001**	0.652	0.498-0.852	**0.002**
	Rectum (n=115)	0.844	0.608-1.1.72	0.311	0.933	0.666-1.308	0.689
Combined Cytoplasmic SOX9 and BSG2020 Guidelines	Low Risk (n=227)	1.0	-	-	1.0	-	-
	Intermediate Risk (n=477)	1.830	1.449-2.312	**<0.001**	1.752	1.386-2.215	**<0.0001**
	High Risk (n=64)	2.678	1.881-3.813	**<0.001**	2.448	1.712-3.501	**<0.0001**

HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. N=768. Site information was not available for 1 patient, hence information on site was included for n=767. Cox regression analysis was carried out to obtain hazard ratios.

This was reflected in the test dataset (Table 6.5), where left colonic polyps (p=0.005), and both intermediate risk (p=0.011) and high risk (p=0.001) were independent predictors of the detection of metachronous polyps or CRC.

Table 6.5. Cytoplasmic SOX9 Expression Combined with BSG2020 Guidelines in Relation to Clinicopathological Characteristics, and the Detection of Metachronous Polyps or CRC in the INCISE Test Dataset

		Univariate			Multivariate		
					Model 3: Combined Cytoplasmic SOX9 and BSG2020 Guidelines		
		HR	95% CI	P	HR	95% CI	P
Site	Right Colon (n=54)	1.0	-	-	1.0	-	-
	Left Colon (n=246)	0.596	0.424-0.837	**0.003**	0.595	0.416-0.851	**0.005**
	Rectum (n=35)	0.676	0.402-1.137	0.140	0.643	0.374-1.105	0.110
Combined Cytoplasmic SOX9 and BSG2020 Guidelines	Low Risk (n=89)	1.0	-	-	1.0	-	-
	Intermediate Risk (n=219)	1.614	1.129-2.307	**0.009**	1.590	1.112-2.275	**0.011**
	High Risk (n=28)	2.473	1.452-4.209	**0.001**	2.413	1.413-4.123	**0.001**

HR=Hazard Ratio. 95% CI=95% confidence interval. BSG2020=British Society of Gastroenterology. Significant p-values are depicted in bold. N=336. Site information was not available for 1 patient, hence information on site was included for n=335. Cox regression analysis was carried out to obtain hazard ratios.

6.2.3.2 SOX9 in Combination with BSG2020 Guidelines and Clinical Characteristics

The high-risk group in the training dataset was associated with male patients (p=0.022), rectal polyps (p<0.0001), 5+ index polyps (p<0.0001), and the detection of metachronous polyps or CRC (p<0.0001), according to X^2 analysis. Two of these associations were reflected in the test dataset, where the high-risk group was associated with multiple index polyps (p<0.0001) and the detection of metachronous polyps or CRC (p=0.004; Table 6.6, Figure 6.17).

Table 6.6. Relationship between the Final Risk Groups and Clinical Characteristics in the Training and Test Datasets

		Training N=768							Test N=336						
		Low Risk		Intermediate Risk		High Risk		P	Low Risk		Intermediate Risk		High Risk		P
		227	30%	477	62%	64	8%		89	27%	219	65%	28	8%	
Sex	Female	8	35%	123	26%	15	23%	**0.022**	30	34%	50	23%	4	14%	0.053
	Male	147	65%	354	74%	49	77%		59	66%	169	77%	24	86%	
Age	50-74 (Screening Age)	214	94%	460	96%	61	95%	0.412	86	97%	210	96%	25	89%	0.237
	75+ (Above Screening Age)	13	6%	17	4%	3	5%		3	3%	9	4%	3	11%	
Site	Right Colon	20	9%	66	14%	12	19%	**<0.0001**	11	12%	36	17%	7	25%	0.268
	Left Colon	176	77%	346	73%	32	50%		70	79%	160	73%	16	57%	
	Rectum	31	14%	64	13%	20	31%		8	9%	22	10%	5	18%	
High Grade Dysplasia	Absent	188	83%	406	85%	49	77%	0.200	74	83%	188	86%	19	68%	0.053
	Present	39	17%	71	15%	15	23%		15	17%	31	14%	9	32%	
Number of Polyps	1	225	99%	35	7%	0	0%	**<0.0001**	87	98%	10	5%	0	0%	**<0.0001**
	2-4	2	1%	365	77%	48	75%		2	2%	162	74%	21	75%	
	5+	0	0%	77	16%	16	25%		0	0%	47	22%	7	25%	
Adenoma Type	Tubular	91	40%	222	46%	29	45%	0.352	35	39%	86	39%	14	50	0.512
	Tubulovillous	116	51%	227	48%	29	45%		48	54%	126	58%	13	46%	
	Villous	20	9%	28	6%	6	10%		6	7%	7	3%	1	4%	
Metachronous Polyps or CRC Detection	No	133	59%	192	40%	18	28%	**<0.0001**	50	56%	87	40%	7	25%	**0.004**
	Yes	94	41%	285	60%	46	72%		39	44%	132	60%	21	75%	

Chi-squared to obtain correlations. P-values were generated by Pearson's Correlation test. Site information was not available for 1 patient in either dataset, hence information on site was included for n=767 and n=335 patients in the training and test dataset, respectively.

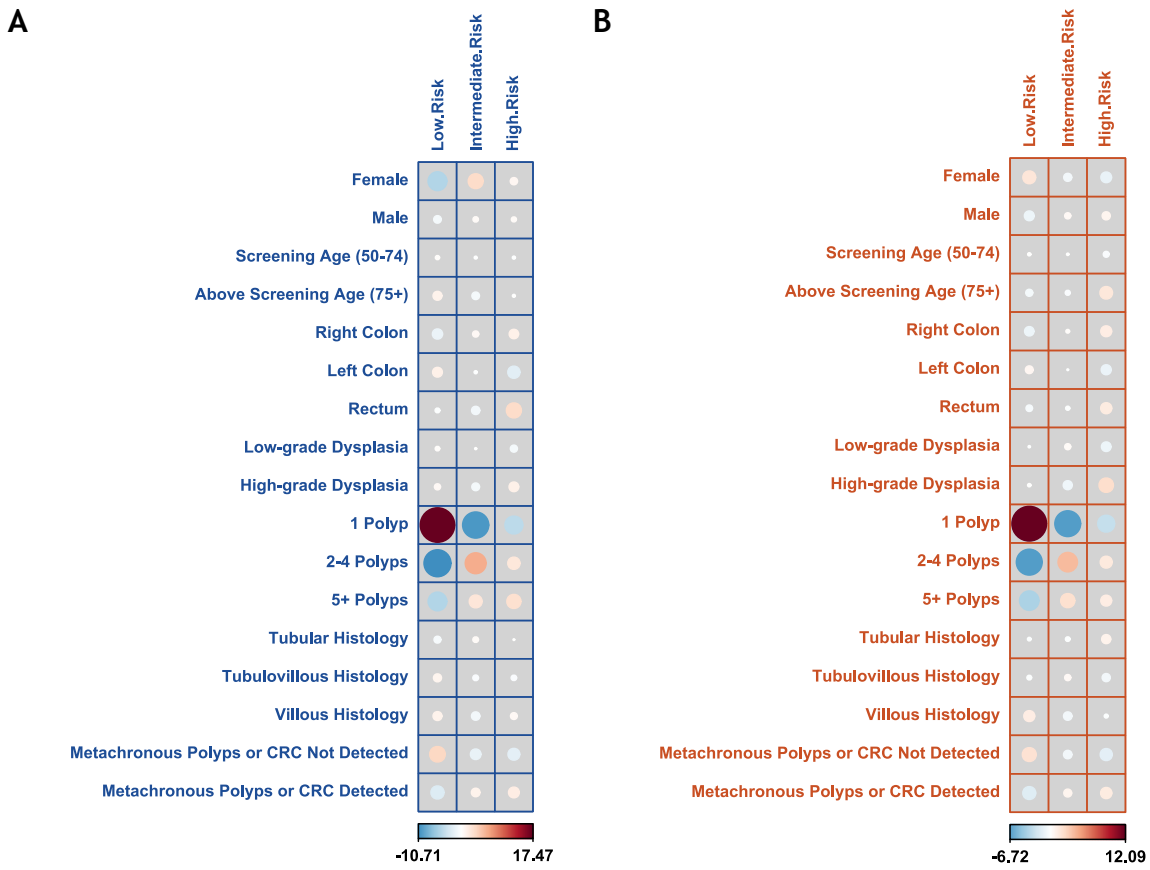


Figure 6.17. Corrplots of Associations Between the Final Risk Groups Against Clinical Characteristics.
Risk groups in the **[A]** training (n=768) and **[B]** test (n=336) dataset. Dot size suggests a high correlation between variables, while and colour intensity indicates negative (blue) or positive (red) associations.

As shown in Table 6.7, the high-risk group was able to accurately detect 55% of patients who develop metachronous polyps 3 years post-index endoscopy, as well as 72% of patients who develop metachronous polyps 6 years post-index endoscopy, in the training dataset. Current BSG2020 Guidelines detect only 41% of patients who develop metachronous polyps 3 years post-index endoscopy, and 61% of patients who develop metachronous polyps 6 years post-index endoscopy, in the training set. This was reflected in the test dataset. This novel high-risk group detected 34% of patients who developed advanced adenomas or CRC 3 years after the index endoscopy, more than double that of the current BSG2020 guidelines, which only detected 15%, in the training dataset. This is not only reflected in the test dataset for the 3-year mark, but also in both datasets for the 6-year mark.

Table 6.7. Cumulative Proportion of Patients Developing Metachronous Polyps and Developing Advanced Adenomas or CRC at 3- and 6-Years Post Index Endoscopy

	Patients Developing a Metachronous Polyp at 3- and 6-Years Post Index Endoscopy				Patients Developing Advanced Adenomas or CRC at 3- and 6-Years Post Index Endoscopy			
	3 Years (%)		6 Years (%)		3 Years (%)		6 Years (%)	
	Training	Test	Training	Test	Training	Test	Training	Test
Cytoplasmic SOX9								
Low	31	32	54	55	10	10	20	19
High	48	44	68	74	28	20	42	37
BSG2020 Guidelines								
Low Risk	19	22	45	45	7	7	17	16
High Risk	41	40	61	62	15	14	25	24
Combined Cytoplasmic SOX9 and BSG2020 Guidelines Individual Groupings								
Low SOX9 Low BSG2020	16	22	42	44	6	7	16	15
Low SOX9 High BSG2020	39	36	60	60	13	12	22	21
High SOX9 Low BSG2020	37	27	60	73	16	11	33	34
High SOX9 High BSG2020	55	50	72	75	28	24	47	37
Combined Cytoplasmic SOX9 and BSG2020 Guidelines								
Low Risk	16	22	42	44	6	7	16	15
Intermediate Risk	39	36	60	60	13	12	23	22
High Risk	55	50	72	75	34	24	47	37

BSG2020=British Society of Gastroenterology.

6.3 Transcriptomic Analysis of SOX9 Expression Groups

As patients with high cytoplasmic SOX9 expression in index polyps were found to be associated with the detection of metachronous polyps or CRC, and the addition of SOX9 expression to the BSG2020 Guidelines significantly improved risk stratification, transcriptomics analysis was used to investigate the underlying biological processes that may result in such effects.

The patients in the lowest and highest quartiles for SOX9 expression were used to represent extreme expression levels, thus 553 patients with index polyps sequenced by TempO-Seq were included. Based on the raw expression scores compared to the dichotomisation cut-off described in 6.2.2.3, the patients were grouped into “Low” and “High”. RStudio was used for all subsequent analysis, and the high SOX9 group was compared with the low SOX9 group. Because of the differences seen at the protein level, it was hypothesised that these differences would be reflected on the transcriptomic level.

6.3.1 Principal Component Analysis and Differential Gene Expression

Principal component analysis (PCA) was performed to assess the transcriptomic variation between the two SOX9 extreme expression groups. As shown in Figure 6.18A, there was no distinct clustering of the data.

A volcano plot was constructed to display the differential expression of genes in the High group when compared to the Low group. After implementing the cut-off criteria of pAdj<0.05 and Log2 Fold-change of -1 and 1, 49 differentially expressed genes (DEGs) were identified, with 10 genes being upregulated and 39 being downregulated in the High group (Figure 6.18B). A heatmap was used to visualise how each DEG behaves across samples, revealing a pattern of expression consistent with the volcano plot (Figure 6.18C). This was further supported in a heatmap of the top 10 most upregulated DEGs and the top 10 most downregulated DEGs (Figure 6.18D).

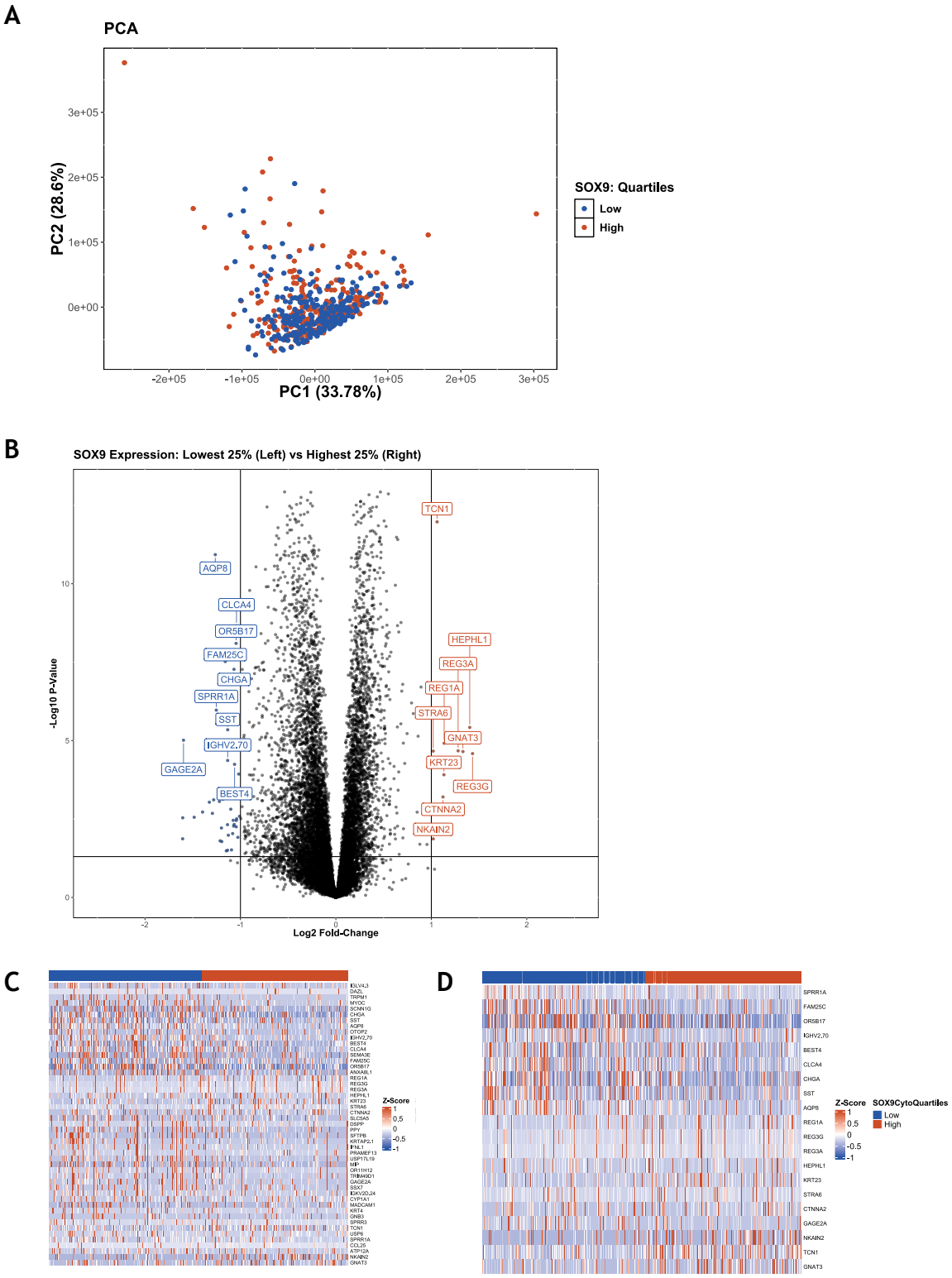


Figure 6.18. Transcriptomic Analysis of Patients in Lowest Quartile and Highest Quartile of Cytoplasmic SOX9 Expression Patients.
[A] Principal component analysis (PCA) of the cytoplasmic SOX9 extreme expression groups. **[B]** Volcano plot of differentially expressed genes (DEGs). The High group is compared to the Low group. The top 10 most downregulated genes (blue) and top 10 most upregulated genes (orange) in the high expression group are shown in boxes. Log2 Fold-change $\leq$ -1 and $\geq$ 1 and pAdj<0.05. **[C]** Heatmap of all 49 DEGs and **[D]** heatmap of top 10 upregulated and downregulated DEGs stratified by SOX9 expression group. N=553, lowest quartile n=283, highest quartile n=270.

To illustrate the expression patterns of the individual genes, violin plots for the top 10 most upregulated and downregulated genes were generated. As shown in Figure 6.19, the genes all show a significant expression pattern (pAdj<0.05) consistent with the previous analysis.

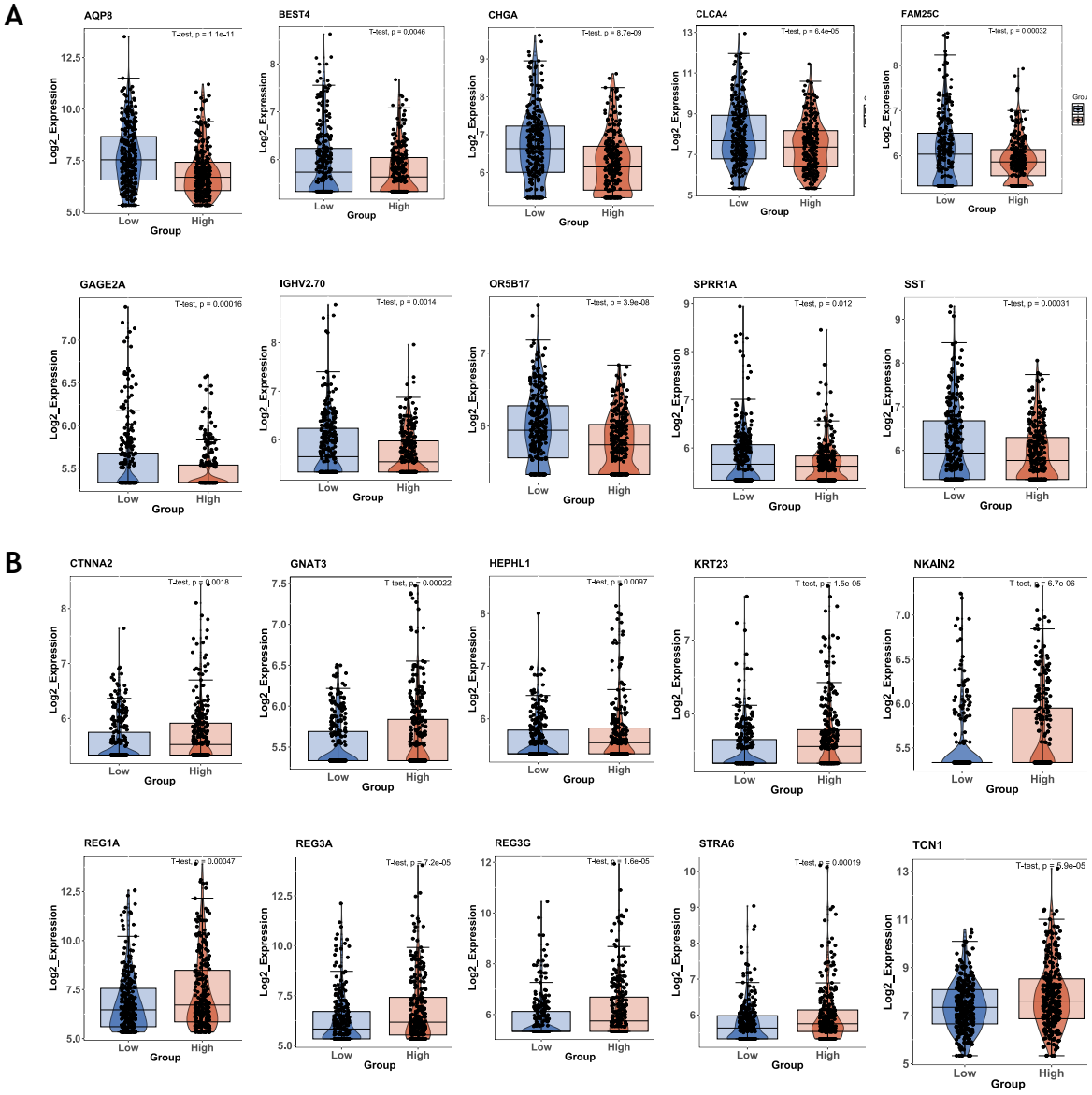


Figure 6.19. Violin Plots of the Top 20 DEGs.
[A] The top 10 most downregulated genes in the High SOX9 extreme expression group. **[B]** The top 10 most upregulated genes in the High SOX9 extreme expression group. N=553, lowest quartile n=283, highest quartile n=270.

6.3.2 Gene Set Enrichment Analysis

Pairwise gene set enrichment analysis (GSEA) was performed using three biological resources: GO, MSigDb, and KEGG.

6.3.2.1 Gene Ontology

Across all three ontologies, 273 GO terms had a pAdj<0.05. A dot plot (Figure 6.20A) was generated to visualise the top 10 most negatively enriched and top 10 most positively enriched statistically significant GO terms in the High SOX9 group.

Of the top 10 most positively enriched terms, ones that suggest anti-microbial functions such as “disruption of cell wall in another organism”, “disruption of cellular anatomical structure in another organism”, and “oligosaccharide binding” were activated in the High SOX9 group. Additionally, terms relating to cell cycle regulation, “maintenance of sister chromatid cohesion”, “maintenance of mitotic sister chromatid cohesion”, “mitotic sister chromatid cohesion” and “negative regulation of double-strand break repair via homologous recombination” were activated in the same group. Of the top 10 most negatively enriched terms were ones that pertain to functions of collagen and extracellular matrix such as “collagen trimer”, “extracellular matrix structural constituent”, “collagen fibril organization”, and “collagen-containing extracellular matrix”. Among the suppressed GO terms was “G protein-coupled receptor activity” which includes four members of the Frizzled family.

A bar plot (Figure 6.20B) was constructed for the top 10 most positively and negatively enriched GO terms ordered by normalised enrichment score (NES <-1 and >1; pAdj<0.05). Of the terms highlighted in the dot plot as activated, six were also positively enriched in the bar plot, with the addition of another cell cycle term “centriole replication” and the addition of “antimicrobial humoral immune response mediated by antimicrobial peptide”. Seven of the 10 terms highlighted as supressed in the dot plot, were also negatively enriched in the bar plot, with the addition of more olfaction terms “sensory perception of smell”, “olfactory receptor activity”, and “detection of chemical stimulus involved in sensory perception of smell”.

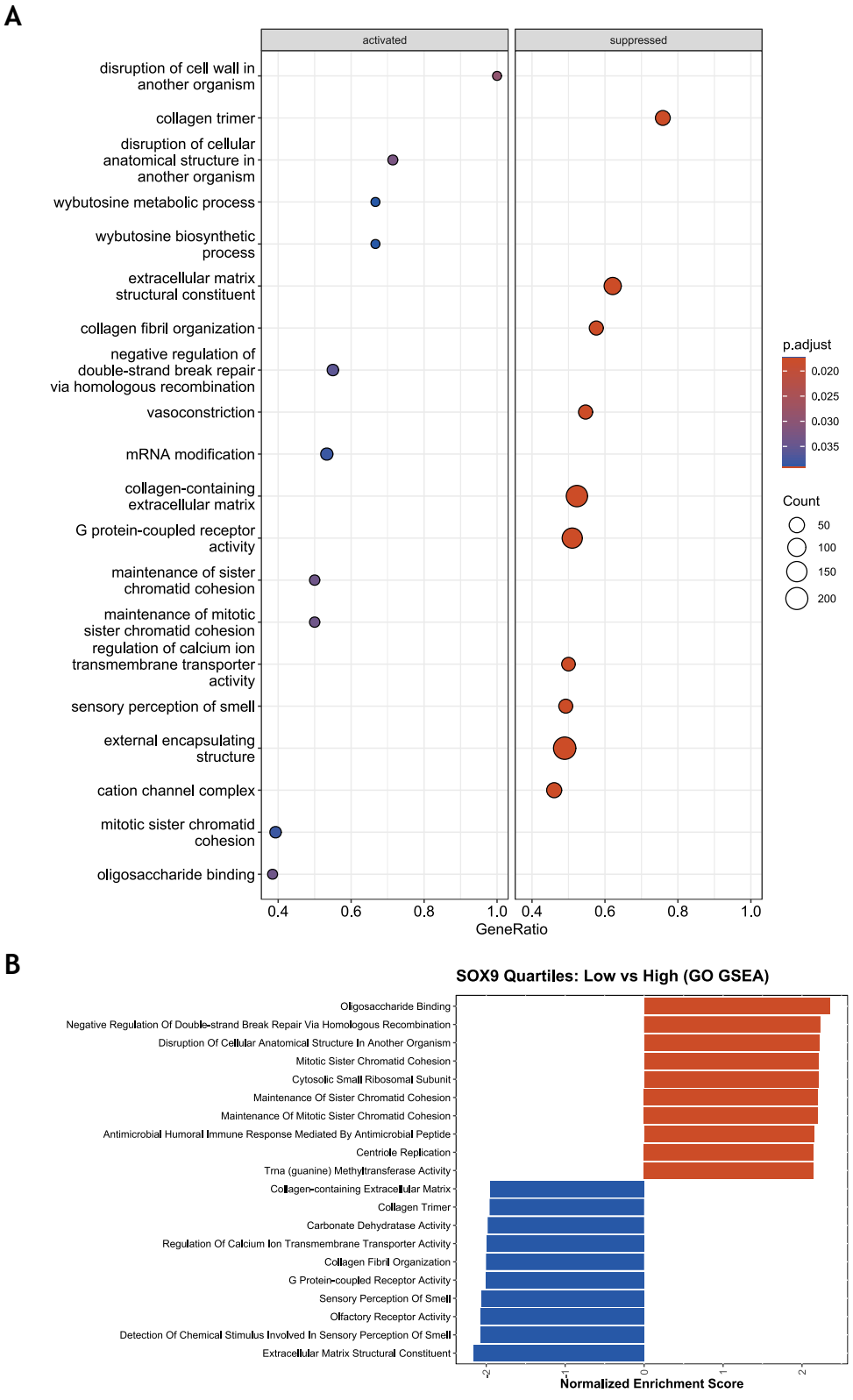


Figure 6.20. GO GSEA Summary.
[A] Dot plot of top 10 activated (left) and supressed (right) GO terms ranked by gene ratio. The x-axis shows the calculated gene ratio of input genes corresponding to the gene set in each term. Dot size refers to gene counts while colour refers to pAdj. [B] Bar plot of top 10 positively enriched (orange) and top 10 negatively enriched (blue) GO terms ranked by NES. The High group is compared to the Low group. N=553, lowest quartile n=283, highest quartile n=270.

Some of these terms share leading edge genes especially the ones with similar functions. UpSet plots (Figure 6.21A) were generated to visualise possible overlaps between terms in the same enrichment trend. As expected within the positively enriched terms, the three pertaining to sister chromatid maintenance and cohesion had six overlapping genes between them (*SLF1*, *SMC5*, *DSCC1*, *RB1*, *BUB1*, and *SLF2*; Figure 6.21B), and the three pertaining to external organisms and antimicrobial activity had four overlapping genes (*REG3G*, *REG3A*, *REG1A*, *and REG1B*). Interestingly, only "antimicrobial humoral immune response mediated by antimicrobial peptide" and "disruption of cellular anatomical structure in another organism" had the *DEFA5* gene overlap between them.

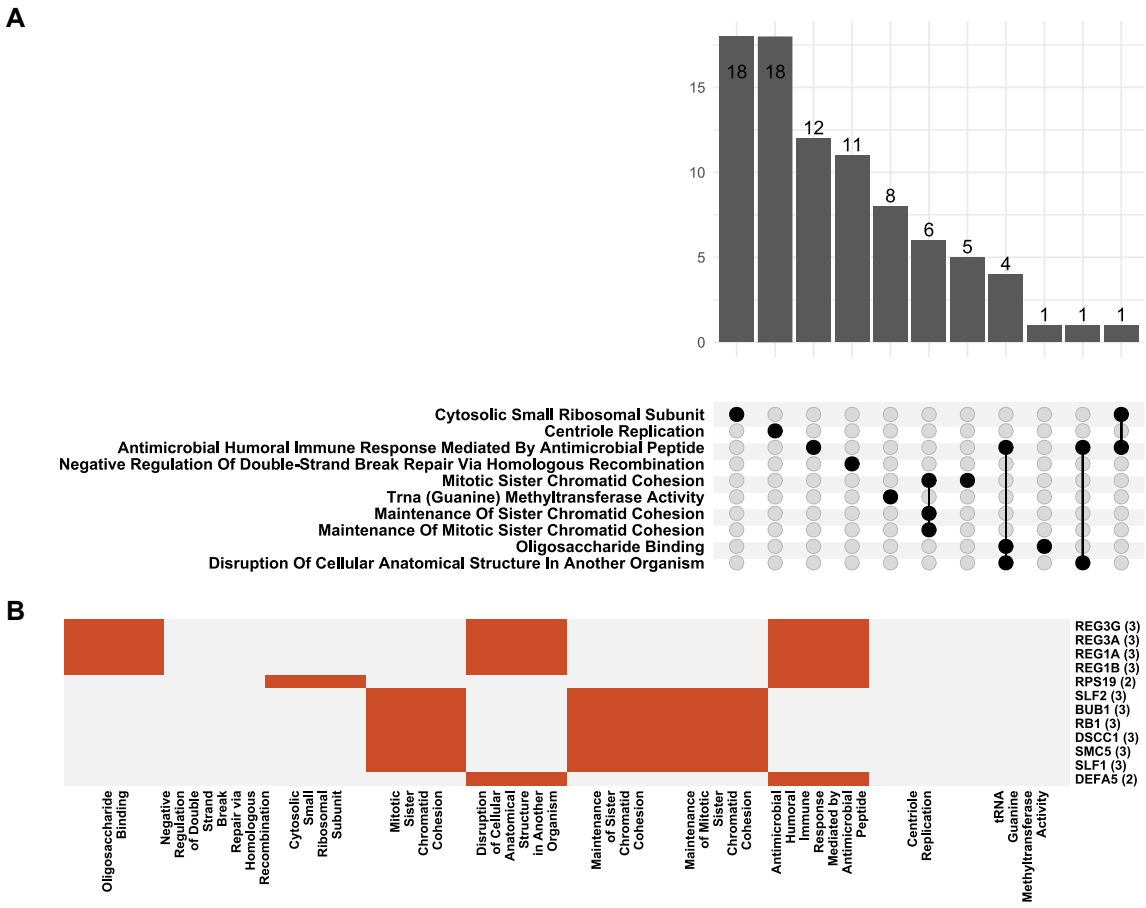


Figure 6.21. Overlapping Leading Edge Genes in Positively Enriched GO Terms.
[A] UpSet plot showing number of overlapping genes as bars across intersecting GO terms (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least two GO terms.

Similarly, negatively enriched terms sharing a function had the greatest number of overlapping genes (Figure 6.22A), with olfactory terms and “G protein-coupled receptor activity” sharing 28 genes of the Olfactory Receptor superfamily (Figure 6.22B). Twelve genes were shared among all four collagen and extracellular matrix terms (10 members of the collagen gene family, and *EMILIN1* and *LUM*; Figure 6.22B).

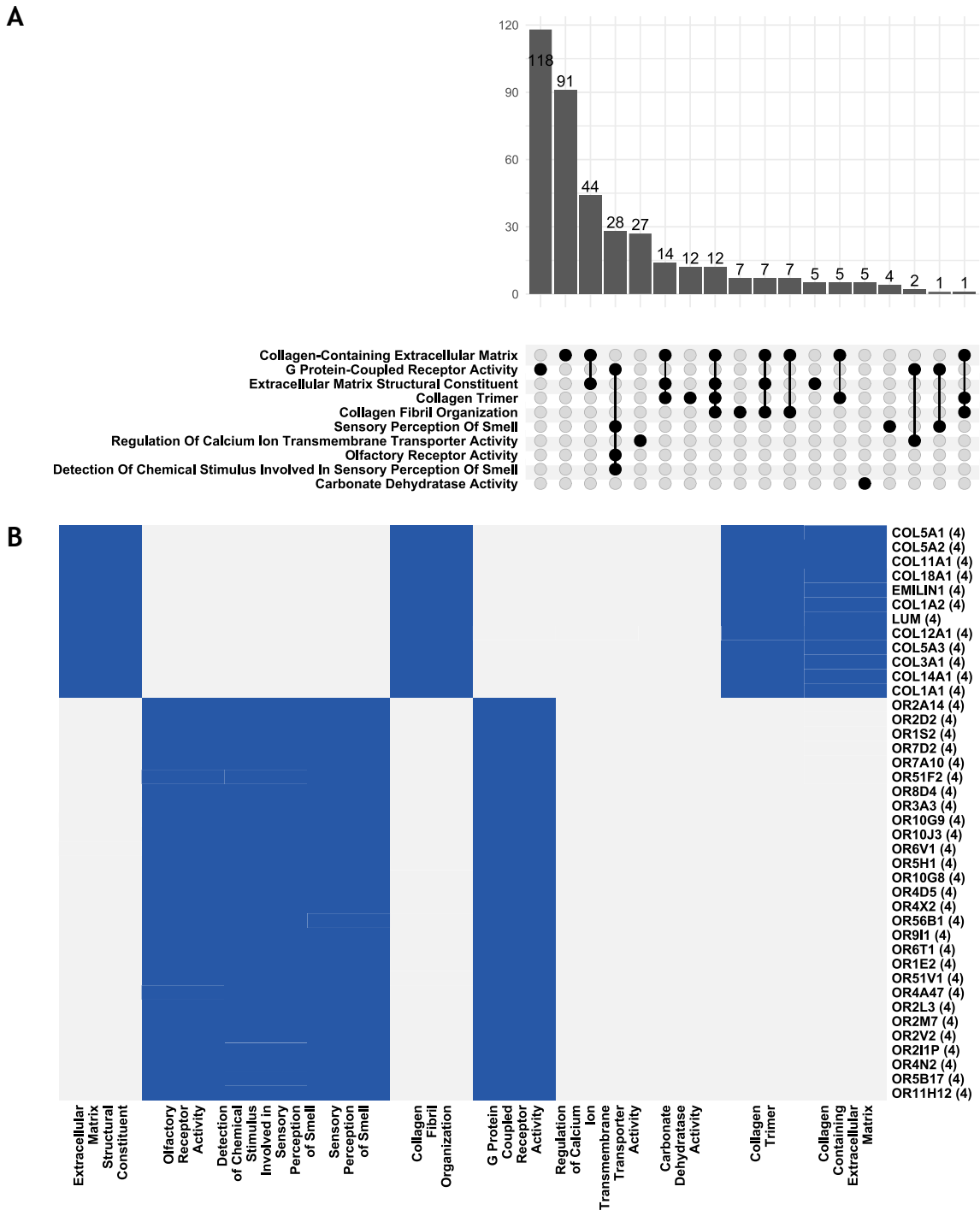


Figure 6.22. Overlapping Leading Edge Genes in Negatively Enriched GO Terms. **[A]** UpSet plot showing number of overlapping genes as bars across intersecting GO terms (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least four GO terms.

GSEA enrichment score plots, that represent the position of enriched genes within GO terms of interest, reveal that the positively enriched terms relating to anti-bacterial functions and cell cycle regulation have genes that are highly ranked (Figure 6.23A-B), while negatively enriched terms like those relating to extra-cellular matrix maintenance and cell membrane receptor activity have low ranking genes (Figure 6.23C-D).

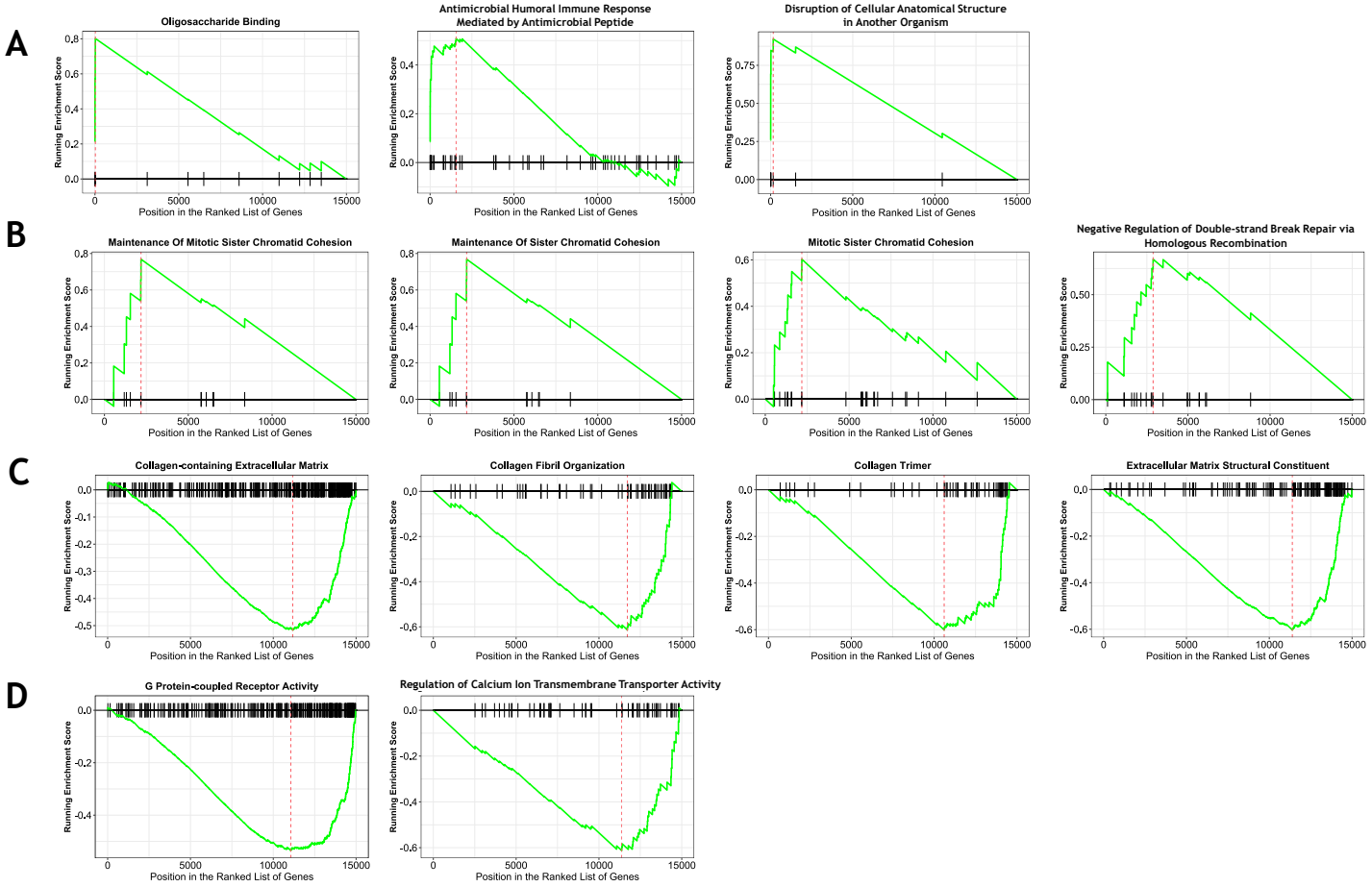


Figure 6.23. GSEA Plots of Selected GO Terms.
[A] Positively enriched GO Terms relating to antimicrobial functions. **[B]** Positively enriched GO terms relating to cell cycle regulation. **[C]** Negatively enriched GO terms relating to collagen functions and extra-cellular matrix maintenance. **[D]** Negatively enriched GO terms relating to cell membrane receptor activity. The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The vertical red line represents the point where gene enrichment is most positive or negative.

6.3.2.2 Molecular Signature Database

The Hallmarks human collections of the MSigDb was then selected for GSEA. When analysing Hallmarks, 15 gene sets had a pAdj<0.05. As shown in the dot plot in Figure 6.24A, five gene sets were activated, “MYC Targets V2”, “MYC Targets V1”, “G2M Checkpoint”, “Mitotic Spindle”, and “E2F Targets”, all of which were categorised as “proliferation”(Liberzon *et al.*, 2015b). In contrast, eight gene sets were suppressed, four of which were categorised as “development”: “Epithelial Mesenchymal Transition”, “Adipogenesis”, “Myogenesis”, and “Pancreas Beta Cells”, two were categorised as “signalling”: “KRAS Signalling Down” and “KRAS Signalling Up”, one was categorised as “cellular component”: “Apical Junction”, and “xenobiotic metabolism” was categorised as “metabolic”.

The positively enriched side of bar plot (Figure 6.24B) shares all but one (“Mitotic Spindle”) activated gene sets observed in the dot plot, while the negatively enriched side and the dot plot shares three gene sets categorised as “development”: “Epithelial Mesenchymal Transition”, “Adipogenesis”, “Myogenesis”, the “cellular component” gene set, “Apical Junction”, and the “metabolic” gene set “Xenobiotic Metabolism”. In addition, another “metabolic” gene set was negatively enriched, "Heme Metabolism", “immune” gene sets: “Coagulation” and “Complement”, one “pathway” gene set: “Hypoxia”, and one “signalling” gene set: “MTORC1 Signalling” (Figure 6.24B).

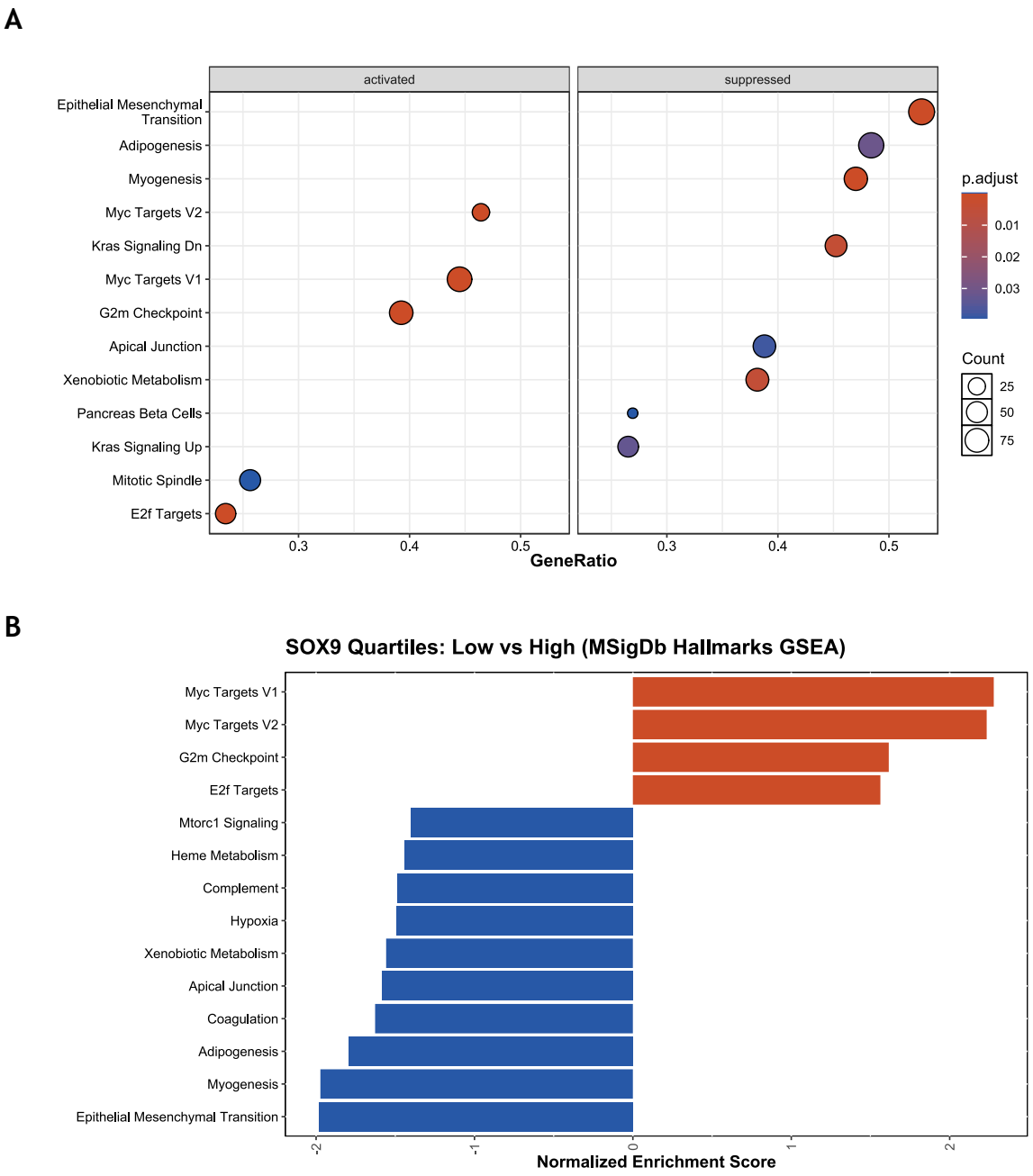


Figure 6.24. MSigDb Hallmarks GSEA Summary.
[A] Dot plot of top 10 activated (left) and supressed (right) Hallmarks gene sets ranked by gene ratio. Dot size refers to gene counts while colour refers to pAdj. **[B]** Bar plot of top 10 positively enriched (orange) and top 10 negatively enriched (blue) Hallmarks gene sets ranked by NES. The High group is compared to the Low group. N=553, lowest quartile n=283, highest quartile n=270.

UpSet plots revealed several gene sets with overlapping genes. All four positively enriched (Figure 6.25A) gene sets which relate to proliferation had three shared genes between them, *CDK4*, *MYC*, *NOLC1* (Figure 6.25B). Interestingly, all but one positively enriched gene set shared the gene *PLK4* (Figure 6.25B).

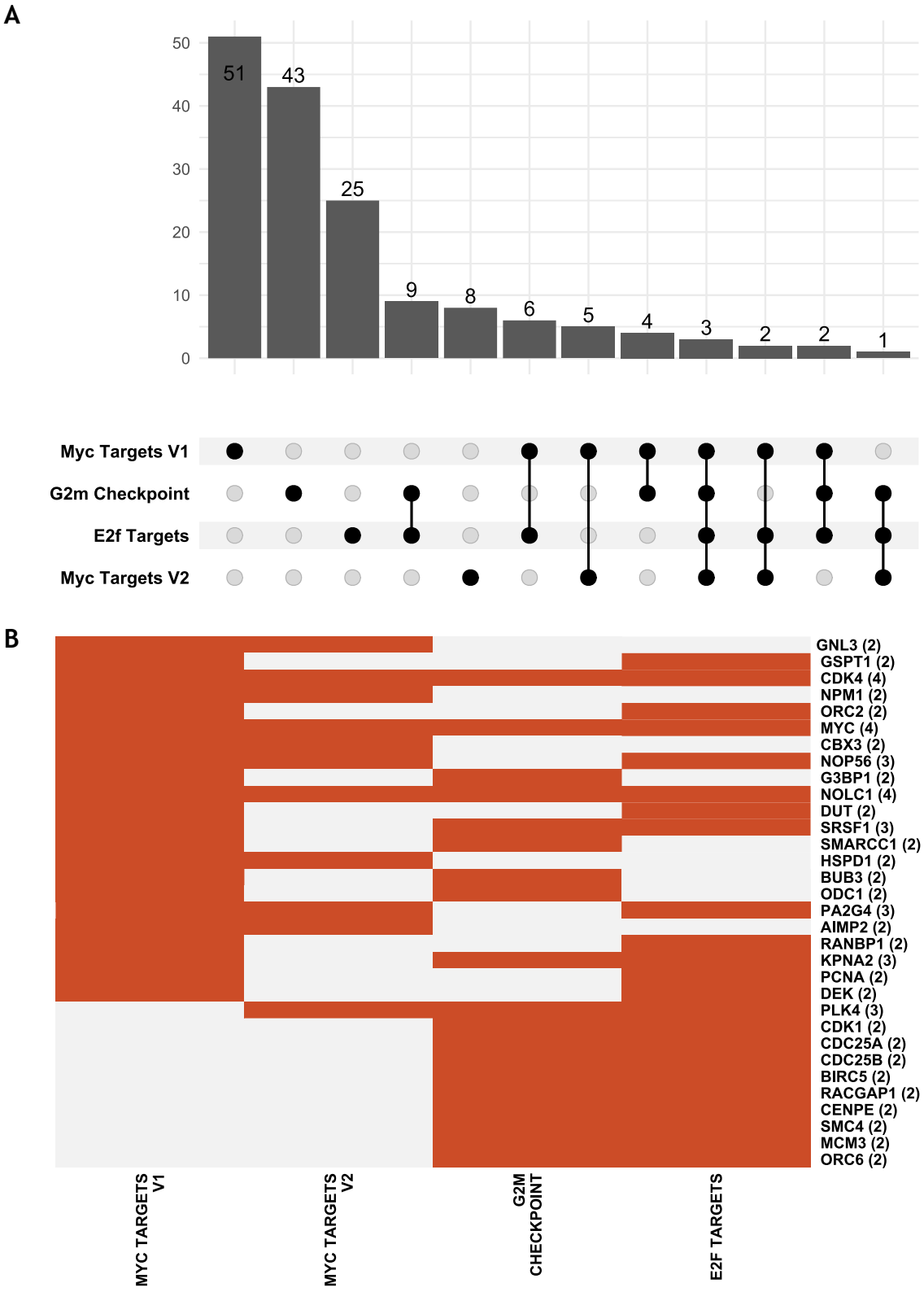


Figure 6.25. Overlapping Leading Edge Genes in Positively Enriched Hallmarks Gene Sets.
[A] UpSet plot showing number of overlapping genes as bars across intersecting Hallmarks gene sets (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between two or more gene sets.

The three negatively enriched “development gene sets” shared only one gene, *MYLK*. The two negatively enriched “metabolic” gene sets shared three genes, *BLVRB, LPIN2,* and *RAP1GAP*. The 2 “immune” gene sets shared 11 genes, *C1QA, C1R, CFH, CSRP1, CTSO, DPP4, GNB2, MMP15, PLAT,* and *SERPINE1*. *SERPINE 1* was shared between “Epithelial Mesenchymal Transition”, “Hypoxia”, “Xenobiotic Metabolism”, “Complement”, and “Coagulation”, while *CD36* was shared between “Adipogenesis”, “Myogenesis”, “Xenobiotic Metabolism”, and “Complement”. This is summarised in Figure 6.26. “Epithelial Mesenchymal Transition” and “Apical Junction” shared four genes, *CDH6, SLIT2, THY1,* and *VCAN*, all four of which have been shown to be involved in cell adhesion in some capacity.

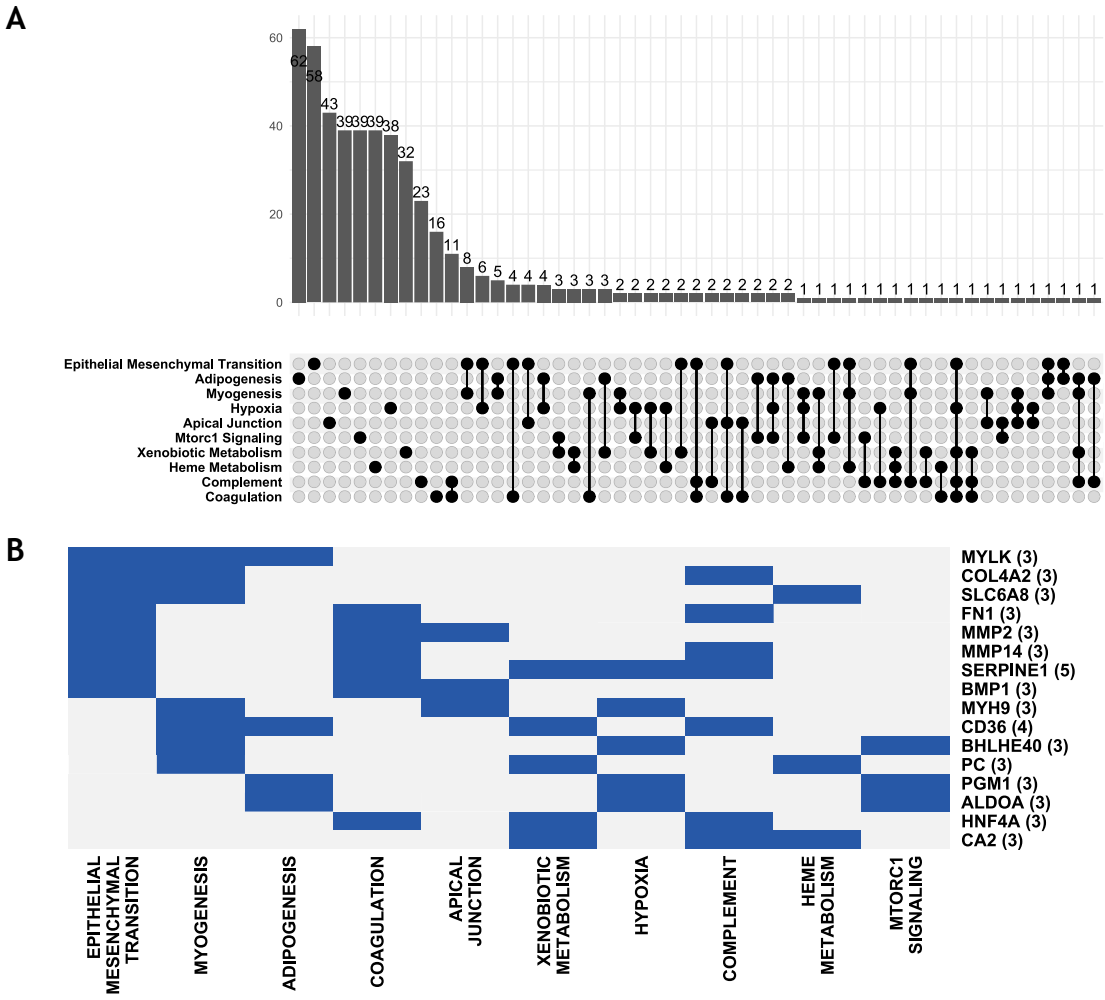


Figure 6.26. Overlapping Leading Edge Genes in Negatively Enriched Hallmarks Gene Sets.
[A] UpSet plot showing number of overlapping genes as bars across Hallmarks gene sets (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least three Hallmarks gene sets.

GSEA enrichment score plots corroborate the enrichment trends of these gene sets, where the positively enriched have high ranked genes (Figure 6.27A), and the negatively enriched have low ranked genes (Figure 6.27B-E).

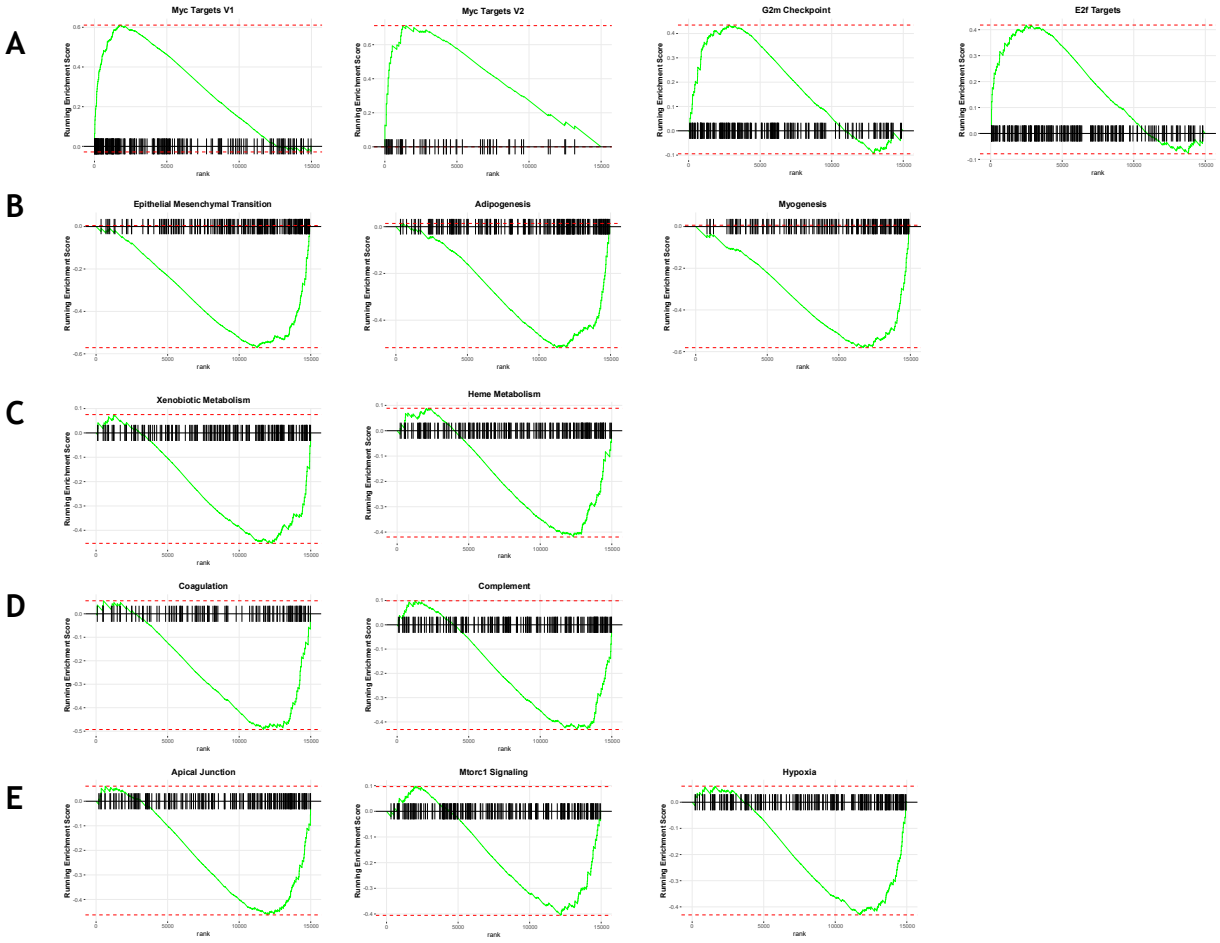


Figure 6.27. GSEA Plots of Hallmarks Gene Sets.
[A] Positively enriched “proliferation” Hallmarks gene sets. **[B]** Negatively enriched “development” Hallmarks gene sets. **[C]** Negatively enriched “metabolism” Hallmarks gene sets. **[D]** Negatively enriched “immune” Hallmarks gene sets. **[E]** Negatively enriched “cellular component”, “signalling”, and “pathway” Hallmarks gene sets, respectively. The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The horizontal red lines represent the point highest or lowest enrichment scores for genes within the set.

6.3.2.3 Kyoto Encyclopaedia of Genes and Genomes

Finally, GSEA was performed using KEGG to study the high-level cellular functions and pathways. The *Homo sapiens* genome was used, and inclusion criteria were set at NES<-1 and >1 and pAdj<0.05.

The dot plot (Figure 6.28A) displays 10 activated pathways with three categorised as “Human Disease – Infectious Disease: Viral”: “Epstein-Barr Virus Infection”, “Human T-cell Leukemia Virus 1 Infection”, and “Human Immunodeficiency Virus 1 Infection”, two categorised as “Organismal Systems – Immune System”: “Neutrophil Extracellular Trap Formation” and “Chemokine Signalling Pathway”, two categorised as “Cellular Processes – Cell Growth and Death”: “Cell Cycle” and “Cellular Senescence”, in addition to “Cytokine-Cytokine Receptor Interaction”, “Viral Carcinogenesis”, “Tuberculosis”. , “Hormone Signalling”, “Focal Adhesion”, “Axon Guidance”, “PI3K-AKT Signalling Pathway”, and “Cytoskeleton in Muscle Cells” were suppressed.

The bar plot (Figure 6.28B) shows the same suppressed pathways to be negatively enriched, and although, 9 out 10 activated pathways were positively enriched, “Viral Carcinogenesis” was lost and “Measles” gained.

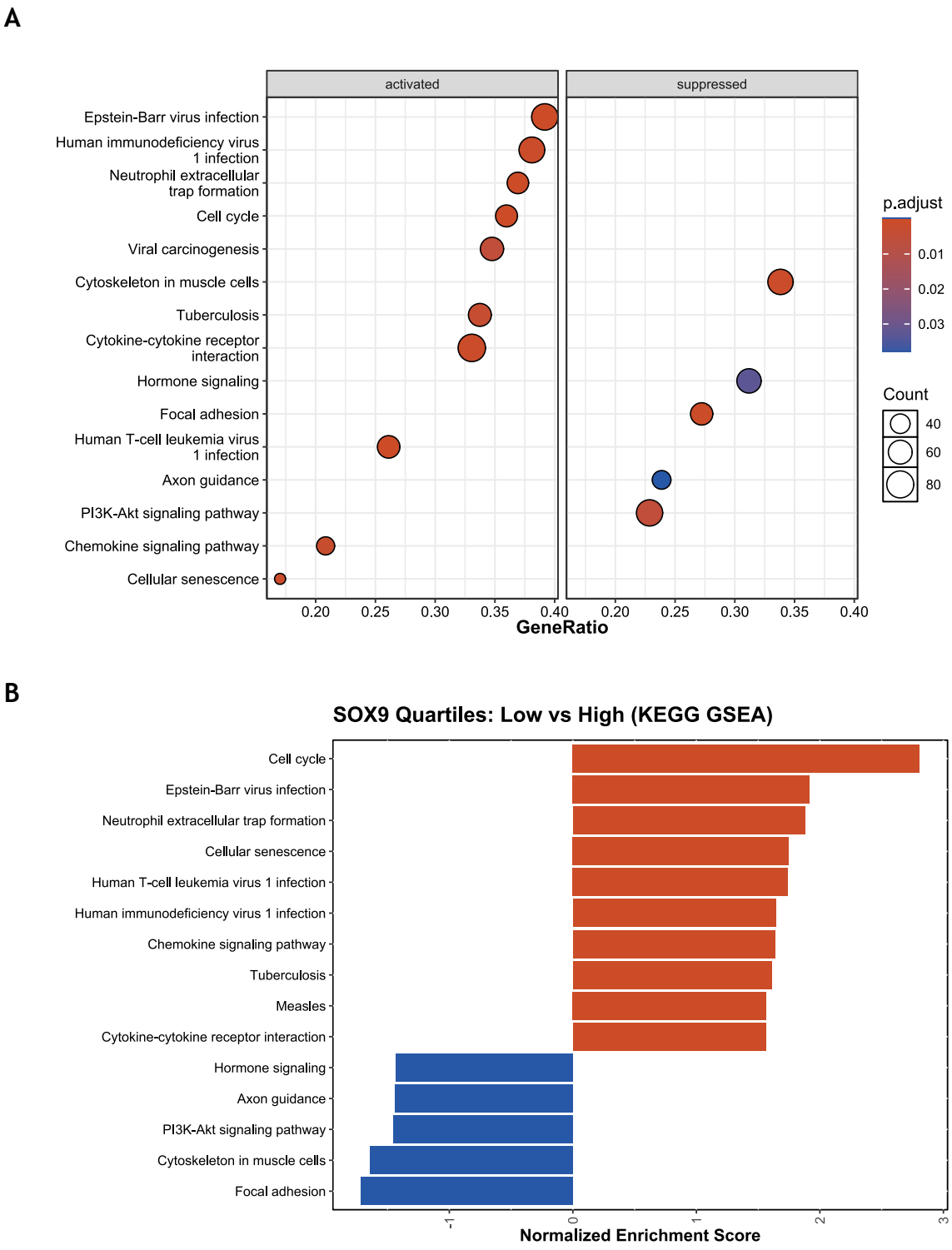


Figure 6.28. KEGG GSEA Summary.
[A] Dot plot of top 10 activated (left) and supressed (right) pathways ranked by gene ratio. Dot size refers to gene counts while colour refers to pAdj. **[B]** Bar plot of top 10 positively enriched (orange) and top 10 negatively enriched (blue) pathways ranked by NES. The High group is compared to the Low group. N=553, lowest quartile n=283, highest quartile n=270.

UpSet plots for the positively enriched pathways (Figure 6.29A) revealed that three genes, *CCNE1*, *CCNE2*, and *CDK4* were all shared by “Epstein-Barr Virus Infection”, “Human T-cell Leukemia Virus 1 Infection”, “Cell Cycle”, “Cellular Senescence”, and “Measles” (Figure 6.29B). *IFNA21* was common between

"Epstein-Barr Virus Infection", "Human Immunodeficiency Virus 1 Infection", "Tuberculosis", "Measles", and "Cytokine-Cytokine Receptor Interaction". *IL6* was shared between "Epstein-Barr Virus Infection", "Human T-cell Leukemia Virus 1 Infection", "Tuberculosis", "Measles", and "Cytokine-Cytokine Receptor Interaction". *TLR2* overlapped between "Epstein-Barr Virus Infection", "Human Immunodeficiency Virus 1 Infection", "Tuberculosis", "Measles", and "Neutrophil Extracellular Trap Formation". *TNF* was shared between "Epstein-Barr Virus Infection", "Human Immunodeficiency Virus 1 Infection", "Tuberculosis", "Human T-cell Leukemia Virus 1 Infection", and "Cytokine-Cytokine Receptor Interaction".

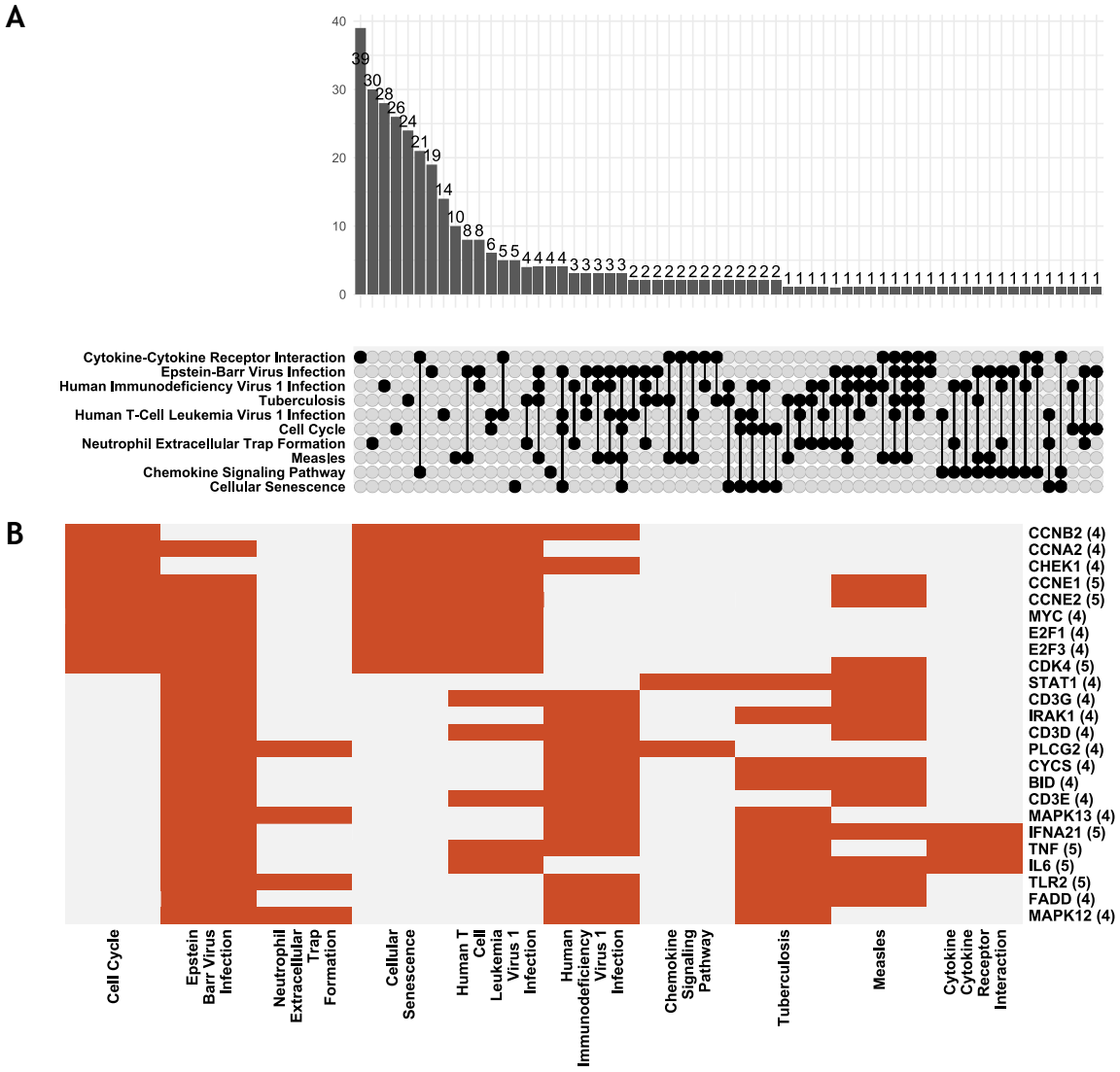


Figure 6.29. Overlapping Leading Edge Genes in Positively Enriched KEGG Pathways. [A] UpSet plot showing number of overlapping genes as bars across intersecting KEGG pathways (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least four KEGG pathways.

Negatively enriched pathways (Figure 6.30A) “Hormone Signalling”, “Axon Guidance”, “PI3K-AKT Signalling Pathway”, and “Focal Adhesion” all shared *PIK3R1*. *ITGAV* was common between “Hormone Signalling”, “Focal Adhesion”, “PI3K-AKT Signalling Pathway”, and “Cytoskeleton in Muscle Cells”, and both *MYL5* and *MYL9* were shared between “Focal Adhesion”, “Axon Guidance”, and “Cytoskeleton in Muscle Cells” (Figure 6.30B).

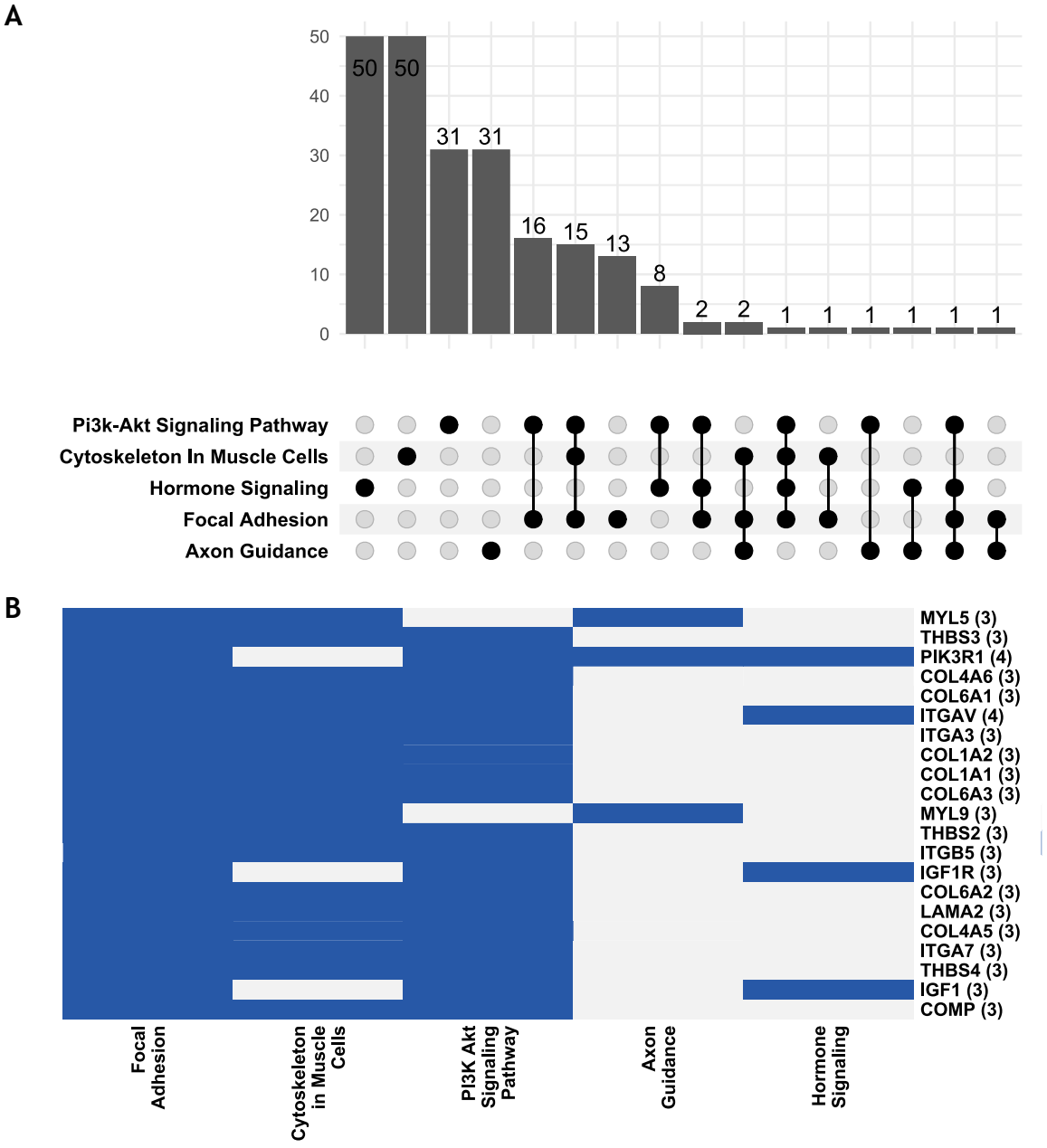


Figure 6.30. Overlapping Leading Edge Genes in Negatively Enriched KEGG Pathways. **[A]** UpSet plot showing number of overlapping genes as bars across KEGG pathways (represented by black circles connected by black lines). **[B]** Heatmap of the genes common between at least three KEGG pathways.

GSEA enrichment score plots corroborate the enrichment trends of these pathways, where the positively enriched have genes that are highly ranked (Figure 6.31A), and the negatively enriched have low ranked genes (Figure 6.31B).

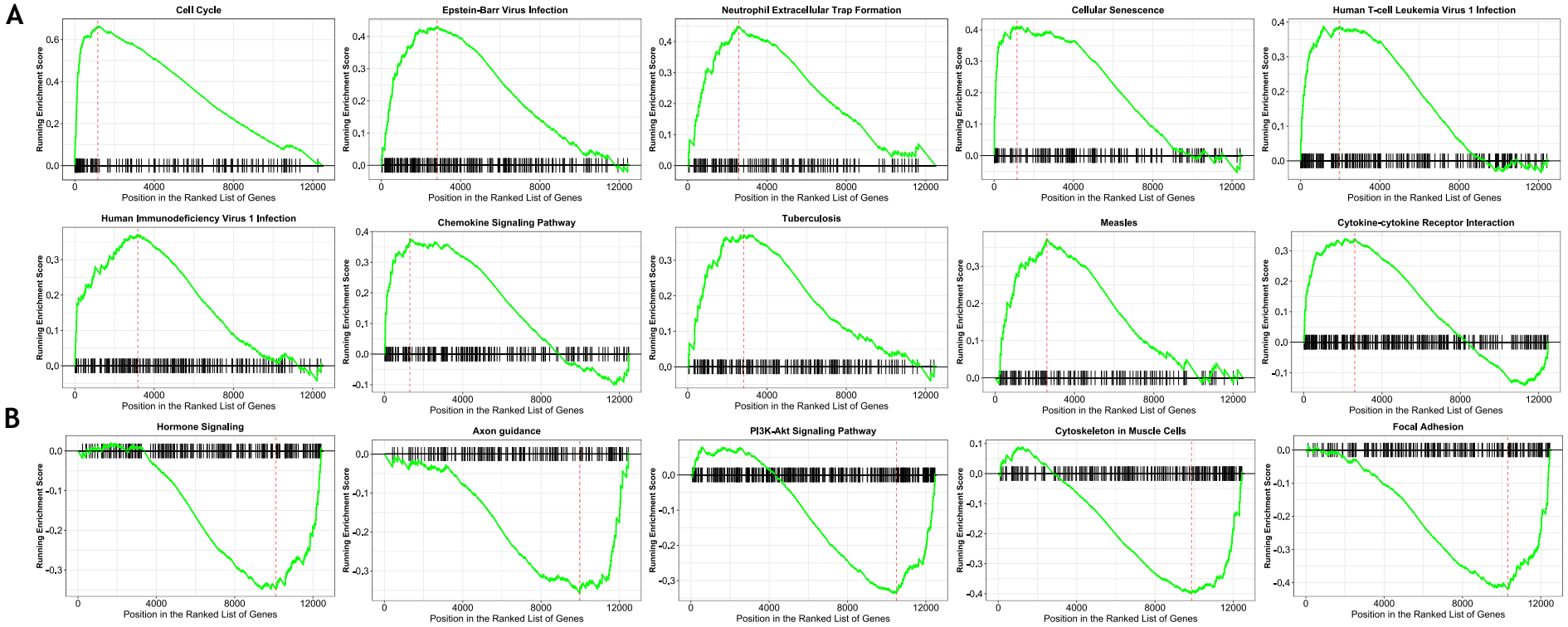


Figure 6.31. GSEA Plots of KEGG Pathways.
[A] Positively enriched KEGG pathways. **[B]** Negatively enriched KEGG pathways. The green curve corresponds to the location of the enriched genes, with positively enriched genes peaking near the start of the ranked gene list, and the negatively enriched ones dipping towards the end. The vertical red line represents the point where gene enrichment is most positive or negative.

6.7 Discussion

This chapter identified cytoplasmic SOX9 protein expression in index polyps to be a predictor of the development of metachronous polyps or CRC, and a suitable addition to current BSG2020 Guidelines, which not only lack a biomarker, but also depend on important yet insufficient polyp properties such as number, size, and dysplasia (Rutter *et al.*, 2020).

The *SOX9* gene was flagged in mutational analysis discussed in Chapter 4, and at the protein level, cytoplasmic expression was validated as significantly associated with metachronous polyps or CRC (Figure 6.10). Previous work on SOX9 is extensive, and its function with regards to tumorigenesis is context-dependent, with many studies suggesting an oncogenic effect based on associations of high SOX9 with invasion and metastasis, advanced T-stage, and involvement in cancer cell stemness (Shen *et al.*, 2015; Carrasco-Garcia *et al.*, 2016; Zhou *et al.*, 2020; Lizárraga-Verdugo *et al.*, 2021). The results presented here are not only evidence of the involvement of SOX9 in the earliest steps of CRC development but also present the possibility of using SOX9 to improve risk stratification of patients undergoing surveillance colonoscopies for detection of metachronous polyps or CRC.

Cytoplasmic SOX9 expression was an independent predictor of metachronous polyp or CRC detection when tested with confounding factors using multivariate Cox regression, and X^2 analysis revealed further associations with the detection of metachronous polyps, which were validated in the test dataset. The addition of SOX9 expression to BSG2020 Guidelines and the generated risk groups, is further evidence of the value of a biomarker in the prediction of metachronous polyps or CRC. The newly formed high-risk group, which includes patients deemed high-risk based on BSG2020 Guidelines and high levels of cytoplasmic SOX9, outperformed current BSG2020 Guidelines alone in both the time to detection of metachronous polyps or CRC (Figure 6.16), and when investigating the percentage of patients who develop metachronous polyps or advanced adenomas or CRC 3 years and 6 years post-index endoscopy.

The significance on SOX9 protein expression on the detection of metachronous polyps, along with the nuance of the effects of SOX9 on the machinations of the

cells in CRC, led to transcriptomic analysis that revealed differentially expressed genes in the SOX9 high group. Interestingly, of the upregulated genes, *TCN1*, a cobalamin transporter, was shown to be an oncogene in CRC (Chu *et al.*, 2014), and a negative prognostic factor associated with poor pathological features (Liu *et al.*, 2020). Furthermore, an upregulation in three members of the *REG* family was observed. Upregulation of *REG* family members has been associated with various stages of cancer progression in many organs, including in CRC (Zheng *et al.*, 2011). Additionally, *STRA6*, which has been shown to initiate the oncogenic JAK2-STAT3 cascade (Berry, Levi and Noy, 2014) was also upregulated in the high SOX9 group.

CRC-inhibitors *AQP8* (Wu *et al.*, 2018) and *CLCA4* (Wei *et al.*, 2020), were downregulated in the high SOX9 group, which is not surprising since this group was associated with metachronous polyp/CRC detection. Interestingly, *OR5B17*, an olfactory receptor and a member of the olfactory receptor (OR) superfamily was downregulated. Similarly, GO GSEA revealed various OR members to be negatively enriched within multiple GO terms relating to olfaction. This is of note, because OR members have been shown to be ectopically expressed in various cancers (Kalra *et al.*, 2020).

The extracellular matrix is a crucial component of the tumour microenvironment, and its remodelling in diseased states supports processes ranging from cell proliferation to invasion (Karamanos, 2019; M. S. Kim *et al.*, 2021). However, it appears that adenomas high in SOX9 might be more resistant to such physical changes. The negative enrichment of collagen genes seen in both the collagen-related GO terms and in the KEGG-based pathway “Focal Adhesion” is supported by the negative enrichment of genes such as *MMP2* and *MMP14* in the “Epithelial Mesenchymal Transition” gene set from MSigDb Hallmarks. While collagen in the extracellular matrix of cancer cells is required to constantly undergo remodelling, to keep up with the demands of a changing environment (Kehlet *et al.*, 2016), metalloproteases like MMP2 aid in the degradation of the extracellular matrix to facilitate various pro-cancer processes like invasion and angiogenesis (Shoari *et al.*, 2024). The opposing results observed in colorectal adenomas high in SOX9 could elucidate some of the changes that take place within cells as they move from normalcy to malignancy. For example, *COL11A1* was found to be overexpressed in CRC but not in

colorectal adenomas or normal colon (Fischer *et al.*, 2001), and *MMP2* levels were found to increase with increasing malignant features from normal tissue to non-dysplastic adenomas, to adenomas with high-grade dysplasia and finally cancer (Murnane *et al.*, 2011). This not only suggests a more change-resistant environment within these adenomas, but also a difference between the benign state and malignancy. Furthermore, as SOX9 is involved in chondrogenesis and has been shown to be highly expressed in *COL11A1*-overexpressing cartilage organoids (Sun *et al.*, 2024), the trend of negative-enrichment of these genes in the high SOX9 group could suggest a tissue-dependent effect, rather than a global one.

The SOX9 high group was positively enriched for GO Terms associated with antimicrobial response, and those terms all shared the same three *REG* family members that were found to be upregulated in the high SOX9 group (Figure 6.21). This family of regenerative proteins was first discovered in regenerating islets cells in the pancreas (Terazono *et al.*, 1988), and the role of *REG1A* has recently been extended to protection of intestinal epithelium during inflammatory injury (H. Mao *et al.*, 2021). Both *REG3A* and *REG1A* were shown to be upregulated in *Fusobacteria*-rich CRC (Lennard, Goosen and Blackburn, 2016), and *REG1A* levels were found to be increased in adenomas (Astrosini *et al.*, 2008). In addition, *DEFA5*, a potent anti-microbial that controls the levels of pathogenic bacteria in the gut, was shared between two positively enriched antimicrobial response GO Terms. While low levels of *DEFA5* have been linked to pro-tumour effects (Nagao *et al.*, 2023), it has been shown to be elevated in some CRC samples in a protective capacity (Zhao *et al.*, 2023). The upregulation of *REG* genes in the high SOX9 group, which was associated with the detection of metachronous polyps or CRC, suggests a link between bacterial-influenced host changes and the development of metachronous polyps or CRC.

A proliferative signature was observed in GSEA from GO, MSigDb, and KEGG. Multiple cell cycle regulation terms were positively enriched in GO, with leading edge genes such as the spindle assembly checkpoint gene *BUB1* and the G_1-S phase regulator and tumour-suppressor *RB1* overlapping with multiple terms. While elevated levels of *BUB1* have been linked to aggressive CRC (Zeng *et al.*, 2024), loss of *RB1* activity has been shown to not only support cancer progression, but also to mediate resistance to chemotherapy (Yu *et al.*, 2024).

As both these genes are leading edge for proliferative terms, *BUB1* could contribute to the increased proliferation associated with adenomas, while *RB1* works to block malignant feature development within the tissue. *MYC* was in positively enriched gene sets and pathways, such as “G2M Checkpoint”, and “Cell Cycle”. *MYC* is a master regulator of cellular processes and a main driver of cell proliferation (Miller *et al.*, 2012) and as previously mentioned, a downstream target of Wnt/β-Catenin signalling. Although it was not in the top 10 most upregulated genes, its presence in positively enriched, proliferative signatures is expected in the group that was most associated with metachronous polyps or CRC detection. *CDK4*, an important regulator of G_1-S phase transition during the cell cycle (Baker and Reddy, 2012), was common between multiple positively enriched gene sets and pathways in the SOX9 high group. In fact, an increase in CDK4 levels was found to be important in the proliferation of cells with dysfunctional *APC* (Cole *et al.*, 2010). As shown in 3.7.2, 87% of sequenced patients had *APC* mutations. Thus, the positive enrichment of *CDK4*-rich pathways follows the expected trend in these adenomas.

Only *SERPINE1* appeared to be shared across multiple negatively enriched gene sets and pathways. While the main function of *SERPINE1* is as a fibrinolysis inhibitor (Iwaki, Urano and Umemura, 2012), it has been linked to CRC (Y. Wang *et al.*, 2023) and elevated levels have been linked to poor survival (Su *et al.*, 2024b). However, no apparent links have been made in the literature regarding its effects on colorectal adenomas. The analysis presented here may further add to the molecular differences between the stages of tissue transition into malignancy previously discussed.

In conclusion, high SOX9 expression in index polyps from bowel screening colonoscopy patients was shown to be clinically valuable for the prediction of metachronous polyps or CRC. The addition of SOX9 expression to current BSG2020 guidelines improved the risk stratification of patients. Furthermore, a difference at the transcriptomics level of the patients with the highest SOX9 expression when compared to those with the lowest SOX9 expression was clear, with an enrichment of several proliferative and defensive biological signatures.

Chapter 7 Discussion

7.1 General Discussion

CRC's incidence and mortality rates have been steadily increasing (Chung *et al.*, 2019; Xi and Xu, 2021) and, with the age of incidence decreasing (Meester *et al.*, 2019), more CRC cases are being recorded every year. Although advances in therapy are achieved regularly, including surgical advancements such as laparoscopic surgery and robotic surgery (Mirkin *et al.*, 2018), repurposed agents for chemotherapy like the hypertension drug sildenafil (Cruz-Burgos *et al.*, 2021), and even targeted therapy for specific CRC mutations such as encorafenib for $BRAF^{V600E}$ metastatic CRC (Tabernero *et al.*, 2021), early detection remains the basis for the best patient outcome.

The long-term goal of CRC screening and early detection is to improve outcomes prevent the increasing incidence and poor outcomes associated with CRC. In fact, bowel screening programmes and surveillance pathways have been used to trial compounds that might have a preventative effect on the growth of metachronous polyps or CRC. For instance, an ongoing clinical trial (COLO-PREVENT) is testing the effects of compounds like aspirin, metformin, and resveratrol on patients with high-risk findings which would currently necessitate ongoing colonoscopic surveillance (Brown and Thomas, 2025; Fan *et al.*, 2025).

However, the coverage rates of national bowel screening programmes vary worldwide (Y. Chen *et al.*, 2025) and, within Europe, the participation rates range from 22% to 88% with variables like first line tests, age, education, and marital status affecting utilisation of available screening services (Ola *et al.*, 2024). Nevertheless, the impact of screening and early detection on CRC mortality rates is significant, with a systematic review on the impact of CRC screening on cancer-specific mortality in Europe finding a 36-41% decrease in CRC mortality in patients who completed a FIT compared to those who did not (Gini *et al.*, 2020).

In the UK, if a polyp is found and resected during a screening colonoscopy, the basic properties of the removed polyp determine risk for the development of a metachronous polyp or CRC. Current BSG2020 Guidelines use the number of polyps at the screening colonoscopy, alongside polyp size and dysplasia as the only determinants of risk (Rutter *et al.*, 2020). This thesis aimed to identify

additional tools to be used alongside current guidelines, since post-polypectomy rates of metachronous lesion development remain at 20-50% (Hao *et al.*, 2020).

As previously described in this thesis, the microbiome has emerged as an enabling characteristic of cancer development (Hanahan, 2022), with a heavy focus on the gut microbiome's role in modulating carcinogenesis, with the role of *F. nucleatum* being of special interest. This is based on foundational work that observed higher levels of *F. nucleatum* in CRC tissue than in surrounding normal tissue, and that patients with higher *F. nucleatum* load were more likely to suffer from lymph node involvement (Castellarin *et al.*, 2012). Similarly, the role of *F. nucleatum* in colonic tumour development was clarified when APC-deficient mice exhibited a higher colonic tumour load upon infection with human isolates of *F. nucleatum* (Kostic *et al.*, 2013). With more abundant *F. nucleatum* being associated with poorly differentiated tumours, low cancer-specific survival, low overall survival, and low disease-free survival (Mima, Nishihara, *et al.*, 2016; Wei *et al.*, 2016), it was logical to investigate *F. nucleatum*'s role in colorectal adenomas.

7.2 *F. nucleatum* and Colorectal Polyps

7.2.1 Detection Techniques

Work on *F. nucleatum* in colorectal polyps is limited and, while it was previously reported that a higher localisation of *F. nucleatum* is present in colonic adenomas when compared to normal tissue (McCoy *et al.*, 2013), there is limited literature on colorectal polyps, with most of the current work being focused on the relationship between *F. nucleatum* and serrated polyps, showing a higher bacterial abundance in the proximal colon (Ito *et al.*, 2015). Furthermore, the methods of detection of *F. nucleatum* were mostly limited to methods like fluorescent *in situ* hybridisation and qPCR.

Chapter 3 of this thesis described a new method for *F. nucleatum* detection in FFPE tissue using bespoke probes for TempO-Seq™. These probes were successfully able to detect *F. nucleatum* RNA in an environment with high levels of human RNA. The advantage of being able to use TempO-Seq™ lies in its ability to detect trace RNA transcripts on FFPE tissue, allowing for retrospective

analysis, as well as in its unique gene region targeting strategy, reducing the cost and processing requirements associated with traditional RNA sequencing techniques (Yeakley *et al.*, 2017).

No associations were made between the patients who were assigned as *F. nucleatum*+ and the detection of metachronous polyps; this could be attributed to the low number of patients in the *F. nucleatum*+ group. However, the association between *F. nucleatum*+ patients and the non-traditional adenomas (i.e., serrated polyps) observed in the INCISE samples reflects previously described literature (Ito *et al.*, 2015). This detection method was accompanied by visualisation of bacteria in FFPE full section tissue from the *F. nucleatum* cohort. RNAscope® is a validated method of detecting target RNA transcripts (Wang *et al.*, 2012) and the *F. nucleatum* probes used here have been previously validated in cancer tissue (Galeano Niño *et al.*, 2022). Although this analysis was only performed on a small subset of the *F. nucleatum* cohort due to limited resources, and the quantity of the bacteria meant the staining was too weak to be digitally quantified, this did confirm the presence of small amounts of *F. nucleatum* on FFPE tissue from bowel screening patients, in agreement with results from TempO-Seq™.

7.2.2 Genomic and Transcriptomic Profiling

When the *F. nucleatum* cohort was split into *F. nucleatum*+ and *F. nucleatum*- patients, it was of interest to explore the genomic and transcriptomic profiles. Interesting phenomena highlighted by transcriptomics analysis like anti-ROS genes (*OXR1*) being downregulated close to the bacteria, and the "Reactive Oxygen Species Pathway" being positively enriched, link *F. nucleatum*-mediated ROS generation to NETs formation not only in CRC (Kong *et al.*, 2023), but also in adenomas. It would be interesting to see if these observations hold in mechanistic studies. *OXR1* levels measured in adenoma PDOs infected with *F. nucleatum*, as well as co-cultures with neutrophils under NETs-forming conditions could allow for a more direct link between the effects of *F. nucleatum* and these transcriptomic trends.

7.2.3 The Immune Response and *F. nucleatum*

However, there was disparity in the immune response between bulk transcriptomics and the spatial transcriptomic profile of areas where *F. nucleatum* was present. While an immune-rich profile was presented through the DEG analysis and the GSEA, this was not reflected in the spatial transcriptomic profile. Various immune pathways were positively enriched in bulk transcriptomics for the *F. nucleatum*+ samples, however, neighbourhood analysis through spatial transcriptomic analysis revealed an absence of adaptive immune-type cells close to the bacteria. *F. nucleatum* was shown to modulate the immune response to cultivate a pro-tumour environment through mechanisms including influencing M2 macrophage polarisation (Hu *et al.*, 2021), which supports carcinogenesis, and NET formation. This all supports the results presented here, as the positively enriched pathways in GSEA all pertain to non-lymphocytic immune cells, and the immune-rich neighbourhoods that are absent near the *F. nucleatum*+ areas are mainly B cells, NK cells, and T cells. Although the number of patients with *F. nucleatum*+ status who had mutational information available was very small, it was interesting to see genetic alterations in both *HLA-A* and *HLA-B*, which are involved in the adaptive immune response, especially in association with CD4+ T cells and CD8+ T cells (Meuer, Schlossman and Reinherz, 1982). *F. nucleatum* has been associated with decreased T cell populations in multiple studies (Mima *et al.*, 2015; Borowsky *et al.*, 2021), and mechanisms of this include the release of inhibitory signals through the binding of bacterial Fap2 to immune cell surface TIGIT (as described in Figure 1.9). This could suggest that *F. nucleatum* effects are directed towards anti-tumour responses only and could imply that these effects begin with genetic alterations. However, whether these mutations are intrinsic to the host or are the effects of the presence of *F. nucleatum* remains to be seen.

The immune response and immune-mediated inflammation are important cancer responses, and immunity accounts for two hallmarks of cancer (Hanahan, 2022). The adaptive immune response was associated with cancer-specific survival irrespective of CRC stage (Thomas *et al.*, 2025). As the effects of *F. nucleatum* on the immune response were different between spatial and bulk transcriptomics and, since the mutational profile of *F. nucleatum*+ patients suggested a genetically altered immune response, the cell subsets that confer

the basic protective effects on tissue against cancer development were of interest.

None of the T cell subsets were correlated with *F. nucleatum*+ status and, although their role in CRC is significant, survival analyses showed no association between them and the detection of metachronous polyps. Despite the number of cases included being relatively small, this does not exclude the possibility that this effect is *F. nucleatum*-mediated. Previous work by Kostic et al. (2013) revealed that T lymphocytic cells that were CD3+, CD4+, and CD8+ were not different between tumour tissue and controls in mice exposed to *F. nucleatum*. Furthermore, T cell subpopulations had an inverse relationship with *F. nucleatum* abundance in human tissue (Mima *et al.*, 2015). These studies used mouse models and human CRC tissue, and since the results in this thesis are from human adenoma tissue, this suggests that this effect is indeed *F. nucleatum*-driven, regardless of tissue type.

Within the immune setting, it was of interest to understand if any immune-mediated effects could influence cellular growth in adenomas. The cytokines used in 2.2.9.4.1 led to an increase in cell viability of FAP PDOs. These cytokines are known to be correlated with local inflammation and, interestingly, levels of IL-6 and TNF-α were correlated with *Fusobacterium* in inflamed tissue within colorectal adenomas (McCoy *et al.*, 2013). Cytokines have been shown to promote tumorigenesis (Borowczak *et al.*, 2022) and the selected cytokines are known to be secreted by lymphocytes and macrophages, which were identified on this tissue using spatial transcriptomics (Yang et al., 2018; Kang et al., 2020; Jang et al., 2021). This study suggests that the immune cells that secrete these cytokines foster a growth-supportive environment for hyperproliferating adenoma cells.

The mutational and transcriptomic profiles combined with the protein level study and backed by the literature, conclude that *F. nucleatum* modulates immunity in colorectal polyps. However, these avenues of investigation did not yield a tool for the detection of metachronous polyps or CRC.

7.3 The Mutational Profile of Colorectal Adenomas

Since the microbiome and the immune response in the *F. nucleatum* cohort did not offer a tool for the prediction of metachronous polyps or CRC, this thesis generated a mutational profile of a larger cohort of INCISE patients to identify possible targets of investigation.

APC and *KRAS* mutations are initiating events in adenoma development (Fearon and Vogelstein, 1990), so it was expected that they were the top two most mutated genes within this cohort. The trend was upheld regardless of polyp location, patient risk for metachronous polyps or CRC, and outcome. Furthermore, *APC* and *KRAS* were co-mutating in the full cohort as well as in the patients who had metachronous polyps or CRC detected. This cements the importance of these two genes in the initiation of adenomas.

This analysis also revealed a highly mutated genome with multiple major pathways affected. The oncogenic pathway most affected by mutations is the WNT/ß-Catenin signalling pathway. The importance of Wnt signalling in CRC and its involvement in cell proliferation signalling through gene transcription are well-established (Mao *et al.*, 2001; Zeng *et al.*, 2007; Shah *et al.*, 2015; Kleeman *et al.*, 2020). Furthermore, activation of Wnt signalling by *F. nucleatum* makes it of special interest (Rubinstein et al., 2019; Li et al., 2020). Two Wnt pathway effectors were in the top 10 most mutated genes: *APC* and *SOX9*. These two genes were significantly co-mutated in left colonic polyps, in high-risk patients, and the co-mutation trend approached significance in patients with metachronous polyps or CRC detected. This suggests that not only do the WNT pathway genes provide favourable candidates for metachronous polyp predictors, but also that the role this pathway plays in CRC is equally as crucial as in adenomas.

Similarly, the RTK-RAS pathway was also highly affected by mutations, with *KRAS* mutations being the second most common among the genetic alterations. Since the discovery of RAS genes and their functions with the MAP/ERK pathway (Scolnick *et al.*, 1973; Kyriakis *et al.*, 1992; Troppmair, Hartkamp and Rapp, 1998), to their role in targeted therapy development(Singhal *et al.*, 2024; Li, Zhao and Li, 2025; Zhu *et al.*, 2025), *KRAS* has been an important point of

research in CRC. Interestingly, *SOX9* co-mutated with *KRAS* in left colonic polyps and in patients stratified as high-risk according to current guidelines, suggesting that SOX9 mutations play an important role in adenomas.

7.4 Wnt Signalling in Colorectal Polyps

Throughout this work, Wnt signalling has been a recurring theme; its established connection with *F. nucleatum*, its support and contribution to CRC development, and the mutations with its associated genes, led to it emerging as a viable target for further investigations.

As shown in Figure 1.10, *F. nucleatum* binds to E-Cadherin (Rubinstein *et al.*, 2013, 2019), leading to increased levels of cytosolic β-Catenin, which is the major factor in Wnt signalling as translocation of β-Catenin to the nucleus and binding to TCF/LEF initiates gene transcription. Chapter 5 explored the value of using E-Cadherin and β-Catenin protein expression levels on tissue from index polyps to predict the development of metachronous polyps or CRC.

As the overarching aim of both this project and the INCISE collaborative, was to identify clinically viable methods of risk stratification, the methods used were selected for their clinical relevance. This study used a large patient cohort, split into a training and test dataset, to build mathematically reproducible models and to adhere to the REMARK criteria (Altman *et al.*, 2012). As per these guidelines, results from the training dataset required to be validated in the test dataset to allow for more confident conclusions to be drawn. While the antibodies used to detect the expression of the proteins of interest were research-grade, their sensitivity and specificity was validated by western blotting of lysates from appropriate cell lines and cell pellet staining of appropriate cell lines.

E-Cadherin was detected in the membrane, which was expected as it is involved in cell-cell adhesion (Maître and Heisenberg, 2013). The expression of E-Cadherin in this compartment was not associated with the detection of metachronous polyps or CRC in either dataset and although low cytoplasmic E-Cadherin expression was associated with the metachronous polyp detection, this was not validated in the test dataset. This suggests that this protein might be

too upstream of the effects induced by Wnt target gene transcription. Within the same pathway, β-Catenin expression was detected in all three cellular compartments, as expected, and while high membranous and low nuclear β-Catenin expression were significantly associated with the detection of metachronous polyps or CRC, these trends were again not validated in the test dataset. This supports the importance of validating such results on a test dataset. The lack of correlation here was interesting, as multiple studies stressed the link of E-Cadherin expression to the prognosis of CRC (Karamitopoulou *et al.*, 2011; Bruun *et al.*, 2021; Lee *et al.*, 2025). Likewise, β-Catenin has been shown to be associated with CRC prognosis (Gao *et al.*, 2014), with a recent meta-analysis revealing that nuclear β-Catenin expression is a possible predictor of survival in CRC (Matly *et al.*, 2021). This divergence from the literature implies that the effects of E-Cadherin and β-Catenin expression on established CRC differ when compared to colorectal adenomas and suggests that the involvement of these proteins is important towards the later stages of disease development.

Regardless of the lack of associations observed with protein expression, the Wnt signalling pathway remained of interest. Hence, FAP PDOs were treated with Wnt pathway stimulating agents to assess the involvement of this pathway in adenoma growth. While there was not a statistically significant difference in the cell viability between treated and untreated PDOs, it was visually discernible that Wnt-stimulated PDOs were larger and more numerous. This was interesting as FAP PDOs have a constitutively active Wnt signalling pathway due to their hallmark *APC* mutation (Leoz *et al.*, 2015), yet proliferation still increased in this saturated model.

7.5 SOX9 Expression Improved Risk Stratification for Post-polypectomy Surveillance

Since E-Cadherin and β-Catenin were not associated with the detection of metachronous polyps or CRC, another target of investigation was sought. Returning to the mutational profile of these adenomas, the mutational pattern of *SOX9* piqued interest, as it was in the top 10 most mutated genes and was co-mutating with *APC* and *KRAS*. Additionally, SOX9 was flagged as a Wnt-associated

protein due to its regulatory effect on the Wnt signalling pathway (Prévostel *et al.*, 2016; Sinha *et al.*, 2021).

SOX9 protein was expressed in the cytoplasm and the nucleus, as expected. Following the stringent criteria of validation in the training and test datasets, high cytoplasmic SOX9 expression was shown to be associated with the detection of metachronous polyps or CRC. SOX9 was also independently predictive of metachronous polyp or CRC detection. When added to current BSG2020 Guidelines, patients who were high-risk and had high cytoplasmic SOX9 expression had a median time to detection of metachronous polyps or CRC almost two years sooner than low-risk patients according to current guidelines.

To our knowledge, this is the first evidence that suggests that, not only can BSG2020 Guidelines for post-polypectomy surveillance be improved upon, but also this improvement can come from the addition of a simple IHC-based biomarker test.

The transcriptomic profile of the patients with the highest SOX9 expression revealed the underlying biological processes that led to its association with the detection of metachronous polyps or CRC. High SOX9 protein expression was associated with metachronous polyps or CRC, suggesting that these patients have a proliferatively driven biology. This was confirmed by GSEA, which identified positive enrichment in mitotic signatures and cell proliferation pathways. SOX9 is a stemness marker that has been shown to aid in CRC progression by promoting stemness and inhibiting differentiation in adenoma cells, as well as being a requirement for CRC cell proliferation (Liang *et al.*, 2022). Increased expression of SOX9 was also revealed to be a product of biallelic inactivation of *APC*, and the subsequent generation of Paneth-like cells (Feng *et al.*, 2013). This, combined with the co-mutation pattern with *APC* observed in Chapter 4**,** supports the idea that SOX9 supports cell growth in adenomas.

However, the most interesting and unexpected result was the anti-bacterial transcriptomic signature exhibited by the high SOX9 group. A study by Liu et al. (2022) suggests that *F. nucleatum*-infected CRC cells gain stemness features by bacterial upregulation of stemness markers, including SOX9. Multiple protective terms, gene sets, and pathways were positively enriched in the high SOX9 group.

Coupled with the upregulation of multiple *REG* gene family members ,which were shown to confer protective effects on epithelial tissue (Lennard, Goosen and Blackburn, 2016; H. Mao *et al.*, 2021), these data suggest that the association of high SOX9 expression with the detection of metachronous polyps might be due to the presence of *F. nucleatum*, and the anti-microbial transcriptomic profile might be a defence mechanism against this bacterially-driven stemness profile.

As a downstream target of Wnt signalling (Blache *et al.*, 2004), SOX9 regulates the Wnt pathway through a physical interaction between the C-terminus of SOX9 and ß-Catenin (Prévostel *et al.*, 2016; Sinha *et al.*, 2021). A SOX9 isoform truncated before the C-terminus (miniSOX9) was identified and shown to stimulate Wnt signalling effects by interacting with ß-Catenin (Abdel-Samad *et al.*, 2011). The pro-growth effects seen in these data could be due to the SOX9 detected being miniSOX9 and not the full-length protein.

7.6 Limitations and Future Work

While the results presented throughout this thesis are novel, this work faced limitations and will require additional studies to further validate the conclusions. All the materials used in this work originated from a single polyp from patients within the Scottish bowel screening program. This could lead to missed associations in patients with multiple polyps, as patient heterogeneity was not captured. This makes any findings presented here lesion-specific (in patients with multiple polyps). Future investigations should be designed to consider the number of polyps present at the initial colonoscopy in order to improve generalisability. There was a trend of underpowered sample pools. This could be seen in the low number of *F. nucleatum*+ patients identified by TempO-Seq™, the number of cases included in the RNAscope® validation, the number of *F. nucleatum*+ cases for which mutational data was available, and in the adaptive immune response study on the *F. nucleatum* cohort. While this is largely a pilot study that was performed using novel techniques, it supports future analysis in a larger cohort. Furthermore, all these tissue-based studies must be repeated and validated in more diverse cohorts in international settings, as this group of patients was localised to a narrow geographical location.

While TempO-Seq™ was validated by RNAscope®, both these techniques are relatively new, and so further validation of the *F. nucleatum*+ and *F. nucleatum*- groups generated by TempO-Seq™ would be required through a more conventional, established method such as qPCR. Furthermore, the external validation of the TempO-Seq™ probes used here is pertinent to prove this method is generalisable and reliable to different patient populations. Additionally, the RNAscope probes used to detect *F. nucleatum* were not validated in-house; and their validation on cancer tissue and not on adenoma was reported in the literature. Infection of adenoma tissue with *F. nucleatum*, followed by fixing and probing the tissue using the RNAscope® probes would further confirm the reliability of the results.

While the focus of this work was *F. nucleatum,* due to its established role in CRC initiation and progression, it would be valuable to examine the effects of other microbiome constituents such as pks+ *E. coli* on adenoma development. Alternatively, a more global approach could be to examine the biofilm on this tissue or use wider profiling techniques such as 16s RNA sequencing and shotgun metagenomic sequencing which would allow the generation of validated metrics such as alpha and beta diversity.

Alongside the *F. nucleatum* survival analysis, this thesis focussed on colorectal adenomas. A comparison of the results observed here, within and outwith the scope of *F. nucleatum*, between adenomas and serrated polyps would be interesting, and would allow for the differentiation between trends intrinsic to the tissue, and trends influenced by factors like the presence of *F. nucleatum*. This is especially the case for transcriptomic profiling, and the adaptive immune response study.

Since the immune response is a major part of CRC research, it is crucial to understand the association between T cell densities and adenoma tissue in a larger cohort, disregarding *F. nucleatum* status. Additionally, a protein level study on other immune cell populations like macrophages and NK cells could offer valuable insights into the relationship between *F. nucleatum* and the innate immune response in colorectal polyps, as well as the general relationship between these cell subsets and colorectal polyps. Additionally, a follow-up *in vitro* study with PDOs co-cultured with various immune cell subpopulations

would allow for a deeper understanding of the effects of inflammation on the growth of non-malignant cells.

As mentioned above, the number of *F. nucleatum*+ cases for which mutational data was available was very low. More *F. nucleatum*+ patients with mutational data would verify whether the unexpected results (like *KRAS* not being among the top 10 most mutated genes in this patient group) are true observations or the result of an underpowered study. It would also be valuable to include patients with serrated polyps as those are likely to confer a different mutational profile. A comparison between the mutational profiles of adenomas and serrated polyps would help to determine if a mutation is crucial for polyp development or if it is more exclusive to a specific pathway of progression.

The *in vitro* studies described in this thesis used FAP PDOs. While this was beneficial as the effects were tested on non-cancer tissue, the constitutive *APC* mutation intrinsic to these PDOs clouded the conclusions drawn from the Wnt stimulation experiment. This study must be replicated on non-FAP non-CRC PDOs to achieve a proper measure of the effect of Wnt stimulation on adenomas. Furthermore, the selected compounds for the study both act to increase ß-Catenin levels by blocking the effects of the ß-Catenin destruction complex. It would be beneficial to visualise the cellular locations of ß-Catenin of stimulated PDOs to confirm it is translocating to the nucleus. Furthermore, it would be imperative to confirm that ß-Catenin associates with TCF/LEF in the nucleus, as well as associating with cell proliferation markers like Ki67. These physical interactions could be visualised through dual or multiplex immunofluorescence studies. Additionally, the effects of the PDO Wnt stimulation on the SOX9-ß-Catenin interaction would be interesting to observe, as well as the visualisation of whether SOX9 co-localises with proliferation markers like Ki67 using dual immunofluorescence.

Since SOX9 is protective against uncontrolled tissue growth but was associated with the detection of metachronous polyps or CRC, it would be imperative to identify the SOX9 isoform present within these adenoma tissue samples. Polyp tissue should be stained with a SOX9 antibody that binds to the C-terminus (unlike the one used in this study which binds to the N-terminus, thus detecting all isoforms with the N-terminus intact), to reveal if the associations observed

with the detection of metachronous polyps or CRC are isoform driven. Although the addition of SOX9 to current BSG2020 Guidelines led to a more accurate low- and high-risk stratification, it did generate an “intermediate-risk” group which will require further stratification, and methods to do this could involve further IHC-based markers. The tissue-based studies within this thesis were all done on TMAs, and while an effort was made to ensure that the TMAs reflect the full tissue section (by including cores from 4 different locations), it would be imperative to verify these findings in whole tissue. Furthermore, these findings need to be verified on multiple large international cohorts.

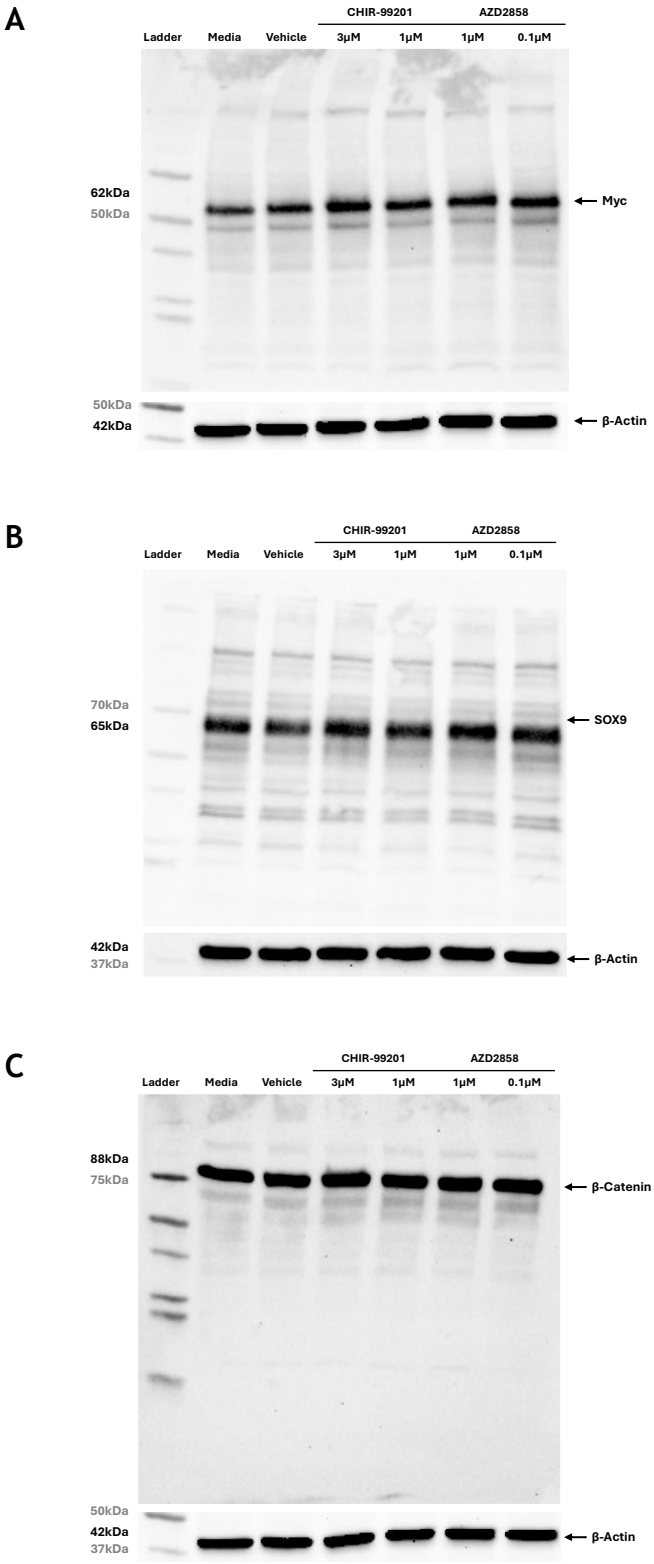
Should these results be validated, feasibility analysis on the clinical integration of SOX9 as a predictive biomarker will be required, in the context of required resources and technical expertise. Furthermore, a consensus will need to be reached between clinicians, scientists, regulators and patients in relation to clinically acceptable rates of false-positive and false negative results from biomarker testing within the post-polypectomy surveillance setting. Additionally, mechanistic studies on adenomas, such as the generation of SOX9 knock-outs, will be required to confirm that the effects seen are directly due to SOX9.

7.7 Conclusion

In conclusion, this work identified a novel technique for detecting *F. nucleatum* in FFPE tissue. Furthermore, this study is the first to describe a difference between the full polyp transcriptome and the spatial transcriptome between polyps where *F. nucleatum* was present and polyps where *F. nucleatum* was absent. However, the presence of *F. nucleatum* within adenomas removed at bowel screening polypectomy was not associated with metachronous lesion risk. While the adaptive immune response in relation to *F. nucleatum* was not significant, this study demonstrated that immune-related cytokines can influence adenoma growth in *in vitro* 3D models. This work investigated a large cohort with a mutational profile that revealed multiple targets for investigation, as well as repeating trends. Through this mutational profile, Wnt signalling was identified as a possible influencer of adenoma growth and, while E-Cadherin and ß-Catenin protein expression were not significantly associated with the detection of metachronous polyps or CRC, blocking the ß-Catenin destruction complex *in vitro* led to a visible increase in cell viability regardless of the

constitutively activated Wnt pathway that characterised the PDOs used. Finally, the stemness marker SOX9 achieved the overarching aim of this study; the addition of cytoplasmic SOX9 protein expression via IHC to the current BSG2020 Guidelines significantly improved patient risk stratification, further discriminating between high-risk and low-risk patients. Although these results are promising, it will be worthwhile to expand this work to include other stemness biomarkers, to further improve risk stratification. If validated in external cohorts, the adoption of a simple IHC marker into the stratification guidelines has the potential to reduce patient risk while improving resource management.

Appendices



Supplementary Figure 1. Full Membranes of Protein Quantification of Wnt Signalling Proteins and Targets Downstream of Stimulatory Effects Presented in Figure 5.22. [A] Myc, **[B]** SOX9, and **[C]** β-Catenin.

List of References

Abdel-Samad, R., Zalzali, H., Rammah, C., Giraud, J., Naudin, C., Dupasquier, S., Poulat, F., Boizet-Bonhoure, B., Lumbroso, S., Mouzat, K., Bonnans, C., Pignodel, C., Raynaud, P., Fort, P., Quittau-Prévostel, C. and Blache, P. (2011) "MiniSOX9, a dominant-negative variant in colon cancer cells," *Oncogene*, 30(22), pp. 2493–2503.

Abed, J., Emgård, J.E.M., Zamir, G., Faroja, M., Almogy, G., Grenov, A., Sol, A., Naor, R., Pikarsky, E., Atlan, K.A., Mellul, A., Chaushu, S., Manson, A.L., Earl, A.M., Ou, N., Brennan, C.A., Garrett, W.S. and Bachrach, G. (2016) "Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc," *Cell host & microbe*, 20(2), p. 215.

Abed, J., Maalouf, N., Manson, A.L., Earl, A.M., Parhi, L., Emgård, J.E.M., Klutstein, M., Tayeb, S., Almogy, G., Atlan, K.A., Chaushu, S., Israeli, E., Mandelboim, O., Garrett, W.S. and Bachrach, G. (2020) "Colon Cancer-Associated Fusobacterium nucleatum May Originate From the Oral Cavity and Reach Colon Tumors via the Circulatory System," *Frontiers in Cellular and Infection Microbiology*, 10, p. 400.

Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., Kirfel, J., Tafazzoli, A., Horpaopan, S., Uhlhaas, S., Stienen, D., Friedrichs, N., Altmüller, J., Laner, A., Holzapfel, S., Peters, S., Kayser, K., Thiele, H., Holinski-Feder, E., Marra, G., Kristiansen, G., Nöthen, M.M., Büttner, R., Möslein, G., Betz, R.C., Brieger, A., Lifton, R.P. and Aretz, S. (2016) "Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis," *American Journal of Human Genetics*, 99(2), p. 337.

Aghabozorgi, A.S., Bahreyni, A., Soleimani, A., Bahrami, A., Khazaei, M., Ferns, G.A., Avan, A. and Hassanian, S.M. (2019) "Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives," *Biochimie*, 157, pp. 64–71.

Agudo, J., Park, E.S., Rose, S.A., Alibo, E., Sweeney, R., Dhainaut, M., Kobayashi, K.S., Sachidanandam, R., Baccarini, A., Merad, M. and Brown, B.D. (2018) "Quiescent Tissue Stem Cells Evade Immune Surveillance," *Immunity*, 48(2), pp. 271-285.e5.

Akiyama, H., Lyons, J.P., Mori-Akiyama, Y., Yang, X., Zhang, R., Zhang, Z., Deng, J.M., Taketo, M.M., Nakamura, T., Behringer, R.R., McCrea, P.D. and Crombrugghe, B. de (2004) "Interactions between Sox9 and B-catenin control chondrocyte differentiation," *Genes & Development*, 18, pp. 1072–1087.

Allison, J.E., Feldman, R. and Tekawa, I.S. (1990) “Hemoccult Screening in Detecting Colorectal Neoplasm: Sensitivity, Specificity, and Predictive Value,” *Annals of Internal Medicine*, 112(5), pp. 328–333.

Altman, D.G., McShane, L.M., Sauerbrei, W. and Taube, S.E. (2012) “Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration,” *BMC Medicine*, 10, p. 51.

Anand, U., Dey, A., Chandel, A.K.S., Sanyal, R., Mishra, A., Pandey, D.K., De Falco, V., Upadhyay, A., Kandimalla, R., Chaudhary, A., Dhanjal, J.K., Dewanjee, S., Vallamkondu, J. and Pérez de la Lastra, J.M. (2023) “Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics,” *Genes & Diseases*, 10(4), pp. 1367–1401.

Aptsiauri, N., Ruiz-Cabello, F. and Garrido, F. (2018) “The transition from HLA-I positive to HLA-I negative primary tumors: the road to escape from T-cell responses,” *Current Opinion in Immunology*, 51, pp. 123–132.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., Butte, A.J. and Bhattacharya, M. (2019) “Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage,” *Nature Immunology*, 20(2), pp. 163–172.

Aronson, M., Colas, C., Shuen, A., Hampel, H., Foulkes, W.D., Feldman, H.B., Goldberg, Y., Muleris, M., Schneider, K.W., McGee, R.B., Jasperson, K., Rangaswami, A., Brugieres, L. and Tabori, U. (2022) “Diagnostic criteria for constitutional mismatch repair deficiency (CMMRD): recommendations from the international consensus working group,” *Journal of Medical Genetics*, 59(4), pp. 318–327.

Arredondo, J., Pastor, C., Baixauli, J., Rodríguez, J., González, I., Vigil, C., Chopitea, A. and Hernández-Lizoáin, J.L. (2013) “Preliminary outcome of a treatment strategy based on perioperative chemotherapy and surgery in patients with locally advanced colon cancer,” *Colorectal Disease*, 15(5), pp. 552–557.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000) “Gene ontology: Tool for the unification of biology,” *Nature Genetics*, 25(1), pp. 25–29.

Ashktorab, H., Brim, H., Hassan, S., Nouraie, M., Gebreselassie, A., Laiyemo, A.O., Kibreab, A., Aduli, F., Latella, G., Brant, S.R., Sherif, Z. and Habtezion, A. (2020) “Inflammatory polyps occur more frequently in inflammatory bowel disease than other colitis patients,” *BMC Gastroenterology*, 20(1), pp. 1–6.

Astrosini, C., Roeefzaad, C., Dai, Y.Y., Dieckgraefe, B.K., Jöns, T. and Kemmner, W. (2008) “REG1A expression is a prognostic marker in colorectal cancer and associated with peritoneal carcinomatosis,” *International Journal of Cancer*, 123(2), pp. 409–413.

Atkin, W., Wooldrage, K., Brenner, A., Martin, J., Shah, U., Perera, S., Lucas, F., Brown, J.P., Kralj-Hans, I., Greliak, P., Pack, K., Wood, J., Thomson, A., Veitch, A., Duffy, S.W. and Cross, A.J. (2017) “Adenoma surveillance and colorectal cancer incidence: a retrospective, multicentre, cohort study,” *The Lancet. Oncology*, 18(6), p. 823.

Atkin, W.S. and Saunders, B.P. (2002) “Surveillance guidelines after removal of colorectal adenomatous polyps,” *Gut*, 51(Suppl 5), p. v6.

Azzouz, D., Khan, M.A. and Palaniyar, N. (2021) “ROS induces NETosis by oxidizing DNA and initiating DNA repair,” *Cell Death Discovery*, 7(1), pp. 1–10.

Azzouz, L.L. and Sharma, S. (2023) “Physiology, Large Intestine,” *StatPearls* [Preprint].

Badia-I-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C.H., Ramirez Flores, R.O. and Saez-Rodriguez, J. (2022) “decoupleR: ensemble of computational methods to infer biological activities from omics data,” *Bioinformatics Advances*, 2(1).

Baker, S.J. and Reddy, E.P. (2012) “CDK4: A Key Player in the Cell Cycle, Development, and Cancer,” *Genes & Cancer*, 3(11–12), p. 658.

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., James, J.A., Salto-Tellez, M. and Hamilton, P.W. (2017) “QuPath: Open source software for digital pathology image analysis,” *Scientific Reports*, 7(1), p. 16878.

Barker, N., Van Es, J.H., Kuipers, J., Kujala, P., Van Den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J. and Clevers, H. (2007) “Identification of stem cells in small intestine and colon by marker gene Lgr5,” *Nature*, 449(7165), pp. 1003–1007.

Benedix, F., Kube, R., Meyer, F., Schmidt, U., Gastinger, I. and Lippert, H. (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 and Rectum*, 53(1), pp. 57–64.

Berg, K.C.G., Eide, P.W., Eilertsen, I.A., Johannessen, B., Bruun, J., Danielsen, S.A., Bjørnslett, M., Meza-Zepeda, L.A., Eknæs, M., Lind, G.E., Myklebost, O., Skotheim, R.I., Sveen, A. and Lothe, R.A. (2017) “Multi-omics of 34 colorectal

cancer cell lines - a resource for biomedical studies," *Molecular Cancer*, 16(1), pp. 1–16.

Berry, D.C., Levi, L. and Noy, N. (2014) "Holo-retinol-binding protein and its receptor STRA6 drive oncogenic transformation," *Cancer research*, 74(21), p. 6341.

Bjerre, R.D., Holm, J.B., Palleja, A., Sølberg, J., Skov, L. and Johansen, J.D. (2021) "Skin dysbiosis in the microbiome in atopic dermatitis is site-specific and involves bacteria, fungus and virus," *BMC Microbiology*, 21(1), p. 256.

Blache, P., Van De Wetering, M., Duluc, I., Domon, C., Berta, P., Freund, J.N., Clevers, H. and Jay, P. (2004) "SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes," *The Journal of Cell Biology*, 166(1), p. 37.

Borowczak, J., Szczerbowski, K., Maniewski, M., Kowalewski, A., Janiczek-Polewska, M., Szylberg, A., Marszałek, A. and Szylberg, Ł. (2022) "The Role of Inflammatory Cytokines in the Pathogenesis of Colorectal Carcinoma—Recent Findings and Review," *Biomedicines*, 10(7), p. 1670.

Borowsky, J., Haruki, K., Lau, M.C., Costa, A.D., Väyrynen, J.P., Ugai, T., Arima, K., Silva, A. da, Felt, K.D., Zhao, M., Gurjao, C., Twombly, T.S., Fujiyoshi, K., Väyrynen, S.A., Hamada, T., Mima, K., Bullman, S., Harrison, T.A., Phipps, A.I., Peters, U., Ng, K., Meyerhardt, J.A., Song, M., Giovannucci, E.L., Wu, K., Zhang, X., Freeman, G.J., Huttenhower, C., Garrett, W.S., Chan, A.T., Leggett, B.A., Whitehall, V.L.J., Walker, N., Brown, I., Bettington, M., Nishihara, R., Fuchs, C.S., Lennerz, J.K., Giannakis, M., Nowak, J.A. and Ogino, S. (2021) "Association of Fusobacterium nucleatum with Specific T-cell Subsets in the Colorectal Carcinoma Microenvironment," *Clinical Cancer Research*, 27(10), pp. 2816–2826.

Bowel cancer incidence statistics | Cancer Research UK (no date). Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#heading-One.

Bowel Cancer Screening | Cancer Research UK (2025). Available at: https://www.cancerresearchuk.org/health-professional/cancer-screening/bowel-cancer-screening.

Bowel cancer statistics | Cancer Research UK (no date). Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I. and Jemal, A. (2024) "Global cancer statistics 2022: GLOBOCAN estimates of

incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, 74(3), pp. 229–263.

Bretthauer, M., Løberg, M., Wieszczy, P., Kalager, M., Emilsson, L., Garborg, K., Rupinski, M., Dekker, E., Spaander, M., Bugajski, M., Holme, Ø., Zauber, A.G., Pilonis, N.D., Mroz, A., Kuipers, E.J., Shi, J., Hernán, M.A., Adami, H.-O., Regula, J., Hoff, G. and Kaminski, M.F. (2022) "Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death," *New England Journal of Medicine*, 387(17), pp. 1547–1556.

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y. and Zychlinsky, A. (2004) "Neutrophil Extracellular Traps Kill Bacteria," *Science*, 303(5663), pp. 1532–1535.

Brown, K. and Thomas, A. (2025) "COLO-PREVENT – do simple drugs (aspirin or aspirin plus metformin) or food supplements (resveratrol) reduce the occurrence of bowel polyps (small growths on the bowel lining), which in turn reduce bowel cancer risk?," *http://isrctn.com/* [Preprint].

de Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T.N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., Abeshouse, A., Baiceanu, D., Ciftci, E., Dogrusoz, U., Dufilie, A., Erkoc, Z., Lara, E.G., Fu, Z., Gross, B., Haynes, C., Heath, A., Higgins, D., Jagannathan, P., Kalletla, K., Kumari, P., Lindsay, J., *et* al. (2023) "Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal," *Cancer Research*, 83(23), p. 3861.

Bruneau, E and Bruneau, Eve (2017) "Basic Anatomy and Physiology of the Gastrointestinal Tract," *Passing the Certified Bariatric Nurses Exam*, pp. 19–25.

Brüning-richardson, A., Shaw, G.C., Tams, D., Brend, T., Sanganee, H., Barry, S.T., Hamm, G., Goodwin, R.J.A., Swales, J.G., King, H., Steele, L., Morton, R., Widyadari, A., Ward, T.A., Esteves, F., Boissinot, M., Droop, A., Lawler, S.E. and Short, S.C. (2021) "GSK-3 Inhibition Is Cytotoxic in Glioma Stem Cells through Centrosome Destabilization and Enhances the Effect of Radiotherapy in Orthotopic Models," *Cancers*, 13(23), p. 5939.

Bruun, J., Eide, P.W., Bergsland, C.H., Bruck, O., Svindland, A., Arjama, M., Välimäki, K., Bjørnslett, M., Guren, M.G., Kallioniemi, O., Nesbakken, A., Lothe, R.A. and Pellinen, T. (2021) "E-cadherin is a robust prognostic biomarker in colorectal cancer and low expression is associated with sensitivity to inhibitors of topoisomerase, aurora, and HSP90 in preclinical models," *Molecular Oncology*, 16(12), p. 2312.

Bueno-Fortes, S., Muenzner, J.K., Berral-Gonzalez, A., Hampel, C., Lindner, P., Berninger, A., Huebner, K., Kunze, P., Bäuerle, T., Erlenbach-Wuensch, K.,

Sánchez-Santos, J.M., Hartmann, A., De Las Rivas, J. and Schneider-Stock, R. (2022) “A gene signature derived from the loss of cdkn1a (P21) is associated with CMS4 colorectal cancer,” *Cancers*, 14(1), p. 136.

Bullman, S., Pedamallu, C.S., Sicinska, E., Clancy, T.E., Zhang, X., Cai, D., Neuberg, D., Huang, K., Guevara, F., Nelson, T., Chipashvili, O., Hagan, T., Walker, M., Ramachandran, A., Diosdado, B., Serna, G., Mulet, N., Landolfi, S., Cajal, S.R. y, Fasani, R., Aguirre, A.J., Ng, K., Élez, E., Ogino, S., Tabernero, J., Fuchs, C.S., Hahn, W.C., Nuciforo, P. and Meyerson, M. (2017) “Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer,” *Science*, 358(6369), pp. 1443–1448.

Burnett-Hartman, A.N., Lee, J.K., Demb, J. and Gupta, S. (2021) “An update on the epidemiology, molecular characterization, diagnosis, and screening strategies for early-onset colorectal cancer,” *Gastroenterology*, 160(4), p. 1041.

Cairns, S.R., Scholefield, J.H., Steele, R.J., Dunlop, M.G., Thomas, H.J.W., Evans, G.D., Eaden, J.A., Rutter, M.D., Atkin, W.P., Saunders, B.P., Lucassen, A., Jenkins, P., Fairclough, P.D. and Woodhouse, C.R.J. (2010) “Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002),” *Gut*, 59(5), pp. 666–689.

Campbell, B.B., Light, N., Fabrizio, D., Zatzman, M., Fuligni, F., de Borja, R., Davidson, S., Edwards, M., Elvin, J.A., Hodel, K.P., Zahurancik, W.J., Suo, Z., Lipman, T., Wimmer, K., Kratz, C.P., Bowers, D.C., Laetsch, T.W., Dunn, G.P., Johanns, T.M., Grimmer, M.R., Smirnov, I. V., Larouche, V., Samuel, D., Bronsema, A., Osborn, M., Stearns, D., Raman, P., Cole, K.A., Storm, P.B., Yalon, M., Opocher, E., Mason, G., Thomas, G.A., Sabel, M., George, B., Ziegler, D.S., Lindhorst, S., Issai, V.M., Constantini, S., Toledano, H., Elhasid, R., Farah, R., Dvir, R., Dirks, P., Huang, A., Galati, M.A., Chung, J., Ramaswamy, V., Irwin, M.S., Aronson, M., Durno, C., Taylor, M.D., Rechavi, G., Maris, J.M., Bouffet, E., Hawkins, C., Costello, J.F., Meyn, M.S., Pursell, Z.F., Malkin, D., Tabori, U. and Shlien, A. (2017) “Comprehensive Analysis of Hypermutation in Human Cancer,” *Cell*, 171(5), p. 1042.

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., *et* al. (2020) “Pan-cancer analysis of whole genomes,” *Nature 2020 578:7793*, 578(7793), pp. 82–93.

Candel, I., Wetwittayakhlang, P., Bessissow, T. and Lakatos, P.L. (2025) “The Importance of Post-Inflammatory Polyps (PIPs) in Colorectal Cancer Surveillance in Inflammatory Bowel Diseases,” *Journal of Clinical Medicine*, 14(2), p. 333.

Capuano, E. (2017) “The behavior of dietary fiber in the gastrointestinal tract determines its physiological effect,” *Critical Reviews in Food Science and Nutrition*, 57(16), pp. 3543–3564.

Carethers, J.M. and Jung, B.H. (2015) “Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer,” *Gastroenterology*, 149(5), p. 1177.

Carr, P.R., Alwers, E., Bienert, S., Weberpals, J., Kloor, M., Brenner, H. and Hoffmeister, M. (2018) “Lifestyle factors and risk of sporadic colorectal cancer by microsatellite instability status: a systematic review and meta-analyses,” *Annals of Oncology*, 29(4), pp. 825–834.

Carrasco-Garcia, E., Lopez, L., Aldaz, P., Arevalo, S., Aldaregia, J., Eganã, L., Bujanda, L., Cheung, M., Sampron, N., Garcia, I. and Matheu, A. (2016) “SOX9-regulated cell plasticity in colorectal metastasis is attenuated by rapamycin,” *Scientific Reports*, 6, p. 32350.

Carroll, M.R.R., Seaman, H.E. and Halloran, S.P. (2014) “Tests and investigations for colorectal cancer screening,” *Clinical Biochemistry*, 47(10–11), pp. 921–939.

Castellarin, M., Warren, R.L., Freeman, J.D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R.A. and Holt, R.A. (2012) “Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma,” *Genome Research*, 22(2), p. 299.

Catlow, J., Beaton, D., Beintaris, I., Shaw, T., Broughton, R., Healey, C., Penman, I., Coleman, M. and Rutter, M. (2021) “JAG/BSG national survey of UK endoscopy services: impact of the COVID-19 pandemic and early restoration of endoscopy services,” *Frontline Gastroenterology*, 12(4), pp. 272–278.

Cauchin, E., Touchefeu, Y. and Matysiak-Budnik, T. (2015) “Hamartomatous Tumors in the Gastrointestinal Tract,” *Gastrointestinal Tumors*, 2(2), p. 65.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C. and Schultz, N. (2012) “The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data,” *Cancer discovery*, 2(5), p. 401.

Chan, D.S.M., Lau, R., Aune, D., Vieira, R., Greenwood, D.C., Kampman, E. and Norat, T. (2011) “Red and Processed Meat and Colorectal Cancer Incidence: Meta-Analysis of Prospective Studies,” *PLoS ONE*, 6(6), p. e20456.

Chen, L., Qiu, X., Wang, X. and He, J. (2017) “FAP positive fibroblasts induce immune checkpoint blockade resistance in colorectal cancer via promoting

immunosuppression," *Biochemical and Biophysical Research Communications*, 487(1), pp. 8–14.

Chen, X., Yamaguchi, K., Rodgers, B., Goehrig, D., Vindrieux, D., Lahaye, X., Nolot, M., Ferry, L., Lanciano, S., Martin, N., Dubus, P., Miura, F., Ito, T., Cristofari, G., Manel, N., Kanemaki, M.T., Bernard, D. and Defossez, P.A. (2025) "DNA methylation protects cancer cells against senescence," *Nature Communications 2025 16:1*, 16(1), pp. 1–22.

Chen, Y., Zhang, Y., Yan, Y., Han, J., Zhang, L., Cheng, X., Lu, B., Li, N., Luo, C., Zhou, Y., Song, K., Iwasaki, M., Dai, M., Wu, D. and Chen, H. (2025) "Global colorectal cancer screening programs and coverage rate estimation: an evidence synthesis," *Journal of Translational Medicine*, 23(1), pp. 1–12.

Chen, Y., Zheng, Y. and Liu, S. (2024) "KRAS mutation promotes the colonization of Fusobacterium nucleatum in colorectal cancer by down-regulating SERTAD4," *Journal of Cellular and Molecular Medicine*, 28(20), p. e70182.

Chu, B., Hong, Z. and Zheng, X. (2021) "Acylglycerol Kinase-Targeted Therapies in Oncology," *Frontiers in Cell and Developmental Biology*, 9, p. 659158.

Chu, C.M., Yao, C.T., Chang, Y.T., Chou, H.L., Chou, Y.C., Chen, K.H., Terng, H.J., Huang, C.S., Lee, C.C., Su, S.L., Liu, Y.C., Lin, F.G., Wetter, T. and Chang, C.W. (2014) "Gene Expression Profiling of Colorectal Tumors and Normal Mucosa by Microarrays Meta-Analysis Using Prediction Analysis of Microarray, Artificial Neural Network, Classification, and Regression Trees," *Disease Markers*, 2014(1), p. 634123.

Chung, R.Y.N., Tsoi, K.K.F., Kyaw, M.H., Lui, A.R., Lai, F.T.T. and Sung, J.J.Y. (2019) "A population-based age-period-cohort study of colorectal cancer incidence comparing Asia against the West," *Cancer Epidemiology*, 59, pp. 29–36.

Ciocalteu, A., Gheonea, D.I., Saftoiu, A., Streba, L., Dragoescu, N.A. and Tenea-Cojan, T.S. (2018) "Current strategies for malignant pedunculated colorectal polyps," *World Journal of Gastrointestinal Oncology*, 10(12), p. 465.

Clark, S.R., Ma, A.C., Tavener, S.A., McDonald, B., Goodarzi, Z., Kelly, M.M., Patel, K.D., Chakrabarti, S., McAvoy, E., Sinclair, G.D., Keys, E.M., Allen-Vercoe, E., DeVinney, R., Doig, C.J., Green, F.H.Y. and Kubes, P. (2007) "Platelet TLR4 activates neutrophil extracellular traps to ensnare bacteria in septic blood," *Nature Medicine*, 13(4), pp. 463–469.

Cohen, S., Hyer, W., Mas, E., Auth, M., Attard, T.M., Spalinger, J., Latchford, A. and Durno, C. (2019) "Management of juvenile polyposis syndrome in children

and adolescents: A position paper from the espghan polyposis working group," *Journal of Pediatric Gastroenterology and Nutrition*, 68(3), pp. 453–462.

Cole, A.M., Myant, K., Reed, K.R., Ridgway, R.A., Athineos, D., Van Den Brink, G.R., Muncan, V., Clevers, H., Clarke, A.R., Sicinski, P. and Sansom, O.J. (2010) "Cyclin D2-CDK4/6 is required for efficient proliferation and tumorigenesis following Apc loss," *Cancer research*, 70(20), p. 8149.

Consortium, T.G.O., Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., *et* al. (2023) "The Gene Ontology knowledgebase in 2023," *Genetics*, 224(1).

Conti, B.A., Ruiz, P.D., Broton, C., Blobel, N.J., Kottemann, M.C., Sridhar, S., Lach, F.P., Wiley, T.F., Sasi, N.K., Carroll, T. and Smogorzewska, A. (2024) "RTF2 controls replication repriming and ribonucleotide excision at the replisome," *Nature Communications 2024 15:1*, 15(1), pp. 1–17.

Cooper, M.D. (2015) "The early history of B cells," *Nature Reviews Immunology*, 15(3), pp. 191–197.

Corley, D.A., Jensen, C.D., Marks, A.R., Zhao, W.K., Lee, J.K., Doubeni, C.A., Zauber, A.G., de Boer, J., Fireman, B.H., Schottinger, J.E., Quinn, V.P., Ghai, N.R., Levin, T.R. and Quesenberry, C.P. (2014) "Adenoma Detection Rate and Risk of Colorectal Cancer and Death," *The New England journal of medicine*, 370(14), p. 1298.

Cornish, A.J., Gruber, A.J., Kinnersley, B., Chubb, D., Frangou, A., Caravagna, G., Noyvert, B., Lakatos, E., Wood, H.M., Thorn, S., Culliford, R., Arnedo-Pac, C., Househam, J., Cross, W., Sud, A., Law, P., Leathlobhair, M.N., Hawari, A., Woolley, C., Sherwood, K., Feeley, N., Gül, G., Fernandez-Tajes, J., Zapata, L., Alexandrov, L.B., Murugaesu, N., Sosinsky, A., Mitchell, J., Lopez-Bigas, N., Quirke, P., Church, D.N., Tomlinson, I.P.M., Sottoriva, A., Graham, T.A., Wedge, D.C. and Houlston, R.S. (2024) "The genomic landscape of 2,023 colorectal cancers," *Nature 2024 633:8028*, 633(8028), pp. 127–136.

Cristofaro, M., Contursi, A., D'Amore, S., Martelli, N., Spaziante, A.F., Moschetta, A. and Villani, G. (2015) "Adenomatous polyposis coli (APC)-induced apoptosis of HT29 colorectal cancer cells depends on mitochondrial oxidative metabolism," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1852(9), pp. 1719–1728.

Cruz-Burgos, M., Losada-Garcia, A., Cruz-Hernández, C.D., Cortés-Ramírez, S.A., Camacho-Arroyo, I., Gonzalez-Covarrubias, V., Morales-Pacheco, M., Trujillo-Bornios, S.I. and Rodríguez-Dorantes, M. (2021) "New Approaches in Oncology for Repositioning Drugs: The Case of PDE5 Inhibitor Sildenafil," *Frontiers in Oncology*, 11, p. 627229.

Davidson, K.W., Barry, M.J., Mangione, C.M., Cabana, M., Caughey, A.B., Davis, E.M., Donahue, K.E., Doubeni, C.A., Krist, A.H., Kubik, M., Li, L., Ogedegbe, G., Owens, D.K., Pbert, L., Silverstein, M., Stevermer, J., Tseng, C.W. and Wong, J.B. (2021) "Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement," *JAMA - Journal of the American Medical Association*, 325(19), pp. 1965–1977.

Dekker, E., Bleijenberg, A., Balaguer, F., IJspeert, J.E.G., Bleijenberg, A.G.C., Pellisé, M., Carballal, S., Rivero, L. and Latchford, A. (2020) "Update on the World Health Organization Criteria for Diagnosis of Serrated Polyposis Syndrome," *Gastroenterology*, 158(6), pp. 1520–1523.

Delon, C., Brown, K.F., Payne, N.W.S., Kotrotsios, Y., Vernon, S. and Shelton, J. (2022) "Differences in cancer incidence by broad ethnic group in England, 2013–2017," *British Journal of Cancer*, 126(12), pp. 1765–1773.

Dickey, J.S., Redon, C.E., Nakamura, A.J., Baird, B.J., Sedelnikova, O.A. and Bonner, W.M. (2009) "H2AX: functional roles and potential applications," *Chromosoma*, 118(6), p. 683.

Dikovskaya, D., Schiffmann, D., Newton, I.P., Oakley, A., Kroboth, K., Sansom, O., Jamieson, T.J., Meniel, V., Clarke, A. and Näthke, I.S. (2007) "Loss of APC induces polyploidy as a result of a combination of defects in mitosis and apoptosis," *The Journal of Cell Biology*, 176(2), p. 183.

Ding, X., Ting, N.L.N., Wong, C.C., Huang, P., Jiang, L., Liu, C., Lin, Y., Li, S., Liu, Y., Xie, M., Liu, W., Yuan, K., Wang, L., Zhang, X., Ding, Y., Li, Q., Sun, Y., Miao, Y., Ma, L., Gao, X., Li, W., Wu, W.K.K., Sung, J.J.Y., Wong, S.H. and Yu, J. (2025) "Bacteroides fragilis promotes chemoresistance in colorectal cancer, and its elimination by phage VA7 restores chemosensitivity," *Cell Host & Microbe*, 33(6), pp. 941-956.e10.

Dong, D., Zheng, L., Lin, J., Zhang, B., Zhu, Y., Li, N., Xie, S., Wang, Y., Gao, N. and Huang, Z. (2019) "Structural basis of assembly of the human T cell receptor–CD3 complex," *Nature*, 573(7775), pp. 546–552.

Dow, L.E., O'Rourke, K.P., Simon, J., Tschaharganeh, D.F., Van Es, J.H., Clevers, H. and Lowe, S.W. (2015) "Apc restoration promotes cellular differentiation and reestablishes crypt homeostasis in colorectal cancer," *Cell*, 161(7), p. 1539.

Dröge, L.H., Weber, H.E., Guhlich, M., Leu, M., Conradi, L.C., Gaedcke, J., Hennies, S., Herrmann, M.K., Rave-Fränk, M. and Wolff, H.A. (2015) "Reduced toxicity in the treatment of locally advanced rectal cancer: a comparison of volumetric modulated arc therapy and 3D conformal radiotherapy," *BMC Cancer*, 15(1), p. 750.

Du, X., Pan, Z., Li, Qiqi, Liu, H. and Li, Qifa (2018) “SMAD4 feedback regulates the canonical TGF-ß signaling pathway to control granulosa cell apoptosis article,” *Cell Death and Disease*, 9(2), pp. 1–12.

Dunne, P.D. and Arends, M.J. (2024) “Molecular pathological classification of colorectal cancer—an update,” *Virchows Archiv*, 484(2), pp. 273–285.

Dy, G.K., Govindan, R., Velcheti, V., Falchook, G.S., Italiano, A., Wolf, J., Sacher, A.G., Takahashi, T., Ramalingam, S.S., Dooms, C., Kim, D.W., Addeo, A., Desai, J., Schuler, M., Tomasini, P., Hong, D.S., Lito, P., Tran, Q., Jones, S., Anderson, A., Hindoyan, A., Snyder, W., Skoulidis, F. and Li, B.T. (2023) “Long-Term Outcomes and Molecular Correlates of Sotorasib Efficacy in Patients With Pretreated KRAS G12C-Mutated Non–Small-Cell Lung Cancer: 2-Year Analysis of CodeBreaK 100,” *Journal of Clinical Oncology*, 41(18), p. 3311.

East, J.E., Atkin, W.S., Bateman, A.C., Clark, S.K., Dolwani, S., Ket, S.N., Leedham, S.J., Phull, P.S., Rutter, M.D., Shepherd, N.A., Tomlinson, I. and Rees, C.J. (2017) “British Society of Gastroenterology position statement on serrated polyps in the colon and rectum,” *Gut*, 66(7), pp. 1181–1196.

Eskandari-Malayeri, F., Rezeai, M., Narimani, T., Esmaeil, N. and Azizi, M. (2024) “Investigating the effect of Fusobacterium nucleatum on the aggressive behavior of cancer-associated fibroblasts in colorectal cancer,” *Discover Oncology*, 15(1), p. 292.

Eurofins (no date) *Oligo Analysis Tool*. Available at: https://eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/%0A.

Evers, D.L., He, J., Kim, Y.H., Mason, J.T. and O’Leary, T.J. (2011) “Paraffin Embedding Contributes to RNA Aggregation, Reduced RNA Yield, and Low RNA Quality,” *The Journal of Molecular Diagnostics : JMD*, 13(6), p. 687.

Facciorusso, A., Di Maso, M., Serviddio, G., Vendemiale, G., Spada, C., Costamagna, G. and Muscatiello, N. (2016) “Factors Associated With Recurrence of Advanced Colorectal Adenoma After Endoscopic Resection,” *Clinical Gastroenterology and Hepatology*, 14(8), pp. 1148-1154.e4.

Fairley, K.J., Li, J., Komar, M., Steigerwalt, N. and Erlich, P. (2014) “Predicting the Risk of Recurrent Adenoma and Incident Colorectal Cancer Based on Findings of the Baseline Colonoscopy,” *Clinical and Translational Gastroenterology*, 5(12), p. e64.

Fan, F., Verma, A., Hull, M., Langley, R., Quirke, P., Rutter, M., Rees, C., Boota, N., Brookes, C., Brown, L., Howells, L., Thomas, A., Sandwell, B. and

Brown, K. (2025) “COLO-PREVENT: A platform for developing colorectal cancer prevention therapies,” *Cancer Research* [AACR Abstract CT252].

Fearon, E. and Vogelstein, B. (1990) “A genetic model for colorectal tumorigenesis,” *Cell*, 61(5), pp. 759–767.

Fedirko, V., Romieu, I., Aleksandrova, K., Pischon, T., Trichopoulos, D., Peeters, P.H., Romaguera-Bosch, D., Bueno-De-Mesquita, H.B., Dahm, C.C., Overvad, K., Chirlaque, M.D., Johansen, C., Bidstrup, P.E., Dalton, S.O., Gunter, M.J., Wark, P.A., Norat, T., Halkjær, J., Tjønneland, A., Dik, V.K., Siersema, P.D., Boutron-Ruault, M.C., Dossus, L., Bastide, N., Kühn, T., Kaaks, R., Boeing, H., Trichopoulou, A., Klinaki, E., Katsoulis, M., Pala, V., Panico, S., Tumino, R., Palli, D., Vineis, P., Weiderpass, E., Skeie, G., González, C.A., Sánchez, M.J., Barricarte, A., Amiano, P., Quiros, J.R., Manjer, J., Jirström, K., Ljuslinder, I., Palmqvist, R., Khaw, K.T., Wareham, N., Bradbury, K.E., Stepien, M., Duarte-Salles, T., Riboli, E. and Jenab, M. (2014) “Pre-diagnostic anthropometry and survival after colorectal cancer diagnosis in Western European populations,” *International Journal of Cancer*, 135(8), pp. 1949–1960.

Feng, Y., Sentani, K., Wiese, A., Sands, E., Green, M., Bommer, G.T., Cho, K.R. and Fearon, E.R. (2013) “Sox9 Induction, Ectopic Paneth Cells, and Mitotic Spindle Axis Defects in Mouse Colon Adenomatous Epithelium Arising From Conditional Biallelic Apc Inactivation,” *The American Journal of Pathology*, 183(2), p. 493.

Fennell, L.J., Jamieson, S., McKeone, D., Corish, T., Rohdmann, M., Furner, T., Bettington, M., Liu, C., Kawamata, F., Bond, C., Van De Pols, J., Leggett, B. and Whitehall, V. (2018) “MLH1–93 G/a polymorphism is associated with MLH1 promoter methylation and protein loss in dysplastic sessile serrated adenomas with BRAFV600E mutation,” *BMC Cancer*, 18(1), pp. 1–7.

Figuero, E., Sánchez-Beltrán, M., Cuesta-Frechoso, S., Tejerina, J.M., Castro, J.A. del, Gutiérrez, J.M., Herrera, D. and Sanz, M. (2011) “Detection of Periodontal Bacteria in Atheromatous Plaque by Nested Polymerase Chain Reaction,” *Journal of Periodontology*, 82(10), pp. 1469–1477.

Fischer, H., Stenling, R., Rubio, C. and Lindblom, A. (2001) “Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2,” *Carcinogenesis*, 22(6), pp. 875–878.

Flanagan, L., Schmid, J., Ebert, M., Soucek, P., Kunicka, T., Liska, V., Bruha, J., Neary, P., Dezeeuw, N., Tommasino, M., Jenab, M., Prehn, J.H.M. and Hughes, D.J. (2014) “Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome,” *European Journal of Clinical Microbiology and Infectious Diseases*, 33(8), pp. 1381–1390.

Fu, X., Wang, X., Duanmu, J., Li, T. and Jiang, Q. (2020) “KRAS mutations are negatively correlated with immunity in colon cancer,” *Aging*, 13(1), pp. 750–768.

Galasso, L., Termite, F., Mignini, I., Esposto, G., Borriello, R., Vitale, F., Nicoletti, A., Paratore, M., Ainora, M.E., Gasbarrini, A. and Zocco, M.A. (2025) “Unraveling the Role of Fusobacterium nucleatum in Colorectal Cancer: Molecular Mechanisms and Pathogenic Insights,” *Cancers*, 17(3), p. 368.

Galeano Niño, J.L., Wu, H., LaCourse, K.D., Kempchinsky, A.G., Baryiames, A., Barber, B., Futran, N., Houlton, J., Sather, C., Sicinska, E., Taylor, A., Minot, S.S., Johnston, C.D. and Bullman, S. (2022) “Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer,” *Nature*, 611(7937), p. 810.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoué, F., Bruneval, P., Cugnenc, P.H., Trajanoski, Z., Fridman, W.H. and Pagès, F. (2006) “Type, density, and location of immune cells within human colorectal tumors predict clinical outcome,” *Science*, 313(5795), pp. 1960–1964.

Galuppini, F., Fassan, M., Mastracci, L., Gafà, R., Lo Mele, M., Lazzi, S., Remo, A., Parente, P., D’Amuri, A., Mescoli, C., Tatangelo, F. and Lanza, G. (2021) “The histomorphological and molecular landscape of colorectal adenomas and serrated lesions,” *Pathologica*, 113(3), p. 218.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C. and Schultz, N. (2013) “Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal,” *Science signaling*, 6(269), p. pl1.

Gao, Y., Zou, T., Xu, P., Wang, Y., Jiang, Y., Chen, Y.X., Chen, H., Hong, J. and Fang, J.Y. (2023) “Fusobacterium nucleatum stimulates cell proliferation and promotes PD-L1 expression via IFIT1-related signal in colorectal cancer,” *Neoplasia*, 35, p. 100850.

Gao, Z.H., Lu, C., Wang, M.X., Han, Y. and Guo, L.J. (2014) “Differential ß-catenin expression levels are associated with morphological features and prognosis of colorectal cancer,” *Oncology Letters*, 8(5), p. 2069.

Gerstung, M., Beisel, C., Rechsteiner, M., Wild, P., Schraml, P., Moch, H. and Beerenwinkel, N. (2012) “Reliable detection of subclonal single-nucleotide variants in tumour cell populations,” *Nature Communications 2012 3:1*, 3(1), pp. 1–8.

Gerstung, M., Papaemmanuil, E. and Campbell, P.J. (2014) “Subclonal variant calling with multiple samples and prior knowledge,” *Bioinformatics*, 30(9), p. 1198.

Gilad, O., Rosner, G., Fliss-Isakov, N., Aharon-Kaspi, S., Strul, H., Gluck, N. and Kariv, R. (2019) “Clinical and Histologic Overlap and Distinction Among Various Hamartomatous Polyposis Syndromes,” *Clinical and Translational Gastroenterology*, 10(5), p. e00035.

Gini, A., Jansen, E.E.L., Zielonke, N., Meester, R.G.S., Senore, C., Anttila, A., Segnan, N., Mlakar, D.N., de Koning, H.J., Lansdorp-Vogelaar, I., Veerus, P., Heinävaara, S., Sarkeala, T., Csanádi, M., Pitter, J., Széles, G., Vokó, Z., Minozzi, S., van Ballegooijen, M., Driesprong - de Kok, I., Heijnsdijk, E., Jansen, E., van Ravesteyn, N., Ivanus, U., Jarm, K., Primic-Žakelj, M., McKee, M. and Priaulx, J. (2020) “Impact of colorectal cancer screening on cancer-specific mortality in Europe: A systematic review,” *European Journal of Cancer*, 127, pp. 224–235.

Goto, N., Westcott, P.M.K., Goto, S., Imada, S., Taylor, M.S., Eng, G., Braverman, J., Deshpande, V., Jacks, T., Agudo, J. and Yilmaz, Ö.H. (2024) “SOX17 enables immune evasion of early colorectal adenomas and cancers,” *Nature 2024 627:8004*, 627(8004), pp. 636–645.

Grazzini, G., Ventura, L., Zappa, M., Ciatto, S., Confortini, M., Rapi, S., Rubeca, T., Visioli, C.B. and Halloran, S.P. (2010) “Influence of seasonal variations in ambient temperatures on performance of immunochemical faecal occult blood test for colorectal cancer screening: observational study from the Florence district,” *Gut*, 59(11), pp. 1511–1515.

Greegor, D.H. (1967) “Diagnosis of Large-Bowel Cancer in the Asymptomatic Patient,” *JAMA*, 201(12), pp. 943–945.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., Sargeant, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J.P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M. and White, R. (1991) “Identification and characterization of the familial adenomatous polyposis coli gene,” *Cell*, 66(3), pp. 589–600.

Gruden, Š. and Ulrih, N.P. (2021) “Diverse mechanisms of antimicrobial activities of lactoferrins, lactoferricins, and other lactoferrin-derived peptides,” *International Journal of Molecular Sciences*, 22(20), p. 11264.

Gu, Z. (2022) “Complex heatmap visualization,” *iMeta*, 1(3), p. e43.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990) “A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes,” *Nature*, 346(6281), pp. 245–250.

Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B.M., Morris, J.S., Simon, I.M., Gerster, S., Fessler, E., De Sousa .E Melo, F., Missiaglia, E., Ramay, H., Barras, D., Homicsko, K., Maru, D., Manyam, G.C., Broom, B., Boige, V., Perez-Villamil, B., Laderas, T., Salazar, R., Gray, J.W., Hanahan, D., Tabernero, J., Bernards, R., Friend, S.H., Laurent-Puig, P., Medema, J.P., Sadanandam, A., Wessels, L., Delorenzi, M., Kopetz, S., Vermeulen, L. and Tejpar, S. (2015) “The Consensus Molecular Subtypes of Colorectal Cancer,” *Nature medicine*, 21(11), p. 1350.

Guo, P., Tian, Z., Kong, X., Yang, L., Shan, X., Dong, B., Ding, X., Jing, X., Jiang, C., Jiang, N. and Yu, Y. (2020) “FadA promotes DNA damage and progression of Fusobacterium nucleatum-induced colorectal cancer through up-regulation of chk2,” *Journal of Experimental & Clinical Cancer Research : CR*, 39(1), p. 202.

Gupta, S., Lieberman, D., Anderson, J.C., Burke, C.A., Dominitz, J.A., Kaltenbach, T., Robertson, D.J., Shaukat, A., Syngal, S. and Rex, D.K. (2020) “Recommendations for Follow-Up After Colonoscopy and Polypectomy: A Consensus Update by the US Multi-Society Task Force on Colorectal Cancer,” *Gastroenterology*, 158(4), pp. 1131-1153.e5.

Gur, C., Ibrahim, Y., Isaacson, B., Yamin, R., Abed, J., Gamliel, M., Enk, J., Bar-On, Y., Stanietsky-Kaynan, N., Coppenhagen-Glazer, S., Shussman, N., Almogy, G., Cuapio, A., Hofer, E., Mevorach, D., Tabib, A., Ortenberg, R., Markel, G., Miklić, K., Jonjic, S., Brennan, C.A., Garrett, W.S., Bachrach, G. and Mandelboim, O. (2015) “Binding of the Fap2 Protein of Fusobacterium nucleatum to Human Inhibitory Receptor TIGIT Protects Tumors from Immune Cell Attack,” *Immunity*, 42(2), p. 344.

Hafemeister, C. and Satija, R. (2019) “Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression,” *Genome Biology*, 20(1), pp. 1–15.

Haggar, F.A. and Boushey, R.P. (2009) “Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors,” *Clinics in Colon and Rectal Surgery*, 22(4), p. 191. Available at: https://doi.org/10.1055/S-0029-1242458.
Hahn, S.A., Schutte, M., Shamsul Hoque, A.T.M., Moskaluk, C.A., Da Costa, L.T., Rozenblum, E., Weinstein, C.L., Fischer, A., Yeo, C.J., Hruban, R.H. and Kern, S.E. (1996) “DPC4, A Candidate Tumor Suppressor Gene at Human Chromosome 18q21.1,” *Science*, 271(5247), pp. 350–353.

Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, Hanhae, Cho, A., Kim, E., Lee, T., Kim, Hyojin, Kim, K., Yang, S., Bae, D., Yun, A., Kim, S., Kim, C.Y., Cho, H.J., Kang, B., Shin, S. and Lee, I. (2015) "TRRUST: A reference database of human transcriptional regulatory interactions," *Scientific Reports*, 5(1), pp. 1–11.

Han, Y., Gao, S., Muegge, K., Zhang, W. and Zhou, B. (2015) "Advanced Applications of RNA Sequencing and Challenges," *Bioinformatics and Biology Insights*, 9(Suppl 1), p. 29. Available at: https://doi.org/10.4137/BBI.S28991.
Han, Y.W. (2015) "Fusobacterium nucleatum: a commensal-turned pathogen," *Current opinion in microbiology*, 0, p. 141.

Han, Y.W., Redline, R.W., Li, M., Yin, L., Hill, G.B. and McCormick, T.S. (2004) "Fusobacterium nucleatum Induces Premature and Term Stillbirths in Pregnant Mice: Implication of Oral Bacteria in Preterm Birth," *Infection and Immunity*, 72(4), p. 2272. Available at: https://doi.org/10.1128/IAI.72.4.2272-2279.2004.
Hanahan, D. (2022) "Hallmarks of Cancer: New Dimensions," *Cancer Discovery*, 12(1), pp. 31–46. Available at: https://doi.org/10.1158/2159-8290.CD-21-1059.
Hanahan, D. and Weinberg, R.A. (2011) "Hallmarks of cancer: the next generation.," *Cell*, 144(5), pp. 646–74.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E.P., Jain, J., Srivastava, A., Stuart, T., Fleming, L.M., Yeung, B., Rogers, A.J., McElrath, J.M., Blish, C.A., Gottardo, R., Smibert, P. and Satija, R. (2021) "Integrated analysis of multimodal single-cell data," *Cell*, 184(13), pp. 3573-3587.e29.

Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C. and Satija, R. (2024) "Dictionary learning for integrative, multimodal and scalable single-cell analysis," *Nature Biotechnology*, 42(2), pp. 293–304.

Hao, Y., Wang, Y., Qi, M., He, X., Zhu, Y. and Hong, J. (2020) "Risk Factors for Recurrent Colorectal Polyps," *Gut and Liver*, 14(4), pp. 399–411.

Hassan, C., Antonelli, G., Dumonceau, J.-M., Regula, J., Bretthauer, M., Chaussade, S., Dekker, E., Ferlitsch, M., Gimeno-Garcia, A., Jover, R., Kalager, M., Pellise, M., Pox, C., Ricciardiello, L., Rutter, M., Helsingen, L.M., Bleijenberg, A., Senore, C., van Hooft, J.E., Dinis-Ribeiro, M. and Quintero, E. (2020) "Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2020," *Endoscopy*, 52, pp. 1–14.

Hata, S., Abe, M., Suzuki, H., Kitamura, F., Toyama-Sorimachi, N., Abe, K., Sakimura, K. and Sorimachi, H. (2010) “Calpain 8/nCL-2 and Calpain 9/nCL-4 Constitute an Active Protease Complex, G-Calpain, Involved in Gastric Mucosal Defense,” *PLoS Genetics*, 6(7), p. e1001040.

Hothorn, T. (2017) “maxstat: Maximally Selected Rank Statistics.” Available at: https://CRAN.R-project.org/package=maxstat (Accessed: February 24, 2024).

Hu, L., Liu, Y., Kong, X., Wu, R., Peng, Q., Zhang, Y., Zhou, L. and Duan, L. (2021) “Fusobacterium nucleatum Facilitates M2 Macrophage Polarization and Colorectal Carcinoma Progression by Activating TLR4/NF-κB/S100A9 Cascade,” *Frontiers in Immunology*, 12, p. 658681.

Hu, Y., Rong, J., Xu, Y., Xie, R., Peng, J., Gao, L. and Tan, K. (2024) “Unsupervised and supervised discovery of tissue cellular neighborhoods from cell phenotypes,” *Nature Methods*, 21(2), pp. 267–278.

Huang, F., Wang, Y., Shao, Y., Zhang, R., Li, M., Liu, L. and Zhao, Q. (2024) “M2 Macrophage Classification of Colorectal Cancer Reveals Intrinsic Connections with Metabolism Reprogramming and Clinical Characteristics,” *Pharmacogenomics and Personalized Medicine*, 17, p. 383.

Huck, M.B. and Bohl, J.L. (2016) “Colonic Polyps: Diagnosis and Surveillance,” *Clinics in Colon and Rectal Surgery*, 29(4), p. 296.

Hunt, K.K., Fleming, J.B., Abramian, A., Zhang, L., Evans, D.B. and Chiao, P.J. (1998) “Overexpression of the tumor suppressor gene Smad4/DPC4 induces p21waf1 expression and growth inhibition in human carcinoma cells,” *Cancer Research*, 58(24), pp. 5656–5661.

Huo, Y., Zhang, H., Sa, L., Zheng, W., He, Y., Lyu, H., Sun, M., Zhang, L., Shan, L., Yang, A. and Wang, T. (2023) “M1 polarization enhances the antitumor activity of chimeric antigen receptor macrophages in solid tumors,” *Journal of Translational Medicine*, 21(1), pp. 1–15.

Ito, M., Kanno, S., Nosho, K., Sukawa, Y., Mitsuhashi, K., Kurihara, H., Igarashi, H., Takahashi, T., Tachibana, M., Takahashi, H., Yoshii, S., Takenouchi, T., Hasegawa, T., Okita, K., Hirata, K., Maruyama, R., Suzuki, H., Imai, K., Yamamoto, H. and Shinomura, Y. (2015) “Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway,” *International Journal of Cancer*, 137(6), pp. 1258–1268.

Iwaki, T., Urano, T. and Umemura, K. (2012) “PAI-1, progress in understanding the clinical problem and its aetiology,” *British Journal of Haematology*, 157(3), pp. 291–298.

Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., Becker, M., Cuadrado, P.R., Davis, H., Kaur, K., Heinimann, K., Howarth, K., East, J., Taylor, J., Thomas, H. and Tomlinson, I. (2012) “Hereditary mixed polyposis syndrome is caused by a 40kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1,” *Nature genetics*, 44(6), p. 699.

Jamieson, C., Sharma, M. and Henderson, B.R. (2011) “Regulation of ß-catenin nuclear dynamics by GSK-3ß involves a LEF-1 positive feedback loop,” *Traffic*, 12(8), pp. 983–999.

Jang, D.I., Lee, A.H., Shin, H.Y., Song, H.R., Park, J.H., Kang, T.B., Lee, S.R. and Yang, S.H. (2021) “The Role of Tumor Necrosis Factor Alpha (TNF-α) in Autoimmune Disease and Current TNF-α Inhibitors in Therapeutics,” *International Journal of Molecular Sciences*, 22(5), p. 2719.

Jänne, P.A., Riely, G.J., Gadgeel, S.M., Heist, R.S., Ou, S.-H.I., Pacheco, J.M., Johnson, M.L., Sabari, J.K., Leventakos, K., Yau, E., Bazhenova, L., Negrao, M. V., Pennell, N.A., Zhang, J., Anderes, K., Der-Torossian, H., Kheoh, T., Velastegui, K., Yan, X., Christensen, J.G., Chao, R.C. and Spira, A.I. (2022) “ Adagrasib in Non–Small-Cell Lung Cancer Harboring a KRAS G12C Mutation ,” *New England Journal of Medicine*, 387(2), pp. 120–131.

Jasmin, M., Ahn, E.H., Voutilainen, M.H., Fombonne, J., Guix, C., Viljakainen, T., Kang, S.S., Yu, L., Saarma, M., Mehlen, P. and Ye, K. (2020) “Netrin-1 and its receptor DCC modulate survival and death of dopamine neurons and Parkinson’s disease features,” *The EMBO Journal*, 40(3), p. e105537.

Javier, B.M., Yaeger, R., Wang, L., Sanchez-Vega, F., Zehir, A., Middha, S., Sadowska, J., Vakiani, E., Shia, J., Klimstra, D., Ladanyi, M., Iacobuzio-Donahue, C.A. and Hechtman, J.F. (2016) “Recurrent, truncating SOX9 mutations are associated with SOX9 overexpression, KRAS mutation, and TP53 wild type status in colorectal carcinoma,” *Oncotarget*, 7(32), p. 50875.

Jia, Y., Huang, X., Shi, H., Wang, M.M., Chen, J., Zhang, H., Hou, D., Jing, H., Du, J., Han, H. and Zhang, J. (2023) “ADAMDEC1 induces EMT and promotes colorectal cancer cells metastasis by enhancing Wnt/ß-catenin signaling via negative modulation of GSK-3ß,” *Experimental Cell Research*, 429(2), p. 113629.

Jo, A., Denduluri, S., Zhang, B., Wang, Z., Yin, L., Yan, Z., Kang, R., Shi, L.L., Mok, J., Lee, M.J. and Haydon, R.C. (2014) “The versatile functions of Sox9 in development, stem cells, and human diseases,” *Genes & Diseases*, 1(2), pp. 149–161.

Johnson, C., Burkhart, D.L. and Haigis, K.M. (2022) “Classification of KRAS activating mutations and the implications for therapeutic intervention,” *Cancer discovery*, 12(4), p. 913.

Johnstone, M.S., Lynch, G., Park, J., McSorley, S. and Edwards, J. (2021) “Novel Methods of Risk Stratifying Patients for Metachronous, Pre-Malignant Colorectal Polyps: A Systematic Review,” *Critical Reviews in Oncology/Hematology*, 164, p. 103421.

Johnstone, M.S., Stoops, R., Lynch, G., Hay, J., Jawny, J., Sloan, W., Edwards, J. and McSorley, S.T. (2023) “Risk stratification for the detection of metachronous polyps after bowel screening polypectomy: clinical outcomes from the Integrated Technologies for Improved Polyp Surveillance (INCISE) study cohort,” *BJS Open*, 7(3).

Juul, F.E., Cross, A.J., Schoen, R.E., Senore, C., Pinsky, P.F., Miller, E.A., Segnan, N., Wooldrage, K., Wieszczy-Szczepanik, P., Armaroli, P., Garborg, K.K., Adami, H.O., Hoff, G., Kalager, M., Bretthauer, M., Holme, Ø. and Løberg, M. (2024) “Effectiveness of Colonoscopy Screening vs Sigmoidoscopy Screening in Colorectal Cancer,” *JAMA Network Open*, 7(2), pp. e240007–e240007.

Kalra, S., Mittal, A., Gupta, K., Singhal, V., Gupta, A., Mishra, T., Naidu, S., Sengupta, D. and Ahuja, G. (2020) “Analysis of single-cell transcriptomes links enrichment of olfactory receptors with cancer cell differentiation status and prognosis,” *Communications Biology 2020 3:1*, 3(1), pp. 1–10.

Kambara, T., Simms, L.A., Whitehall, V.L.J., Spring, K.J., Wynter, C.V.A., Walsh, M.D., Barker, M.A., Arnold, S., McGivern, A., Matsubara, N., Tanaka, N., Higuchi, T., Young, J., Jass, J.R. and Leggett, B.A. (2004) “BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum,” *Gut*, 53(8), p. 1137.

Kanehisa, M. (2019) “Toward understanding the origin and evolution of cellular organisms,” *Protein Science : A Publication of the Protein Society*, 28(11), p. 1947.

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. and Ishiguro-Watanabe, M. (2024) “KEGG: biological systems database as a model of the real world,” *Nucleic Acids Research*, 53(D1), p. D672.

Kanehisa, M. and Goto, S. (2000) “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Research*, 28(1), p. 27.

Kang, S., Narazaki, M., Metwally, H. and Kishimoto, T. (2020) “Historical overview of the interleukin-6 family cytokine,” *The Journal of Experimental Medicine*, 217(5), p. e20190347.

Kang, X., Zhang, R., Kwong, T.N., Lui, R.N., Wu, W.K., Sung, J.J., Yu, J. and Wong, S.H. (2021) “Serrated neoplasia in the colorectum: gut microbiota and molecular pathways,” *Gut Microbes*, 13(1), pp. 1–12.

Kapałczyńska, M., Kolenda, T., Przybyła, W., Zajączkowska, M., Teresiak, A., Filas, V., Ibbs, M., Bliźniak, R., Łuczewski, Ł. and Lamperska, K. (2016) “2D and 3D cell cultures – a comparison of different types of cancer cell cultures,” *Archives of Medical Science : AMS*, 14(4), p. 910.

Karaca, S. and Kırlı Bölükbaş, M. (2025) “Time Matters: A Review of Current Radiotherapy Practices and Efficiency Strategies,” *Technology in Cancer Research & Treatment*, 24, p. 15330338251345376.

Karamanos, N.K. (2019) “Extracellular matrix: key structural and functional meshwork in health and disease,” *FEBS Journal*, 286(15), pp. 2826–2829.

Karamitopoulou, E., Zlobec, I., Patsouris, E.S., Peros, G. and Lugli, A. (2011) “Loss of E-cadherin independently predicts the lymph node status in colorectal cancer,” *Pathology*, 43(2), pp. 133–137.

Kasi, P.M., Afable, M.G., Herting, C., Lukanowski, M. and Jin, Z. (2023) “Anti-EGFR Antibodies in the Management of Advanced Colorectal Cancer,” *The Oncologist*, 28(12), pp. 1034–1048.

Kassambara, A., Kosinski, M. and Biecek, P. (2021) “survminer: Drawing Survival Curves using ‘ggplot2.’” Available at: https://CRAN.R-project.org/package=survminer.

Kasurinen, J., Hagström, J., Kaprio, T., Beilmann-Lehtonen, I., Haglund, C. and Böckelman, C. (2022) “Tumor-associated CD3- and CD8-positive immune cells in colorectal cancer: The additional prognostic value of CD8+-to-CD3+ ratio remains debatable,” *Tumor Biology*, 44(1), pp. 37–52.

Katzir, A. (2003) “Optical Fiber Techniques for Medical Applications,” *Encyclopedia of Physical Science and Technology*, pp. 315–333.

Kehlet, S.N., Sanz-Pamplona, R., Brix, S., Leeming, D.J., Karsdal, M.A. and Moreno, V. (2016) “Excessive collagen turnover products are released during colorectal cancer progression and elevated in serum from metastatic colorectal cancer patients,” *Scientific Reports*, 6(1), pp. 1–7.

Kent, W.J. (2002) “BLAT—The BLAST-Like Alignment Tool,” *Genome Research*, 12(4), pp. 656–664. Available at: https://doi.org/10.1101/GR.229202.

Kim, H.J., Lee, H.N., Jeong, M.S. and Jang, S.B. (2021) “Oncogenic KRAS: Signaling and Drug Resistance,” *Cancers*, 13(22), p. 5599.

Kim, H.S., Kim, C.G., Kim, W.K., Kim, K.A., Yoo, J., Min, B.S., Paik, S., Shin, S.J., Lee, H., Lee, K., Kim, H., Shin, E.C., Kim, T.M. and Ahn, J.B. (2023) "Fusobacterium nucleatum induces a tumor microenvironment with diminished adaptive immunity against colorectal cancers," *Frontiers in Cellular and Infection Microbiology*, 13, p. 1101291.

Kim, J., Koo, B.K. and Knoblich, J.A. (2020) "Human organoids: model systems for human biology and medicine," *Nature Reviews Molecular Cell Biology 2020 21:10*, 21(10), pp. 571–584.

Kim, M.S., Ha, S.E., Wu, M., Zogg, H., Ronkon, C.F., Lee, M.Y. and Ro, S. (2021) "Extracellular Matrix Biomarkers in Colorectal Cancer," *International Journal of Molecular Sciences*, 22(17), p. 9185.

Kim, S.A., Inamura, K., Yamauchi, M., Nishihara, R., Mima, K., Sukawa, Y., Li, T., Yasunari, M., Morikawa, T., Fitzgerald, K.C., Fuchs, C.S., Wu, K., Chan, A.T., Zhang, X., Ogino, S. and Qian, Z.R. (2016) "Loss of CDH1 (E-cadherin) expression is associated with infiltrative tumour growth and lymph node metastasis," *British Journal of Cancer 2016 114:2*, 114(2), pp. 199–206.

Kim, S.Y., Kwak, M.S., Yoon, S.M., Jung, Y., Kim, J.W., Boo, S.J., Oh, E.H., Jeon, S.R., Nam, S.J., Park, S.Y., Park, S.K., Chun, J., Baek, D.H., Choi, M.Y., Park, S., Byeon, J.S., Kim, H.K., Cho, J.Y., Lee, M.S. and Lee, O.Y. (2022) "Korean guidelines for postpolypectomy colonoscopic surveillance: 2022 revised edition," *Clinical Endoscopy*, 55(6), pp. 703–725.

Kim, W.K., Kwon, Y., Jang, M., Park, M., Kim, J., Cho, S., Jang, D.G., Lee, W. Bin, Jung, S.H., Choi, H.J., Min, B.S., Il Kim, T., Hong, S.P., Paik, Y.K. and Kim, H. (2019) "β-catenin activation down-regulates cell-cell junction-related genes and induces epithelial-to-mesenchymal transition in colorectal cancers," *Scientific Reports*, 9(1), pp. 1–15.

Kirkegaard, T., Edwards, J., Tovey, S., McGlynn, L.M., Krishna, S.N., Mukherjee, R., Tam, L., Munro, A.F., Dunne, B. and Bartlett, J.M.S. (2006) "Observer variation in immunohistochemical analysis of protein expression, time for a change?," *Histopathology*, 48, pp. 787–794.

Kleeman, S.O., Koelzer, V.H., Jones, H.J.S., Vazquez, E.G., Davis, H., East, J.E., Arnold, R., Koppens, M.A.J., Blake, A., Domingo, E., Cunningham, C., Beggs, A.D., Pestinger, V., Loughrey, M.B., Wang, L.M., Lannagan, T.R.M., Woods, S.L., Worthley, D., Consortium, S., Tomlinson, I., Dunne, P.D., Maughan, T. and Leedham, S.J. (2020) "Exploiting differential Wnt target gene expression to generate a molecular biomarker for colorectal cancer stratification," *Gut*, 69(6), pp. 1092–1103.

Klimeck, L., Heisser, T., Hoffmeister, M. and Brenner, H. (2023) “Colorectal cancer: A health and economic problem,” *Best Practice & Research Clinical Gastroenterology*, 66, p. 101839.

Koh, C.H., Lee, S., Kwak, M., Kim, B.S. and Chung, Y. (2023) “CD8 T-cell subsets: heterogeneity, functions, and therapeutic potential,” *Experimental & Molecular Medicine 2023 55:11*, 55(11), pp. 2287–2299.

Komiya, Y., Shimomura, Y., Higurashi, T., Sugi, Y., Arimoto, J., Umezawa, S., Uchiyama, S., Matsumoto, M. and Nakajima, A. (2019) “Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity,” *Gut*, 68(7), pp. 1335–1337.

Kong, X., Zhang, Y., Xiang, L., You, Y., Duan, Y., Zhao, Y., Li, S., Wu, R., Zhang, J., Zhou, L. and Duan, L. (2023) “Fusobacterium nucleatum-triggered neutrophil extracellular traps facilitate colorectal carcinoma progression,” *Journal of Experimental & Clinical Cancer Research : CR*, 42(1), p. 236.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru and Raychaudhuri, S. (2019) “Fast, sensitive and accurate integration of single-cell data with Harmony,” *Nature Methods*, 16(12), pp. 1289–1296.

Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., Clancy, T.E., Chung, D.C., Lochhead, P., Hold, G.L., El-Omar, E.M., Brenner, D., Fuchs, C.S., Meyerson, M. and Garrett, W.S. (2013) “Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor immune microenvironment,” *Cell host & microbe*, 14(2), p. 207.

Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R.A., Ogino, S., Birren, B.W., Huttenhower, C., Garrett, W.S. and Meyerson, M. (2012) “Genomic analysis identifies association of Fusobacterium with colorectal carcinoma,” *Genome Research*, 22(2), p. 292.

Krieg, A., Kolbe, E.W., Kaspari, M., Krieg, S., Loosen, S.H., Roderburg, C. and Kostev, K. (2024) “Trends and outcomes in colorectal cancer surgery: a multicenter cross-sectional study of minimally invasive versus open techniques in Germany,” *Surgical Endoscopy*, 38(11), p. 6338.

Kriegl, L., Neumann, J., Vieth, M., Greten, F.R., Reu, S., Jung, A. and Kirchner, T. (2011) “Up and downregulation of p16Ink4a expression in BRAF-mutated polyps/adenomas indicates a senescence barrier in the serrated route to colon cancer,” *Modern Pathology 2011 24:7*, 24(7), pp. 1015–1022.

Kuwahara, T., Hazama, S., Suzuki, N., Yoshida, S., Tomochika, S., Nakagami, Y., Matsui, H., Shindo, Y., Kanekiyo, S., Tokumitsu, Y., Iida, M., Tsunedomi, R., Takeda, S., Yoshino, S., Okayama, N., Suehiro, Y., Yamasaki, T., Fujita, T., Kawakami, Y., Ueno, T. and Nagano, H. (2019) "Intratumoural-infiltrating CD4 + and FOXP3 + T cells as strong positive predictive markers for the prognosis of resectable colorectal cancer," *British Journal of Cancer*, 121(8), pp. 659–665.

Kyriakis, J.M., App, H., Zhang, X.F., Banerjee, P., Brautigan, D.L., Rapp, U.R. and Avruch, J. (1992) "Raf-1 activates MAP kinase-kinase," *Nature*, 358(6385), pp. 417–421.

Laish, I., Seregeev, I., Naftali, T. and Konikoff, F.M. (2017) "Surveillance after positive colonoscopy based on adenoma characteristics," *Digestive and Liver Disease*, 49(10), pp. 1115–1120.

Lane, D.P. and Benchimol, S. (1990) "P53: Oncogene or anti-oncogene?," *Genes and Development*, 4(1), pp. 1–8.

Langfelder, P. and Horvath, S. (2008) "WGCNA: An R package for weighted correlation network analysis," *BMC Bioinformatics*, 9(1), pp. 1–13.

Langfelder, P. and Horvath, S. (2012) "Fast R Functions for Robust Correlations and Hierarchical Clustering," *Journal of Statistical Software*, 46(11), pp. 1–17.

Larsen, N., Vogensen, F.K., Van Den Berg, F.W.J., Nielsen, D.S., Andreasen, A.S., Pedersen, B.K., Al-Soud, W.A., Sørensen, S.J., Hansen, L.H. and Jakobsen, M. (2010) "Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults," *PLoS ONE*, 5(2), p. e9085.

Lee, J.G., Park, I., Lee, H., Nam, S., Kim, J., Lee, W.S., Kang, M. and Kim, J.H. (2025) "Integrating E-cadherin expression levels with TNM staging for enhanced prognostic prediction in colorectal cancer patients," *BMC Cancer*, 25(1), pp. 1–10.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. (2012) "The sva package for removing batch effects and other unwanted variation in high-throughput experiments," *Bioinformatics*, 28(6), pp. 882–883.

Lennard, K.S., Goosen, R.W. and Blackburn, J.M. (2016) "Bacterially-Associated Transcriptional Remodelling in a Distinct Genomic Subtype of Colorectal Cancer Provides a Plausible Molecular Basis for Disease Development," *PLoS ONE*, 11(11), p. e0166282.

Leoz, M.L., Carballal, S., Moreira, L., Ocaña, T. and Balaguer, F. (2015) "The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management," *The Application of Clinical Genetics*, 8, p. 95.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. and Pfister, H. (2014) "UpSet: Visualization of intersecting sets," *IEEE Transactions on Visualization and Computer Graphics*, 20(12), pp. 1983–1992.

Li, B., Zhang, G. and Xu, X. (2023) "APC mutation correlated with poor response of immunotherapy in colon cancer," *BMC Gastroenterology*, 23(1), pp. 1–10. Available at: https://doi.org/10.1186/S12876-023-02725-3/FIGURES/7.
Li, G.M. (2007) "Mechanisms and functions of DNA mismatch repair," *Cell Research 2008 18:1*, 18(1), pp. 85–98.

Li, H. and Durbin, R. (2009) "Fast and accurate short read alignment with Burrows–Wheeler transform," *Bioinformatics*, 25(14), pp. 1754–1760.

Li, H., Zhao, Y. and Zheng, F. (2019) "Prognostic significance of elevated preoperative neutrophil-to-lymphocyte ratio for patients with colorectal cancer undergoing curative surgery: A meta-analysis," *Medicine*, 98(3), p. e14126.

Li, J., Yi, C.H., Hu, Y.T., Li, J.S., Yuan, Y., Zhang, S.Z., Zheng, S. and Ding, K.F. (2016) "TNM Staging of Colorectal Cancer Should be Reconsidered According to Weighting of the T Stage: Verification Based on a 25-Year Follow-Up," *Medicine*, 95(6), p. e2711.

Li, X., Huang, J., Yu, T., Fang, X., Lou, L., Xin, S., Ji, L., Jiang, F. and Lou, Y. (2020) "Fusobacterium nucleatum Promotes the Progression of Colorectal Cancer Through Cdk5-Activated Wnt/ß-Catenin Signaling," *Frontiers in Microbiology*, 11.

Li, Y., Li, X., Qu, J., Luo, D. and Hu, Z. (2020) "Cas9 Mediated Correction of ß-catenin Mutation and Restoring the Expression of Protein Phosphorylation in Colon Cancer HCT-116 Cells Decrease Cell Proliferation in vitro and Hamper Tumor Growth in Mice in vivo," *OncoTargets and therapy*, 13, p. 17.

Li, Ying, Zhao, J. and Li, Yintao (2025) "New exploration of KRASG12D inhibitors and the mechanisms of resistance," *Experimental Hematology and Oncology*, 14(1).

Liang, P.S., Chen, T.Y. and Giovannucci, E. (2009) "Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis," *International Journal of Cancer*, 124(10), pp. 2406–2415.

Liang, X., Duronio, G.N., Yang, Y., Bala, P., Hebbar, P., Spisak, S., Sahgal, P., Singh, H., Zhang, Y., Xie, Y., Cejas, P., Long, H.W., Bass, A.J. and Sethi, N.S. (2022) "An Enhancer-Driven Stem Cell–Like Program Mediated by SOX9 Blocks Intestinal Differentiation in Colorectal Cancer," *Gastroenterology*, 162(1), pp. 209–222.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. (2015) “The Molecular Signatures Database (MSigDB) hallmark gene set collection,” *Cell systems*, 1(6), p. 417.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P. (2011) “Molecular signatures database (MSigDB) 3.0,” *Bioinformatics*, 27(12), pp. 1739–1740.

Lim, J.C., Kania, K.D., Wijesuriya, H., Chawla, S., Sethi, J.K., Pulaski, L., Romero, I.A., Couraud, P.O., Weksler, B.B., Hladky, S.B. and Barrand, M.A. (2008) “Activation of ß-catenin signalling by GSK-3 inhibition increases p-glycoprotein expression in brain endothelial cells,” *Journal of neurochemistry*, 106(4), p. 1855.

Lin, X., Wang, W., Yang, M., Damseh, N., de Sousa, M.M.L., Jacob, F., Lång, A., Kristiansen, E., Pannone, M., Kissova, M., Almaas, R., Kuśnierczyk, A., Siller, R., Shahrour, M., Al-Ashhab, M., Abu-Libdeh, B., Tang, W., Slupphaug, G., Elpeleg, O., Bøe, S.O., Eide, L., Sullivan, G.J., Rinholm, J.E., Song, H., Ming, G. li, van Loon, B., Edvardson, S., Ye, J. and Bjørås, M. (2023) “A loss-of-function mutation in human Oxidation Resistance 1 disrupts the spatial–temporal regulation of histone arginine methylation in neurodevelopment,” *Genome Biology*, 24(1), pp. 1–34.

Linhares, M.M., Affonso, R.J., De Souza Viana, L., Silva, S.R.M., Denadai, M.V.A., De Toledo, S.R.C. and Matos, D. (2015) “Genetic and Immunohistochemical Expression of Integrins ITGAV, ITGA6, and ITGA3 As Prognostic Factor for Colorectal Cancer: Models for Global and Disease-Free Survival,” *PLoS ONE*, 10(12), p. e0144333.

Liu, G. jie, Wang, Y. jie, Yue, M., Zhao, L. mei, Guo, Y.D., Liu, Y. ping, Yang, H. chai, Liu, F., Zhang, X., Zhi, L. hui, Zhao, J., Sun, Y.H. and Wang, G. ying (2020) “High expression of TCN1 is a negative prognostic biomarker and can predict neoadjuvant chemosensitivity of colon cancer,” *Scientific Reports*, 10(1), pp. 1–11.

Liu, H., Du, J., Chao, S., Li, S., Cai, H., Zhang, H., Chen, G., Liu, P. and Bu, P. (2022) “Fusobacterium nucleatum Promotes Colorectal Cancer Cell to Acquire Stem Cell-Like Features by Manipulating Lipid Droplet-Mediated Numb Degradation,” *Advanced Science*, 9(12), p. 2105222.

Liu, J., Pan, S., Hsieh, M.H., Ng, N., Sun, F., Wang, T., Kasibhatla, S., Schuller, A.G., Li, A.G., Cheng, D., Li, J., Tompkins, C., Pferdekamper, A.M., Steffy, A., Cheng, J., Kowal, C., Phung, V., Guo, G., Wang, Y., Graham, M.P., Flynn, S., Brenner, J.C., Li, C., Villarroel, M.C., Schultz, P.G., Wu, X., McNamara, P., Sellers, W.R., Petruzzelli, L., Boral, A.L., Seidel, H.M., McLaughlin, M.E., Che, J., Carey, T.E., Vanasse, G. and Harris, J.L. (2013) “Targeting Wnt-driven cancer

through the inhibition of Porcupine by LGK974," *Proceedings of the National Academy of Sciences*, 110(50), pp. 20224–20229.

Liu, J., Wu, X., Mitchell, B., Kintner, C., Ding, S. and Schultz, P.G. (2005) "A small-molecule agonist of the Wnt signaling pathway," *Angewandte Chemie - International Edition*, 44(13), pp. 1987–1990.

Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., Zhou, Z., Shu, G. and Yin, G. (2022) "Wnt/ß-catenin signalling: function, biological mechanisms, and therapeutic opportunities," *Signal Transduction and Targeted Therapy 2021 7:1*, 7(1), pp. 1–23.

Liu, P., Liu, Y., Wang, J., Guo, Y., Zhang, Y. and Xiao, S. (2014) "Detection of Fusobacterium Nucleatum and fadA Adhesin Gene in Patients with Orthodontic Gingivitis and Non-Orthodontic Periodontal Inflammation," *PLoS ONE*, 9(1), p. e85280.

Liu, S. cheng, Gong, L. lan, Huang, F. chang, Xu, N., Yang, K. xin, Liu, X. hong and Li, W. liang (2023) "RNF114 facilitates the proliferation, stemness, and metastasis of colorectal cancer," *Pathology - Research and Practice*, 248, p. 154716.

Liu, Z.L., Chen, H.H., Zheng, L.L., Sun, L.P. and Shi, L. (2023) "Angiogenic signaling pathways and anti-angiogenic therapy for cancer," *Signal Transduction and Targeted Therapy 2023 8:1*, 8(1), pp. 1–39.

Lizárraga-Verdugo, E., Carmona, T.G., Ramos-Payan, R., Avendaño-Félix, M., Bermúdez, M., Parra-Niebla, M., López-Camarillo, C., Fernandez-Figueroa, E., Lino-Silva, L., Saavedra, H.A., Vela-Sarmiento, I., Ovando, R.C., Ruíz-García, E. and Aguilar-Medina, M. (2021) "SOX9 is associated with advanced T-stages of clinical stage II colon cancer in young Mexican patients," *Oncology Letters*, 22(1).

Llabata, P., Mitsuishi, Y., Choi, P.S., Cai, D., Francis, J.M., Torres-Diz, M., Udeshi, N.D., Golomb, L., Wu, Z., Zhou, J., Svinkina, T., Aguilera-Jimenez, E., Liu, Y., Carr, S.A., Sanchez-Cespedes, M., Meyerson, M. and Zhang, X. (2019) "Multi-Omics analysis identifies MGA as a negative regulator of the MYC pathway in lung adenocarcinoma," *Molecular cancer research : MCR*, 18(4), p. 574.

Løberg, M., Kalager, M., Holme, Ø., Hoff, G., Adami, H.-O. and Bretthauer, M. (2014) "Long-Term Colorectal-Cancer Mortality after Adenoma Removal," *New England Journal of Medicine*, 371(9), pp. 799–807.

Lopez-Bergami, P. and Barbero, G. (2020) "The emerging role of Wnt5a in the promotion of a pro-inflammatory and immunosuppressive tumor

microenvironment," *Cancer and Metastasis Reviews 2020 39:3*, 39(3), pp. 933–952.

Loree, J.M., Pereira, A.A.L., Lam, M., Willauer, A.N., Raghav, K., Dasari, A., Van Morris, K., Advani, S., Menter, D.G., Eng, C., Shaw, K., Broaddus, R., Routbort, M.J., Liu, Y., Morris, J.S., Luthra, R., Meric-Bernstam, F., Overman, M.J., Maru, D. and Kopetz, S. (2017) "Classifying colorectal cancer by tumor location rather than sidedness highlights a continuum in mutation profiles and Consensus Molecular Subtypes," *Clinical cancer research : an official journal of the American Association for Cancer Research*, 24(5), p. 1062.

Love, M.I., Huber, W. and Anders, S. (2014) "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, 15(12), pp. 1–21.

Ludwig, R.L., Bates, S. and Vousden, K.H. (1996) "Differential activation of target cellular promoters by p53 mutants with impaired apoptotic function," *Molecular and Cellular Biology*, 16(9), p. 4952.

MacDonald, B.T., Tamai, K. and He, X. (2009) "Wnt/β-catenin Signaling: Components, Mechanisms, and Diseases," *Developmental Cellell*, 17, p. 9.

Magnúsdóttir, S., Ravcheev, D., De Crécy-Lagard, V. and Thiele, I. (2015) "Systematic genome assessment of B-vitamin biosynthesis suggests cooperation among gut microbes," *Frontiers in Genetics*, 6(MAR), p. 148.

Maître, J.L. and Heisenberg, C.P. (2013) "Three Functions of Cadherins in Cell Adhesion," *Current Biology*, 23(14), p. R626.

Mao, H., Jia, J., Sheng, J., Zhang, S., Huang, K., Li, H. and He, F. (2021) "Protective and anti-inflammatory role of REG1A in inflammatory bowel disease induced by JAK/STAT3 signaling axis," *International Immunopharmacology*, 92, p. 107304.

Mao, J., Wang, J., Liu, B., Pan, W., Farr, G.H., Flynn, C., Yuan, H., Takada, S., Kimelman, D., Li, L. and Wu, D. (2001) "Low-Density Lipoprotein Receptor-Related Protein-5 Binds to Axin and Regulates the Canonical Wnt Signaling Pathway," *Molecular Cell*, 7(4), pp. 801–809.

Mao, X., Xu, J., Wang, W., Liang, C., Hua, J., Liu, J., Zhang, B., Meng, Q., Yu, X. and Shi, S. (2021) "Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives," *Molecular Cancer 2021 20:1*, 20(1), pp. 1–30.

Marchesi, J.R., Dutilh, B.E., Hall, N., Peters, W.H.M., Roelofs, R., Boleij, A. and Tjalsma, H. (2011) “Towards the Human Colorectal Cancer Microbiome,” *PLOS ONE*, 6(5), p. e20447.

Marei, H.E., Althani, A., Afifi, N., Hasan, A., Caceci, T., Pozzoli, G., Morrione, A., Giordano, A. and Cenciarelli, C. (2021) “p53 signaling in cancer progression and therapy,” *Cancer Cell International*, 21(1), pp. 1–15.

Marsell, R., Sisask, G., Nilsson, Y., Sundgren-Andersson, A.K., Andersson, U., Larsson, S., Nilsson, O., Ljunggren, Ö. and Jonsson, K.B. (2012) “GSK-3 inhibition by an orally active small molecule increases bone mass in rats,” *Bone*, 50(3), pp. 619–627.

Mathews, A.A., Draganov, P. V and Yang, D. (2021) “Endoscopic management of colorectal polyps: From benign to malignant polyps.,” *World journal of gastrointestinal endoscopy*, 13(9), pp. 356–370.

Matly, A., Quinn, J.A., McMillan, D.C., Park, J.H. and Edwards, J. (2021) “The relationship between ß-catenin and patient survival in colorectal cancer systematic review and meta-analysis,” *Critical Reviews in Oncology/Hematology*, 163, p. 103337.

Matly, A.A.M. (2024) *Investigating the association between epithelial-mesenchymal transition pathway related proteins and patients’ outcomes in colorectal cancer*. University of Glasgow. [Thesis].

Matthews, H.K., Bertoli, C. and de Bruin, R.A.M. (2021) “Cell cycle control in cancer,” *Nature Reviews Molecular Cell Biology 2021 23:1*, 23(1), pp. 74–88.

Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C. and Koeffler, H.P. (2018) “Maftools: efficient and comprehensive analysis of somatic variants in cancer,” *Genome Research*, 28(11), pp. 1747–1756.

McCoy, A.N., Araújo-Pérez, F., Azcárate-Peril, A., Yeh, J.J., Sandler, R.S. and Keku, T.O. (2013) “Fusobacterium Is Associated with Colorectal Adenomas,” *PLoS ONE*, 8(1), p. e53653.

Meester, R.G.S., Mannalithara, A., Lansdorp-Vogelaar, I. and Ladabaum, U. (2019) “Trends in Incidence and Stage at Diagnosis of Colorectal Cancer in Adults Aged 40 Through 49 Years, 1975-2015,” *JAMA*, 321(19), p. 1933.

Meng, M., Zhong, K., Jiang, T., Liu, Z., Kwan, H.Y. and Su, T. (2021) “The current understanding on the impact of KRAS on colorectal cancer,” *Biomedicine & Pharmacotherapy*, 140, p. 111717.

Menon, A.P., Moreno, B., Meraviglia-Crivelli, D., Nonatelli, F., Villanueva, H., Barainka, M., Zheleva, A., van Santen, H.M. and Pastor, F. (2023) “Modulating T Cell Responses by Targeting CD3,” *Cancers*, 15(4), p. 1189.

Meuer, S.C., Schlossman, S.F. and Reinherz, E.L. (1982) “Clonal analysis of human cytotoxic T lymphocytes: T4+ and T8+ effector T cells recognize products of different major histocompatibility complex regions,” *Proceedings of the National Academy of Sciences of the United States of America*, 79(14), p. 4395.

Milburn, Michael V, Tong, Liang, Devos, A.M., Brunger, A., Yamaizumi, Ziro, Nishimura, Susumu, Kim, S., Milburn, M V, Tong, L, deVos, A.M., Kim are, S., Yamaizumi, Z and Nishimura, S (1990) “Molecular Switch for Signal Transduction: Structural Differences Between Active and Inactive Forms of Protooncogenic ras Proteins,” *Science*, 247(4945), pp. 939–945.

Miller, D.M., Thomas, S.D., Islam, A., Muench, D. and Sedoris, K. (2012) “c-Myc and Cancer Metabolism,” *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18(20), p. 5546.

Mima, K., Cao, Y., Chan, A.T., Qian, Z.R., Nowak, J.A., Masugi, Y., Shi, Y., Song, M., Da Silva, A., Gu, M., Li, W., Hamada, T., Kosumi, K., Hanyuda, A., Liu, L., Kostic, A.D., Giannakis, M., Bullman, S., Brennan, C.A., Milner, D.A., Baba, H., Garraway, L.A., Meyerhardt, J.A., Garrett, W.S., Huttenhower, C., Meyerson, M., Giovannucci, E.L., Fuchs, C.S., Nishihara, R. and Ogino, S. (2016) “Fusobacterium nucleatum in Colorectal Carcinoma Tissue According to Tumor Location,” *Clinical and Translational Gastroenterology*, 7(11), p. e200.

Mima, K., Nishihara, R., Qian, Z.R., Cao, Y., Sukawa, Y., Nowak, J.A., Yang, J., Dou, R., Masugi, Y., Song, M., Kostic, A.D., Giannakis, M., Bullman, S., Milner, D.A., Baba, H., Giovannucci, E.L., Garraway, L.A., Freeman, G.J., Dranoff, G., Garrett, W.S., Huttenhower, C., Meyerson, M., Meyerhardt, J.A., Chan, A.T., Fuchs, C.S. and Ogino, S. (2016) “Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis,” *Gut*, 65(12), pp. 1973–1980.

Mima, K., Sukawa, Y., Nishihara, R., Qian, Z.R., Yamauchi, M., Inamura, K., Kim, S.A., Masuda, A., Nowak, J.A., Nosho, K., Kostic, A.D., Giannakis, M., Watanabe, H., Bullman, S., Milner, D.A., Harris, C.C., Giovannucci, E., Garraway, L.A., Freeman, G.J., Dranoff, G., Chan, A.T., Garrett, W.S., Huttenhower, C., Fuchs, C.S. and Ogino, S. (2015) “Fusobacterium nucleatum and T-cells in Colorectal Carcinoma,” *JAMA oncology*, 1(5), p. 653.

Mimura, K., Teh, J.L., Okayama, H., Shiraishi, K., Kua, L.F., Koh, V., Smoot, D.T., Ashktorab, H., Oike, T., Suzuki, Y., Fazreen, Z., Asuncion, B.R., Shabbir, A., Yong, W.P., So, J., Soong, R. and Kono, K. (2017) “PD-L1 expression is mainly regulated by interferon gamma associated with JAK-STAT pathway in gastric cancer,” *Cancer Science*, 109(1), p. 43.

Mirkin, K.A., Kulaylat, A.S., Hollenbeak, C.S. and Messaris, E. (2018) “Robotic versus laparoscopic colectomy for stage I–III colon cancer: oncologic and long-term survival outcomes,” *Surgical Endoscopy*, 32(6), pp. 2894–2901.

Mizutani, S., Yamada, T. and Yachida, S. (2020) “Significance of the gut microbiome in multistep colorectal carcinogenesis,” *Cancer Science*, 111(3), p. 766.

Monreal-Robles, R., Jáquez-Quintana, J.O., Benavides-Salgado, D.E. and González-González, J.A. (2021) “Serrated polyps of the colon and rectum: a concise review,” *Revista de Gastroenterología de México (English Edition)*, 86(3), pp. 276–286.

Moreira, L., Balaguer, F., Lindor, N., De La Chapelle, A., Hampel, H., Aaltonen, L.A., Hopper, J.L., Le Marchand, L., Gallinger, S., Newcomb, P.A., Haile, R., Thibodeau, S.N., Gunawardena, S., Jenkins, M.A., Buchanan, D.D., Potter, J.D., Baron, J.A., Ahnen, D.J., Moreno, V., Andreu, M., De Leon, M.P., Rustgi, A.K. and Castells, A. (2012) “Identification of Lynch Syndrome Among Patients With Colorectal Cancer,” *JAMA : the journal of the American Medical Association*, 308(15), p. 10.1001/jama.2012.13088.

Morgan, E., Arnold, M., Gini, A., Lorenzoni, V., Cabasag, C.J., Laversanne, M., Vignat, J., Ferlay, J., Murphy, N. and Bray, F. (2023) “Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN,” *Gut*, 72(2), pp. 338–344.

Morgan, M. and Ramos, M. (2023) “BiocManager: Access the Bioconductor Project Package Repository.” Available at: https://CRAN.R-project.org/package=BiocManager.

Morin, P.J., Vogelstein, B. and Kinzler, K.W. (1996) “Apoptosis and APC in colorectal tumorigenesis,” *Proceedings of the National Academy of Sciences of the United States of America*, 93(15), p. 7950.

Moy, B.T., Forouhar, F., Kuo, C.L. and Devers, T.J. (2018) “Endoscopic Features of Mucous Cap Polyps: A Way to Predict Serrated Polyps,” *Clinical Endoscopy*, 51(4), p. 368.

Münz, M., Ruark, E., Renwick, A., Ramsay, E., Clarke, M., Mahamdallie, S., Cloke, V., Seal, S., Strydom, A., Lunter, G. and Rahman, N. (2015) “CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting,” *Genome Medicine*, 7(1), p. 76.

Murnane, M.J., Cai, J., Shuja, S., McAneny, D. and Willett, J.B. (2011) “Active MMP-2 Activity Discriminates Colonic Mucosa, Adenomas with and without High Grade Dysplasia and Cancers,” *Human pathology*, 42(5), p. 688.

Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., *et* al. (2012) “Comprehensive molecular characterization of human colon and rectal cancer,” *Nature*, 487(7407), p. 330.

Naccarati, A., Polakova, V., Pardini, B., Vodickova, L., Hemminki, K., Kumar, R. and Vodicka, P. (2012) “Mutations and polymorphisms in TP53 gene—an overview on the role in colorectal cancer,” *Mutagenesis*, 27(2), pp. 211–218.

Nagao, S., Takahashi, Y., Denda, T., Tanaka, Y., Miura, Y., Mizutani, H., Ohki, D., Sakaguchi, Y., Yakabi, S., Tsuji, Y., Niimi, K., Kakushima, N., Yamamichi, N., Ota, Y., Koike, K. and Fujishiro, M. (2023) “Reduced DEFA5 Expression and STAT3 Activation Underlie the Submucosal Invasion of Early Gastric Cancers,” *Digestion*, 104(6), p. 480.

Negishi, M., Chiba, T., Saraya, A., Miyagi, S. and Iwama, A. (2009) “Dmap1 plays an essential role in the maintenance of genome integrity through the DNA repair process,” *Genes to Cells*, 14(11), pp. 1347–1357.

Ni, C., Wang, X., Liu, S., Zhang, J., Luo, Z. and Xu, B. (2024) “KMT2C mutation as a predictor of immunotherapeutic efficacy in colorectal cancer,” *Scientific Reports*, 14(1), pp. 1–10.

Nishimura, H., Fukui, H., Wang, X., Ebisutani, N., Nakanishi, T., Tomita, T., Oshima, T., Hirota, S. and Miwa, H. (2021) “Role of the ß-Catenin/REG Iα Axis in the Proliferation of Sessile Serrated Adenoma/Polyps Associated with Fusobacterium nucleatum,” *Pathogens*, 10(4), p. 434.

Novellasdemunt, L., Antas, P. and Li, V.S.W. (2015) “Targeting Wnt signaling in colorectal cancer. A Review in the Theme: Cell Signaling: Proteins, Pathways and Mechanisms,” *American Journal of Physiology - Cell Physiology*, 309(8), p. C511.

Nunes, L., Li, F., Wu, M., Luo, T., Hammarström, K., Torell, E., Ljuslinder, I., Mezheyeuski, A., Edqvist, P.H., Löfgren-Burström, A., Zingmark, C., Edin, S., Larsson, C., Mathot, L., Osterman, E., Osterlund, E., Ljungström, V., Neves, I., Yacoub, N., Guðnadóttir, U., Birgisson, H., Enblad, M., Ponten, F., Palmqvist, R., Xu, X., Uhlén, M., Wu, K., Glimelius, B., Lin, C. and Sjöblom, T. (2024) “Prognostic genome and transcriptome signatures in colorectal cancers,” *Nature 2024 633:8028*, 633(8028), pp. 137–146.

Nusse, R., Van Ooyen, A., Cox, D., Fung, Y.K.T. and Varmus, H. (1984) “Mode of proviral activation of a putative mammary oncogene (int-1) on mouse chromosome 15,” *Nature*, 307(5947), pp. 131–136.

Obuch, J.C., Pigott, C.M. and Ahnen, D.J. (2015) “Sessile Serrated Polyps: Detection, Eradication, and Prevention of the Evil Twin,” *Current treatment options in gastroenterology*, 13(1), p. 156.

Ogobuiro, I., Gonzales, J., Shumway, K.R. and Tuma, F. (2023) “Physiology, Gastrointestinal,” *StatPearls* [Preprint].

Ola, I., Cardoso, R., Hoffmeister, M. and Brenner, H. (2024) “Utilization of colorectal cancer screening tests across European countries: a cross-sectional analysis of the European health interview survey 2018–2020,” *The Lancet Regional Health - Europe*, 41, p. 100920.

Olkinuora, A.P., Peltomäki, P.T., Aaltonen, L.A. and Rajamäki, K. (2021) “From APC to the genetics of hereditary and familial colon cancer syndromes,” *Human Molecular Genetics*, 30(R2), p. R206.

Orange, S.T., Leslie, J., Ross, M., Mann, D.A. and Wackerhage, H. (2023) “The exercise IL-6 enigma in cancer,” *Trends in Endocrinology & Metabolism*, 34(11), pp. 749–763.

Ozawa, M., Barbault, H. and Kemler, R. (1989) “The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species,” *The EMBO Journal*, 8(6), p. 1711.

Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., *et* al. (2012) “Germline mutations in the proof-reading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas,” *Nature genetics*, 45(2), p. 136.

Park, J.H., Kim, M.J., Park, S.C., Hong, C.W., Sohn, D.K., Han, K.S. and Oh, J.H. (2015) “Difference in Time to Locoregional Recurrence between Patients with Right-Sided and Left-Sided Colon Cancers,” *Diseases of the Colon and Rectum*, 58(9), pp. 831–837.

Parker, T.W. and Neufeld, K.L. (2020) “APC controls Wnt-induced ß-catenin destruction complex recruitment in human colonocytes,” *Scientific Reports*, 10, p. 2957.

Participants in the Paris Workshop (2003) “The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002,” *Gastrointestinal Endoscopy*, 58(6), pp. S3–S43.

Patel, R., McGinty, P., Cuthill, V., Hawkins, M., Clark, S.K. and Latchford, A. (2021) “Risk of colorectal adenomas and cancer in monoallelic carriers of MUTYH pathogenic variants: a single-centre experience,” *International Journal of Colorectal Disease*, 36(10), p. 2199.

Pedersen, A.M.L., Sørensen, C.E., Proctor, G.B. and Carpenter, G.H. (2018) “Salivary functions in mastication, taste and textural perception, swallowing and initial digestion,” *Oral Diseases*, 24(8), pp. 1399–1416.

Pellatt, A.J., Bhamidipati, D. and Subbiah, V. (2024) “Ready, Set, Go: Setting Off on the Mission to Target KRAS in Colorectal Cancer,” *JCO Oncology Practice*, 20(10), pp. 1289–1292.

Peng, H., Ying, J., Zang, J., Lu, H., Zhao, X., Yang, P., Wang, X., Li, J., Gong, Z., Zhang, D. and Wang, Z. (2023) “Specific Mutations in APC, with Prognostic Implications in Metastatic Colorectal Cancer,” *Cancer Research and Treatment*, 55(4), pp. 1270–1280.

Perne, C., Peters, S., Cartolano, M., Horpaopan, S., Grimm, C., Altmüller, J., Sommer, A.K., Hillmer, A.M., Thiele, H., Odenthal, M., Möslein, G., Adam, R., Sivalingam, S., Kirfel, J., Schweiger, M.R., Peifer, M., Spier, I. and Aretz, S. (2021) “Variant profiling of colorectal adenomas from three patients of two families with MSH3-related adenomatous polyposis,” *PLoS ONE*, 16(11), p. e0259185.

Pickhardt, P.J., Hain, K.S., Kim, D.H. and Hassan, C. (2010) “Low Rates of Cancer or High-Grade Dysplasia in Colorectal Polyps Collected From Computed Tomography Colonography Screening,” *Clinical Gastroenterology and Hepatology*, 8(7), pp. 610–615.

Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H.M.,*et* al. (2020) “Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli,” *Nature*, 580(7802), p. 269.

Politis, D.S., Katsanos, K.H., Tsianos, E. V. and Christodoulou, D.K. (2017) “Pseudopolyps in inflammatory bowel diseases: Have we learned enough?,” *World Journal of Gastroenterology*, 23(9), p. 1541.

Ponder, B.A. and Wilkinson, M.M. (1986) “Direct examination of the clonality of carcinogen-induced colonic epithelial dysplasia in chimeric mice,” *Journal of National Cancer Institute*, 77(4), pp. 967–976.

Prévostel, C., Rammah-Bouazza, C., Trauchessec, H., Canterel-Thouennon, L., Busson, M., Ychou, M. and Blache, P. (2016) “SOX9 is an atypical intestinal tumor suppressor controlling the oncogenic Wnt/ß-catenin signaling,” *Oncotarget*, 7(50), p. 82228.

Principe, D.R. (2022) “Patients deriving long-term benefit from immune checkpoint inhibitors demonstrate conserved patterns of site-specific mutations,” *Scientific Reports*, 12(1), pp. 1–13.

Prior, I.A., Hood, F.E. and Hartley, J.L. (2020) “The frequency of Ras mutations in cancer,” *Cancer research*, 80(14), p. 2969.

Puré, E. and Blomberg, R. (2018) “Pro-tumorigenic roles of fibroblast activation protein in cancer: back to the basics,” *Oncogene*, 37(32), p. 4343.

Pyysalo, M.J., Pyysalo, L.M., Pessi, T., Karhunen, P.J. and Öhman, J.E. (2013) “The connection between ruptured cerebral aneurysms and odontogenic bacteria,” *Journal of Neurology, Neurosurgery & Psychiatry*, 84(11), pp. 1214–1218.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Yanfei, Shao, L., Guo, J., Le Chatelier, E., Yao, J., Wu, L., Zhou, J., Ni, S., Liu, L., Pons, N., Batto, J.M., Kennedy, S.P., Leonard, P., Yuan, C., Ding, W., Chen, Yuanting, Hu, X., Zheng, B., Qian, G., Xu, W., Ehrlich, S.D., Zheng, S. and Li, L. (2014) “Alterations of the human gut microbiome in liver cirrhosis.,” *Nature*, 513(7516), pp. 59–64.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2013) “The SILVA ribosomal RNA gene database project: improved data processing and web-based tools,” *Nucleic Acids Research*, 41(D1), pp. D590–D596.

R Core Team (2021) “R: A language and environment for statistical computing.” Vienna, Austria: R Foundation for Statistical Computing.

Rai, V. and Mishra, N. (2017) “Colonoscopic Perforations,” *Clinics in Colon and Rectal Surgery*, 31(1), p. 41.

Rajan, A., Kim, C., Heery, C.R., Guha, U. and Gulley, J.L. (2016) “Nivolumab, anti-programmed death-1 (PD-1) monoclonal antibody immunotherapy: Role in advanced cancers,” *Human Vaccines & Immunotherapeutics*, 12(9), p. 2219.

Raji, R.J., Sasikumar, R. and Jacob, E. (2018) “Multiple roles of Adenomatous Polyposis Coli gene in Wnt Signalling – a Computational Model,” *Biosystems*, 172, pp. 26–36.

Raskov, H., Orhan, A., Christensen, J.P. and Gögenur, I. (2020) “Cytotoxic CD8+ T cells in cancer and cancer immunotherapy,” *British Journal of Cancer 2020 124:2*, 124(2), pp. 359–367.

Ravindran, S., Bassett, P., Shaw, T., Dron, M., Broughton, R., Johnston, D., Healey, C.J., Green, J., Ashrafian, H., Darzi, A., Coleman, M. and Thomas-Gibson, S. (2021) “National census of UK endoscopy services in 2019,” *Frontline Gastroenterology*, 12(6), pp. 451–460.

Ravindran, S., Thomas-Gibson, S., Bano, M., Robinson, E., Jenkins, A., Marshall, S., Ashrafian, H., Darzi, A., Coleman, M. and Healey, C. (2022) “National census of UK endoscopy services 2021,” *Frontline Gastroenterology*, 13(6), p. 463.

Renckens, R., Roelofs, J.J.T.H., Bonta, P.I., Florquin, S., De Vries, C.J.M., Levi, M., Carmeliet, P., Van’t Veer, C. and Van Der Poll, T. (2007) “Plasminogen activator inhibitor type 1 is protective during severe Gram-negative pneumonia,” *Blood*, 109(4), pp. 1593–1601.

Reuter, J.S. and Mathews, D.H. (2010) “RNAstructure: software for RNA secondary structure prediction and analysis,” *BMC Bioinformatics*, 11, p. 129.

Rhoades, N.S., Pinski, A.N., Monsibais, A.N., Jankeel, A., Doratt, B.M., Cinco, I.R., Ibraim, I. and Messaoudi, I. (2021) “Acute SARS-CoV-2 infection is associated with an increased abundance of bacterial pathogens, including Pseudomonas aeruginosa in the nose,” *Cell Reports*, 36(9), p. 109637.

Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D. and Nusse, R. (1987) “The Drosophila homology of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless,” *Cell*, 50(4), pp. 649–657.

Ring, D.B., Johnson, K.W., Henriksen, E.J., Nuss, J.M., Goff, D., Kinnick, T.R., Ma, S.T., Reeder, J.W., Samuels, I., Slabiak, T., Wagman, A.S., Hammond, M.E.W. and Harrison, S.D. (2003) “Selective Glycogen Synthase Kinase 3 Inhibitors Potentiate Insulin Activation of Glucose Transport and Utilization In Vitro and In Vivo,” *Diabetes*, 52(3), pp. 588–595.

Rodriguez, M.S., Egaña, I., Lopitz-Otsoa, F., Aillet, F., Lopez-Mato, M.P., Dorronroso, A., Lobato-Gil, S., Sutherland, J.D., Barrio, R., Trigueros, C. and Lang, V. (2014) “The RING ubiquitin E3 RNF114 interacts with A20 and modulates NF-κB activity and T-cell activation,” *Cell Death & Disease 2014 5:8*, 5(8), pp. e1399–e1399.

Romero, P. and Coulie, P.G. (2014) “Adaptive T-cell immunity and tumor antigen recognition,” *Tumor Immunology and Immunotherapy*, pp. 1–14.

Van Roon, A.H.C., Hol, L., Van Vuuren, A.J., Francke, J., Ouwendijk, M., Heijens, A., Nagtzaam, N., Reijerink, J.C.I.Y., Van Der Togt, A.C.M., Van Ballegooijen, M., Kuipers, E.J. and Van Leerdam, M.E. (2012) “Are fecal immunochemical test characteristics influenced by sample return time a

population-based colorectal cancer screening trial," *American Journal of Gastroenterology*, 107(1), pp. 99–107.

Rosen, Ryan.D. and Sapra, A. (2023) *TNM Classification - StatPearls - NCBI Bookshelf*, *Treasure Island*.

Rubinstein, M.R., Baik, J.E., Lagana, S.M., Han, R.P., Raab, W.J., Sahoo, D., Dalerba, P., Wang, T.C. and Han, Y.W. (2019) "Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1," *EMBO Reports*, 20(4).

Rubinstein, M.R., Wang, X., Liu, W., Hao, Y., Cai, G. and Han, Y.W. (2013) "Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin," *Cell host & microbe*, 14(2), p. 195.

Ruijtenberg, S. and van den Heuvel, S. (2016) "Coordinating cell proliferation and differentiation: Antagonism between cell cycle regulators and cell type-specific gene expression," *Cell Cycle*, 15(2), p. 196.

Rutter, M.D., East, J., Rees, C.J., Cripps, N., Docherty, J., Dolwani, S., Kaye, P. V., Monahan, K.J., Novelli, M.R., Plumb, A., Saunders, B.P., Thomas-Gibson, S., Tolan, D.J.M., Whyte, S., Bonnington, S., Scope, A., Wong, R., Hibbert, B., Marsh, J., Moores, B., Cross, A. and Sharp, L. (2020) "British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines," *Gut*, 69, pp. 201–223.

Rutter, M.D., Ranjan, R., Westwood, C., Barbour, J., Biran, A., Blackett, H., Burr, N.E., Carlisle, J., Clare, B., Cripps, N., Coyne, P., Dolwani, S., Hodson, R., Holtham, S., Mohammed, N., Morris, E.J.A., Neilson, L., Oliphant, R., Painter, J., Prakash, A., Pullan, R., Sarkar, S., Sloan, M., Swart, M., Thomas-Gibson, S., Trudgill, N.J., Vance, M., Yeadon, K. and Sharp, L. (2025) "BSG/ACPGBI guidance on the management of colorectal polyps in patients with limited life expectancy," *Gut*, 0, pp. 1–10.

De Ruysscher, D., Niedermann, G., Burnet, N.G., Siva, S., Lee, A.W.M. and Hegi-Johnson, F. (2019) "Radiotherapy toxicity," *Nature Reviews Disease Primers 2019 5:1*, 5(1), pp. 1–20.

Săftoiu, A., Hassan, C., Areia, M., Bhutani, M.S., Bisschops, R., Bories, E., Cazacu, I.M., Dekker, E., Deprez, P.H., Pereira, S.P., Senore, C., Capocaccia, R., Antonelli, G., Van Hooft, J., Messmann, H., Siersema, P.D., Dinis-Ribeiro, M. and Ponchon, T. (2020) "Role of gastrointestinal endoscopy in the screening of digestive tract cancers in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement," *Endoscopy*, 52(4), pp. 293–304.

Samadder, N.J., Baffy, N., Giridhar, K. V., Couch, F.J. and Riegert-Johnson, D. (2019) “Hereditary Cancer Syndromes—A Primer on Diagnosis and Management, Part 2: Gastrointestinal Cancer Syndromes,” *Mayo Clinic Proceedings*, 94(6), pp. 1099–1116.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. and Regev, A. (2015) “Spatial reconstruction of single-cell gene expression data,” *Nature Biotechnology*, 33(5), pp. 495–502.

Sawicki, T., Ruszkowska, M., Danielewicz, A., Niedźwiedzka, E., Arłukowicz, T. and Przybyłowicz, K.E. (2021) “A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms and Diagnosis,” *Cancers*, 13(9), p. 2025.

Saygun, I., Nizam, N., Keskiner, I., Bal, V., Kubar, A., Açikel, C., Serdar, M. and Slots, J. (2011) “Salivary infectious agents and periodontal disease status,” *Journal of Periodontal Research*, 46(2), pp. 235–239.

Schmöle, A.C., Brennführer, A., Karapetyan, G., Jaster, R., Pews-Davtyan, A., Hübner, R., Ortinau, S., Beller, M., Rolfs, A. and Frech, M.J. (2010) “Novel indolylmaleimide acts as GSK-3β inhibitor in human neural progenitor cells,” *Bioorganic & Medicinal Chemistry*, 18(18), pp. 6785–6795.

Schneikert, J. and Behrens, J. (2007) “The canonical Wnt signalling pathway and its APC partner in colon cancer development,” *Gut*, 56(3), p. 417.

Scolnick, E.M., Rands, E., Williams, D. and Parks, W.P. (1973) “Studies on the Nucleic Acid Sequences of Kirsten Sarcoma Virus: a Model for Formation of a Mammalian RNA-Containing Sarcoma Virus,” *Journal of Virology*, 12(3), p. 458.

Scott, A.J., Alexander, J.L., Merrifield, C.A., Cunningham, D., Jobin, C., Brown, R., Alverdy, J., O’Keefe, S.J., Gaskins, H.R., Teare, J., Yu, J., Hughes, D.J., Verstraelen, H., Burton, J., O’Toole, P.W., Rosenberg, D.W., Marchesi, J.R. and Kinross, J.M. (2019) “International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis,” *Gut*, 68(9), p. 1624.

Segev, L., Kalady, M.F. and Church, J.M. (2018) “Left-sided dominance of early-onset colorectal cancers: A rationale for screening flexible sigmoidoscopy in the young,” *Diseases of the Colon and Rectum*, 61(8), pp. 897–902.

Serra, D.O., Richter, A.M. and Hengge, R. (2013) “Cellulose as an Architectural Element in Spatially Structured Escherichia coli Biofilms,” *Journal of Bacteriology*, 195(24), p. 5540.

Shah, M., Rennoll, S.A., Raup-Konsavage, W.M. and Yochum, G.S. (2015) “A dynamic exchange of TCF3 and TCF4 transcription factors controls MYC expression in colorectal cancer cells,” *Cell Cycle*, 14(3), p. 323.

Shen, Z., Deng, H., Fang, Y., Zhu, X., Ye, G., Yan, L., Liu, H. and Li, G. (2015) “Identification of the interplay between SOX9 and S100P in the metastasis and invasion of colon carcinoma,” *Oncotarget*, 6(24), p. 20672.

Shoari, A., Ashja Ardalan, A., Dimesa, A.M. and Coban, M.A. (2024) “Targeting Invasion: The Role of MMP-2 and MMP-9 Inhibition in Colorectal Cancer Therapy,” *Biomolecules*, 15(1), p. 35.

Shussman, N. and Wexner, S.D. (2014) “Colorectal polyps and polyposis syndromes,” *Gastroenterology Report*, 2(1), p. 1.

Sidebottom, A.M. (2023) “A Brief History of Microbial Study and Techniques for Exploring the Gastrointestinal Microbiome,” *Clinics in Colon and Rectal Surgery*, 36(2), p. 98.

Simon, K. (2016) “Colorectal cancer development and advances in screening,” *Clinical Interventions in Aging*, 11, pp. 967–976.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R. and Goodfellow, P.N. (1990) “A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif,” *Nature*, 346(6281), pp. 240–244.

Singh, H. and Gallier, S. (2014) “Food Structures in the Gastrointestinal Tract and Physiological Responses,” in M. Boland, M. Golding, and H. Singh (eds.) *Food Structures, Digestion and Health*. Palmerston North: Academic Press, pp. 51–81.

Singhal, S., Schokrpur, S., Gandara, D. and Riess, J.W. (2024) “KRAS G12C Inhibitors: New Drugs, A New Hope,” *Journal of Thoracic Oncology*, 19(12), pp. 1594–1598.

Sinha, A., Fan, V.B., Ramakrishnan, A.B., Engelhardt, N., Kennell, J. and Cadigan, K.M. (2021) “Repression of Wnt/β-catenin signaling by SOX9 and Mastermind-like transcriptional coactivator 2,” *Science Advances*, 7, p. eabe0849.

Sparks Stein, P., Steffen, M.J., Smith, C., Jicha, G., Ebersole, J.L., Abner, E. and Dawson, D. (2012) “Serum antibodies to periodontal pathogens are a risk factor for Alzheimer’s disease,” *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*, 8(3), p. 196.

Srivastava, N.S. and Srivastava, R.A.K. (2019) “Curcumin and quercetin synergistically inhibit cancer cell proliferation in multiple cancer cells and modulate Wnt/β-catenin signaling and apoptotic pathways in A375 cells,” *Phytomedicine*, 52, pp. 117–128.

Stagg, B., Foster-Smith, E. and Tam, P.C.K. (2024) “The diagnostic accuracy of Gram stain on formalin-fixed paraffin-embedded sections of skin tissue in the diagnosis of bacterial skin infection,” *Journal of Cutaneous Pathology*, 51(3), pp. 246–250.

Stoffel, E.M., Koeppe, E., Everett, J., Ulintz, P., Kiel, M., Osborne, J., Williams, L., Hanson, K., Gruber, S.B. and Rozek, L.S. (2018) “Germline Genetic Features of Young Individuals With Colorectal Cancer,” *Gastroenterology*, 154(4), pp. 897-905.e1.

Stothard, P. (2000) “The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences,” *BioTechniques*, 28(6), pp. 1102–1104.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019) “Comprehensive Integration of Single-Cell Data,” *Cell*, 177(7), pp. 1888-1902.e21.

Su, H., Cai, T., Zhang, S., Yan, X., Zhou, L., He, Z., Xue, P., Li, J., Zheng, M., Yang, X. and Feng, B. (2021) “Identification of hub genes associated with neutrophils infiltration in colorectal cancer,” *Journal of Cellular and Molecular Medicine*, 25(7), p. 3371.

Su, X., Wang, X., Lai, J., Mao, S. and Li, H. (2024a) “Unraveling a novel hippo-associated immunological prognostic signature: The contribution of SERPINE1 in facilitating colorectal cancer progression via the notch signaling pathway,” *Genomics*, 116(2), p. 110794.

Su, X., Wang, X., Lai, J., Mao, S. and Li, H. (2024b) “Unraveling a novel hippo-associated immunological prognostic signature: The contribution of SERPINE1 in facilitating colorectal cancer progression via the notch signaling pathway,” *Genomics*, 116(2), p. 110794.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. (2005) “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15545–15550.

Sullivan, K.D., Galbraith, M.D., Andrysik, Z. and Espinosa, J.M. (2017) “Mechanisms of transcriptional regulation by p53,” *Cell Death & Differentiation 2018 25:1*, 25(1), pp. 133–143.

Sun, F., Wei, Y., Liu, Z., Jie, Q., Yang, X., Long, P., Wang, J., Xiong, Y., Li, Q., Quan, S. and Ma, Y. (2022) “Acylglycerol kinase promotes ovarian cancer progression and regulates mitochondria function by interacting with ribosomal protein L39,” *Journal of Experimental and Clinical Cancer Research*, 41(1), pp. 1–18.

Sun, Y., You, Y., Wu, Q., Hu, R. and Dai, K. (2024) “Genetically inspired organoids prevent joint degeneration and alleviate chondrocyte senescence via Col11a1–HIF1α-mediated glycolysis–OXPHOS metabolism shift,” *Clinical and Translational Medicine*, 14(2), p. e1574.

Świątkowski, F., Górnicki, T., Bułdyś, K. and Chabowski, M. (2022) “The Quality of Life of Patients with Surgically Treated Colorectal Cancer: A Narrative Review,” *Journal of Clinical Medicine*, 11(20), p. 6211.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J. and Von Mering, C. (2022) “The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest,” *Nucleic Acids Research*, 51(D1), p. D638.

Tabernero, J., Grothey, A., van Cutsem, E., Yaeger, R., Wasan, H., Yoshino, T., Desai, J., Ciardiello, F., Loupakis, F., Hong, Y.S., Steeghs, N., Guren, T.K., Arkenau, H.T., Garcia-Alfonso, P., Elez, E., Gollerkeri, A., Maharry, K., Christy-Bittel, J. and Kopetz, S. (2021) “Encorafenib Plus Cetuximab as a New Standard of Care for Previously Treated BRAF V600E–Mutant Metastatic Colorectal Cancer: Updated Survival Results and Subgroup Analyses from the BEACON Study,” *Journal of Clinical Oncology*, 39(4), p. 273.

Tang, B., Wang, K., Jia, Y.P., Zhu, P., Fang, Y., Zhang, Z.J., Mao, X., Li, Q. and Zeng, D.Z. (2016) “Fusobacterium nucleatum-Induced Impairment of Autophagic Flux Enhances the Expression of Proinflammatory Cytokines via ROS in Caco-2 Cells,” *PLOS ONE*, 11(11), p. e0165701.

Terazono, K., Yamamoto, H., Takasawa, S., Shiga, K., Yonemura, Y., Tochino, Y. and Okamoto, H. (1988) “A novel gene activated in regenerating islets.,” *Journal of Biological Chemistry*, 263(5), pp. 2111–2114.

Thiagalingam, S., Lengauer, C., Leach, F.S., Schutte, M., Hahn, S.A., Overhauser, J., Willson, J.K.V., Markowitz, S., Hamilton, S.R., Kern, S.E., Kinzler, K.W. and Vogelstein, B. (1996) “Evaluation of candidate tumour

suppressor genes on chromosome 18 in colorectal cancers," *Nature Genetics*, 13(3), pp. 343–346.

Thomas, C.E., Takashima, Y., Buchanan, D.D., Wesselink, E., Qu, C., Hsu, L., Costa, A.D., Gallinger, S., Grant, R.C., Huyghe, J.R., Thomas, S., Ugai, S., Zhong, Y., Matsuda, K., Ugai, T., Peters, U., Ogino, S., Nowak, J.A. and Phipps, A.I. (2025) "Density of T-cell Subsets in Colorectal Cancer in Relation to Disease-Specific Survival," *Cancer Epidemiology Biomarkers and Prevention*, 34(7), pp. 1122–1133.

Topol, L., Chen, W., Song, H., Day, T.F. and Yang, Y. (2009) "Sox9 Inhibits Wnt Signaling by Promoting β-Catenin Phosphorylation in the Nucleus," *The Journal of Biological Chemistry*, 284(5), pp. 3323–333.

Trejo, C.L., Babić, M., Imler, E., Gonzalez, M., Bibikov, S.I., Shepard, P.J., VanSteenhouse, H.C., Yeakley, J.M. and Seligmann, B.E. (2019) "Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue," *PLoS ONE*, 14(2), p. e0212031.

Troppmair, J., Hartkamp, J. and Rapp, U.R. (1998) "Activation of NF-κB by oncogenic Raf in HEK 293 cells occurs through autocrine recruitment of the stress kinase cascade," *Oncogene*, 17(6), pp. 685–690.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., Meyers, R.M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W.F.J., Howell, S., Merkel, E., Ghandi, M., Garraway, L.A., Root, D.E., Golub, T.R., Boehm, J.S. and Hahn, W.C. (2017) "Defining a Cancer Dependency Map," *Cell*, 170(3), p. 564.

Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., Ölbei, M., Gábor, A., Theis, F., Módos, D., Korcsmáros, T. and Saez-Rodriguez, J. (2021) "Integrated intra- and intercellular signaling knowledge for multicellular omics analysis," *Molecular Systems Biology*, 17(3), p. e9923.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R. and Gordon, J.I. (2007) "The human microbiome project: exploring the microbial part of ourselves in a changing world," *Nature*, 449(7164), p. 804.

Turnbull, A.K., Selli, C., Martinez-Perez, C., Fernando, A., Renshaw, L., Keys, J., Figueroa, J.D., He, X., Tanioka, M., Munro, A.F., Murphy, L., Fawkes, A., Clark, R., Coutts, A., Perou, C.M., Carey, L.A., Dixon, J.M. and Sims, A.H. (2020) "Unlocking the transcriptomic potential of formalin-fixed paraffin embedded clinical tissues: comparison of gene expression profiling approaches," *BMC Bioinformatics*, 21(1), p. 30.

Udayasuryan, B., Zhou, Z., Ahmad, R.N., Sobol, P., Deng, C., Nguyen, T.T.D., Kodikalla, S., Morrison, R., Goswami, I., Slade, D.J., Verbridge, S.S. and Lu, C. (2024) “Fusobacterium nucleatum infection modulates the transcriptome and epigenome of HCT116 colorectal cancer cells in an oxygen-dependent manner,” *Communications Biology*, 7(1), p. 551.

Vuik, F.E.R., Nieuwenburg, S.A.V., Bardou, M., Lansdorp-Vogelaar, I., Dinis-Ribeiro, M., Bento, M.J., Zadnik, V., Pellisé, M., Esteban, L., Kaminski, M.F., Suchanek, S., Ngo, O., Májek, O., Leja, M., Kuipers, E.J. and Spaander, M.C.W. (2019) “Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years,” *Gut*, 68(10), p. 1820.

Waclawiková, B. and El Aidy, S. (2018) “Role of Microbiota and Tryptophan Metabolites in the Remote Effect of Intestinal Inflammation on Brain and Depression,” *Pharmaceuticals*, 11(3), p. 63.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., Wolf, U., Tommerup, N., Schempp, W. and Scherer, G. (1994) “Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9,” *Cell*, 79(6), pp. 1111–1120.

Wang, F., Flanagan, J., Su, N., Wang, L.C., Bui, S., Nielson, A., Wu, X., Vo, H.T., Ma, X.J. and Luo, Y. (2012) “RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues,” *The Journal of Molecular Diagnostics : JMD*, 14(1), p. 22.

Wang, H., Guo, M., Wei, H. and Chen, Y. (2023) “Targeting p53 pathways: mechanisms, structures and advances in therapy,” *Signal Transduction and Targeted Therapy 2023 8:1*, 8(1), pp. 1–35.

Wang, X., Buhimschi, C.S., Temoin, S., Bhandari, V., Han, Y.W. and Buhimschi, I.A. (2013) “Comparative Microbial Analysis of Paired Amniotic Fluid and Cord Blood from Pregnancies Complicated by Preterm Birth and Early-Onset Neonatal Sepsis,” *PLoS ONE*, 8(2), p. e56131.

Wang, Y., Wang, J., Gao, J., Ding, M. and Li, H. (2023) “The expression of SERPINE1 in colon cancer and its regulatory network and prognostic value,” *BMC Gastroenterology*, 23(1), p. 33.

Wang, Z., Gerstein, M. and Snyder, M. (2009) “RNA-Seq: a revolutionary tool for transcriptomics,” *Nature reviews. Genetics*, 10(1), p. 57.

Wei, L., Chen, W., Zhao, J., Fang, Y. and Lin, J. (2020) “Downregulation of CLCA4 expression is associated with the development and progression of colorectal cancer,” *Oncology Letters*, 20(1), p. 631.

Wei, Z., Cao, S., Liu, S., Yao, Z., Sun, T., Li, Y., Li, J., Zhang, D. and Zhou, Y. (2016) "Could gut microbiota serve as prognostic biomarker associated with colorectal cancer patients' survival? A pilot study on relevant mechanism," *Oncotarget*, 7(29), p. 46158.

Weir, T.L., Manter, D.K., Sheflin, A.M., Barnett, B.A., Heuberger, A.L. and Ryan, E.P. (2013) "Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults," *PLoS ONE*, 8(8), p. e70803.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer International Publishing (Use R!).

Wilschut, J.A., Habbema, J.D.F., Van Leerdam, M.E., Hol, L., Lansdorp-Vogelaar, I., Kuipers, E.J. and Van Ballegooijen, M. (2011) "Fecal Occult Blood Testing When Colonoscopy Capacity is Limited," *JNCI: Journal of the National Cancer Institute*, 103(23), pp. 1741–1751.

Wilson, M.R., Jiang, Y., Villalta, P.W., Stornetta, A., Boudreau, P.D., Carrá, A., Brennan, C.A., Chun, E., Ngo, L., Samson, L.D., Engelward, B.P., Garrett, W.S., Balbo, S. and Balskus, E.P. (2019) "The human gut bacterial genotoxin colibactin alkylates DNA," *Science (New York, N.Y.)*, 363(6428), p. eaar7785.

Witold, K., Anna, K., Maciej, T. and Jakub, J. (2018) "Adenomas – Genetic factors in colorectal cancer prevention," *Reports of Practical Oncology & Radiotherapy*, 23(2), pp. 75–83.

Woo, H.D., Kim, K. and Kim, J. (2015) "Association between preoperative C-reactive protein level and colorectal cancer survival: a meta-analysis," *Cancer Causes and Control*, 26(11), pp. 1661–1670.

Wu, D.Q., Yang, Z.F., Wang, K.J., Feng, X.Y., Lv, Z.J., Li, Y. and Jian, Z.X. (2018) "AQP8 inhibits colorectal cancer growth and metastasis by down-regulating PI3K/AKT signaling and PCDH7 expression," *American Journal of Cancer Research*, 8(2), p. 266.

Xi, Y. and Xu, P. (2021) "Global colorectal cancer burden in 2020 and projections to 2040," *Translational Oncology*, 14(10), p. 101174.

Xie, L., Fang, J., Yu, J., Zhang, W., He, Z., Ye, L. and Wang, H. (2023) "The role of CD4+ T cells in tumor and chronic viral immune responses," *MedComm*, 4(5), p. e390.

Xu, W., Zhang, Y., Chen, D., Huang, D., Zhao, Y., Hu, W., Lin, L., Liu, Y., Wang, S., Zeng, J., Xie, C., Chan, H., Li, Q., Chen, H., Liu, X., Wong, S.H., Yu, J., Chan, F.K.L., Chan, M.T.V., Ng, S.C., Wu, W.K.K. and Zhang, L. (2024)

"Elucidating the genotoxicity of Fusobacterium nucleatum-secreted mutagens in colorectal cancer carcinogenesis," *Gut Pathogens*, 16(1), p. 50.

Xu, Yun, Li, C., Zhang, Y., Guo, T., Zhu, C., Xu, Ye and Liu, F. (2020) "Comparison Between Familial Colorectal Cancer Type X and Lynch Syndrome: Molecular, Clinical, and Pathological Characteristics and Pedigrees," *Frontiers in Oncology*, 10, p. 1603.

Yang, K., Liang, Y., Sun, Z., Liu, L., Liao, J., Xu, H., Zhu, M., Fu, Y.X. and Peng, H. (2018) "T cell-derived lymphotoxin limits Th1 response during HSV-1 infection," *Scientific Reports*, 8(1), pp. 1–9.

Yang, L., Tang, Z., Zhang, H., Kou, W., Lu, Z., Li, X., Li, Q. and Miao, Z. (2013) "PSMA7 Directly Interacts with NOD1 and Regulates its Function," *Cellular Physiology and Biochemistry*, 31(6), pp. 952–959.

Yang, M., Lin, X., Rowe, A., Rognes, T., Eide, L. and Bjørås, M. (2015) "Transcriptome analysis of human OXR1 depleted cells reveals its role in regulating the p53 signaling pathway," *Scientific Reports 2015 5:1*, 5(1), pp. 1–12.

Yang, N.-Y., Zhang, Q., Li, J.-L., Yang, S.-H. and Shi, Q. (2014) "Progression of periodontal inflammation in adolescents is associated with increased number of Porphyromonas gingivalis, Prevotella intermedia, Tannerella forsythensis, and Fusobacterium nucleatum," *International Journal of Paediatric Dentistry*, 24(3), pp. 226–233.

Yang, Y., Weng, W., Peng, J., Hong, L., Yang, L., Toiyama, Y., Gao, R., Liu, M., Yin, M., Pan, C., Li, H., Guo, B., Zhu, Q., Wei, Q., Moyer, M.P., Wang, P., Cai, S., Goel, A., Qin, H. and Ma, Y. (2016) "Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating TLR4 Signaling to NFκB, Upregulating Expression of microRNA-21," *Gastroenterology*, 152(4), p. 851.

Yano, J.M., Yu, K., Donaldson, G.P., Shastri, G.G., Ann, P., Ma, L., Nagler, C.R., Ismagilov, R.F., Mazmanian, S.K. and Hsiao, E.Y. (2015) "Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis," *Cell*, 161(2), pp. 264–276.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. (2009) "Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads," *Bioinformatics*, 25(21), p. 2865.

Yeakley, J.M., Shepard, P.J., Goyena, D.E., VanSteenHouse, H.C., McComb, J.D. and Seligmann, B.E. (2017) "A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling," *PLoS ONE*, 12(5).

Yehia, L., Plitt, G., Tushar, A.M., Liu, D., Joo, J., Ni, Y., Patil, S. and Eng, C. (2025) “Extended spectrum of cancers in PTEN hamartoma tumor syndrome,” *npj Precision Oncology*, 9(1), pp. 1–8.

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F.O. (2014) “The SILVA and ‘All-species Living Tree Project (LTP)’ taxonomic frameworks,” *Nucleic Acids Research*, 42(D1), pp. D643–D648.

Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012) “clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters,” *OMICS : a Journal of Integrative Biology*, 16(5), p. 284.

Yu, G., Wang, L.G., Yan, G.R. and He, Q.Y. (2015) “DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis,” *Bioinformatics*, 31(4), pp. 608–609.

Yu, J., Chen, Y., Fu, X., Zhou, X., Peng, Y., Shi, L., Chen, T. and Wu, Y. (2016) “Invasive Fusobacterium nucleatum may play a role in the carcinogenesis of proximal colon cancer through the serrated neoplasia pathway,” *International Journal of Cancer*, 139(6), pp. 1318–1326.

Yu, Z., Deng, P., Chen, Y., Lin, D., Liu, S., Hong, J., Guan, P., Chen, Jianfeng, Zhong, M.E., Chen, Jinghong, Chen, X., Sun, Y., Wang, Y., Wang, P., Cai, Z., Chan, J.Y., Huang, Y., Xiao, R., Guo, Y., Zeng, X., Wang, W., Zou, Y., Yu, Q., Lan, P., Teh, B.T., Wu, X. and Tan, J. (2024) “Pharmacological modulation of RB1 activity mitigates resistance to neoadjuvant chemotherapy in locally advanced rectal cancer,” *Proceedings of the National Academy of Sciences of the United States of America*, 121(6), p. e2304619121.

Zeng, Q., Zhang, S., He, L., Fu, Q., Liao, L., Chen, L. and Ding, X. (2024) “Knockdown of BUB1B Inhibits the Proliferation, Migration, and Invasion of Colorectal Cancer by Regulating the JNK/c-Jun Signaling Pathway,” *Cancer Biotherapy and Radiopharmaceuticals*, 39(3), pp. 236–246.

Zeng, X., Huang, H., Tamai, K., Zhang, X., Harada, Y., Yokota, C., Almeida, K., Wang, J., Doble, B., Woodgett, J., Wynshaw-Boris, A., Hsieh, J.C. and He, X. (2007) “Initiation of Wnt signaling: control of Wnt coreceptor Lrp6 phosphorylation/activation via frizzled, dishevelled and axin functions,” *Development (Cambridge, England)*, 135(2), p. 367.

Zhang, D., Ni, Q.Q., Wang, S.Y., He, W.F., Hong, Z.X., Liu, H.Y., Chen, X.H., Chen, L.J., Han, F.Y., Zhang, L.J., Li, X. ming, Ding, Y. qing, Jiao, H. li and Ye, Y. ping (2024) “APC mutations disrupt ß-catenin destruction complex

condensates organized by Axin phase separation," *Cellular and Molecular Life Sciences: CMLS*, 81(1), p. 57.

Zhang, L., Theodoropoulos, P.C., Eskiocak, U., Wang, W., Moon, Y.A., Posner, B., Williams, N.S., Wright, W.E., Kim, S.B., Nijhawan, D., De Brabander, J.K. and Shay, J.W. (2016) "Selective targeting of mutant adenomatous polyposis coli (APC) in colorectal cancer," *Science translational medicine*, 8(361), p. 361ra140.

Zhang, X., Liu, S. and Zhou, Y. (2016) "Circulating levels of C-reactive protein, interleukin-6 and tumor necrosis factor-α and risk of colorectal adenomas: a meta-analysis," *Oncotarget*, 7(39), p. 64371.

Zhang, Y., Parmigiani, G. and Johnson, W.E. (2020) "ComBat-seq: batch effect adjustment for RNA-seq count data," *NAR Genomics and Bioinformatics*, 2(3).

Zhao, H., Ming, T., Tang, S., Ren, S., Yang, H., Liu, M., Tao, Q. and Xu, H. (2022) "Wnt signaling in colorectal cancer: pathogenic role and therapeutic target," *Molecular Cancer 2022 21:1*, 21(1), pp. 1–34.

Zhao, L., Zhang, X., Zhou, Y., Fu, K., Lau, H.C.H., Chun, T.W.Y., Cheung, A.H.K., Coker, O.O., Wei, H., Wu, W.K.K., Wong, S.H., Sung, J.J.Y., To, K.F. and Yu, J. (2022) "Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients," *Oncogene*, 41(36), pp. 4200–4210.

Zhao, S., Wang, S., Pan, P., Xia, T., Chang, X., Yang, X., Guo, L., Meng, Q., Yang, F., Qian, W., Xu, Z., Wang, Y., Wang, Z., Gu, L., Wang, R., Jia, F., Yao, J., Li, Z. and Bai, Y. (2019) "Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis," *Gastroenterology*, 156(6), pp. 1661-1674.e11.

Zhao, X., Lu, M., Liu, Z., Zhang, M., Yuan, H., Dan, Z., Wang, D., Ma, B., Yang, Y., Yang, F., Sun, R., Li, L. and Dang, C. (2023) "Comprehensive analysis of alfa defensin expression and prognosis in human colorectal cancer," *Frontiers in Oncology*, 12, p. 974654.

Zheng, H.C., Sugawara, A., Okamoto, H., Takasawa, S., Takahashi, H., Masuda, S. and Takano, Y. (2011) "Expression Profile of the REG Gene Family in Colorectal Carcinoma," *Journal of Histochemistry and Cytochemistry*, 59(1), p. 106.

Zhou, C., Gao, Y., Ding, P., Wu, T. and Ji, G. (2023) "The role of CXCL family members in different diseases," *Cell Death Discovery 2023 9:1*, 9(1), pp. 1–12.

Zhou, T., Wu, L., Ma, N., Tang, F., Yu, Z., Jiang, Z., Li, Y., Zong, Z. and Hu, K. (2020) "SOX9-activated FARSA-AS1 predetermines cell growth, stemness, and

metastasis in colorectal cancer through upregulating FARSA and SOX9," *Cell Death & Disease*, 11(12), p. 1071.

Zhu, G., Pei, L., Xia, H., Tang, Q. and Bi, F. (2021) "Role of oncogenic KRAS in the prognosis, diagnosis and treatment of colorectal cancer," *Molecular Cancer 2021 20:1*, 20(1), pp. 1–17.

Zhu, G., Pei, L., Ye, D., Tang, Q., Xu, H. and Bi, F. (2025) "Targeting KRAS in colorectal cancer immunotherapy: rationale, challenges and future directions," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1880(4), p. 189382.

Zhu, H., Li, M., Bi, D., Yang, H., Gao, Y., Song, F., Zheng, J., Xie, R., Zhang, Y., Liu, H., Yan, X., Kong, C., Zhu, Y., Xu, Q., Wei, Q. and Qin, H. (2024) "Fusobacterium nucleatum promotes tumor progression in KRAS p.G12D-mutant colorectal cancer by binding to DHX15," *Nature Communications 2024 15:1*, 15(1), pp. 1–15.