Tumour growth patterns at the invasive front of colorectal cancer through integrated histopathological, transcriptomic and spatial-omic analysis

Walaiphorn Woraharn
BSc, MSc

Thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

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

July 2026

Abstract

Colorectal cancer (CRC) is one of the most common cancers worldwide, yet patients with the same TNM stage often have very different outcomes, indicating that staging does not capture all of the biology that drives prognosis. Much of this missing information is concentrated at the invasive front. Here, the tumour growth pattern (TGP), which describes whether a tumour advances as a cohesive pushing border or dissects into the stroma as an infiltrative front, can be read on routine haematoxylin and eosin (H&E) sections and predicts survival. Why the infiltrative pattern carries a poorer prognosis and how cells at this interface are arranged in space have remained unclear.

The central hypothesis of this thesis is that the spatial arrangement of cells at the invasive front, rather than their abundance alone, distinguishes the growth patterns and carries independent prognostic information. This was tested across two cohorts, the Glasgow Royal Infirmary test cohort (n = 538) and the TransSCOT validation cohort (n = 1,037), using histological TGP scoring, bulk transcriptomics (TempO-Seq), multiplex immunofluorescence with density-corrected proximity analysis (the Nearest Distance Index, NDI), and compartment-resolved spatial transcriptomics (GeoMx) with cell-state and ligand-receptor inference.

A three-tier TGP classification was reproducible (weighted $\kappa = 0.78$) and was an independent prognostic factor for colon cancer-specific survival (HR = 1.65, $p < 0.001$), validated in TransSCOT. Bulk transcriptomics revealed stromal, epithelial-mesenchymal transition, and inflammatory programmes in infiltrative tumours, and proliferative programmes in pushing tumours. Importantly, cellular composition and density did not differ between patterns. Instead, spatial analysis showed that FAP+ cancer-associated fibroblasts (CAFs) lay closer to tumour cells and further from T cells in infiltrative tumours, and closer proximity of αSMA+FAP+ CAFs to tumour cells was independently associated with worse survival (HR = 2.46, $p = 0.05$). Spatial transcriptomics further linked the separation of CAFs from T cells to candidate inhibitory axes involving TIGIT-PVR/NECTIN2 and hyaluronan (HAS2-CD44).

These findings reframe the TGP from a morphological descriptor into a spatial phenotype in which prognosis is determined by where cells sit rather than how many are present. The TGP offers a practical, H&E-based entry point into the

spatial biology of the invasive front. As the spatial findings come from a single cohort and rely on inferred mechanisms, they are hypothesis-generating and require external validation.

Publications and Presentations

Publications relating to this thesis

Woraharn, W., McCulloch, A., Bigley, C., Hatthakarnkul, P., Pennel, K., Alexander, P., van Wyk, H., Roseweir, A., Hay, J., Maka, N., Park, J., Jamieson, N.B., Edwards, J. and Roxburgh, C.S. (2025), Investigation of three alternative histopathological scoring methods at the invasive tumour front in colorectal cancer. J Pathol Clin Res, 11: e70031.

Poster presentations

Investigation of two alternative histopathological scoring methods at the tumour margin in colorectal cancer. CRUK Scotland Colorectal Cancer Annual Theme Day Symposium. Institute of Genetics and Cancer, Edinburgh, United Kingdom. 12 February 2024.

Single-cell spatial proteomics decodes the architecture of CAFs at the primary colorectal cancer invasive front. 6th European Symposium on Single Cell Proteomics. Vienna, Austria. 26-27 August 2025.

Opposing architectural roles of cancer-associated fibroblasts at the invasive front characterise primary colorectal cancer growth patterns. Spatial Biology: The Melting Pot (EMBL Conference). EMBL Heidelberg, Germany. 14-17 October 2025.

Spatial proteomics reveals how distinct fibroblast populations shape the tumour-immune ecosystem in colorectal cancer. CancersScape: Spatial Biology of the Tumour Ecosystem. Barcelona, Spain. 5-7 November 2025.

Table of Contents

List of Tables

List of Figures

Acknowledgement

First and foremost, I would like to thank my primary supervisor, Professor Campbell Roxburgh, for giving me, a student from Thailand, the opportunity to undertake this PhD. Thank you for your patience through my endless re-analyses, for trusting me with the freedom to follow my own ideas, and for shaping the way I think into that of a true researcher. I could not have asked for a better mentor to guide me through this journey, nor for a kinder one to put up with my many (sticker-laden) WhatsApp messages along the way.

I am also grateful to my second supervisor, Professor Nigel Jamieson, for his guidance, kindness, and support during my hospital stay, which meant a great deal to me. I would also like to thank everyone in his lab for their help with all things spatial.

There are not enough words to thank Professor Joanne Edwards. Although not officially my supervisor, she looked after me as closely as if I were one of her own and supported me in every possible way, somehow always replying to my emails faster than anyone I know. I will treasure the time we spent at the EMBL conference in Heidelberg, which was as rewarding as it was fun, and I will certainly never forget the party on the very first night at the hotel. Thank you, Joanne, for everything.

Thank you to Ashley McCulloch, the first person I met here. I still remember that day clearly, and I have never forgotten the first cup of coffee you bought me. Thank you for being there to help me with everything. To everyone in the Edwards lab: even though I was not always around in person, you supported me throughout, and you all mean a great deal to me. There are far too many of you to name, but please know that not one of you has been forgotten.

To the Roxburgh group, I truly could not have done it without you, and I will always be grateful, not least because I do love a good party. I would also like to thank the Le Quesne lab for teaching the imaging.

This work would not have been possible without the financial support of the Thai government scholarship, to which I am sincerely grateful. I am also thankful to the patients who generously donated the tissue that made this research possible.

On a more personal note, thank you to Ajarn Gap, Ajarn Thaval, and Ajarn Aem back home, who patiently listened to me complain about absolutely

everything. Thank you, too, to all the Thai friends I have made here, for your kindness and for always being a source of encouragement. You made this place feel like home.

Thank you, my father, for your endless encouragement and unwavering support throughout this journey, and for always believing in me, even when I doubted myself. Without you, I would not be where I am today. And to my younger sister, thank you for being by my side every single day.

Finally, this thesis is for my mother. Although she is no longer here to see it, I hope I have made her proud. She could never have imagined that this day would come.

Author’s Declaration

The work presented in this thesis was performed by the author, except where acknowledged. This thesis has not been submitted for a degree or diploma at this or any other institution.

Walaiphorn Woraharn

July 2026

List of Abbreviations

αSMA (SMA)	alpha-smooth muscle actin
AJCC	American Joint Committee on Cancer
APP	Application Protocol Package (GeoMx)
BH	Benjamini-Hochberg
BP	biological process (Gene Ontology)
BRAF	B-Raf proto-oncogene
CAF	cancer-associated fibroblast
CD3	cluster of differentiation 3 (pan T-cell marker)
CD4	cluster of differentiation 4
CD8	cluster of differentiation 8
CD44	cluster of differentiation 44 (hyaluronan receptor)
CIMP	CpG island methylator phenotype
CIN	chromosomal instability
CMS	consensus molecular subtype
CRC	colorectal cancer
CSR	complete spatial randomness
CSS	cancer-specific survival
DBSCAN	density-based spatial clustering of applications with noise
DFS	disease-free survival
DGE	differential gene expression
dMMR	deficient mismatch repair
DSP	Digital Spatial Profiler (GeoMx)
ECM	extracellular matrix
EcoTyper	cell-state deconvolution framework
EMT	epithelial-mesenchymal transition
FAP	fibroblast activation protein
FDR	false discovery rate
FFPE	formalin-fixed, paraffin-embedded
GeoMx	GeoMx Digital Spatial Profiler (spatial transcriptomics platform)
GMS	Glasgow Microenvironment Score
GO	Gene Ontology
GRI	Glasgow Royal Infirmary
GSEA	gene set enrichment analysis

GTRF	Glasgow Tissue Research Facility
H&E	haematoxylin and eosin
HAS2	hyaluronan synthase 2
HR	hazard ratio
HRP	horseradish peroxidase
ICC	intraclass correlation coefficient
IHC	immunohistochemistry
IL6	interleukin 6
ITBCC	International Tumour Budding Consensus Conference
JAK	Janus kinase
KM	Klintrup-Mäkinen (grade)
LIANA	ligand-receptor analysis framework
log2FC	log2 fold change
LOQ	limit of quantification
LR	ligand-receptor
MaxStat	maximally selected rank statistics
mIF	multiplex immunofluorescence
MMR	mismatch repair
MRC1	mannose receptor C-type 1 (CD206)
MSI	microsatellite instability
MSigDB	Molecular Signatures Database
myCAF	myofibroblast-like cancer-associated fibroblast
NDI	Nearest Distance Index
NECTIN2	nectin cell adhesion molecule 2 (CD112)
NES	normalised enrichment score
NF-κB	nuclear factor kappa B
NND	nearest-neighbour distance
OS	overall survival
Pan-CK	pan-cytokeratin
PC1 / PC2	principal component 1 / principal component 2
PCA	principal component analysis
PD-1	programmed cell death protein 1
PD-L1	programmed death-ligand 1
PDPN	podoplanin
pMMR	proficient mismatch repair
PTPRC	protein tyrosine phosphatase receptor type C (CD45)
PVR	poliovirus receptor (CD155)

RNA-seq	RNA sequencing
ROI	region of interest
S100A4	S100 calcium-binding protein A4
SARIFA	Stroma AReactive Invasion Front Area
SCOT	Short Course Oncology Therapy (trial)
SD	standard deviation
ssGSEA	single-sample gene set enrichment analysis
STAT3	signal transducer and activator of transcription 3
TAN	tumour-associated neutrophil
TempO-Seq	Templated Oligo assay with Sequencing readout (targeted transcriptomics)
TGF-β	transforming growth factor beta
TGP	tumour growth pattern
TIGIT	T-cell immunoreceptor with immunoglobulin and ITIM domains
TLS	tertiary lymphoid structure
TMA	tissue microarray
TME	tumour microenvironment
TNF-α	tumour necrosis factor alpha
TNM	tumour-node-metastasis (staging)
TransSCOT	translational SCOT cohort (validation cohort)
TSP	tumour stroma percentage
UICC	Union for International Cancer Control
UV	ultraviolet
WHU	weighted histoscore units
WTA	Whole Transcriptome Atlas (GeoMx)

Chapter 1 Introduction

1.1 Incidence, demographics and risk factors

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second most common cause of cancer death worldwide, with approximately 1.9 million new cases and 900,000 deaths recorded in 2022 (Bray et al., 2024). The risk of CRC increases with age. Men are affected more often than women. A notable recent trend is the rising incidence of early-onset CRC, diagnosed before the age of 50, which has continued to increase in many high-income countries even as overall incidence has fallen in older, screened populations, marking CRC as a growing global health burden (Cercek et al., 2026; Sung et al., 2025).

Most CRC is sporadic. Around 5% of cases are attributable to well-defined hereditary syndromes. The two best characterised inherited syndromes are Lynch syndrome, caused by germline mutation of DNA mismatch repair genes (MLH1, MSH2, MSH6 and PMS2), and familial adenomatous polyposis, caused by germline mutation of the APC gene (Dekker et al., 2019). The majority of sporadic cases are associated with modifiable lifestyle and environmental factors, including obesity, physical inactivity, smoking, high alcohol intake, and diets high in red and processed meat; chronic inflammatory bowel disease also increases risk (Dekker et al., 2019).

1.2 Development and pathogenesis

CRC develops through the gradual accumulation of genetic and epigenetic alterations that allow normal epithelial cells to progress towards invasive carcinoma. These alterations are commonly described through several major molecular pathways. The best-established model is the adenoma-carcinoma sequence, in which normal mucosa progresses through a benign adenoma before becoming carcinoma. This process is associated with a characteristic order of molecular events, including early loss of the tumour suppressor gene APC, followed by activating mutation of KRAS, and later loss of TP53, which together promote malignant transformation (Fearon, 2011; Fearon & Vogelstein, 1990).

This classical sequence is most closely linked to the chromosomal instability (CIN) pathway, the most common molecular pathway in CRC. This pathway is characterised by large-scale gains and losses of chromosomal material and commonly involves the APC-KRAS-TP53 alterations described above. A second major pathway is the microsatellite instability pathway (MSI), which results from defective DNA mismatch repair. This may occur through inherited germline mutations, as seen in Lynch syndrome, or more commonly through epigenetic silencing of MLH1 in sporadic tumours. Loss of mismatch repair leads to the accumulation of multiple mutations and results in tumours with a high mutational burden (Muzny et al., 2012).

A third pathway is the CpG island methylator phenotype (CIMP), characterised by widespread promoter hypermethylation. This pathway is associated with MLH1 silencing and BRAF mutation (Fearon, 2011). Tumours with these features often arise through the serrated pathway, which begins in serrated polyps rather than conventional adenomas and accounts for many sporadic microsatellite-unstable colorectal cancers (Leggett & Whitehall, 2010). Altogether, these molecular pathways explain much of the biological heterogeneity observed in CRC and provide the basis for the molecular classifications discussed in the following section.

1.3 Pathology, staging and molecular subtyping

CRC can be described at three complementary levels: its microscopic appearance, the anatomical extent of disease, and its molecular profile. Each of these provides prognostic information, and the most established features are summarised here. Most colorectal cancers are adenocarcinomas that arise from the glandular epithelium. Less common histological variants include mucinous and signet-ring cell carcinomas (Korphaisarn et al., 2019). Tumour location is also clinically relevant. Right-sided, or proximal, cancers differ from left-sided, or distal, cancers in both molecular features and clinical behaviour. Right-sided tumours more frequently show microsatellite instability and BRAF mutation and are often associated with a less favourable prognosis (Dekker et al., 2019).

The anatomical extent of CRC is described using the tumour-node-metastasis (TNM) classification of the American Joint Committee on Cancer and the Union for

International Cancer Control. This system records the depth of tumour invasion through the bowel wall (T), the involvement of regional lymph nodes (N), and the presence or absence of distant metastasis (M) (Amin et al., 2017). TNM stage remains the main determinant of prognosis and treatment selection and provides the standard framework against which additional prognostic markers are assessed.

In addition to TNM stage, several histopathological features are routinely reported because they provide additional prognostic information. Tumour grade is based on the degree of gland formation, with poorly differentiated tumours associated with a poorer prognosis. Resection margin status is also important, particularly the circumferential resection margin (CRM), as margin involvement is associated with an increased risk of local recurrence and poorer survival. Venous invasion, particularly extramural venous invasion, is a well-established adverse prognostic feature and reflects the ability of tumour cells to spread beyond the primary site through blood vessels. Peritoneal involvement occurs when tumour extends through the bowel wall to involve the serosal surface and is classified as pT4a disease. Tumour perforation at the site of the primary tumour is also associated with poorer prognosis and, when it communicates with peritoneal cavity, is classified as a pT4a. These features are assessed alongside TNM stage in routine pathological reporting of CRC (Loughrey, 2023).

In addition to anatomical stage, transcriptomic profiling has identified four consensus molecular subtypes (CMS), which capture important aspects of the biological heterogeneity of CRC (Guinney et al., 2015). CMS1, the microsatellite instability-immune subtype, is characterised by microsatellite instability and strong immune activation. CMS2, the canonical subtype, shows epithelial features with activation of WNT and MYC signalling. CMS3, the metabolic subtype, is defined by metabolic dysregulation. CMS4, the mesenchymal subtype, is characterised by abundant stromal content, transforming growth factor-β (TGF-β) activation, and angiogenesis. Among these subtypes, CMS4 has the poorest prognosis, consistent with its stromal-rich and mesenchymal biology (Guinney et al., 2015). Although the CMS classification has improved understanding of CRC biology, it relies on bulk transcriptomic data and has not yet been incorporated into routine pathological reporting.

1.4 Treatment

Treatment for CRC is guided by disease stage and, increasingly, by the molecular features of the tumour. The main treatment approaches and the factors used to select them are summarised here. For localized CRC, including stages I-III, surgical removal of the primary tumour and regional lymph nodes remains the main treatment with curative intent (Dekker et al., 2019). Adjuvant chemotherapy, usually using a fluoropyrimidine together with oxaliplatin, is recommended for stage III disease because it improves survival. It may also be considered for stage II disease when high-risk features are present (André et al., 2004). The duration of adjuvant chemotherapy has been refined to balance clinical benefit against treatment-related toxicity, and shorter courses are suitable for many patients (Grothey et al., 2018). In rectal cancer, radiotherapy or chemoradiotherapy is often given before surgery to reduce the risk of local recurrence (Dekker et al., 2019).

In metastatic disease, systemic chemotherapy is combined with targeted agents selected according to the tumour's molecular profile. Antibodies against vascular endothelial growth factor (VEGF), such as bevacizumab, inhibit tumour angiogenesis and are used across molecular subgroups. Antibodies against the epidermal growth factor receptor (EGFR), such as cetuximab and panitumumab, are effective only in tumours that are wild-type for RAS, because RAS mutation activates the pathway downstream of the receptor and confers resistance (Karapetis et al., 2008). Tumours carrying the BRAF V600E mutation have a particularly poor prognosis and are treated with combinations that target this mutation directly (Kopetz et al., 2019). These examples illustrate how molecular features increasingly guide treatment selection.

Immunotherapy has emerged as an effective option for the subset of tumours with mismatch repair deficiency. Because these tumours carry a high mutational burden and are strongly immunogenic, they respond to immune checkpoint blockade (Le et al., 2015). In metastatic disease, the anti-PD-1 antibody pembrolizumab has become a first-line standard of care for mismatch repair-deficient or microsatellite instability-high tumours, producing more durable responses than chemotherapy (André et al., 2020). The same principle has been extended to early rectal cancer, where checkpoint blockade in mismatch repair-

deficient tumours has achieved high rates of complete clinical response and may allow surgery to be avoided in selected patients (Cercek et al., 2022). These benefits are, however, confined to the minority of tumours with mismatch repair deficiency, and most colorectal cancers remain unresponsive to current immunotherapies.

This stage- and molecular-based framework determines treatment for the majority of patients. Nevertheless, patients with the same stage and molecular profile can follow different outcomes, which has motivated the search for additional prognostic features.

1.5 Colorectal cancer and the limitations of TNM staging classification

CRC is one of the most commonly diagnosed cancers worldwide and a leading cause of cancer-related death (Cercek et al., 2026). The TNM staging classification, developed by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC), is still the main framework for assessing CRC (Amin et al., 2017). It is widely used to guide treatment and classify patients for clinical trials. In clinical practice, treatment decisions are influenced by disease stage. For stage I-III CRC, surgical resection is the main treatment. Adjuvant chemotherapy is commonly recommended for stage III disease and may be considered for selected stage II patients with high-risk pathological features. However, TNM stage alone does not fully predict clinical outcomes, as patients within the same stage can show different risks of recurrence and survival (Cuyle & Prenen, 2017; Dienstmann et al., 2017). This variation within the same stage suggests that additional features are needed to better identify patients with more aggressive behaviour, who may benefit from more intensive treatment.

Molecular classifications such as the CMS have improved our understanding of CRC heterogeneity by identifying biologically distinct groups with prognostic relevance (Guinney et al., 2015; Thanki et al., 2017). However, these classifications do not replace TNM staging, and their use in routine reporting remains limited. This highlights the need for additional, accessible tumour features that can capture clinically meaningful biology beyond the conventional stage. One approach is to assess tumour morphology. Surgical resection specimens

are routinely examined on haematoxylin and eosin (H&E)-stained sections, and several histopathological features visible on the sections provide prognostic information beyond TNM stage (Lugli et al., 2012; Rajaganeshan et al., 2007), particularly at the invasive front, where the tumour interacts with the surrounding stroma and immune microenvironment. The histological appearance of this interface may provide a practical means to refine prognosis in CRC. For instance, the infiltrative growth pattern at the invasive front has been shown to be associated with poor cancer-specific survival independently of microsatellite instability, the CpG island methylator phenotype, LINE-1 methylation, KRAS, BRAF, and PIK3CA mutations (Morikawa et al., 2012). In addition, a large study supported the notion that high tumour budding, assessed at the invasive front, is a strong and independent predictor of disease recurrence and mortality in CRC (Mitrovic et al., 2021).

1.6 The invasive front as a tumour-host interface

The invasive front is the advancing edge of the tumour, where malignant cells meet the non-neoplastic tissues of the host: the stroma, blood and lymphatic vessels, and immune cells. It has long been seen as the key interface between tumour and host, the region where the factors driving tumour progression and spread are concentrated (Zlobec et al., 2009). This distinction from the tumour centre is not only conceptual. Within the same tumour, the protein phenotype of cancer cells at the invasive front can differ from that of the main tumour body. For example, the expression of the anti-apoptotic protein Bcl-2 is significantly lower at the invasive front compared to the tumour centre (Karamitopoulou et al., 2010). These findings suggest that cancer cells at the invasive front acquire a phenotype distinct from and more invasive than those within the main tumour body.

The behaviour of tumour cells at the invasive front can explain this difference. At the advancing edge, cancer cells frequently lose part of their epithelial organisation and acquire a more migratory phenotype, a process associated with epithelial-mesenchymal transition (EMT), in which epithelial cells gain mesenchymal characteristics and increased motility (Goossens et al., 2017). Histologically, this process is most clearly seen as tumour budding, defined as single tumour cells or small clusters of dedifferentiated cells that detach from the

main tumour body. Tumour budding is widely considered to represent a partial EMT state (Grigore et al., 2016). It is also closely related to the infiltrative growth pattern and is mainly observed at the invasive front rather than in the tumour centre (Zlobec et al., 2009). Together, these features support the idea that the invasive front is the region where invasive and disseminating behaviour is actively established.

The host tissue at the invasive front is not a passive boundary. Tumour invasion at the invasive front is accompanied by an activation of the surrounding stroma, remodelling of the extracellular matrix, recruitment of immune cells, and changes in the local vasculature. These vascular changes are clinically relevant, as they have been linked to prognosis and to the development of liver metastasis in CRC (Rajaganeshan et al., 2007). Other stromal and immune cells show spatial enrichment at the tumour invasive front. For example, activated cancer-associated fibroblasts, marked by fibroblast activation protein (FAP), are enriched at the invasive front relative to the tumour centre (Sandberg et al., 2019), and these fibroblasts can actively promote cancer cell invasion from the earliest stage of CRC (Dang et al., 2023).

In addition, the invasive front has been viewed as a dynamic balance between pro-tumour factors, such as tumour budding and the infiltrative growth pattern, and anti-tumour factors, including the local immune and inflammatory response (Zlobec et al., 2009). In keeping with this view, the infiltrative growth pattern and a high tumour budding tend to be inversely related to immune response at the invasive front (Dawson et al., 2020; Zlobec et al., 2009).

Because the invasive front is a combination of the invasive behaviour of the tumour and the host response to it, histopathological features assessed at this site can provide prognostic information beyond TNM stage (Graham et al., 2015; Keum et al., 2012; Rajaganeshan et al., 2007). Several features are assessed at the invasive margin, including tumour budding, the growth pattern, and measures of the local inflammatory and stromal reaction, such as the Klintrup-Mäkinen grade and tumour stroma percentage. The latter two are combined in the Glasgow Microenvironment Score, which captures inflammatory and stromal characteristics of the tumour microenvironment (Alexander et al., 2021).

Despite this recognised importance, the invasive front is often assessed in routine practice in relatively broad terms, either through the overall appearance of the tumour border or by measuring individual features at this site separately such as tumour budding, or stromal content. However, the invasive front represents a complex interaction between tumour cells and the host microenvironment; it is important to understand what these histological features actually reflect in terms of biological significance, especially if they are used for prognostication. The following section explains how the invasive front is assessed histologically, with particular focus on the tumour growth pattern (TGP).

1.7 Histological assessment of the invasive front and the tumour growth pattern

The invasive front can be examined histologically at two levels of resolution. At high magnification, the focus is on individual tumour cells and their dissociation from the main tumour mass at the advancing edge, seen histologically as tumour budding (Lugli et al., 2017). At low magnification, the same region is assessed more broadly by examining the configuration of the tumour border, which indicates how the tumour invades the surrounding tissue (Jass et al., 1996). These two levels of resolution are related but not interchangeable. Tumour budding captures the detachment of single tumour cells or small clusters from the main tumour mass, whereas TGP captures the overall shape of the advancing edge and whether the tumour infiltrates the surrounding host tissue. Both features have been studied as prognostic markers in CRC (Basile et al., 2022; Unal Kocabey & Cakir, 2024), and both can be assessed on routine H&E-stained sections. However, they have followed different paths towards clinical standardisation.

Tumour budding refers to a single tumour cell or small clusters of up to four cells located ahead of the main tumour mass at the invasive front and is interpreted as a histological correlate of the partial epithelial-mesenchymal transition. Among histological features in primary CRC, tumour budding has one of the strongest evidence bases for prognostic relevance. Several independent cohorts have shown that a higher tumour bud count is consistently associated with poorer outcomes (Koelzer et al., 2016; Shah et al., 2022; van Wyk et al., 2016). This evidence led to efforts to standardise its assessment, as earlier studies had used different definitions and counting methods, making comparisons between

studies difficult. The International Tumour Budding Consensus Conference (ITBCC) therefore proposed a standardised approach for counting and reporting tumour budding, providing a common framework for the field (Lugli et al., 2017). As a result, tumour budding has been incorporated into routine pathological reporting. In contrast, the architectural TGP of the invasive front has been recognised for many years but has not achieved the same level of standardisation.

The growth pattern at the invasive front was originally described by Jass and colleagues as a distinction between pushing and infiltrative margins, as part of a broader assessment of invasive growth and lymphocytic infiltration (Jass et al., 1996). This binary category was later expanded by Morikawa and colleagues into a three-tier classification, comprising expansile, intermediate, and infiltrative patterns (Morikawa et al., 2012). In that study, the infiltrative growth pattern was associated with poorer survival outcomes in CRC. Importantly, this association remained significant after adjustment for tumour molecular features, suggesting that the growth pattern provides prognostic information beyond a simple surrogate for tumour genotype. More recent studies have also shown that the infiltrative growth pattern is an independent adverse prognostic factor in stage II and III CRC (Tokodai et al., 2016). Taken together, these findings support the view that the TGP is a histomorphological prognostic feature in its own right (Koelzer & Lugli, 2014).

Morphologically, a pushing margin presented a broad and cohesive tumour front with a relatively well-defined boundary. In contrast, an infiltrative margin extends irregularly into the surrounding tissue as small glands or isolated tumour cells, producing a poorly defined margin. Both infiltrative growth pattern and high tumour budding are considered invasive behaviour characteristics at the tumour front and are often associated with one another (Unal Kocabey & Cakir, 2024; Wang et al., 2009). However, they are not the same feature. Tumour budding is assessed at high magnification by counting single cells or small clusters, while TGP is assessed at low magnification and describes the architecture of the invasive front. Whether tumour budding and the growth pattern provide overlapping or independent prognostic information cannot be determined from their definitions alone and is examined in the chapters that follow.

Despite its recognised prognostic value, the TGP approach has been incorporated into routine pathological reporting far less than tumour budding (Koelzer & Lugli, 2014). The main reason is that there is still no single agreed approach for assessing it. The original classification by Jass et al. was categorical and based on the overall appearance of the invasive margin, leaving room for variation between observers (Jass et al., 1996). To make the assessment more quantitative, Karamitopoulou et al. proposed estimating the proportion of the tumour margin showing infiltrative growth, rather than assigning the whole case to a single category (Karamitopoulou et al., 2015).

These approaches differ in how detailed they are and how much they rely on observer interpretation. Because they have been applied in different cohorts using different definitions, it remains difficult to determine which method provides the most reproducible and prognostically informative assessment of the invasive front in CRC. Until this question is answered, the TGP is unlikely to be used widely in clinical practice.

The idea of scoring at the interface between tumour and host tissue is not limited to primary CRC. The clearest example is in liver metastases, where the same principle has been developed into a formal scoring system. Liver metastases show three common histopathological growth patterns, and two rarer patterns, and these are assessed on routine H&E-stained sections at the tumour-liver interface in the same way as the growth pattern of a primary tumour (van Dam et al., 2017). In the same study, among 374 patients with colorectal liver metastases, a predominantly desmoplastic pattern was associated with longer overall survival than the replacement or pushing patterns, and the replacement pattern remained associated with shorter survival after adjustment. The distribution of these patterns, and their prognostic value, appear to depend on the tumour type. Almost all breast cancer liver metastases show a replacement pattern, while colorectal liver metastases can show any of the described patterns (van Dam et al., 2017). In liver metastases from neuroendocrine tumours, the pushing pattern was the most common, seen in about half of the patients, which is in clear contrast to the small proportion reported in colorectal and breast liver metastases, and no association was found between growth pattern and either overall or disease-free survival in that cohort (Meyer et al., 2023). Similar assessment has been applied outside the liver, where the microscopic appearance of the border of resected

lung metastases, described as smooth, blurred or irregular, was closely related to the presence of aggressive local growth and could be assessed on preoperative CT images (Issa et al., 2019). Two points follow from this literature. First, the tumour border configuration is a general property of the tumour-host interface rather than a feature specific to primary CRC, supporting the view that it reflects an underlying biological process rather than an incidental morphological appearance. Second, the liver metastases field presents what primary CRC currently lacks: an agreed definition, a published scoring guideline, and a formal test of agreement between observers. The variation in pattern distribution and prognostic value between tumour types indicates that a scoring method developed in one setting cannot be assumed to transfer directly to another.

Interest in the morphology of the invasive front extends beyond the TGP and tumour budding. More recently, Stroma AReactive Invasion Front Areas (SARIFA), regions at the invasive front where tumour cells are in direct contact with adipocytes, have been identified on routine H&E-stained sections and associated with adverse outcomes in CRC (Offermans et al., 2024). This demonstrates the effort to identify simple and reproducible histological features at the invasive front that can provide prognostic information. The practical advantage of these features is that they can be assessed on routine H&E-stained sections, which are already prepared for every resection specimen, require no additional staining, and are readily scalable to routine practice. This accessibility is one reason why morphological assessment of the invasive front remains an attractive approach to prognostic refinement.

However, histological assessment at the invasive front only describes the appearance but does not explain the biological processes that give rise to that phenotype. The TGP alone does not show which cell populations are present at the tumour front, how abundant they are, or how they are spatially organised. To understand why these morphological patterns are associated with different prognoses, it is necessary to look beyond their appearance on H&E sections and consider the cellular composition and spatial arrangement of the invasive front microenvironment. This is the focus of the following section.

1.8 Cellular components of the invasive front tumour microenvironment

The invasive front contains many types of host cells, including fibroblasts, immune cells, and the cells lining blood and lymphatic vessels. This section focuses on the cell types most relevant to interpreting growth pattern biology and most extensively studied at the invasive front of CRC: cancer-associated fibroblasts (CAFs), immune cells, and also tumour cells themselves. For each component, the aim is to set out what is known about its role at the invasive front and where the open questions remain.

CAFs are a major stromal cell population in the tumour microenvironment. They are not a uniform cell type, but a heterogeneous mixture of subtypes with different markers and functions (Nurmik et al., 2020; Sahai et al., 2020). Because the growth pattern reflects the border configuration between the tumour and the surrounding stroma, the cells that build and reshape that stroma are central to its biology. As the most abundant stromal cells and the principal remodellers of the extracellular matrix, CAFs have been a particular focus in studies of the invasive front (Belhabib et al., 2021).

Identifying CAFs relies on a combination of markers rather than a single one, since no marker labels them all. The most commonly used are fibroblast activation protein (FAP), alpha-smooth muscle actin (αSMA), and podoplanin (PDPN), which mark overlapping but not identical subsets of fibroblasts (Nurmik et al., 2020). αSMA identifies contractile, myofibroblast-like CAFs (myCAFs), whereas FAP is a cell-surface enzyme involved in reshaping the tissue surrounding the tumour (Xin et al., 2021). In addition to remodelling the stroma, CAFs can influence the immune cells, and through these combined effects, they may suppress the local anti-tumour immune response (Monteran & Erez, 2019).

The relationship between CAFs and clinical outcomes is complex. A high CAF component has been linked to poorer survival outcomes in several cancers (Liu et al., 2025; Min et al., 2021), but some CAF subtypes have been associated with favourable outcomes (Cords et al., 2024). PDPN positive CAFs have been reported as favourable in some studies and unfavourable in others (Pula et al., 2013; Yamanashi et al., 2009). This suggests that the prognostic effect of CAFs appears

to depend not only on the amount, but also on the subtype composition. FAP-positive CAFs are of particular interest at the invasive front, as they are more abundant there than in the tumour centre and have been linked to a poorer prognosis in CRC (Sandberg et al., 2019). However, most of these studies measure the number of CAFs rather than their location relative to tumour or immune cells. Whether their prognostic effect is determined by abundance, phenotype, or spatial arrangement remains unclear, highlighting the importance of understanding their organisation at the invasive front.

Immune cells are another major host component at the invasive front. The immune compartment contains many cell types, including macrophages, B cells, neutrophils, and lymphocytes, but T-cell infiltration has the most established association with CRC outcomes. A higher density of T cells, identified by markers such as CD3, is associated with longer survival outcomes (Williams et al., 2024). This immune response tends to be weaker where the growth pattern is infiltrative, and tumour budding is high, as noted in section 1.2.

The spatial distribution of immune cells also varies between tumours. In some cases, immune cells infiltrate the tumour tissue extensively, while in others they remain confined to the tumour edge (Joyce & Fearon, 2015). Mismatch-repair-deficient tumours, which more often display a pushing border pattern, are enriched for immune-cell infiltration (Koelzer & Lugli, 2014). However, immune abundance does not necessarily translate into an effective anti-tumour response. Why some immune-rich tumours continue to progress remains incompletely understood, suggesting that immune-cell function and spatial arrangement may be as important as cell number. Tertiary lymphoid structures (TLS) provide a further example of this principle, as they indicate that immune organisation, rather than immune abundance alone, may carry prognostic information near the invasive front, although their precise role in CRC remains under investigation (Sautès-Fridman et al., 2019).

The tumour cells at the invasive front are also important. As described earlier, tumour cells at the invasive edge can acquire a more mobile, mesenchymal-like phenotype through EMT (Goossens et al., 2017; Grigore et al., 2016). In the context of TGP, these tumour cell changes are important because

they may help explain why some tumour fronts remain broad and cohesive, while others break into irregular, infiltrative strands and clusters.

Across all these cell types, most studies share the same limitation: they measure cell abundance but not spatial arrangement. This shows what the invasive front is made of, but not how it is put together. A pushing and an infiltrative front may contain similar cell types but arranged in different ways. The position of CAFs, immune cells, and tumour cells, which are located next to one another, may therefore matter as much as their abundance. The next section turns to this question of spatial organisation.

1.9 From cellular density to spatial architecture

To understand why the spatial organisation of the invasive front may be clinically relevant, it is first necessary to consider how its cellular composition has been evaluated. For many years, the prognostic significance of the tumour microenvironment has been evaluated by counting cells, through measuring the density or abundance of specific cell populations. The immune response demonstrates an example of this approach. The Immunoscore, which quantifies the densities of CD3+ and CD8+ T cells in the tumour centre and tumour invasive front, is prognostic in CRC and provides information beyond TNM stage (Galon et al., 2006; Pagès et al., 2018). As the first internationally recommended immune-based prognostic measure for colon cancer, its success reflects a broader principle: that the host response carries prognostic value rather than the tumour alone (Pagès et al., 2018). Importantly, the Immunoscore recognises that location matters by separating the tumour centre and the tumour invasive front. However, it captures location at a coarse level rather than the spatial arrangement of cell populations inside those regions.

This highlights a broader limitation of density-based measures. While density provides a summary of the overall number of cells within tissue, it does not capture how those cells are spatially distributed or positioned relative to one another. The same number of cells may be evenly dispersed, arranged in clusters, or restricted to the tumour margin, and these different patterns may reflect different underlying biology despite identical cell counts. This principle is illustrated in Figure 1-1: two fields containing identical numbers of the same two

cell types differ only in their spatial relationship, with the cell types occupying separate zones in one field and lying in close proximity in the other. This distinction is not merely descriptive, but functional. For example, a cytotoxic T cell can exert its anti-tumour effect only if it can reach and interact with tumour cells. Its position relative to the tumour cells, and not its number alone, helps determine whether it can contribute to tumour control. There is no single distance that defines a biologically relevant cell-cell interaction, as this depends on the cell types and mechanism involved. However, spatial studies within the tumour microenvironment can be examined over distances ranging from tens to a few hundred micrometres. For instance, in luminal breast cancer, immune-cell phenotypes were assessed from 10 to 100 um from the nearest tumour cell in 10-um intervals, while in head and neck squamous cell carcinoma, a leading-edge ecosystem was defined within approximately 200 um of the histological tumour border (Hatse et al., 2024; Su et al., 2026). The same reasoning applies to stromal and tumour cells. For this reason, attention in the field has shifted from how many cells are present to where they are positioned, in other words, from density to geometry. In addition, the Immunoscore has been extended to include spatial information, incorporating the proximity between CD8+ and PD-L1+ cells, a concrete example of this shift (Antoniotti et al., 2022; Moretto et al., 2023).

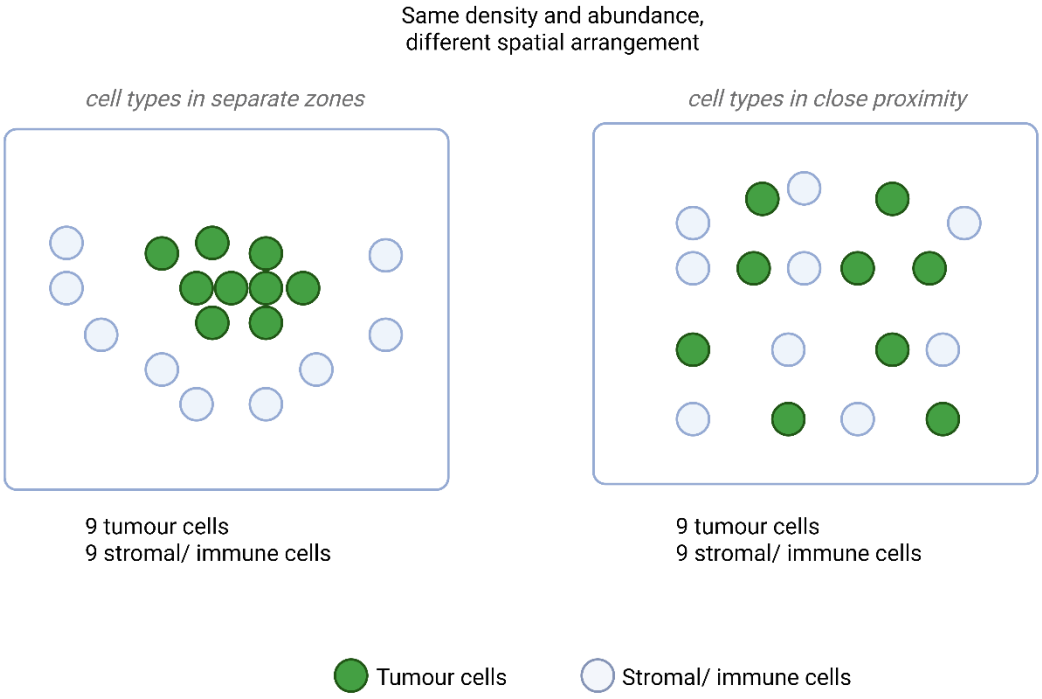


Figure 1-1: Equal density, different spatial organisation.

Two hypothetical tissue fields contain identical numbers of tumour cells and stromal/immune cells (equal composition and density) yet differ in spatial organisation: in the left field, the two cell types occupy separate zones, whereas in the right field, they are interspersed in close proximity. Abundance and density-based analyses return identical results for both fields. Quantifying this spatial dimension, while controlling cell density, is a central aim of this thesis (Chapter 5).

Recent advances in multiplexed imaging technologies and computational methods have enabled detailed investigation of the spatial organisation of cells within intact tissue. These approaches can capture the positions of multiple cell types simultaneously, allowing spatial relationships to be described beyond cell abundance alone. Such an organisation can be quantified, for example, by measuring the proximity between specific cell populations or by identifying cellular neighbourhoods, defined as groups of cell types that are repeatedly found together within tissue.

At the invasive front of CRC, Schürch et al. identified conserved cellular neighbourhoods and found that patient outcomes were related to how these cellular neighbourhoods were connected, separated, and coordinated across the tumour microenvironment, rather than simply to the presence or absence of particular cell types (Schürch et al., 2020). These neighbourhoods are defined at single-cell resolution and differ in scale from the TGP. They demonstrate the broader principle that, at the invasive front, cell arrangement carries prognostic

information. The TGP can be viewed as a low-magnification morphological reading of the same front. Its prognostic value may reflect differences in the spatial arrangement of cells that density-based measurements do not capture.

The invasive front has been examined spatially in other cancer types. In luminal B-like breast cancer, Hatse et al. applied multiplex immunohistochemistry (mIHC) on tissue microarrays (TMA) in which tumour centre and the invasive front were sampled separately, with the invasive front defined as a band extending 500 um on either side of the tumour border, and classified tumours as immune desert, immune excluded or immune inflamed according to where the cytotoxic T cells were located rather than how many cells were present (Hatse et al., 2024). The same study related marker expression to the distance from the nearest tumour cell, and found that, in older patients, the proportion of helper T cells expressing the activation marker OX-40 increased as this distance decreased, indicating a difference that would not have been detected by measuring cell density alone (Hatse et al., 2024). In head and neck squamous cell carcinoma, Su et al. applied spatial transcriptomics to define a leading-edge ecosystem approximately 200 um from the histological border, which showed higher expression of genes related to EMT, invasion and extracellular matrix remodelling than the tumour centre, and in which MMP1-positive fibroblasts were spatially associated with C1QC-positive macrophages, which were in turn spatially associated with cytotoxic T cells, while the fibroblast themselves were not (Su et al., 2026). These studies indicate that the invasive front is treated as a spatially distinct area across cancer types, and that fibroblasts are closely involved in how it is organised. However, none of them related the spatial organisation described to the low magnification growth pattern of the tumour border.

Although spatial studies of the CRC invasive front have provided important insights, most have focused on the immune compartment (Nearchou et al., 2019; Schürch et al., 2020; Zwing et al., 2020). The stromal compartment, particularly CAFs, remains less well characterised spatially, despite its likely importance in shaping tumour-immune interactions and influencing immune exclusion. CAFs are also closely related to extracellular matrix remodelling, mesenchymal-like biology, and the infiltrative growth pattern. The overlap between the infiltrative pattern and stromal enrichment suggests that the spatial organisation of the

stromal compartment at the invasive front is an important but underexplored area of CRC biology.

Altogether, these observations highlight several remaining gaps. First, most studies of the invasive front continue to focus on cellular abundance rather than spatial organisation, despite growing evidence that prognostic information may be encoded in the arrangement of cells within tissue. Second, although spatial analyses have increasingly been adopted, they have predominantly examined the immune cells, while the stromal compartment and the spatial relationship between stromal, immune, and tumour cells remain underexplored. A further gap is directly relevant to this thesis: although the prognostic importance of TGP is well established, it remains unknown whether the spatial architecture of the invasive front differs between pushing and infiltrative patterns. Therefore, it is unclear whether the prognostic differences between these patterns reflect differences in cellular composition, cellular organisation, or specific spatial relationships between tumour, stromal, and immune populations.

These questions cannot be addressed by a single approach alone. Histological assessment can define the TGP, but it does not explain the biological processes that underlie it. Bulk transcriptomic analysis can provide broader molecular information, but it loses spatial information needed to understand how cells are organised within tissue. Conversely, imaging approaches based on a limited number of markers can preserve spatial position but provide a more restricted view of single-cell protein context. Addressing the questions raised in this section requires a multimodal approach that can capture not only which cells are present, but also where they are located and how they may communicate across tumour, stromal, and immune compartments. The following section describes the strategies used in this thesis to integrate these approaches.

1.10 A multimodal approach to the invasive front

The question outlined above encompasses morphology, molecular biology, and tissue organisation, and cannot be addressed by a single technique. This thesis adopts a multimodal, data-driven strategy that links morphological growth pattern to its cellular and molecular basis. The analysis begins with the histological features itself, then examines its transcriptational, spatial organisation, and

potential intercellular communication. Two independent patient cohorts provide the framework for this work: the Glasgow Royal Infirmary (GRI) cohort (n = 538), used as the test cohort, and the TransSCOT cohort (n = 1,037), used to test the reproducibility and prognostic value of the histological findings.

The first component is the histological assessment of the TGP. Because the prognostic question begins with a morphological observation, the growth pattern must first be defined in a reproducible way before its underlying biology can be examined. To this end, two published scoring approaches, proposed by Karamitopoulou et al. and adapted from Taskin et al., are evaluated alongside a three-category growth pattern classification, to identify which approach provides the most reproducible and prognostically informative reading of the invasive front (Karamitopoulou et al., 2015; Morikawa et al., 2012; Taskin et al., 2022). A reproducible classification is also required for the subsequent molecular and spatial analyses, which depend on consistent growth pattern assignment across tumours. This step, therefore, establishes the phenotype that the remainder of the thesis aims to explain. However, as a morphological measure, it describes the pattern of tumour growth without identifying the biological processes that underlie it.

To investigate this biology, bulk transcriptomics is performed using TempO-Seq, a targeted, ligation-based assay developed for the degraded RNA characteristic of archival formalin-fixed, paraffin-embedded tissue (Trejo et al., 2019). This assay provides a transcriptome-wide view of genes and biological programmes that differ between growth patterns, and computational estimation of cell-type composition indicates which cellular compartments contribute to these differences. In this way, it identifies candidate populations and pathways that subsequent spatial analyses can be guided by, rather than relying on prior assumptions. Its principal limitation is that the signal is averaged across all cells, so the spatial information needed to understand how those cells are organised within the tissue is lost.

The main question of this thesis, whether the spatial architecture of the invasive front differs between TGPs, is addressed using the multiplex immunofluorescence (mIF). In contrast to conventional single-marker immunohistochemistry, mIF detects several markers simultaneously on the same

section while preserving the position of every cell, so that the arrangement of tumour, stromal, and immune populations can be quantified rather than only their abundance (Hernandez et al., 2021). The markers are selected on the basis of the preceding transcriptomic findings, ensuring that the populations examined are highlighted by the data. Spatial arrangement is then quantified using the Nearest Distance Index (NDI), a measure of the distance between two cell populations relative to the distance that would be expected if the same cells were distributed at random, allowing two populations to be described as lying closer together or further apart than expected by chance. Because it is corrected for cell density, the NDI separates spatial arrangement from cell abundance, addressing the distinction between how many cells are present and where they are positioned.

The mIF resolves individual cells but is restricted to a small number of markers, and so cannot identify the specific cell states or molecular signals associated with the spatial patterns. To extend the analysis towards mechanism, compartment-resolved spatial transcriptomics was performed using the GeoMx Digital Spatial Profiler, which measures whole-transcriptome expression within defined tissue compartments (the tumour epithelium, the αSMA-enriched stroma, and the remaining microenvironment) (Van & Blank, 2019). Because the platform profiles define regions rather than individual cells, each compartment still represents a mixture of cell types, and computational inference is required to interpret these measurements. Two complementary methods are applied for this purpose. EcoTyper is used to infer the transcriptionally defined cell states present within each compartment (Luca et al., 2021), and LIANA is used to infer candidate ligand–receptor interactions through which these populations may communicate (Dimitrov et al., 2022). Combining these inferences with the spatial measurements links the cellular arrangements observed by imaging to the cell states and signalling that may contribute to these patterns.

Taken together, these approaches move from defining the growth pattern, to its transcriptional basis, to the single-cell spatial architecture of the invasive front, and finally towards the cell states and intercellular signalling that may underlie it. Each addresses a limitation of the one before, and together they intend to determine not only which cells are present at the invasive front, but where they are located and how they may interact. The specific aims and objectives that follow from this strategy are set out in the next section.

1.11 Hypothesis and aims

The preceding sections described the TGP as a histological feature of the invasive front that has been associated with patient outcome and noted that most studies of the invasive front have characterised which cells are present and in what numbers rather than how those cells are arranged in space. It remains unclear whether any prognostic differences between growth patterns reflect differences in cellular composition, cellular organisation, or specific spatial relationships between tumour, stromal, and immune populations.

The central hypothesis of this thesis is that the spatial arrangement of cells at the invasive front, rather than their abundance or composition alone, distinguishes pushing from infiltrative growth patterns and carries prognostic information beyond that captured by cell counts. A related proposition is that the growth pattern is unlikely to be a passive readout of the tumour's underlying genetics alone, but may instead reflect how tumour, stromal, and immune cells are organised relative to one another at the advancing edge. These propositions cannot be examined by a single technique, and the thesis addresses them through the sequential, multimodal strategy described above, moving from the morphological definition of the growth pattern to its cellular composition, spatial organisation, and the cell states and signalling that may accompany it.

To address this hypothesis, the thesis pursues four linked aims, addressed in turn across Chapters 3 to 6:

1. To establish a reproducible and prognostically significant histological assessment of the TGP at the invasive front, by comparing three scoring approaches for their interobserver reproducibility and association with survival and validating the selected method in an independent cohort.

2. To characterise the transcriptional differences between infiltrative and pushing growth patterns.

3. To determine whether the spatial architecture of the invasive front differs between growth patterns, and whether the spatial relationships between

tumour, stromal, and immune populations are associated with survival independently of established clinicopathological features.

4. To explore the cellular states and ligand-receptor signalling that may accompany the spatial organisation of the invasive front, by integrating spatial transcriptomics with the inference of cell states and ligand-receptor interactions.

1.12 Thesis structure

The overall analysis strategy of this thesis is summarised in Figure 1-2. Chapter 2 describes the materials and methods common to the work, including the two patient cohorts on which it is based: the Glasgow Royal Infirmary (GRI) test cohort and the independent TransSCOT validation cohort. Chapters 3 to 6 present the empirical work in turn, progressing from the histological definition of the growth pattern to its transcriptional, spatial proteomic, and spatial transcriptomic characterisation, as set out in the aims above. Chapter 7 draws these findings together and considers their biological and clinical implications, limitations, and directions for future research.

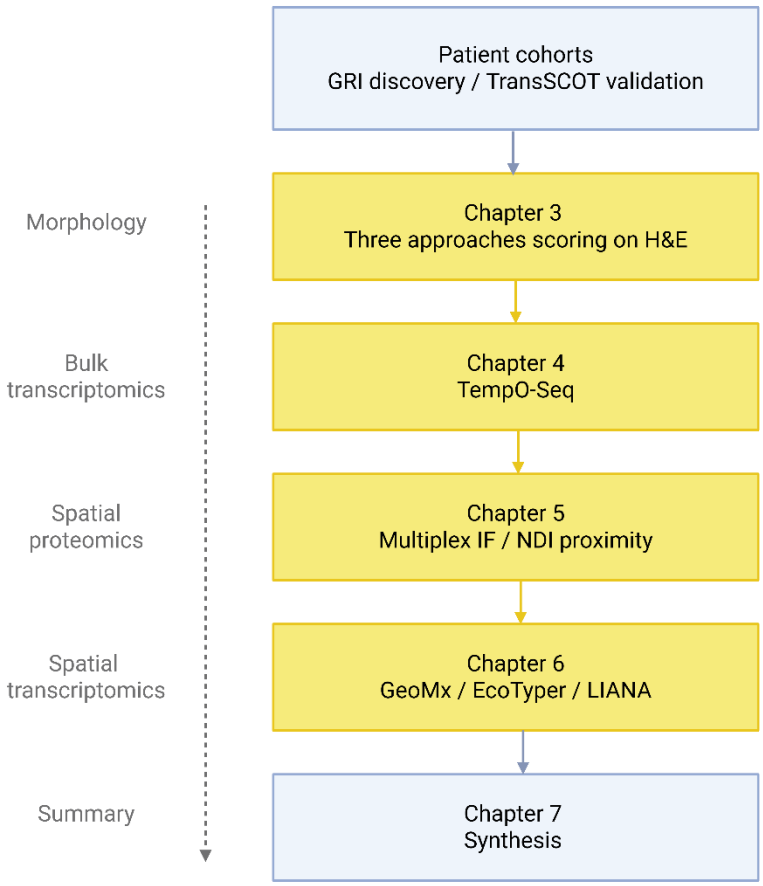


Figure 1-2: The workflow of the overall analysis strategy of this thesis

Chapter 2 Materials and Methods

2.1 Samples and cohorts

2.1.1 Glasgow Royal Infirmary (GRI) test cohort

The test cohort included patients with stage I-III primary CRC who underwent surgical resection at Glasgow Royal Infirmary between 1997 and 2013. A total of 787 patients were identified. Patients who received neoadjuvant radiotherapy were excluded from this study because it can disrupt tissue architecture and may affect tumour border configuration at the invasive front (Hav et al., 2015). Patients who died within 30 days of surgery were excluded to minimise the impact of surgery-related deaths that may reflect operative complications rather than tumour-related prognosis. In total, 538 patients were included in this test cohort. This retrospective study was approved by the West of Scotland Research Ethics Committee (16/WS/0207).

Tumour stage was assessed using a standard report, the 5th edition of the AJCC/UICC TNM staging system. During the long recruitment period, survival outcomes may have been influenced by temporal changes in clinical practice, such as advances in staging, surgical management, and adjuvant treatment. Therefore, the calendar period was assessed as a covariate in the survival analyses, with patients grouped into three predefined periods: 1997-2003, 2004-2008, and 2009-2013.

Clinicopathological features, including age, sex, tumour site, and use of adjuvant therapy, were collected from clinical records. Pathological variables were extracted in line with the RCPath colorectal cancer datasets used at the time of reporting, including tumour differentiation, resection margin involvement, venous invasion, perineural invasion, tumour perforation, and peritoneal involvement.

Tumour budding had previously been scored by Dr Phimmada Hatthakarnkul using the recommendations of the International Tumour Budding Consensus Conference 2016. In brief, tumour budding was evaluated at the invasive front and grouped into low budding (0-9 buds) and high budding (≥10 buds).

Tumour microenvironment assessments in the GRI cohort, including Klintrup–Mäkinen grade (KM), tumour stroma percentage (TSP), and the Glasgow Microenvironment Score (GMS), were scored by Peter G. Alexander and Kathryn A. F. Pennel, as previously described (Alexander et al., 2021).

The KM score was applied to evaluate the inflammatory reaction at the invasive front and was classified as low (absent or weak inflammation) or high (moderate or prominent inflammation).

TSP was assessed as the proportion of tumour-associated stroma within the tumour centre and grouped into low and high categories. Low TSP was defined as ≤50% stroma, while high TSP was defined as >50% stroma.

GMS was a combination of KM and TSP scores. High KM and any TSP scores were assigned to GMS 0. Low KM and low TSP scores were assigned to GMS 1. Low KM and high TSP scores were assigned to GMS 2.

2.1.2 TransSCOT validation cohort

The TransSCOT cohort consisted of patients with high-risk stage II or stage III CRC who received adjuvant chemotherapy as part of the SCOT trial between 2008 and 2013 (ISRCTN No. 59757862). All the eligibility criteria have been described previously (Iveson et al., 2018). Briefly, patients aged 18 years or older who had undergone surgical resection were included. Patients who had received previous long-course chemoradiotherapy were excluded. For rectal cancer patients, total mesorectal excision (TME) surgery was required, with negative resection margins, defined as a clearance of greater than 1mm from tumour. All patients were treated with adjuvant chemotherapy (FOLFOX or CAPOX) for 3 or 6 months, assigned randomly. With 2,912 tissue samples available, whole-section H&E-stained images were reviewed using a selected tumour border configuration method from the GRI test cohort. In total, 1,037 samples were successfully scored.

Clinical data was collected as previously described (Iveson et al., 2018). Tumour stage was assessed according to the 7th edition of the TNM classification. Tumour microenvironment assessments in the TransSCOT cohort, including KM, TSP, and GMS, were assessed using the same approach as described in the GRI

cohort. The selected approach from the GRI cohort (TGP method) was applied to validate in the TransSCOT cohort.

2.2 Histological assessment of tumour border configuration

2.2.1 Haematoxylin and Eosin (H&E) staining

Tumour sections of 5μm thickness were deparaffinised in Histoclear (Agar Scientific, Essex, UK) for three 3-minute washes and rehydrated through graded alcohols: two 3-minute washes in 100% ethanol, followed by 3 minutes each in 95%, 90%, 80%, 70%, 50%, and 30% ethanol. Sections were then washed in deionised water for 5 minutes and stained with Harris haematoxylin for 3 minutes. After rinsing in water for 5 minutes, slides were differentiated in acid alcohol (3% HCl in 70% ethanol) for 3 seconds and rinsed again in water for 2 minutes. Sections were then counterstained with eosin for 30 seconds before dehydration through graded alcohols: 2 minutes each in 30%, 50%, 70%, 80%, 90%, and 95% ethanol, followed by two 2-minute washes in 100% ethanol. Slides were cleared in Histoclear for three 3-minute washes, mounted using DPX mounting medium (distyrene, plasticiser and xylene), and left to dry overnight. Stained sections were scanned at 20× magnification using a Hamamatsu NanoZoomer (Hamamatsu, Hertfordshire, UK), and images were viewed using NDP.view software. Staining was performed by Dr Jennifer Hay at the Glasgow Tissue Research Facility (GTRF).

2.2.2 Assessment of tumour border configuration at the invasive front

The whole H&E-stained section was reviewed to evaluate the tumour border configuration at the invasive front. Three scoring approaches were applied: the Karamitopoulou percentage-based method, the adapted Taskin infiltration scale method, and the TGP classification. The scoring approach for each method is described below.

2.2.2.1 Karamitopoulou method

The invasive front was scored by the infiltrating percentage (Karamitopoulou et al., 2015). A completely pushing border was scored as 0%. An infiltrative

component was estimated in 5% increments, and cases were then categorised into two groups: low (≤50% infiltrative score) and high (>50% infiltrative score). Representative images are shown in Figure 2-1.

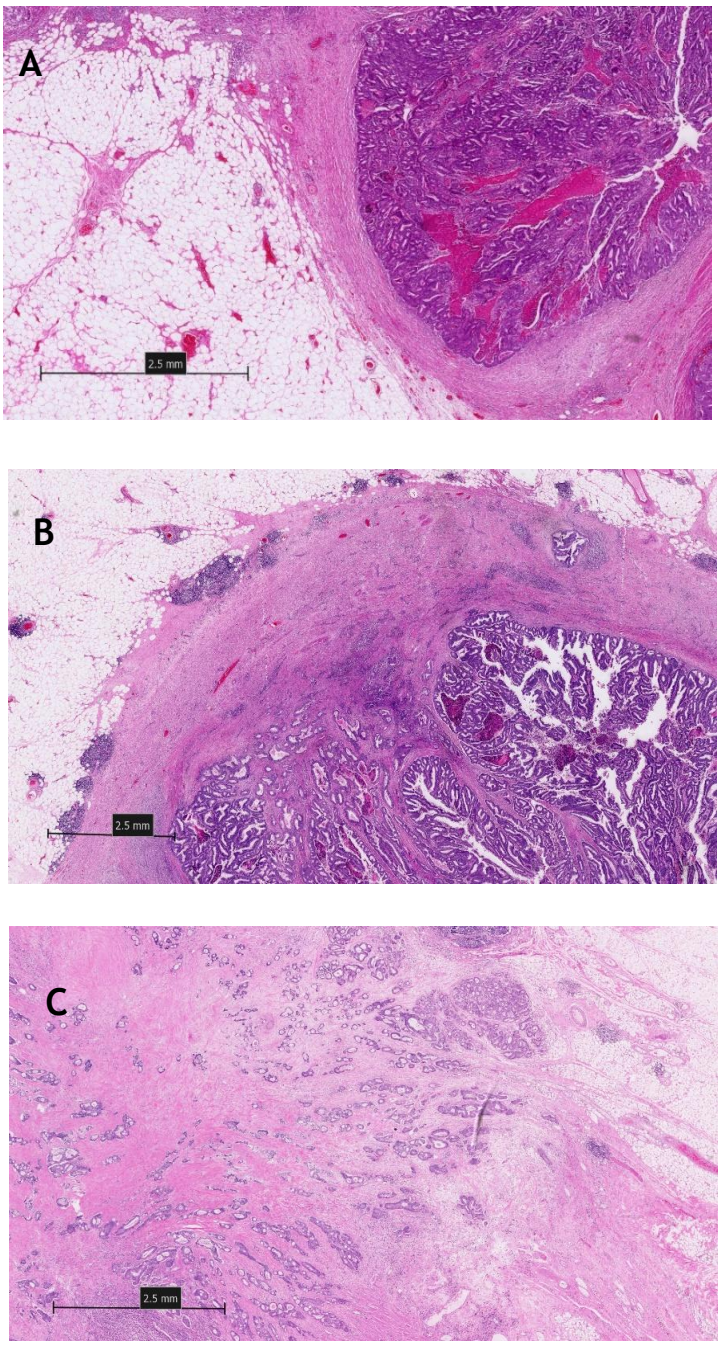


Figure 2-1: Representative H&E images of the Karamitopoulou scoring system at the invasive front of primary colorectal cancer.

The percentage of infiltrative growth is scored in 5% increments from 0% to 100%. (A) 0; (B) 50%; (C) 100%

2.2.2.2 Adapted Taskin method

This method uses a five-point scale to assess the degree of tumour border irregularity, tumour-cell clustering, and spread into the peritumoural tissue (Taskin et al., 2022). Score 1 indicates a well-demarcated tumour border. Score 2 indicates a mildly irregular tumour border, with tumour-cell clusters close to or connected to the main tumour. Score 3 represents larger tumour nodules or irregular tumour borders and small clusters extending near the peritumoural area. Score 4 represents small clusters spread more into the peritumoural area but these remained close to the main tumour. Score 5 represents loss of a clear tumour border, with multiple clusters located further away from the main tumour. Patients were finalised and grouped into two groups: low (score 1-3) and high (score 4-5). Representative images are shown in (Figure 2-2).

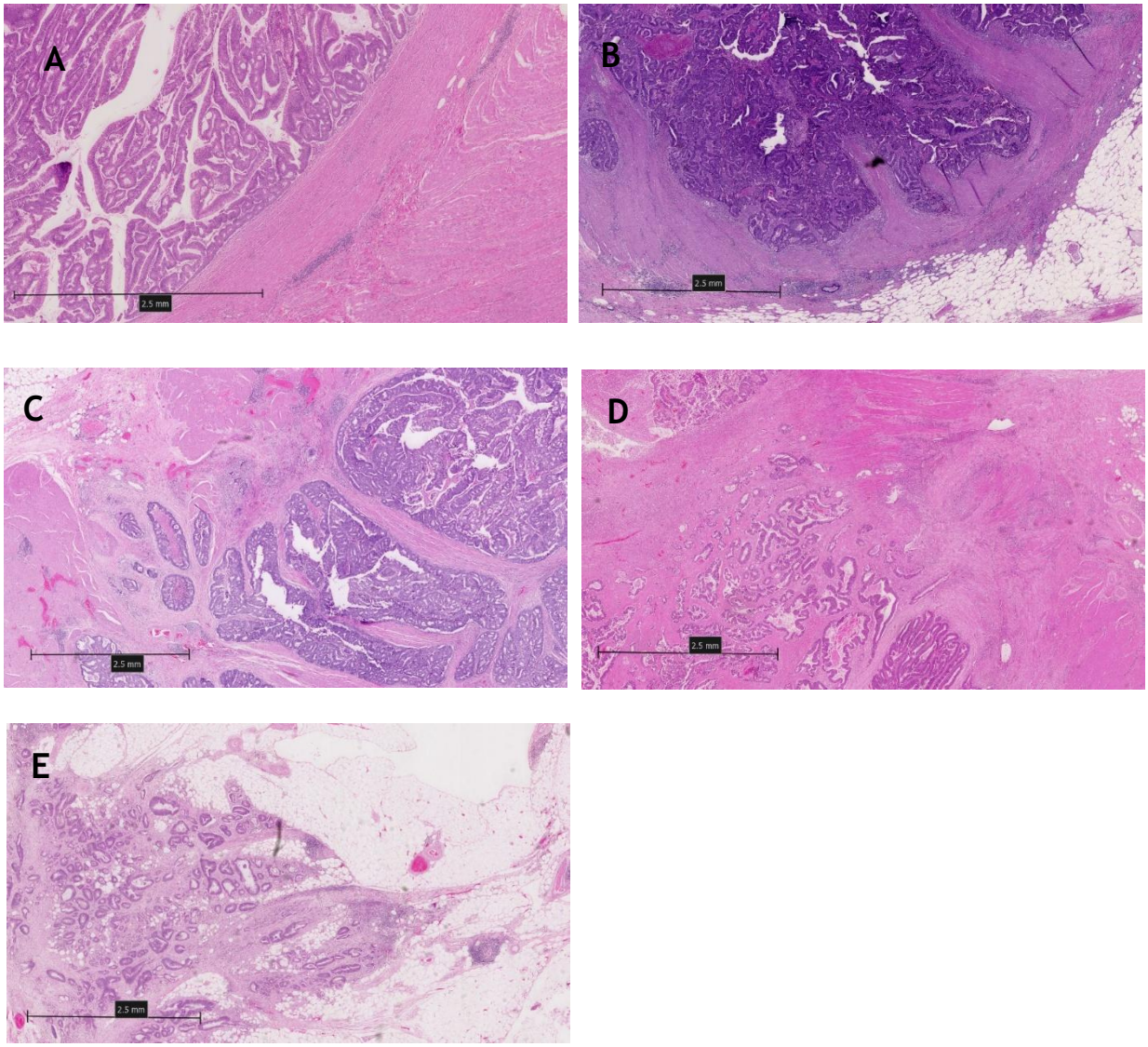


Figure 2-2: Representative H&E images of the invasive front scored using the Adapted Taskin method. (A) Score 1; **(B)** Score 2; **(C)** Score 3; **(D)** Score 4; **(E)** Score 5

2.2.2.3 Tumour growth pattern method

This method is generally classified into three categories: pushing, intermediate, and infiltrative (Morikawa et al., 2012). A pushing growth pattern was assigned when the invasive front was well circumscribed and clearly separated from the surrounding tissue. An intermediate growth pattern was assigned when the invasive front was irregular and showed invasion of medium- or large-sized glands. An infiltrative growth pattern was assigned when small glands or tumour clusters invaded the surrounding tissues diffusely, with no clear tumour border. Where more than one growth pattern was present along the invasive front, the predominant pattern, defined as that occupying the greatest proportion of the front, was recorded. Representative images are shown in (Figure 2-3).

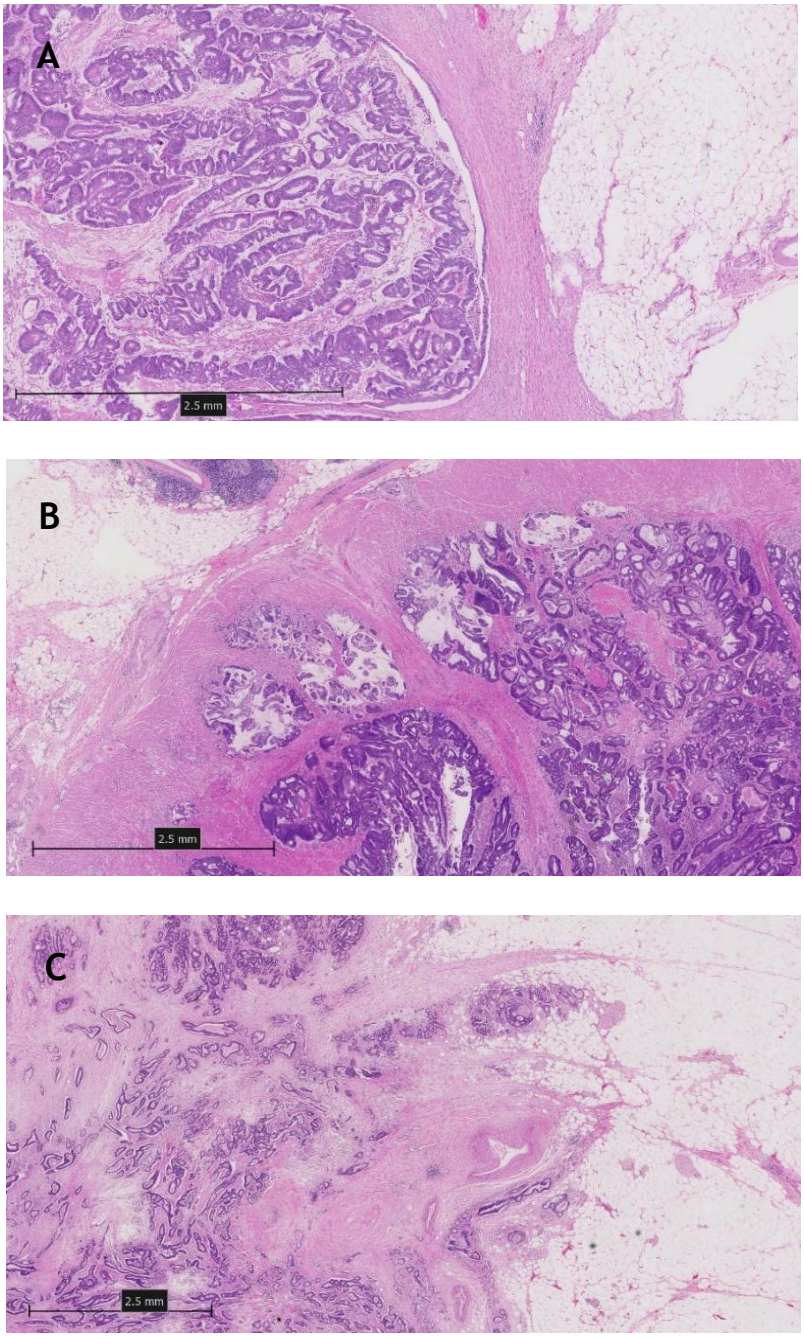


Figure 2-3*:* Representative H&E images of the invasive front scored using the tumour growth pattern (TGP) method. (A) Pushing growth pattern; **(B)** Intermediate growth pattern; **(C)** Infiltrative growth pattern

2.2.3 Interobserver assessment (IOA)

To evaluate the reproducibility of the scoring, H&E-stained sections from 55 patients (approximately 10% of the GRI test cohort) were randomly selected for interobserver assessment by three observers: Christopher Bigley (CB), Ashley McCulloch (AM), and Walaiphorn Woraharn (WW). The selection of a subset of approximately 10% for interobserver assessment was based on approaches used in previous studies, in which a proportion of cases was independently co-scored to

assess scoring reproducibility (Park et al., 2016; van Wyk et al., 2016). The selected cases were initially scored independently by each observer. Cases with discrepant scores were reviewed and discussed to clarify the interpretation of the scoring criteria. Following this discussion, the observers independently rescored the cases, and these final independent scores were used to calculate interobserver agreement. Interobserver agreement for the Karamitopoulou method was assessed using the intraclass correlation coefficient (ICC), as it yields a continuous percentage score (Koo & Li, 2016). Agreement for the adapted Taskin scale and TGP classification was assessed using weighted kappa, as both were considered ordered categorical assessments of increasing tumour-border disruption (Landis & Koch, 1977).

2.3 Study endpoints and statistical analysis

2.3.1 Study endpoints

In the GRI test cohort, overall survival (OS) was measured from the date of surgery until the date of any cause of death or censoring. Cancer-specific survival (CSS) endpoint was measured from the date of surgery to colorectal cancer-related death.

In the TransSCOT validation cohort, disease-free survival (DFS) was used as the endpoint starting from the date of randomisation, or trial registration for patients who were randomised after the first 3 months of therapy, to relapse, development of a new colorectal cancer, or death from any cause.

The endpoints available differed between the two cohorts. In the GRI test cohort, OS, CSS and DFS were available. CSS was selected as the principal outcome because the primary hypothesis concerned the association between tumour border morphology and CRC-related mortality. OS was analysed in parallel, whereas DFS was not prioritised because its composite definition includes death from any cause. In the TransSCOT validation cohort, CSS was not available, and DFS was therefore used as the disease-related outcome. As CSS and DFS capture different clinical events, the TransSCOT analysis was used to assess whether the prognostic association of TGP could be reproduced in an independent cohort rather than as a direct validation using an identical endpoint.

2.3.2 Statistical analysis

Statistical analyses were performed using SPSS version 28.0. Associations between categorical clinicopathological variables and scoring groups were calculated using the Chi-square test. The Kaplan-Meier curves were compared using the log-rank test. Univariable Cox regression was performed to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). The significant variables from univariable analysis were then entered into the multivariable Cox regression analysis using a backward conditional model.

2.4 TempO-Seq transcriptomic profiling

2.4.1 Selection of TGP groups for TempO-Seq transcriptomic analysis

The intermediate cases were excluded from the transcriptomic comparison because they contained mixed tumour-border features. The analysis was restricted to pushing and infiltrative tumours to maximise the biological contrast between the two morphologically distinct ends of the TGP spectrum.

2.4.2 TempO-Seq assay

TempO-Seq library preparation and sequencing was performed using the Human Whole Transcriptome v2.0 panel (BioSpyder Technologies, Carlsbad, CA). The TempO-Seq design removes the need for RNA extraction or reverse transcription and tolerates the degraded RNA typical of FFPE specimens (Trejo et al., 2019).

An FFPE full section was scraped from the slide, and lysis buffer was added. The sample was then overlaid in mineral oil and heated, prior to being lysed with a protease mix and combined with a mixture of detector oligonucleotides (DOs). DOs are designed as pairs that anneal in immediate juxtaposition to one another on the target RNAs, and ligated. Amplification of the ligated oligonucleotides was then performed by PCR using a unique primer set for each sample, which introduced a sample-specific barcode and Illumina sequencing adaptors. Barcoded samples were pooled into a single library and sequenced on an Illumina HiSeq 2500 High Output v4 flow cell. Sequencing reads were demultiplexed using BCL2FASTQ

software (Illumina, USA). FASTQ files were then aligned to the Human Whole Transcriptome v2.0 panel, which consists of 22,537 probes, using STAR (Trejo et al., 2019), allowing up to two mismatches in the 50-nucleotide sequencing read and producing BAM files. Raw gene counts were subsequently generated from the BAM files using featureCounts.

2.4.3 Pre-processing of TempO-Seq data and downstream analysis

Samples with total counts lower than 0.25 standard deviations (SD) below the mean total count across all samples were excluded. At the gene level, probes with variance below the 25th percentile across samples were removed as non-informative, while the more variable probes were retained for downstream analysis. Where multiple probes targeted the same gene, counts were summed prior to differential expression analysis. Batch effects were then addressed using the ComBat function from the SVA package in R version 4.4.2. Batch-corrected counts were then normalised using DESeq2.

2.4.3.1 The principal component analysis (PCA) and differential gene expression analysis (DGE)

The principal component analysis (PCA) plot was performed using normalised counts as input in the prcomp function. The differential gene expression (DGE) analysis was then performed between infiltrative and pushing growth patterns using raw counts as input in the edgeR package. The results were considered as differentially expressed based on the adjusted p-value (FDR) < 0.05, and an absolute log2 fold change ≥ 1 criteria. Finally, a Venn diagram was generated to visualise and identify the intersection between significantly up-regulated genes in infiltrative compared to pushing growth patterns and the five common CAF markers.

2.4.3.2 Gene Set Enrichment Analysis (GSEA) and Cell deconvolution by MCPcounter

Gene Set Enrichment Analysis (GSEA) was conducted using clusterProfiler package for the Hallmark and Gene Ontology Biological Process (GO: BP) gene sets. Then, genes were ranked by log2 fold-change. Normalised enrichment scores (NES) were calculated, and a false discovery rate (FDR) <0.05 was considered

significantly enriched. The MCPcounter package was utilised using normalised counts to estimate cell type in the tumour microenvironment (TME). Differences in cell population scores between infiltrative and pushing growth patterns were evaluated using the non-parametric Wilcoxon test and $p<0.05$ was considered as statistically significant.

2.4.3.3 Consensus molecular subtype classification

CMS classification was performed using CMSclassifier. Samples assigned to one of the four CMS subtypes (CMS1-4) were included in the analysis, while samples reported as not classified were excluded. The proportion of CMS subtypes was then compared between TGPs using Pearson's chi-squared test.

2.4.3.4 Genes of interest comparison

The log2 Counts Per Million (log2CPM) expression of common CAF markers (ACTA2, PDPN, PDGFRA, FAP, and S100A4) and EMT markers (CDH1, CTNNB1, SNAI1, ZEB2, FSCN1, TWIST2, and SOX9) were compared between infiltrative and pushing growth patterns using the Wilcoxon test and $p<0.05$ was defined as statistically significant.

2.5 Immunohistochemistry of EMT markers

Immunohistochemistry (IHC) for E-cadherin, ß-catenin, SOX9, ZEB1, TWIST1 and Snail had previously been performed and scored on the Glasgow Royal Infirmary (GRI) tissue microarrays (TMAs) as part of a separate study by Dr Amna Matly (AM) (Matly, 2024), and the resulting histoscores were used in the present work; the methods are summarised here, with full protocols reported in Matly (2024).

Antibodies were validated by Western blot and optimised prior to staining, which was carried out by AM. Stained TMAs were digitised by Dr Hannah Morgan (GTRF, Queen Elizabeth University Hospital) on a Hamamatsu NanoZoomer scanner and assessed at 20× magnification, with cores included where tumour cellularity exceeded 20% of the core area. Tumour-cell expression was quantified using the weighted histoscore method, which sums the percentage of cells at each staining

intensity, weighted by intensity (0–3+), yielding a continuous score from 0 to 300 weighted histoscore units (WHU).

Scoring was performed digitally using QuPath (ZEB1, TWIST1 and Snail) and Visiopharm Oncotopix (E-cadherin, ß-catenin and SOX9), by AM, Dr Christopher Bigley and Matthew McFarlane. To confirm reliability, 10% of cores were independently co-scored by a second, blinded observer (AM or Dr Jean Quinn), with inter-observer agreement assessed by intraclass correlation coefficient (ICC > 0.6) and visualised using Bland–Altman and scatter plots. For survival analysis, each marker was dichotomised into low and high expression using a cut-point derived from the maximally selected rank statistic for cancer-specific survival (R packages survminer and maxstat).

For the present analysis, the pre-existing scoring data low or high expression categories (derived from the weighted histoscores described above) were re-analysed by TGP. The proportion of cases with high versus low expression was compared between infiltrative and pushing TGPs using the chi-squared test, with $p < 0.05$ considered statistically significant.

2.6 Multiplex immunofluorescence

2.6.1 Staining

All multiplex staining steps were performed by Kara Luckett, a member of the John Le Quesne laboratory. Five different markers of a multiplex immunofluorescent assay were applied to six 4µm thick sections of human colorectal cancer tissue microarrays by Ventana Discovery Ultra autostainer (Roche Tissue Diagnostics, RUO Discovery Universal V21.00.0019). Heat-induced epitope retrieval was performed on sections using CC1 (Roche Tissue Diagnostics, 06414575001) for 32 minutes at 95°C. Then, the antibodies were performed on the following sequential steps, as shown in Table 2-1

Table 2-1*:* All antibodies for multiplex staining

Step	Antibody	Clone	Species	Supplier	Dilution / Time	Secondary	Detection
1	Podoplanin (PDPN)	D2-40	Mouse	Roche	RTU / 32 min	OmniMap Ms HRP (12 min)	CY5 (12 min)
2	FAP	F1A4G	Rabbit	Cell Signaling	1:25 / 60 min	OmniMap Rb HRP (12 min)	Rhodamine 6G (20 min)
3	SMA	1A4	Mouse	Roche	RTU / 32 min	OmniMap Ms HRP (12 min)	FAM (8 min)
4	CD3	2GV6	Rabbit	Roche	RTU / 32 min	OmniMap Rb HRP (16 min)	Red 610 (60 min)
5	Pan-CK	AE1/AE3	Mouse	Leica	1:250 / 28 min	OmniMap Ms HRP (12 min)	Opal 780 (60 min)

Before SMA staining, goat Ig block (Roche Tissue Diagnostics, 07988214001) was applied for 12 minutes to prevent cross-reactivity between same mouse primary antibodies from a previous step. For Pan-CK staining, TSA-DIG (Akoya Bioscience, FP1502001KT) at 1:100 was stained for 12 minutes followed by Opal 780 (Akoya Bioscience, FP1501001KT) at 1:10 for 1 hour. The last step was the DAPI nuclear staining. Before slides scanning, all sections were mounted using ProLong Diamond antifade mountant (Thermo Fisher Scientific, P36970).

For the negative control, the same staining protocol was used on human CRC tissue sections and control CRC TMA sections with isotype-matched negative control antibodies: CONFIRM Negative Control Rabbit Ig (Roche Tissue Diagnostics, 760-1029) and Negative Control Mouse Monoclonal Antibody MOPC-211 (Roche Tissue Diagnostics, 760-2014).

The mIF panel was designed primarily to identify tumour cells, T cells and major CAF phenotypes. Although CD34 is commonly used as a vascular marker in conventional immunohistochemistry, it was not included in the present panel.

2.6.2 Image acquisition

Whole slide images were acquired at 20x magnification using the Akoya PhenoImager HT (Akoya Biosciences, software version HT 2.1.0). Images were spectrally unmixed using inForm software (Akoya Biosciences, version 3.1) to separate the fluorescent signals of each marker channel and remove tissue autofluorescence. Image quality was reviewed by Silvia Martinelli. Unmixed images were then imported into Visiopharm (Oncotopix Discovery, Visiopharm A/S, Hørsholm, Denmark; version 2021.02.5.10297) for downstream image analysis.

2.6.3 Image analysis

The image analysis pipeline consisted of 3 Application Protocol Packages (APPs), tissue segmentation, cell detection, and marker classification, respectively, using Visiopharm software (version 2021.02.5.10297). The six TMA slides comprised 262 TMA cores at the invasive margin of primary CRC from the GRI cohort.

2.6.3.1 Tissue segmentation

The APP was adapted from a previously established APP by Rachel L. Baird of the John Le Quesne laboratory. This APP applied the deep learning algorithm to classify tissue compartments into the following: tumour, stroma, necrosis, and background.

Representative TMA cores were drawn for training. Then, the deep learning model was trained iteratively as follows: different areas were annotated, then the model was retrained. The APP was trained for approximately 1 million iterations. Post-processing steps were applied to remove miscellaneous tissue labels smaller than 3 μm and separate merged nuclei where two cells were detected as a single cell. The final APP was applied to all TMA cores, and approximately 20% were reviewed by a pathologist, Kai Rakovic.

2.6.3.2 Cell detection

The cell detection APP established by Rachel L. Baird was applied. The APP was trained using U-Net deep learning. An iterative training process was performed: annotate, train APP, run on example cores, correct U-Net annotations, and retrain. Post-processing steps included removing objects smaller than 5 µm as they might be artefacts or debris. Then, 20% of all core images were reviewed by a pathologist Kai Rakovic, and the cells were correctly detected.

2.6.3.3 Marker classification and data export

Following cell detection, each cell was labelled based on the fluorescence intensity of each marker channel within the whole-cell label. Intensity thresholds were set for each marker to classify cells as positive or negative, determined by a pathologist, Kai Rakovic, across multiple TMA cores. Thresholds were applied consistently across all images. For each cell, the following data was exported: positivity status for each marker (CK, CD3, SMA, PDPN, FAP), XY coordinates, tissue compartment, and mean marker intensities. Cell phenotyping was then performed in R.

2.6.4 Cell phenotyping

Cells were phenotyped separately within each core using a combination of marker positivity and DBSCAN-based spatial clustering in R. CK-positive cells were classified as tumour cells, while CD3-positive cells were classified as T lymphocytes. These classifications were assigned first and were prioritised over all subsequent stromal categories.

αSMA is expressed by both vascular smooth muscle and fibroblasts, while PDPN is expressed by both lymphatic endothelial cells and fibroblasts. Therefore, marker positivity alone was not sufficient to distinguish vessel-associated cells from fibroblasts. To separate these populations, αSMA-positive cells and PDPN-positive cells were each classified as vessel-associated only if they fulfilled three criteria.

First, the cell had to be located within a dense spatial cluster, identified separately within each core using DBSCAN and required to contain a minimum of

cells as shown in Table 2-2. This criterion reflected the close, contiguous arrangement of cells within vessel walls.

Second, the cluster had to show a narrow, tubular morphology rather than a broad sheet-like arrangement, therefore, its minor-axis width had to fall below predefined threshold as shown in Table 2-2. This helped distinguish thin vessel-wall structures from dense stromal regions. Third, the cell had to be negative for FAP and negative for the alternative vessel-associated marker: αSMA-defined vessel cells had to be PDPN-negative, and PDPN-defined vessel cells had to be αSMA-negative. This was because FAP was used to identify fibroblasts. Cells that did not meet all three criteria were retained within the CAF category.

Table 2-2: Settings applied to identify vessels

Step	Setting	αSMA	Podoplanin (PDPN)
Dense cluster (DBSCAN)	Search radius	40 μm	50 μm
Dense cluster (DBSCAN)	minimum points (minPts)	5	5
Cluster size	smallest cluster kept	Over 10 cells	Over 5 cells
Shape	widest allowed (shorter direction)	180 μm	180 μm

Cluster width was measured along the shorter axis of each cluster using principal component analysis. The width threshold was then selected by comparing the identified clusters with the corresponding mIF images.

The remaining cells positive for αSMA, FAP, or PDPN were assigned to CAF subtypes according to their marker expression profile: αSMA only, FAP only, PDPN only, αSMA+FAP+, αSMA+PDPN+, FAP+PDPN+, or αSMA+FAP+PDPN+. Cells negative for all three markers were classified as other. These phenotype labels were then used in the downstream spatial analyses.

Lacking endothelial and smooth muscle-specific markers, CD31 and desmin, respectively, αSMA-positive vessels and smooth muscle cells could not be distinguished from αSMA-positive fibroblasts. Therefore, vessels within dense

stroma or captured in cross-section were often merged with surrounding fibroblasts and not identified as separate structures. The αSMA-positive, FAP-negative populations (αSMA only CAFs) may therefore include non-fibroblast cells, particularly vascular smooth muscle cells and, where present, smooth muscle from the bowel wall. As these cells are FAP-negative, any misclassification would be limited to the αSMA-only CAF subtype and would not affect the prognostic endpoints based on FAP or combined αSMA+FAP+ expression. Therefore, the αSMA-only CAF subtype was interpreted with caution. One representative core is shown to illustrate the output of the phenotyping pipeline in Figure 2-4.

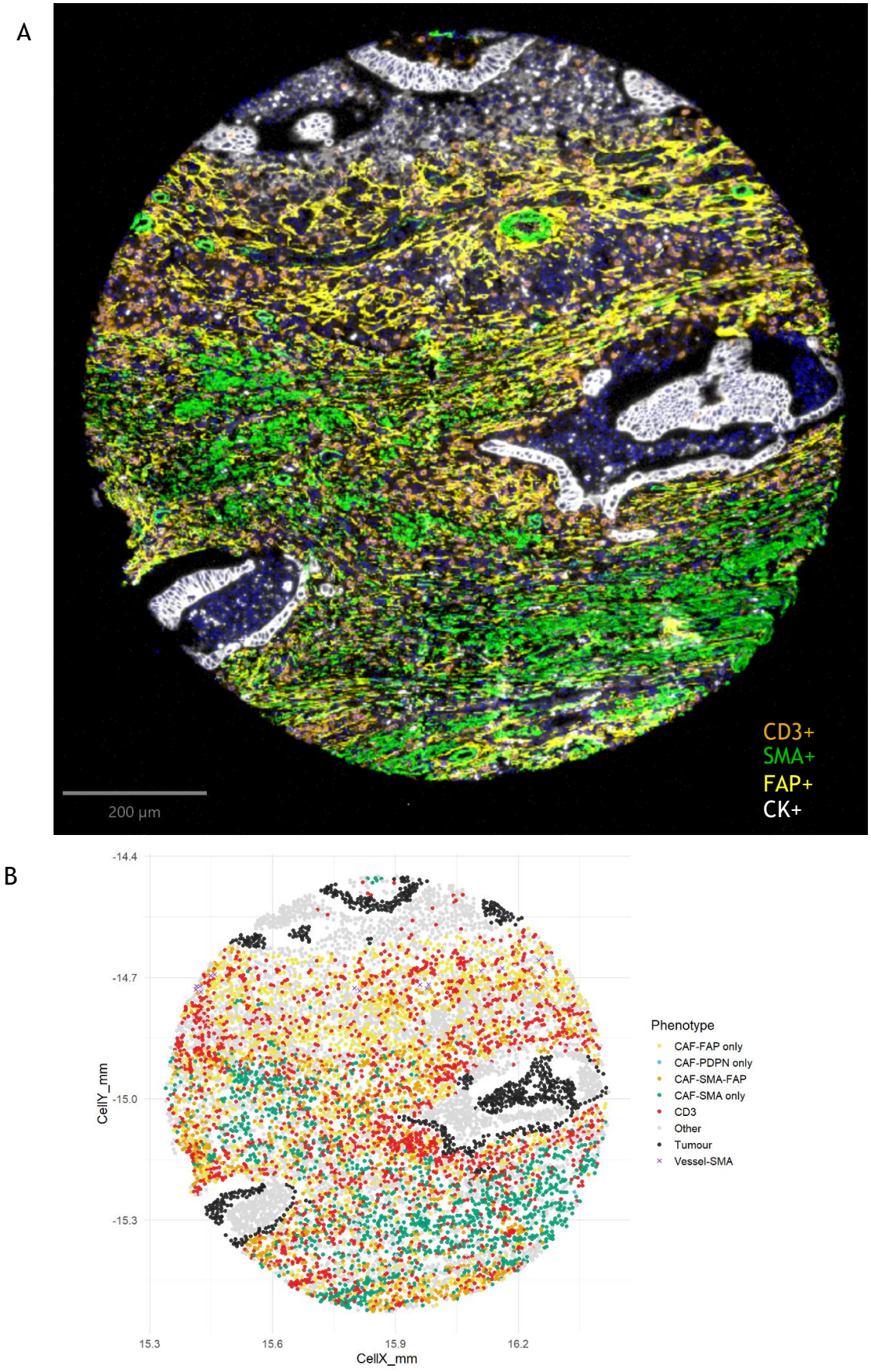


Figure 2-4: Representative example of the cell phenotyping output in a single core. (A) Multiplex immunofluorescence image (markers in the colour key). (B) The same core plotted in R, with its assigned phenotype following the rules in Section 2.6.4.

2.6.5 Downstream analysis

2.6.5.1 Cell composition

Cell compositions were compared between pushing and infiltrative growth patterns. Each patient consisted of 2-3 TMA cores, and the phenotype of each core was calculated as the number of cells of that phenotype divided by the total number of cells within a core. A zero count was assigned when no cells of that phenotype were present.

Cell compositions at the core level were then summarised at the patient level by calculating the median composition across all cores within each patient. The Wilcoxon rank-sum test was applied to test the differences between pushing and infiltrative growth patterns compositions. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg (BH) method, and an adjusted p-value of less than 0.05 was considered statistically significant.

2.6.5.2 Cell density

Cell density was calculated as the total number of cells divided by the total tissue area. The tissue area of each core was estimated from the range of X and Y cell coordinates, calculated by multiplying the horizontal and vertical span, and converted to mm^2. Core-level cell counts and areas were then combined at the patient level by summing the total cell counts and total tissue areas across all cores for each phenotype. A zero count was assigned when no cells of that phenotype were present.

The Wilcoxon rank-sum test was applied to test the differences between pushing and infiltrative growth pattern densities. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg (BH) method, and an adjusted p-value of less than 0.05 was considered statistically significant.

2.6.5.3 Nearest distance index (NDI)

The NDI was calculated using the spatstat package in R. For each TMA core, the median nearest-neighbour distance from cell A to the nearest cell B was calculated (r_obs) and compared to the expected median distance under complete spatial randomness (r_exp), calculated as:

$$r\_exp = \sqrt{\frac{ln2}{\pi \times \lambda\_B}}$$

where λ_B is the density of cell B within the observation window, defined as the number of B cells divided by the total area of the core. The NDI was then calculated per core as:

$$NDI = \log2(\frac{r\_obs}{r\_exp})$$

Core-level NDI values were subsequently summarised at the patient level by taking the median across all cores (2–3 per patient) for each NDI metric.

NDI was used in preference to the raw nearest-neighbour distance, which is strongly affected by the density of the target cell population and so cannot, on its own, distinguish a genuine difference in how closely the two cell types lie to one another from a simple difference in cell numbers (Figure 2-5).

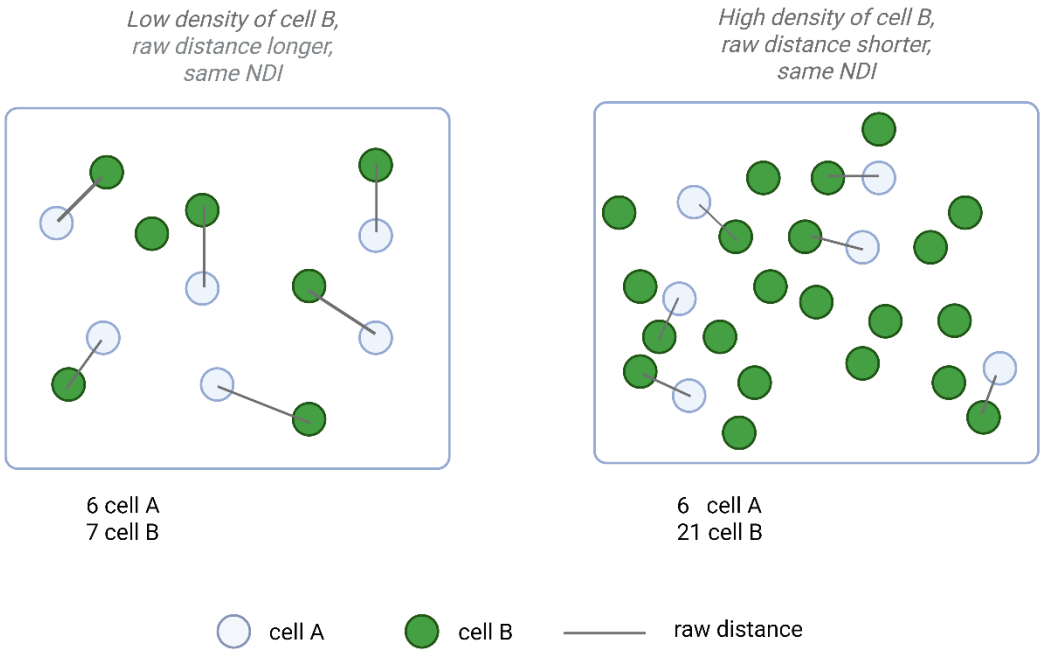


Figure 2-5: Schematic showing how the Nearest Distance Index (NDI) removes the effect of cell density. On the left, cell B is sparse, and the raw nearest-neighbour distances are long; on the right, cell B is dense, and the same distances are short. NDI reflects how closely cells lie to one another, not how many are present. Figure created by BioRender.

For the NDI interpretation, Table 2-3 summarises the meaning of NDI values. A positive NDI indicates that cells A and B are located further apart than expected, suggesting a spatial exclusion. A negative NDI indicates that cells A and B are located closer than expected, suggesting co-localisation. A zero NDI indicates that the spatial relationship between cell A and B does not differ from randomness.

Table 2-3: The NDI interpretation

NDI value	Meaning	Spatial pattern
NDI < 0	Closer than expected	Cell A and B are nearer to each other than random placement would predict
NDI = 0	As expected by random	No spatial preference: distance matches what random distribution predicts
NDI > 0	Further than expected	Cell A and B are further apart than random placement would predict

In this thesis, the NDI is treated as a continuous measure. A higher NDI means greater separation between cells A and B. The TGPs are compared on this scale, so a higher NDI in one pattern means A and B lie further apart in that pattern than in another, even when both NDI values are positive.

2.6.5.4 NDI comparison between tumour growth patterns

NDI values were then compared between infiltrative and pushing growth patterns using linear regression, adjusting for log-transformed target cell density to account for differences in cell density across patients. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg (BH) method, and an adjusted p-value of less than 0.05 was considered statistically significant.

2.6.5.5 Survival analysis

Each NDI metric was separated into two groups, high and low, using the maximally selected rank statistic method (MaxStat). This method identifies the optimal cutpoint that results in the most significant difference in cancer-specific survival between the two groups. A minimum proportion of 20% was required to prevent either group from containing too few patients.

Univariate Cox regression was performed for each NDI metric, including clinicopathological variables such as tumour budding, age, TGP, T stage, and N stage. Variables that were significant ($p < 0.05$) in univariate analysis were then entered into a multivariate Cox regression. Backward stepwise selection was applied. All survival analyses used cancer-specific survival (CSS) as the endpoint.

2.7 GeoMx spatial transcriptomics

2.7.1 Sample selection

GeoMx spatial transcriptomic profiling was performed on formalin-fixed paraffin-embedded (FFPE) CRC tissue microarray (TMA) sections derived from the invasive front of primary tumours. Each TMA block included internal control tissue cores, including a normal colon tissue. A total of 2 TMA slides were included in the analysis, comprising 51 cases, of which 31 were classified as infiltrative and 20 as pushing growth patterns.

2.7.2 GeoMx Whole Transcriptome Atlas (WTA) workflow

The overall GeoMx workflow used in this study is summarised in Figure 2-6, including morphology-guided ROI selection, compartment-specific tissue segmentation, UV-mediated barcode cleavage, oligonucleotide collection, and sequencing-based transcriptomic readout.

FFPE TMA sections were processed according to the manufacturer's protocol for the GeoMx Whole Transcriptome Atlas (WTA) assay. Tissue sections were stained with fluorescent markers for morphology, including pan-cytokeratin (CK) and alpha-smooth muscle actin (αSMA), to visualise tissue architecture and guide spatial profiling. In parallel, barcoded RNA probes were hybridised to RNA targets in situ to enable transcriptomic profiling.

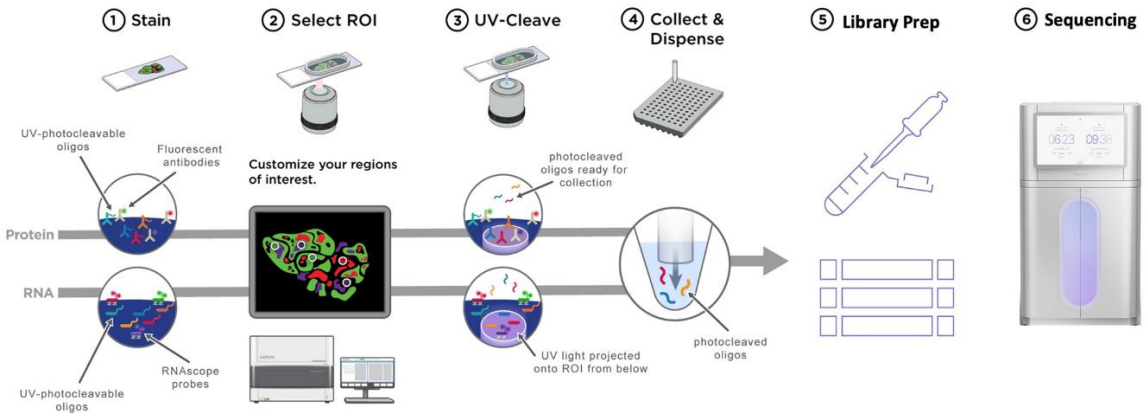


Figure 2-6: Overview of the GeoMx Whole Transcriptome Atlas (WTA) workflow.

FFPE CRC tissue sections were stained with fluorescent morphology markers to guide ROI selection at the invasive front, while barcoded RNA probes were hybridised in situ for transcriptomic profiling. Following ROI selection, UV-mediated cleavage released oligonucleotide barcodes from the selected ROI or segment, which were then collected, prepared as sequencing libraries, and sequenced for downstream spatial transcriptomic analysis. The figure was reproduced from NanoString Technologies, Inc.

2.7.3 Region of interest selection

Regions of interest (ROIs) were selected from the fluorescence image based on tissue morphology at the invasive front. ROI selection was guided by the distribution of CK and αSMA staining to capture biologically relevant tumour–stroma regions.

2.7.4 Tissue segmentation

Following ROI selection, tissue segmentation was performed using Visiopharm® (Visiopharm, Hørsholm, Denmark) to generate compartment-specific masks within each ROI. U-NET deep learning workflow developed by Dr Assya Legrini was trained using the Author™ module on representative CRC TMA images with variable staining intensity and tissue morphology.

Manual annotations were generated to define three tissue compartments: CK-positive (CK+), αSMA-positive (αSMA+), and Rest (CK−SMA−). The CK+ compartment was used to identify tumour-rich epithelial regions, the αSMA+ compartment was used to identify αSMA-enriched stromal regions, and the Rest compartment was defined as tissue negative for both CK and αSMA (CK−SMA−).

Model training was performed iteratively using the following workflow: annotate → train application → apply to representative images → review segmentation output → correct annotations → re-train. Training continued until satisfactory segmentation quality was achieved across representative sections. Segmentation quality was assessed by visual comparison of predicted masks with the corresponding fluorescence images. Post-processing steps were applied to improve mask quality, including the removal of small artefactual objects below 3 µm, filling of minor unlabelled gaps within masks, and refinement of compartment boundaries where required.

The final segmentation masks were exported and used to define tumour, αSMA-enriched stromal, and Rest compartments for downstream spatial profiling.

2.7.5 Oligonucleotide collection and sequencing

Following ROI selection and tissue segmentation, ultraviolet (UV) illumination was applied within the GeoMx platform to specifically cleave photocleavable oligonucleotide barcodes from the selected ROI or segment. The released oligonucleotides were collected by microcapillary aspiration into individual wells, each corresponding to a specific ROI or segment.

Collected oligonucleotides were then subjected to library preparation and sequencing according to the manufacturer's protocol. Sequencing reads were subsequently assigned back to their original ROI or segment, thereby enabling spatially resolved transcriptomic profiling of compartment-specific tissue regions.

2.7.6 GeoMx DSP platform quality control and normalisation

Sequencing-level quality control was first reviewed in the GeoMx DSP platform using the built-in NGS QC metrics. The following thresholds were used to assess assay performance: raw read threshold of 1,000 reads, aligned reads >80%, sequence saturation 50%, negative probe count of 4, no-template control count of 1,000, and minimum nuclei count of 100. Biological probe QC was also assessed within the DSP platform using a geometric mean probe threshold of 0.1, standard deviation from the geometric mean of 2, and a Grubb's outlier test threshold of 20%. Target-level filtering was then performed in the DSP platform using both

frequency filtering (10%) and limit of quantification (LOQ) filtering. Following platform-based quality control and filtering, the LOQ+Q3-normalised expression matrix was exported for downstream analysis.

2.7.7 Principal component analysis (PCA)

Principal component analysis (PCA) was conducted on the exported LOQ+Q3-normalised expression matrix to identify major sources of transcriptomic variation across AOIs. The 2,000 most variable genes, ranked by row variance, were selected for analysis. PCA was performed on the transposed, centred, and scaled expression data using R's prcomp function. The AOIs were plotted based on the first two principal components and labelled by spatial segment (Tumour, αSMA, Rest) and, if relevant, by TGP (Infiltrative and Pushing). Connecting lines between AOIs from the same patient were added to illustrate spatial variation within individuals across different compartments.

2.7.8 Data-driven marker gene identification

To identify genes enriched in each spatial compartment, a data-driven method was used. For each segment (Tumour, αSMA, Rest), the mean normalised expression of each gene was compared between AOIs assigned to that segment and all other AOIs. Genes were ranked by the difference between the mean in-segment and out-of-segment expression, and the top 12 genes with the largest positive differences were chosen for each segment. To prevent redundancy, genes identified for one segment were not allowed in the marker list of another segment. The resulting gene set was visualised as a row-scaled (z-score) heatmap, with AOIs ordered by TGP and segment.

2.7.9 Differential gene expression analysis

Differential gene expression (DGE) analysis was conducted in R using the limma package on the exported LOQ+Q3-normalised expression matrix. For comparisons between spatial compartments, a linear model was fitted with mask (Tumour, αSMA, Rest) as the variable of interest and patient included as a covariate to account for inter-patient variability. All pairwise contrasts between the three compartments were tested. To evaluate transcriptomic differences between infiltrative and pushing TGPs, separate analyses were conducted within

each spatial compartment. Within each mask, a linear model was fitted with TGP as the variable of interest.

To account for repeated measurements from the same patient, in-patient correlation was estimated using the duplicateCorrelation function in limma and incorporated into the model fitting. In both analyses, moderated statistics were calculated using the empirical Bayes method, and p-values were adjusted for multiple testing using the Benjamini–Hochberg (BH) procedure. For visualisation, volcano plots were generated by plotting log2 fold change against −log10(FDR). Genes with FDR < 0.05 and an absolute log2 fold change ≥ 1 were highlighted in the volcano plots.

2.7.10 Pathway analysis

Single-sample gene set enrichment analysis (ssGSEA) was performed in R using the GSVA package on the LOQ+Q3-normalised expression matrix imported from the data. The Hallmark gene set collection was obtained from the Molecular Signatures Database (MSigDB) via the msigdbr package. Only gene sets with at least five overlapping genes in the expression matrix were kept. ssGSEA scores were calculated for each AOI individually using the ssGSEA method with alpha set to 0.25 and score normalisation enabled.

For comparisons between spatial compartments, ssGSEA scores were analysed using limma, with mask as the main variable and patient included as a covariate. All pairwise contrasts between Tumour, αSMA, and Rest were tested. To compare pathway activity between infiltrative and pushing TGPs, separate analyses were performed within each spatial compartment. Within each mask, a linear model was fitted with TGP as the variable of interest, and within-patient correlation was accounted for using duplicateCorrelation. Moderated statistics were calculated with the empirical Bayes method, and p-values were adjusted using the BH procedure. Significant pathway findings were visualised using horizontal bar plots showing the log-fold change in ssGSEA scores.

2.7.11 ssGSEA and Nearest Distance Index (NDI) correlation analysis

Pre-ranked Gene Set Enrichment Analysis (GSEA) and Single Sample Gene Set Enrichment Analysis (ssGSEA), and NDI correlation were constructed by using GSEA as the first step with a pre-ranked gene list. Briefly, for each endpoint per mask, genes were ranked by Spearman's rank correlation between log2 gene expression and NDI values, and enrichment scores were then calculated to test whether genes in each pathway were enriched at the top or bottom of this ranked list. Subsequently, the top five positively enriched (NES >0) and top five negatively enriched (NES <0) pathways (ranked by adjusted p-values) were selected, and the selected pathways were combined (union) as the input pathways for ssGSEA in the next step. For each sample, log2 gene expression was ranked from high to low and scored to generate ssGSEA enrichment scores for each selected pathway. Finally, within each endpoint per mask, Spearman correlation was applied to analyse the association between NDI and ssGSEA scores, with multiple testing adjustment using the Benjamini–Hochberg (BH) procedure (FDR <0.05).

2.7.12 EcoTyper cell-state inference and Nearest Distance Index correlation analysis

To obtain deeper mechanistic insights beyond spatial metrics, cell-state abundances were inferred from GeoMx AOI-level expression profiles using EcoTyper (recovery mode) from carcinoma reference framework (Luca et al., 2021). GeoMx expression values were used on the linear (non-log) scale. EcoTyper recovery mode was performed using the built-in Carcinoma discovery reference, and analyses were run separately for each mask (Rest, Tumour, αSMA). Cell state abundances are normalised to approximately sum to 1 within that cell type. Therefore, state abundances are comparable across samples within the same cell type but are not directly comparable across different cell types.

For downstream analyses, Spearman correlations were computed between spatial metrics (per endpoint) and cell state abundance for each mask. Multiple testing was applied using the Benjamini–Hochberg procedure, and results were considered significant at FDR < 0.05.

2.7.13 Ligand-receptor, cell-state, and Nearest Distance Index correlation analysis

Ligand-receptor pairs were obtained from the LIANA Consensus resource (Dimitrov et al., 2022) using the LIANA R package. Gene symbol columns were used to extract ligand and receptor identifiers. Duplicate pairs and entries with missing gene symbols were removed. Only ligand genes present in the GeoMx expression matrix of the source mask and receptor genes present in the target mask were retained for analysis.

2.7.13.1 Endpoint mapping

Each spatial endpoint was mapped to a source and target GeoMx mask based on the biological identities of the cell populations measured by the NDI. For example, the αSMA+FAP+ to CD3 endpoint was mapped with the αSMA mask as the source (representing SMA/FAP-positive stromal cells) and the Rest mask as the target (representing CD3-positive T cells and other immune populations). An additional mapping using the Rest mask as both source and target was included for this endpoint to capture signalling within the immune/stromal compartment.

2.7.13.2 Ligand pre-screening

For each candidate sender state, ligands whose expression in the source mask correlated with the state abundance were identified. Specifically, for each ligand gene in the LIANA Consensus resource that was also present in the source mask expression matrix, the Spearman correlation between ligand expression and state abundance was computed across specimens sharing data for both measurements. Ligands with a positive correlation (rho > 0) and a nominal p-value below 0.05 were retained, up to a maximum of 50 ligands per sender state.

2.7.13.3 Weighted ligand-receptor scoring

For each retained ligand-receptor pair, a weighted interaction score was computed for each specimen as follows:

Score = (Ligand expression in source mask + Receptor expression in target mask) × √ (State abundance)

The square root of the sender state abundance served as a weight, such that the score is higher in specimens where the sender cell state is more abundant. This weighting approach assumes that ligand-receptor interactions are more biologically relevant when the sender cell state is present at higher levels.

2.7.13.4 Statistical testing

The association between each weighted LR score and the corresponding NDI value was assessed using two complementary approaches. First, Spearman's rank correlation was used to estimate the strength and direction of the association. Second, linear regression of the LR score on the standardised NDI value was performed to obtain a standardised beta coefficient and associated p-value. P-values from the linear regression were adjusted across all tested LR pairs using the Benjamini-Hochberg method. LR pairs were considered significant if the adjusted p-value (FDR) was below 0.05 and the absolute Spearman rho exceeded 0.30. Pairs where the ligand and receptor were the same gene were excluded.

2.7.13.5 Functional annotation

Significant LR pairs were manually annotated into functional themes based on the known biological roles of the ligand and receptor proteins, supported by current literature. Annotations included immune checkpoint signalling, extracellular matrix and barrier formation, growth factor signalling, and immunosuppressive stromal interactions.

Chapter 3 Evaluating the invasive front morphology in primary colorectal cancer: comparison of histological scoring methods and validation of the prognostic value of tumour growth pattern

3.1 Introduction

Colorectal cancer prognostication and therapeutic decisions are currently based on the AJCC/UICC TNM staging system. This classification requires assessment of the tumour, regional lymph nodes, and distant metastasis. Although the TNM system is the gold standard, patients within the same stage group have different clinical outcomes (O'Connell et al., 2004). Additional histopathological features are required to refine prognostic stratification, particularly to identify patients with a higher-risk tumour biology who need more intensive clinical management.

Surgical resection specimens from patients with primary CRC are routinely assessed by using haematoxylin and eosin stained (H&E) tissue sections as part of standard histopathological reporting. The invasive front of the tumour has been recognised as an important tumour-host interface area, where the tumour interacts with the surrounding stroma. Previous studies have shown that histo-morphological features in this area, including tumour budding and tumour border configuration, provide additional prognostic information beyond the conventional TNM staging (Graham et al., 2015; Keum et al., 2012; Lugli et al., 2012; Rajaganeshan et al., 2007). Tumour budding is a widely recognised feature at the invasive front of CRC. However, tumour budding reflects a microscopic view of individual tumour cells or small clusters of up to 4 cells at the invasive front. In contrast, the tumour border configuration, or TGP, focuses on a broader architectural assessment of the invasive front, reflecting a macroscopic view of tumour growth or invasion.

Several approaches have been proposed to assess the tumour border configuration. Jass et al. originally presented the invasive growth pattern in CRC stratified by pushing/expansile or infiltrative growth patterns (Jass et al., 1996). Morikawa et al. subsequently applied a three-category classification that adds an

intermediate group to Jass's original two patterns, capturing tumours with mixed border configurations (Morikawa et al., 2012). The present study adopted this three-category TGP. Karamitopoulou et al. later proposed an alternative scoring method based on the percentage of the infiltrative front component, aiming to provide a more reliable and quantitative approach for assessing tumour border configuration (Karamitopoulou et al., 2015).

While the TGP method provides the three-category classification of pushing, intermediate, and infiltrative invasion, and the Karamitopoulou approach offers a more quantitative estimate of the infiltrative percentage, an alternative morphology-based framework may provide an additional way to grade the architecture of the invasive front.

Although originally proposed in pancreatic neuroendocrine tumours, the Taskin method applies a structured five-point scale based on progressive changes in tumour border demarcation, edge irregularity, tumour cell clustering, and the extension into peri-tumoural area (Taskin et al., 2022). Therefore, in this study, the Taskin method was assessed as an exploratory external comparator, rather than as an established CRC scoring system. This was used to evaluate whether a graded infiltration-based framework could be reproducibly applied to the primary CRC.

Altogether, these approaches capture different levels of invasive front assessment: the three-category TGP, a colorectal cancer-specific percentage-based estimation, and an infiltration grading scale adapted from pancreatic tumours. Therefore, this study aimed to compare these three methods in primary CRC to determine which method provides the most reproducible and prognostically meaningful assessment of the invasive front. Following the comparison, the selected method was further evaluated in an independent CRC cohort, the TransSCOT cohort, to assess whether its prognostic value could be reproduced in an external dataset. The objectives of this study were:

1. To assess the tumour border configuration at the invasive front using three different approaches: the three-category TGP classification, the Karamitopoulou percentage-based method, and the adapted Taskin infiltration scale.

2. To evaluate and compare the interobserver reproducibility of these three scoring methods.

3. To determine the association between each scoring method and survival outcomes.

4. To compare the prognostic value of the three invasive front scoring methods in primary CRC.

5. To validate the prognostic value of the selected scoring method in an external cohort, TransSCOT.

3.2 Results

3.2.1 GRI test cohort

3.2.1.1 Interobserver assessment (IOA) agreement

Interobserver assessment agreement was evaluated for three scoring methods. For the Karamitopoulou method, the intraclass correlation coefficients (ICC) were 0.63 between CB and WW and 0.79 between AM and WW, indicating moderate to good reproducibility between observers. For the adapted Taskin method, weighted kappa values were lower, with agreement of 0.31 between CB and WW, and 0.27 between AM and WW. These values indicate fair agreement. For TGP classification, the weighted kappa value between CB and WW was 0.78, indicating substantial agreement.

Taken together, the interobserver agreement for the Karamitopoulou and TGP methods was higher than for the adapted Taskin method.

3.2.1.2 Clinicopathological characteristics of the GRI test cohort

After applying the inclusion and exclusion criteria, 538 patients with stage II-III primary CRC from the GRI cohort were included in the analysis. Most patients were aged 50-75 years (62.3%). Whereas 305 patients were male (56.7%) and 233 were female (43.3%). Most tumours were in the colon area (69.5%), while 30.5% were in the rectal area. Most patients had stage II or III CRC, accounting for 47.2% and 42.6%, respectively (Table 3-1)

Table 3-1: Associations between clinicopathological characteristics and the three invasive front scoring methods (Karamitopoulou, Taskin, and TGP) in the GRI cohort of 538 patients with stage I-III colorectal cancer. Bold p-values are statistically significant

Features	Total	Karamitopoulou method			Taskin method			TGP method			
		0-50% Infiltrative (n = 190)	51-100% Infiltrative (n = 348)	p-value	Score 1-3 (n = 176)	Score 4-5 (n = 362)	p-value	Expansile /pushing (n=168)	Intermediate (n=134)	Infiltrative (n = 236)	p-value
Age				0.344			0.516				0.776
<50	37 (6.9)	13 (6.8)	24 (6.9)		11 (6.3)	26 (7.2)		11 (6.5)	11 (8.2)	15 (6.4)	
50-75	335 (62.3)	111 (58.4)	224 (64.4)		105 (59.7)	230 (63.5)		100 (59.5)	82 (61.2)	153 (64.8)	
>75	166 (30.9)	66 (34.7)	100 (28.7)		60 (34.1)	106 (29.3)		57 (33.9)	41 (30.6)	68 (28.8)	
Sex				0.677			0.967				0.853
Female	233 (43.3)	80 (42.1)	153 (44.0)		76 (43.2)	157 (43.4)		70 (41.7)	58 (43.3)	105 (44.5)	
Male	305 (56.7)	110 (57.9)	195 (56.0)		100 (56.8)	205 (56.6)		98 (58.3)	76 (56.7)	131 (55.5)	
T stage				**<0.001**			**<0.001**				**<0.001**
pT1	12 (2.2)	10 (5.3)	2 (0.6)		10 (5.7)	2 (0.6)		10 (6.0)	1 (0.7)	1 (0.4)	
pT2	52 (9.7)	33 (17.4)	19 (5.5)		29 (16.5)	23 (6.4)		28 (16.7)	13 (9.7)	11 (4.7)	
pT3	315 (58.6)	112 (58.9)	203 (58.3)		105 (59.7)	210 (58.0)		99 (58.9)	93 (69.4)	123 (52.1)	
pT4	159 (29.6)	35 (18.4)	124 (35.6)		32 (18.2)	127 (35.1)		31 (18.5)	27 (20.1)	101 (42.8)	
N stage				**<0.001**			**<0.001**				**<0.001**
pN0	308 (57.2)	148 (77.9)	160 (46.0)		137 (77.8)	171 (47.2)		129 (76.8)	82 (61.2)	97 (41.1)	
pN1 & pN2	230 (42.8)	42 (22.1)	188 (54.0)		39 (22.2)	191 (52.8)		39 (23.2)	52 (38.8)	139 (58.9)	
Stage				**<0.001**			**<0.001**				**<0.001**
Stage I	55 (10.2)	37 (19.5)	18 (5.2)		33 (18.8)	22 (6.1)		33 (19.6)	12 (9.0)	10 (4.2)	
Stage II	254 (47.2)	111 (58.4)	143 (41.1)		104 (59.1)	150 (41.4)		96 (57.1)	70 (52.2)	88 (37.3)	
Stage III	229 (42.6)	42 (22.1)	187 (53.7)		39 (22.2)	190 (52.5)		39 (23.2)	52 (38.8)	138 (58.5)	
Site				0.987			0.742				0.420
Colon	374 (69.5)	132 (69.5)	242 (69.5)		124 (70.5)	250 (69.1)		116 (69.0)	99 (73.9)	159 (67.4)	
Rectal	164 (30.5)	58 (30.5)	106 (30.5)		52 (29.5)	112 (30.9)		52 (31.0)	35 (26.1)	77 (32.6)	
Differentiation				0.541			0.948				0.060
Moderate/Well	500 (93.3)	178 (94.2)	322 (92.8)		164 (93.2)	336 (93.3)		157 (94.0)	129 (97.0)	214 (90.7)	
Poor	36 (6.7)	11 (5.8)	25 (7.2)		12 (6.8)	24 (6.7)		10 (6.0)	4 (3.0)	22 (9.3)	
Adjuvant therapy				**<0.001**			**<0.001**				**0.005**
No	331 (70.1)	140 (80.0)	191 (64.3)		128 (80.5)	203 (64.9)		121 (78.6)	83 (71.6)	127 (62.9)	
Received	141 (29.9)	35 (20.0)	106 (35.7)		31 (19.5)	110 (35.1)		33 (21.4)	33 (28.4)	75 (37.1)	
Tumour budding				**0.002**			**0.003**				**<0.001**
Absent	373 (70.8)	146 (79.3)	227 (66.2)		135 (79.4)	238 (66.7)		126 (77.3)	108 (81.8)	139 (59.9)	
Present	154 (29.2)	38 (20.7)	116 (33.8)		35 (20.6)	119 (33.3)		37 (22.7)	24 (18.2)	93 (40.1)	

Features	Total	Karamitopoulou method			Taskin method			TGP method			
		0-50% Infiltrative (n = 190)	51-100% Infiltrative (n = 348)	p-value	Score 1-3 (n = 176)	Score 4-5 (n = 362)	p-value	Expansile /pushing (n=168)	Intermediate (n=134)	Infiltrative (n = 236)	p-value
Margin involvement				0.536			0.513				0.520
Absent	502 (93.3)	179 (94.2)	323 (92.8)		166 (94.3)	336 (92.8)		158 (94.0)	127 (94.8)	217 (91.9)	
Involved	36 (6.7)	11 (5.8)	25 (7.2)		10 (5.7)	26 (7.2)		10 (6.0)	7 (5.2)	19 (8.1)	
Venous invasion				**0.008**			**0.006**				**0.025**
Absent	215 (46.1)	94 (54.0)	121 (41.4)		87 (55.1)	128 (41.6)		80 (52.3)	58 (50.4)	77 (38.9)	
Present	251 (53.9)	80 (46.0)	171 (58.6)		71 (44.9)	180 (58.4)		73 (47.7)	57 (49.6)	121 (61.1)	
Peritoneal involvement				**<0.001**			**<0.001**				**<0.001**
Absent	396 (73.6)	161 (84.7)	235 (67.5)		150 (85.2)	246 (68.0)		143 (85.1)	111 (82.8)	142 (60.2)	
Involved	143 (26.4)	29 (15.3)	113 (32.5)		26 (14.8)	116 (32.0)		25 (14.9)	23 (17.2)	94 (39.8)	
Tumour perforation				0.569			0.176				0.964
0	520 (96.7)	184 (96.8)	336 (96.6)		169 (96.0)	351 (97.0)		163 (97.0)	129 (96.3)	228 (96.6)	
1	12 (2.2)	3 (1.6)	9 (2.6)		3 (1.7)	9 (2.5)		3 (1.8)	3 (2.2)	6 (2.5)	
2	6 (1.1)	3 (1.6)	3 (0.9)		4 (2.3)	2 (0.6)		2 (1.2)	2 (1.5)	2 (0.8)	
GMS				**<0.001**			**<0.001**				**<0.001**
0	94 (17.9)	39 (21.2)	55 (16.1)		34 (20.0)	60 (16.9)		33 (20.4)	23 (17.3)	38 (16.5)	
1	301 (57.2)	129 (70.1)	172 (50.3)		120 (70.6)	181 (50.8)		117 (72.2)	80 (60.2)	104 (45.0)	
2	131 (24.9)	16 (8.7)	115 (33.6)		16 (9.4)	115 (32.3)		12 (7.4)	30 (22.6)	89 (38.5)	
KM				0.144			0.378				0.595
Low grade	432 (82.1)	145 (78.8)	287 (83.9)		136 (80.0)	296 (83.1)		129 (79.6)	110 (82.7)	193 (83.5)	
High grade	94 (17.9)	39 (21.2)	55 (16.1)		34 (20.0)	60 (16.9)		33 (20.4)	23 (17.3)	38 (16.5)	
TSP				**<0.001**			**<0.001**				**<0.001**
Low	380 (72.2)	165 (89.7)	215 (62.9)		153 (90.0)	227 (63.8)		148 (91.4)	99 (74.4)	133 (57.6)	
High	146 (27.8)	19 (10.3)	127 (37.1)		17 (10.0)	129 (36.2)		14 (8.6)	34 (25.6)	98 (42.4)	
Time period				0.169			0.225				0.219
1997-2003	138 (29.6)	45 (25.9)	93 (31.8)		43 (27.2)	95 (30.8)		38 (24.8)	35 (30.4)	65 (32.8)	
2004-2008	144 (30.9)	51 (29.3)	93 (31.8)		44 (27.8)	100 (32.5)		45 (29.4)	33 (28.7)	66 (33.3)	
2009-2013	184 (39.5)	78 (44.8)	106 (36.3)		71 (44.9)	113 (36.7)		70 (45.8)	47 (40.9)	67 (33.8)	
MMR status				0.122			**0.048**				0.459
dMMR	149 (32.5)	62 (36.9)	87 (29.9)		59 (38.6)	90 (29.4)		51 (34.7)	32 (27.8)	66 (33.5)	
pMMR	310 (67.5)	106 (63.1)	204 (70.1)		94 (61.4)	216 (70.6)		96 (65.3)	83 (72.2)	131 (66.5)	

In the Karamitopoulou percentage-based method, 190 patients (35.3%) were grouped into the low group (0-50% infiltrative score), and 348 patients (64.7%) were grouped into the high group (51-100% infiltrative score). For the adapted Taskin method, 176 patients (32.7%) were classified as the low group (score 1-3), and 362 patients (67.3%) were classified as the high group (score 4-5). For TGP approach, 168 patients (31.2%) were defined as pushing pattern, 134 patients (24.9%) were defined as intermediate pattern, and 236 patients (43.9%) were defined as infiltrative pattern (Table 3-1).

3.2.1.3 Association between three scoring methods and clinicopathological characteristics

The association between each scoring method and clinicopathological features are shown in (Table 3-1). For the Karamitopoulou method, a high percentage group (51-100%) was significantly associated with higher T stage ($p<0.001$), higher N stage ($p<0.001$), stage III CRC ($p<0.001$), receipt of adjuvant therapy ($p<0.001$), high tumour budding ($p = 0.002$), venous invasion ($p = 0.008$), peritoneal involvement ($p<0.001$), higher GMS ($p<0.001$) and higher TSP ($p<0.001$).

Similarly, for the adapted Taskin method, the high score group (score 4-5) was significantly associated with higher T stage ($p<0.001$), higher N stage ($p<0.001$), stage III CRC ($p<0.001$), receipt of adjuvant therapy ($p<0.001$), high tumour budding ($p = 0.003$), venous invasion ($p = 0.006$), peritoneal involvement ($p<0.001$), higher GMS ($p<0.001$) and higher TSP ($p<0.001$).

The TGP method showed a similar correlation with adverse pathological features. The infiltrative growth pattern was significantly associated with higher T stage ($p<0.001$), higher N stage ($p<0.001$), receipt of adjuvant therapy ($p = 0.005$), high tumour budding ($p<0.001$), venous invasion ($p = 0.025$), peritoneal involvement ($p<0.001$), higher GMS ($p<0.001$) and higher TSP ($p<0.001$). In contrast, age, sex, tumour site, margin involvement, tumour perforation and KM grade were not significantly associated with TGP.

Altogether, these associations between three scoring methods and clinicopathological characteristics indicate that more infiltrative tumour borders were associated with adverse features.

3.2.1.4 Survival analysis

Overall survival (OS) analysis

In univariate Cox regression in all CRC (n=538), several clinicopathological features were significantly associated with OS, including age, T stage, N stage, margin involvement, venous invasion, peritoneal involvement, tumour budding, GMS, KM, TSP, and the time period (Table 3-2). None of the three scoring methods reached significance for OS (Karamitopoulou p = 0.155, Taskin p = 0.217, TGP p = 0.088). Consistent with this, Kaplan–Meier analysis with log-rank testing showed no significant difference in overall survival curves across the three methods (Figure 3-1).

Table 3-2: Univariate and multivariable Cox regression analyses of clinicopathological variables and the three invasive front scoring methods for overall survival in the GRI cohort of 538 patients. Bold p-values are statistically significant. * Not included in the multivariable analysis due to multicollinearity.

Features	Univariate HR (95% CI)	p value	Multivariable HR (95% CI)	p value
Age (<50/ 50-75/ >75)	2.40 (1.96 - 2.93)	**<0.001**	2.76 (2.19 - 3.49)	**<0.001**
Sex (Female/ Male)	1.02 (0.82 - 1.27)	0.857		
T stage (pT1/ pT2/ pT3 /pT4)	1.43 (1.21 - 1.69)	**<0.001**	1.41 (1.16 - 1.72)	**<0.001**
N stage (pN0/ pN1&pN2)	1.30 (1.05 - 1.62)	**0.015**	1.07 (0.83 - 1.38)	0.593
Site (Colon/ Rectum)	0.96 (0.76 - 1.21)	0.711		
Differentiation (Moderate&Well/ Poor)	1.04 (0.67 - 1.60)	0.873		
Margin involvement (Absent/ Involved)	2.21 (1.52 - 3.22)	**<0.001**	2.29 (1.44 - 3.64)	**<0.001**
Venous invasion (Absent/ Present)	1.41 (1.11 - 1.78)	**0.004**	1.21 (0.95 - 1.56)	0.128
Peritoneal involvement (Absent/ Involved)	1.49 (1.18 - 1.88)	**<0.001**	1.04 (0.68 - 1.59)	0.861
Tumour perforation (0/ 1/ 2)	1.32 (0.90 - 1.94)	0.162		
Tumour budding (Absent/ Present)	1.33 (1.06 - 1.67)	**0.015**	1.35 (1.05 - 1.72)	**0.019**
GMS (0/ 1/ 2) *	1.45 (1.22 - 1.72)	**<0.001**		
KM (Low/ High)	0.64 (0.48 - 0.87)	**0.005**	0.79 (0.55 - 1.14)	0.214
TSP (Low/ High)	1.44 (1.14 - 1.82)	**0.003**	1.25 (0.94 - 1.66)	0.121
Karamitopoulou method (0-50%/51-100%)	1.18 (0.94 - 1.48)	0.155		
Taskin method (0-3/4-5)	1.16 (0.92 - 1.45)	0.217		
TGP method (Pushing/ Intermediate/ Infiltrative)	1.12 (0.98 - 1.26)	0.088		
MMR status (dMMR/ pMMR)	0.89 (0.70 - 1.14)	0.357		
Time period (1997-2003/ 2004-2008/ 2009-2013)	0.84 (0.72 - 0.97)	**0.016**	0.88 (0.75 - 1.03)	0.103

In multivariable Cox regression, only age (HR = 2.76 (2.19 - 3.49, p<0.001)), T stage (HR = 1.41 (1.16 – 1.72, p<0.001)), margin involvement (HR = 2.29 (1.44 – 3.64, p <0.001)), and tumour budding (HR = 1.35 (1.05 – 1.72), p = 0.019) were independent prognostic factors for OS.

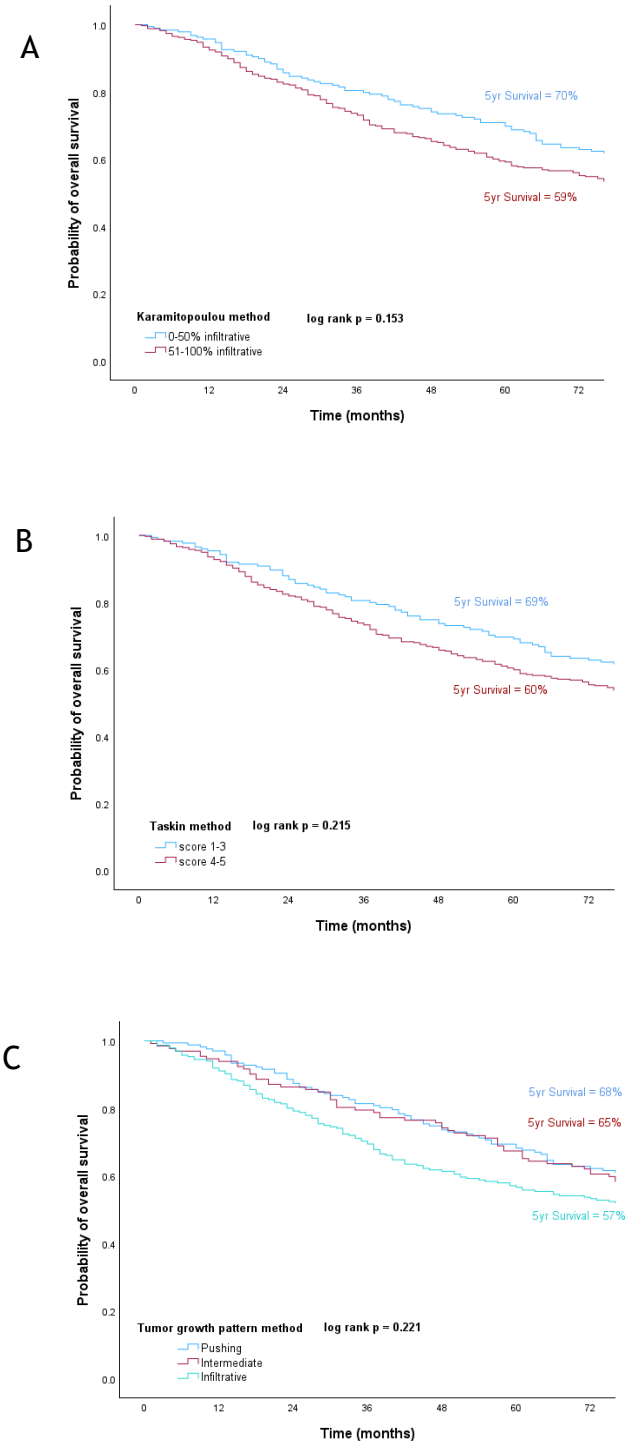


Figure 3-1: Kaplan–Meier curves for overall survival in the GRI cohort (n = 538) stratified by the (A) Karamitopoulou, (B) Taskin, and (C) TGP scoring methods. P-values were calculated using the log-rank test.

Cancer-specific survival (CSS) in all CRC

In contrast to OS, all three scoring methods were significantly associated with CSS in univariate Cox regression (Table 3-3). The Kaplan-Meier curves with the log-rank test showed that higher Karamitopoulou (51-100%) (p<0.001) and Taskin (score 4-5) (p = 0.003) scores were associated with CSS (Figure 3-2). Similarly, the TGP method presented that the infiltrative group had the worst survival outcomes compared to intermediate and pushing growth patterns (p<0.001).

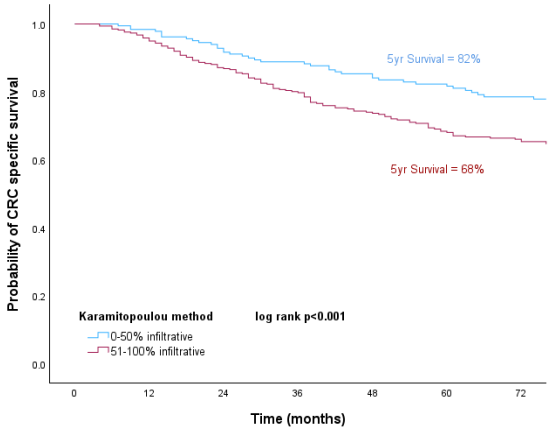
Table 3-3: Univariate and multivariable Cox regression analyses of clinicopathological variables and the three invasive front scoring methods for all colorectal cancer-specific survival in the GRI cohort of 538 patients. Bold p-values are statistically significant. *Not included in the multivariable analysis due to multicollinearity.

Features	Univariate HR (95% CI)	p value	Multivariable HR (95% CI)	p value
Age (<50/ 50-75/ >75)	1.47 (1.12 - 1.94)	**0.006**	1.86 (1.34 - 2.58)	**<0.001**
Sex (Female/ Male)	1.19 (0.87 - 1.62)	0.277		
T stage (pT1/ pT2/ pT3 /pT4)	2.17 (1.68 - 2.80)	**<0.001**	1.48 (1.09 - 2.01)	**0.011**
N stage (pN0/ pN1&pN2)	1.92 (1.42 - 2.60)	**<0.001**	1.54 (1.07 - 2.21)	**0.020**
Site (Colon/ Rectum)	0.79 (0.57 - 1.11)	0.181		
Differentiation (Moderate&Well/ Poor)	1.20 (0.67 - 2.16)	0.545		
Margin involvement (Absent/ Involved)	4.32 (2.86 - 6.54)	**<0.001**	4.20 (2.50 - 7.04)	**<0.001**
Venous invasion (Absent/ Present)	1.60 (1.14 - 2.24)	**0.007**	1.23 (0.84 - 1.82)	0.285
Peritoneal involvement (Absent/ Involved)	2.20 (1.61 - 2.99)	**<0.001**	1.09 (0.59 - 2.02)	0.788
Tumour perforation (0/ 1/ 2)	1.68 (1.06 - 2.67)	**0.028**	0.90 (0.31 - 2.60)	0.850
Tumour budding (Absent/ Present)	1.87 (1.37 - 2.55)	**<0.001**	1.82 (1.28 - 2.58)	**<0.001**
GMS (0/ 1/ 2) *	2.09 (1.63 - 2.67)	**<0.001**		
KM (Low/ High)	0.31 (0.17 - 0.55)	**<0.001**	0.57 (0.29 - 1.09)	0.090
TSP (Low/ High)	1.89 (1.38 - 2.60)	**<0.001**	1.15 (0.77 - 1.70)	0.495
Karamitopoulou method (0-50%/51-100%)	1.85 (1.30 - 2.62)	**<0.001**	0.94 (0.30 - 2.96)	0.913
Taskin method (0-3/4-5)	1.69 (1.19 - 2.41)	**0.004**	0.85 (0.41 - 1.76)	0.667
TGP method (Pushing/ Intermediate/ Infiltrative)	1.42 (1.18 - 1.72)	**<0.001**	1.23 (0.98 - 1.54)	0.079
MMR status (dMMR/ pMMR)	0.93 (0.65 - 1.33)	0.692		
Time period (1997-2003/ 2004-2008/ 2009-2013)	0.76 (0.62 - 0.93)	**0.008**	0.89 (0.72 - 1.10)	0.266

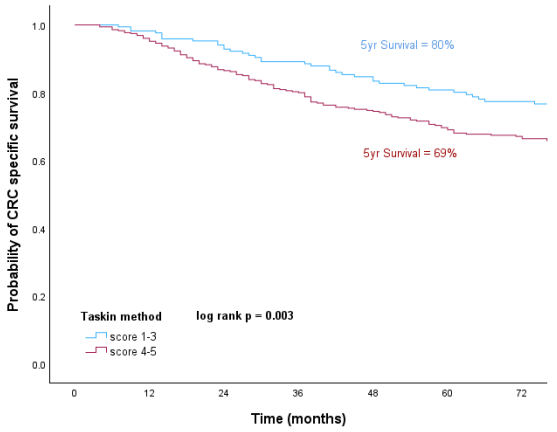
In multivariable Cox regression, none of the three scoring methods remained as independent prognostic factors for CSS. While age (HR = 1.86(1.34-2.58), p<0.001), T stage (HR = 1.48(1.09-2.01), p = 0.011), N stage (HR = 1.54(1.07-2.21), p = 0.020), margin involvement (HR = 4.20 (2.50-7.04), p<0.001), and

tumour budding (HR = 1.82(1.28-2.58), p<0.001) were remained as independent prognostic factors.

A



B



C

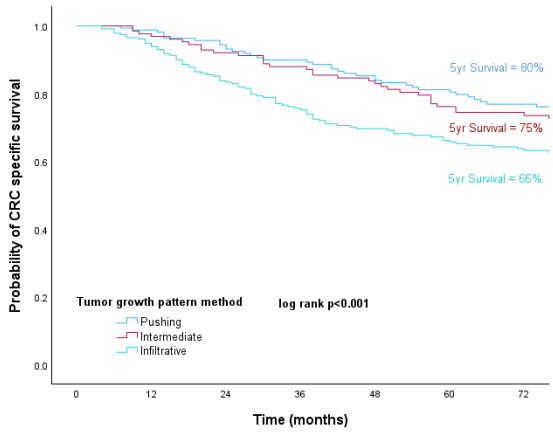


Figure 3-2: Kaplan–Meier curves for CRC-specific survival in the GRI cohort (n = 538) stratified by the (A) Karamitopoulou, (B) Taskin, and (C) TGP scoring methods. P-values were calculated using the log-rank test.

Cancer-specific survival in colon cancer

When the Cox regression analysis was performed in colon cancer only (n=374), all three scoring methods were again significantly associated with CSS on univariate Cox regression (Table 3-4). The Kaplan-Meier curves supported all three scoring methods, showing higher scores associated with shorter survival Figure 3-3. The high Karamitopoulou group (51-100%) had shorter colon CSS than the low group (≤50%) (p<0.001); similarly, the high Taskin group (score 4-5) had shorter colon CSS than the low group (score 1-3) (p<0.001). The infiltrative TGP group had the shortest colon CSS compared to intermediate and pushing TGPs (p<0.001).

Table 3-4: Univariate and multivariable Cox regression analyses of clinicopathological variables and the three invasive front scoring methods for colon cancer-specific survival in the GRI cohort of 374 patients. Bold p-values are statistically significant. *Not included in the multivariable analysis due to multicollinearity.

Features	Univariate HR (95% CI)	p value	Multivariable HR (95% CI)	p value
Age (<50/ 50-75/ >75)	1.83 (1.32 - 2.53)	**<0.001**	2.83 (1.91 - 4.20)	**<0.001**
Sex (Female/ Male)	1.06 (0.74 - 1.52)	0.751		
T stage (pT1/ pT2/ pT3 /pT4)	2.64 (1.92 - 3.63)	**<0.001**	2.11 (1.40 - 3.19)	**<0.001**
N stage (pN0/ pN1&pN2)	2.03 (1.42 - 2.90)	**<0.001**	1.32 (0.86 - 2.04)	0.201
Differentiation (Moderate&Well/ Poor)	1.35 (0.73 - 2.52)	0.339		
Margin involvement (Absent/ Involved)	4.99 (2.94 - 8.47)	**<0.001**	4.60 (2.27 - 9.32)	**<0.001**
Venous invasion (Absent/ Present)	1.60 (1.06 - 2.40)	**0.024**	1.27 (0.79 - 2.03)	0.327
Peritoneal involvement (Absent/ Involved)	2.63 (1.85 - 3.75)	**<0.001**	1.23 (0.56 - 2.71)	0.610
Tumour perforation (0/ 1/ 2)	1.66 (1.05 - 2.63)	**0.031**	0.82 (0.25 - 2.66)	0.740
Tumour budding (Absent/ Present)	1.59 (1.10 - 2.30)	**0.013**	1.32 (0.86 - 2.04)	0.204
GMS (0/ 1/ 2) *	2.15 (1.60 - 2.87)	**<0.001**		
KM (Low/ High)	0.32 (0.16 - 0.64)	**0.001**	0.52 (0.24 - 1.13)	0.099
TSP (Low/ High)	2.11 (1.46 - 3.04)	**<0.001**	0.93 (0.58 - 1.50)	0.772
Karamitopoulou method (0-50%/51-100%)	2.41 (1.55 - 3.73)	**<0.001**	1.18 (0.32 - 4.42)	0.801
Taskin method (0-3/4-5)	2.19 (1.41 - 3.39)	**<0.001**	0.72 (0.30 - 1.72)	0.456
TGP method (Pushing/ Intermediate/ Infiltrative)	1.68 (1.34 - 2.12)	**<0.001**	1.65 (1.24 - 2.18)	**<0.001**
MMR status (dMMR/ pMMR)	1.00 (0.67 - 1.51)	0.985		
Time period (1997-2003/ 2004-2008/ 2009-2013)	0.68 (0.53 - 0.86)	**0.002**	0.78 (0.61 - 1.01)	0.056

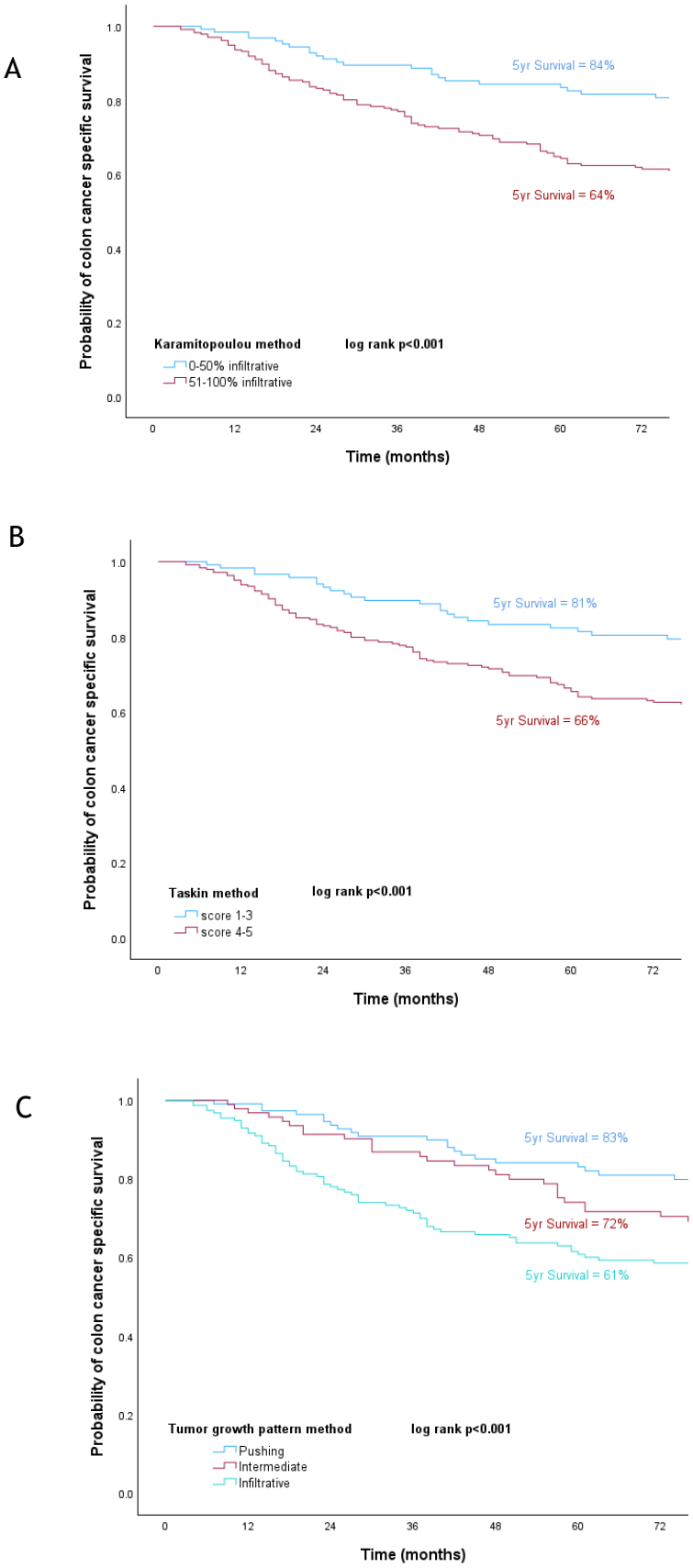


Figure 3-3: Kaplan–Meier curves for colon cancer-specific survival in the GRI cohort (n = 374) stratified by the (A) Karamitopoulou, (B) Taskin, and (C) TGP scoring methods. P-values were calculated using the log-rank test

Cancer-specific survival in rectal cancer

Of 164 rectal cancer patients, none of the three scoring methods were significantly associated with rectal CSS on univariate Cox regression (Table 3-5). The Kaplan-Meier curves showed no differences between score categories (Karamitopoulou p = 0.963, Taskin p = 0.836, TGP p = 0.698) (Figure 3-4). On multivariable Cox regression, margin involvement (HR = 2.80 (1.29-6.07), p =0.009), tumour budding (HR = 2.22(1.19-4.15), p = 0.012), and GMS (HR = 1.69(1.03-2.77), p = 0.038) were retained as independent prognostic factors for rectal CSS. However, with only 45 rectal cancer deaths in this analysis, it may be underpowered to detect a prognostic effect. Therefore, the validation in a larger and independent cohort was necessary to distinguish whether the absence of an independent prognostic effect of TGP in rectal cancer reflects a true biological difference between colon and rectal cancers, or insufficient statistical power in the GRI cohort.

Table 3-5: Univariate and multivariable Cox regression analyses of clinicopathological variables and the three invasive front scoring methods for rectal cancer-specific survival in the GRI cohort of 164 patients. Bold p-values are statistically significant. *Not included in the multivariable analysis due to multicollinearity.

Features	Univariate HR (95% CI)	p value	Multivariable HR (95% CI)	p value
Age (<50/ 50-75/ >75)	0.78 (0.45 - 1.34)	0.368		
Sex (Female/ Male)	1.69 (0.90 - 3.16)	0.103		
T stage (pT1/ pT2/ pT3 /pT4)	1.34 (0.85 - 2.12)	0.214		
N stage (pN0/ pN1&pN2)	1.67 (0.94 - 2.99)	0.082		
Differentiation (Moderate&Well/ Poor)	0.46 (0.06 - 3.37)	0.448		
Margin involvement (Absent/ Involved)	4.34 (2.20 - 8.57)	**<0.001**	2.80 (1.29 - 6.07)	**0.009**
Venous invasion (Absent/ Present)	1.63 (0.87 - 3.05)	0.129		
Peritoneal involvement (Absent/ Involved)	0.82 (0.32 - 2.07)	0.670		
Tumour budding (Absent/ Present)	2.86 (1.58 - 5.18)	**<0.001**	2.22 (1.19 - 4.15)	**0.012**
GMS (0/ 1/ 2)	1.94 (1.22 - 3.09)	**0.005**	1.69 (1.03 - 2.77)	**0.038**
KM (Low/ High)	0.27 (0.08 - 0.88)	**0.029**	0.64 (0.15 - 2.79)	0.549
TSP (Low/ High)	1.47 (0.79 - 2.72)	0.223		
Karamitopoulou method (0-50%/51-100%)	1.01 (0.55 - 1.86)	0.963		
Taskin method (0-3/4-5)	0.94 (0.51 - 1.74)	0.837		
TGP method (Pushing/ Intermediate/ Infiltrative)	0.98 (0.70 - 1.36)	0.896		
MMR status (dMMR/ pMMR)	0.81 (0.39 – 1.65)	0.555		
Time period (1997-2003/ 2004-2008/ 2009-2013)	0.98 (0.68 – 1.40)	0.906		

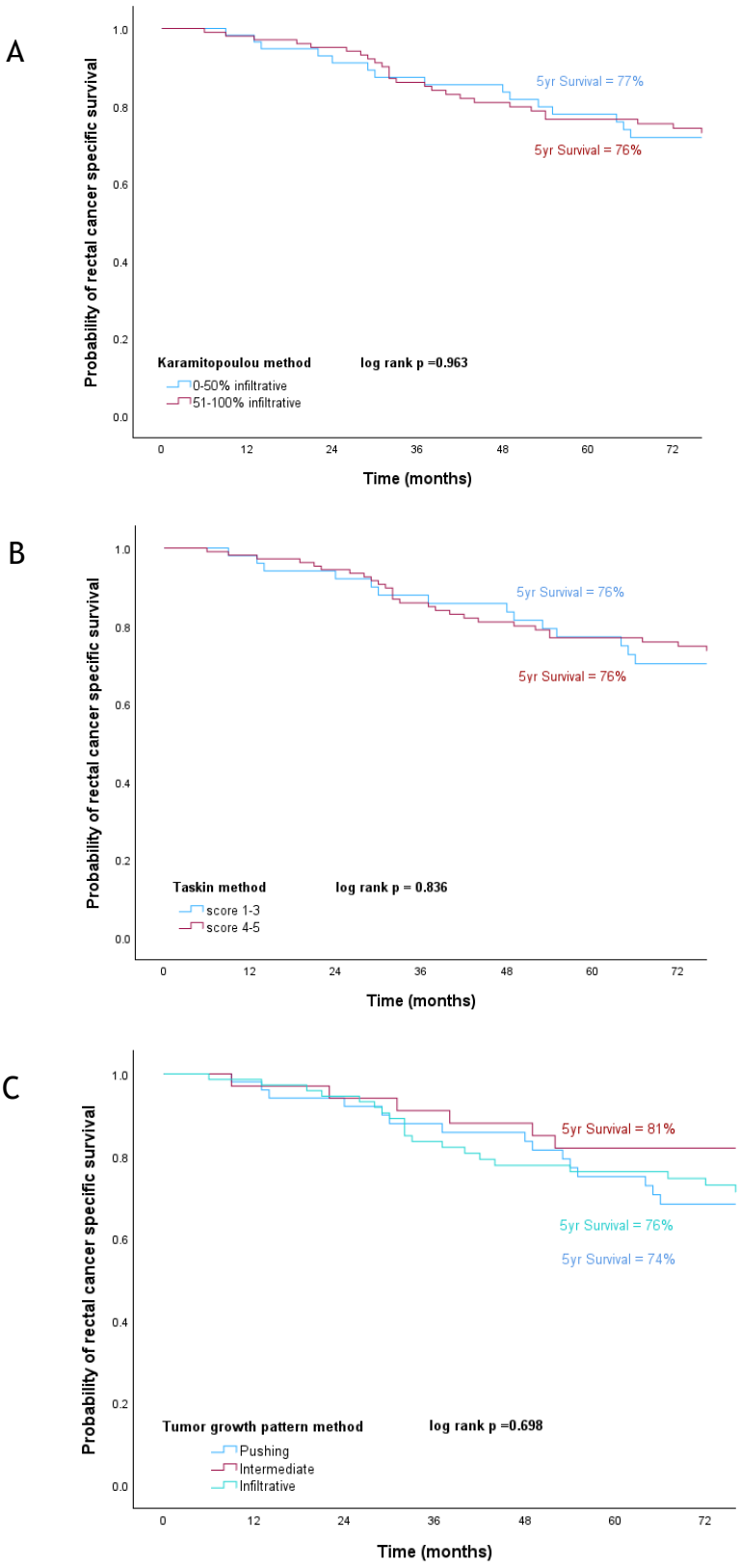


Figure 3-4: Kaplan–Meier curves for rectal cancer-specific survival in the GRI cohort (n = 164) stratified by the (A) Karamitopoulou, (B) Taskin, and (C) TGP scoring methods. P-values were calculated using the log-rank test

3.2.1.5 TGP stratified by tumour budding

The previous multivariable Cox regression analysis revealed an unexpected interplay between two invasive front characteristics. In all CRC and rectal cancer analyses, tumour budding remained an independent prognostic factor, whereas the TGP did not. In contrast, for colon cancer, the results reversed: TGP remained an independent prognostic factor, while tumour budding did not. These multivariable results established which invasive front features better predicted survival outcomes for specific sites, but they did not address a clinically relevant question: how two invasive front features can be interpreted together when assessed on the same H&E slide.

To address this, a Kaplan-Meier analysis was performed for TGP, stratified by tumour budding category (low vs high). This stratification was applied in all CRC and colon cancer, but rectal cancer was not included because TGP did not reach a univariate significance in rectal cancer.

In patients with low tumour budding, TGP showed significant prognostic discrimination. The infiltrative pattern was associated with the shortest survival, while the pushing pattern correlated with the most favourable outcomes (Figure 3-5). Because this stratification divides the cohort into smaller subgroups, the survival estimates are less precise, and the comparison is more likely to be influenced by chance. These stratified results are therefore exploratory. The direction of the effect, with the infiltrative pattern associated with the shortest survival and the pushing pattern with the longest, is consistent with the primary analysis. However, the finding would benefit from confirmation in a larger cohort.

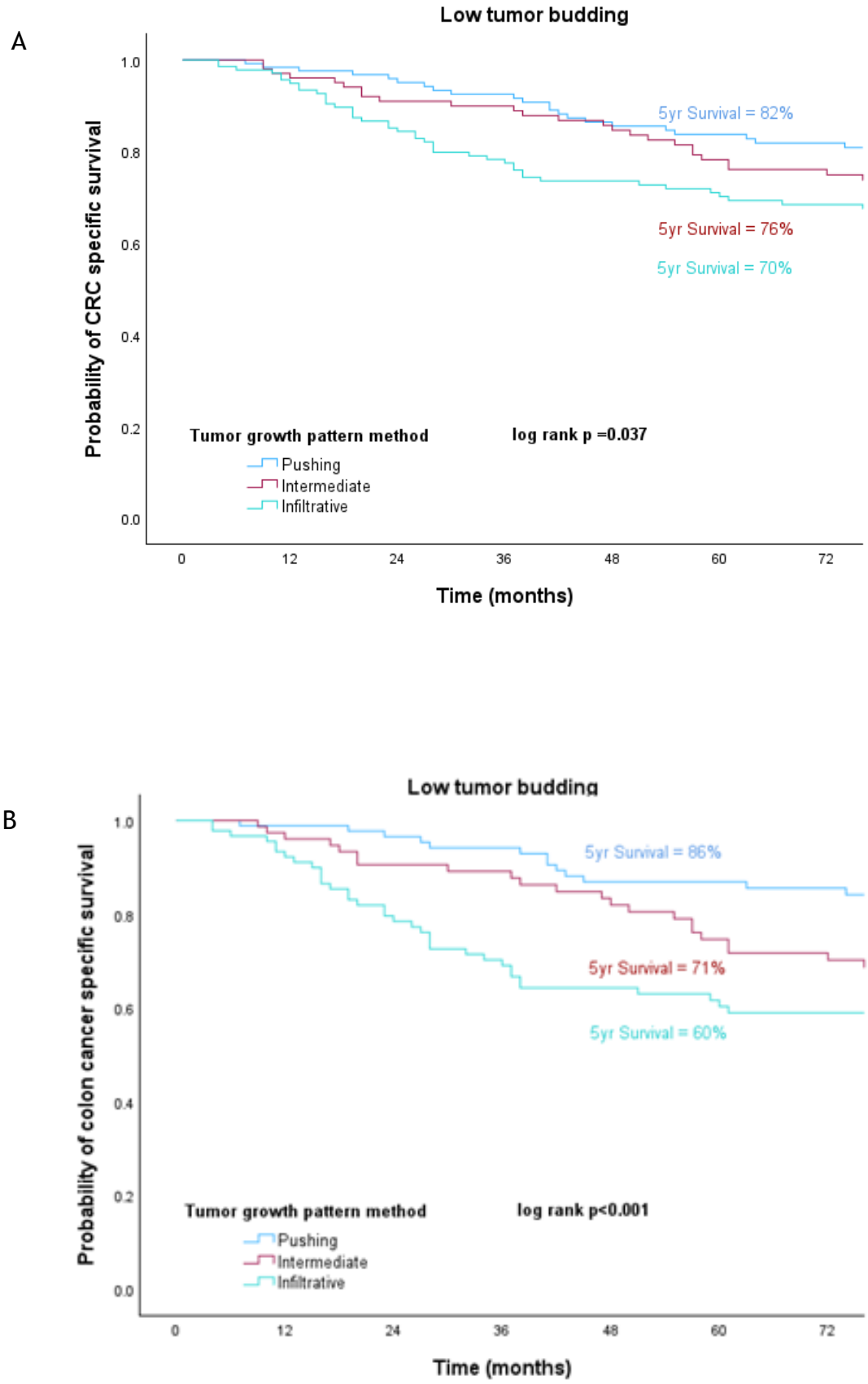
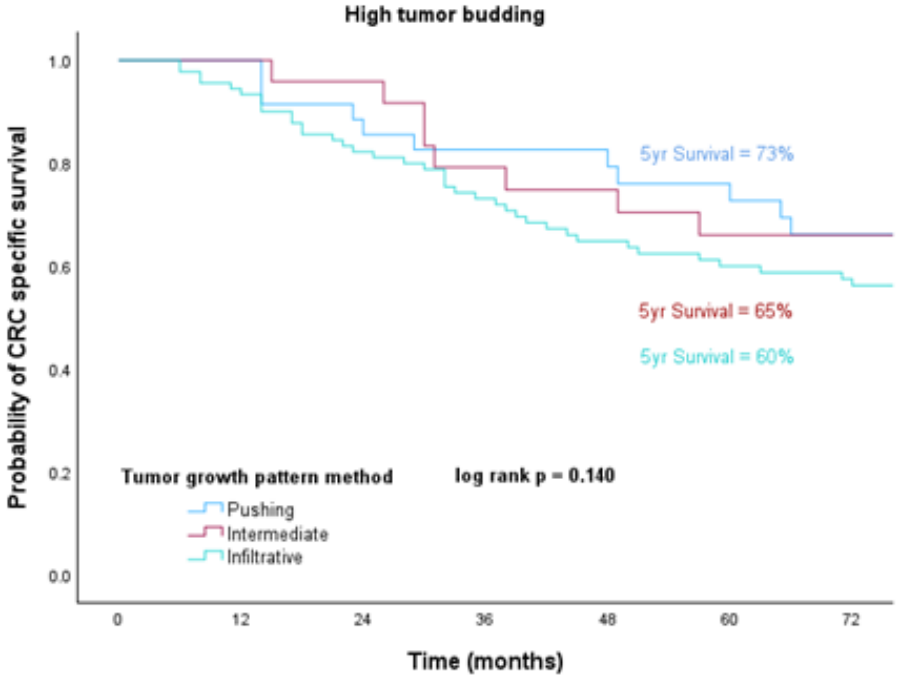


Figure 3-5: Kaplan–Meier curves for cancer-specific survival in the GRI cohort, stratified by the TGP scoring method within low tumour budding subset. (A) All-CRC, low tumour budding subset; (B) Colon cancer, low tumour budding subset

In contrast, patients with high tumour budding, the TGP category, were no longer significantly different. These results were consistent in both CRC and colon cancer (Figure 3-6).

A



B

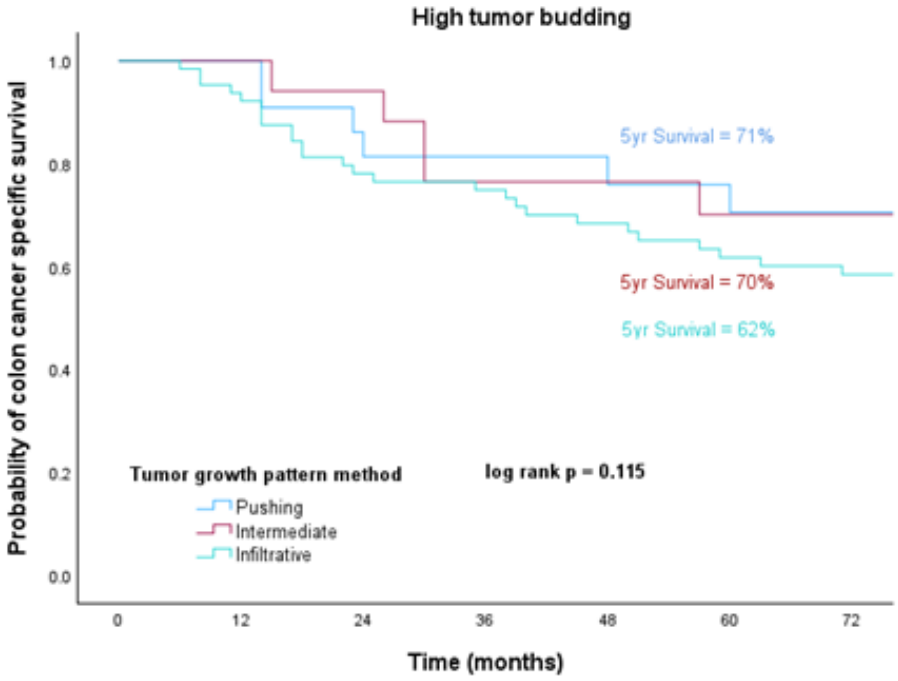


Figure 3-6: Kaplan–Meier curves for cancer-specific survival in the GRI cohort, stratified by the TGP scoring method within high tumour budding subset. (A) All-CRC, high tumour budding subset; (B) Colon cancer, high tumour budding subset

3.2.1.6 GRI cohort summary and rationale for validation in another cohort

For the GRI cohort, the TGP and Karamitopoulou methods showed substantial interobserver agreement, while the Taskin method showed a fair agreement. Importantly, the TGP method was the only invasive front scoring method that remained as an independent prognostic factor, specifically in colon cancer, but not in rectal cancer. These findings required validation of the TGP method in an independent, larger cohort to determine whether the TGP could be reproduced in colon cancer and whether it would be an independent prognostic factor in a larger rectal cancer cohort.

3.2.2 TransSCOT validation cohort

3.2.2.1 Association between TGP method and clinicopathological characteristics

The TransSCOT cohort is drawn from the SCOT adjuvant chemotherapy trial and comprises higher-risk CRC, in contrast to the GRI cohort of unselected curative resections. Its larger size and independent, multi-centre composition make it well-suited to validating the prognostic value of TGP in an external setting. Of 2,912 tissue samples available, 1,037 were included in the analysis to validate the TGP invasive front scoring method.

The TGP categories in the TransSCOT cohort was 444 pushing (42.8%), 210 intermediate (20.3%), and 383 infiltrative (36.9%) (Table 3-6: Associations between clinicopathological characteristics and the TGP scoring method in the TransSCOT cohort of 1,037 patients with colorectal cancer. Bold p-values are statistically significant.

Table 3-6: Associations between clinicopathological characteristics and the TGP scoring method in the TransSCOT cohort of 1,037 patients with colorectal cancer. Bold p-values are statistically significant.

Features	Total	Tumour Growth Pattern			p-value
		Expansile /pushing (n = 444)	Intermediate (n= 210)	Infiltrative (n = 383)	
Age					0.998
<50	70 (6.8)	30 (6.8)	15 (7.1)	25 (6.5)	
50-75	894 (86.2)	382 (86.0)	181 (86.2)	331 (86.4)	
>75	73 (7.0)	32 (7.2)	14 (6.7)	27 (7.0)	
Sex					0.829
Female	391 (37.7)	164 (36.9)	78 (37.1)	149 (38.9)	
Male	646 (62.3)	280 (63.1)	132 (62.9)	234 (61.1)	
T stage					**<0.001**
pT1	9 (0.9)	9 (2.0)	0 (0)	0 (0)	
pT2	74 (7.1)	61 (13.7)	8 (3.8)	5 (1.3)	
pT3	632 (60.9)	286 (64.4)	148 (70.5)	198 (51.7)	
pT4	322 (31.1)	88 (19.8)	54 (25.7)	180 (47.0)	
N stage					**0.010**
pN0	771 (74.3)	348 (78.4)	158 (75.2)	265 (69.2)	
pN1 & pN2	266 (25.7)	96 (21.6)	52 (24.8)	118 (30.8)	
Site					0.957
Colon	854 (82.4)	364 (82.0)	173 (82.4)	317 (82.8)	
Rectal	183 (17.6)	80 (18.0)	37 (17.6)	66 (17.2)	
GMS					**<0.001**
0	107 (10.3)	74 (16.7)	19 (9.0)	14 (3.7)	
1	729 (70.3)	310 (69.8)	159 (75.7)	260 (67.9)	
2	201 (19.4)	60 (13.5)	32 (15.2)	109 (28.5)	
KM					**<0.001**
Low grade	924 (89.1)	365 (82.2)	190 (90.5)	369 (96.3)	
High grade	113 (10.9)	79 (17.8)	20 (9.5)	14 (3.7)	
TSP					**<0.001**
Low	828 (79.8)	382 (86.0)	177 (84.3)	269 (70.2)	
High	209 (20.2)	62 (14.0)	33 (15.7)	114 (29.8)	

3.2.2.2 Survival analysis

Disease-free survival (DFS) in all CRC

In univariate Cox regression for all CRC DFS (n = 1,037), T stage, N stage, GMS, KM, TSP, and TGP were significantly associated with DFS ($p<0.001$) (Table 3-7).

Table 3-7: Univariate and multivariable Cox regression analyses of clinicopathological variables and the TGP scoring method for disease-free survival in the TransSCOT cohort of 1,037 patients with colorectal cancer. Bold p-values are statistically significant. *Not included in multivariable analysis because of multicollinearity.

Features	Univariate HR (95% CI)	p-value	Multivariable HR (95% CI)	p-value
Age (<50, 50-75, >75)	1.09 (0.79 - 1.52)	0.591		
Sex (Female/ Male)	0.82 (0.63 - 1.06)	0.130		
T stage (pT1/pT2/pT3/pT4)	2.00 (1.61 - 2.47)	<0.001	1.58 (1.25 - 1.98)	**<0.001**
N stage (pN0/ pN1&pN2)	2.30 (1.80 - 2.95)	<0.001	2.06 (1.61 - 2.64)	**<0.001**
GMS (0/ 1/ 2) *	1.80 (1.44 - 2.25)	<0.001		
KM (Low/ High)	0.46 (0.27 - 0.77)	0.003	0.60 (0.35 - 1.02)	0.057
TSP (Low/ High)	1.89 (1.45 - 2.46)	<0.001	1.40 (1.06 - 1.83)	**0.017**
Tumour growth pattern method (pushing/ intermediate/ infiltrative)	1.58 (1.38 - 1.82)	<0.001	1.33 (1.14 - 1.54)	**<0.001**

The Kaplan-Meier curve with the log-rank test supported worse survival outcomes in the infiltrative growth pattern when compared to intermediate and pushing growth patterns (p<0.001) (Figure 3-7).

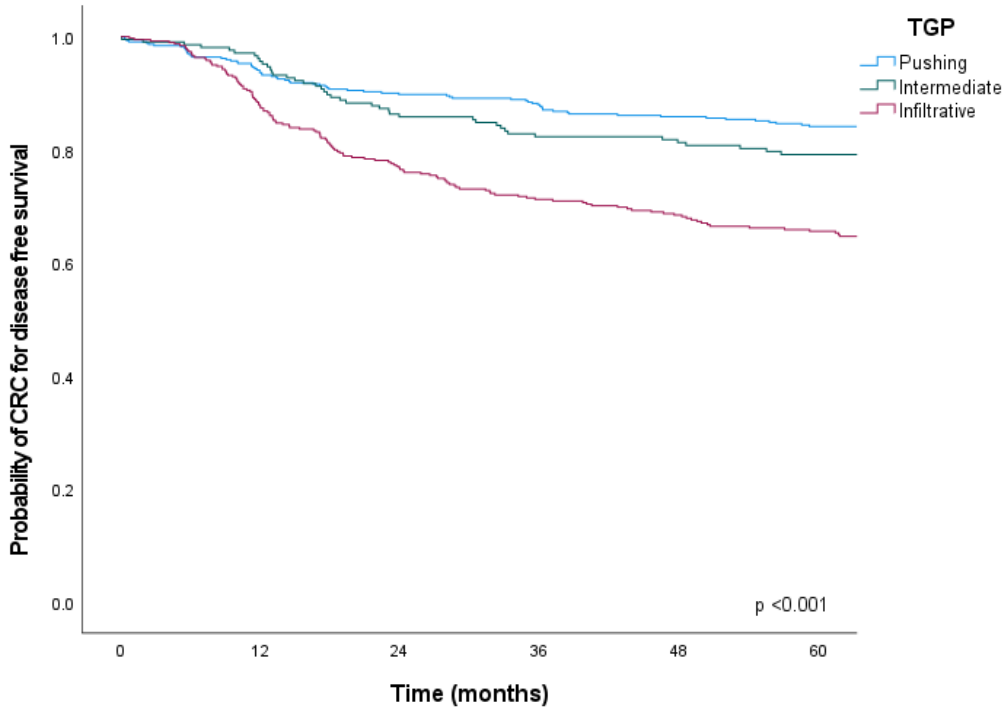


Figure 3-7: Kaplan–Meier curves for disease-free survival of 1,037 patients with colorectal cancer in the TransSCOT cohort, stratified by the three TGP categories (pushing, intermediate, infiltrative). P-value calculated using the log-rank test.

In multivariable Cox regression, T stage (HR = 1.58 (1.25-1.98), p<0.001), N stage (HR = 2.06 (1.61-2.64), p<0.001), TSP (HR = 1.40 (1.06-1.83), p = 0.017), and TGP (HR = 1.33 (1.14-1.54), p<0.001) were retained as independent prognostic factors for colorectal DFS (Table 3-7). Notably, TGP was retained as an independent prognostic feature in all CRCs in this TransSCOT cohort, while in the GRI cohort, TGP was not an independent feature. This finding suggests that the larger sample size and absence of tumour budding as a covariate feature in the TransSCOT cohort might affect the TransSCOT multivariable Cox analysis.

Disease-free survival (DFS) in colon cancer

In univariate Cox regression with 854 colon cancer patients, T stage (p<0.001), N stage (p<0.001), GMS (p<0.001), KM grade (p = 0.011), TSP (p<0.001), and TGP (p<0.001) were significantly associated with colon cancer-free survival (Table 3-8).

Table 3-8: Univariate and multivariable Cox regression analyses of clinicopathological variables and the TGP scoring method for disease-free survival in the TransSCOT cohort of 854 patients with colon cancer. Bold p-values are statistically significant. *Not included in multivariable analysis because of multicollinearity.

Features	Univariate HR (95% CI)	p-value	Multivariable HR (95% CI)	p-value
Age (<50, 50-75, >75)	1.03 (0.73 - 1.45)	0.880		
Sex (Female/ Male)	0.87 (0.67 - 1.15)	0.330		
T stage (pT1/ pT2/ pT3 /pT4)	1.75 (1.39 - 2.21)	**<0.001**	1.36 (1.06 - 1.75)	**0.017**
N stage (pN0/ pN1&pN2)	2.40 (1.84 - 3.13)	**<0.001**	2.20 (1.68 - 2.87)	**<0.001**
GMS (0/ 1/ 2) *	1.74 (1.37 - 2.20)	**<0.001**		
KM (Low/ High)	0.50 (0.29 - 0.85)	**0.011**	0.63 (0.36 - 1.09)	0.097
TSP (Low/ High)	1.85 (1.40 - 2.45)	**<0.001**	1.44 (1.08 - 1.92)	**0.013**
Tumour growth pattern method (pushing/ intermediate/ infiltrative)	1.54 (1.33 - 1.79)	**<0.001**	1.33 (1.14 - 1.57)	**<0.001**

Consistent with the log-rank test for Kaplan-Meier analysis, the infiltrative growth pattern had the shortest colon cancer-free survival (p<0.001) (Figure 3-8).

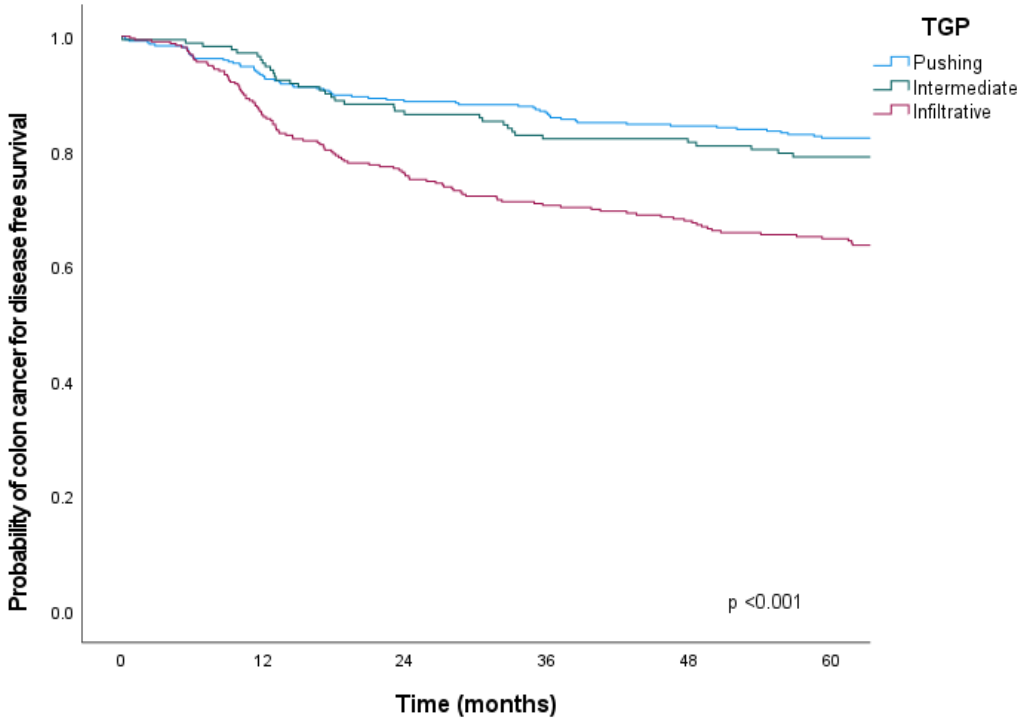


Figure 3-8: Kaplan–Meier curves for disease-free survival of 854 patients with colon cancer in the TransSCOT cohort, stratified by the three TGP categories (pushing, intermediate, infiltrative). P-value calculated using the log-rank test.

In multivariable Cox regression analysis, T stage (HR = 1.36 (1.06-1.75), p = 0.017), N stage (HR = 2.20 (1.68-2.87), p<0.001), TSP (HR = 1.44 (1.08-1.92), p = 0.013), and TGP (HR = 1.33 (1.14-1.57), p<0.001) were retained as independent prognostic factors for colon cancer-free survival (Table 3-8). These findings were similar to those for colon cancer in the GRI cohort.

Disease-free survival (DFS) in locally advanced rectal cancer

With the 183 rectal cancer patients, the TransSCOT cohort had greater statistical power than the GRI cohort to detect a prognostic feature of the TGP. In univariate Cox regression, T stage (p<0.001), GMS (p = 0.037), and TGP (p = 0.002) were significantly associated with locally advanced rectal cancer-free survival (Table 3-9).

Table 3-9: Univariate and multivariable Cox regression analyses of clinicopathological variables and the TGP scoring method for disease-free survival in the TransSCOT cohort of 183 patients with locally advanced rectal cancer. Bold p-values are statistically significant. *Not included in multivariable analysis because of multicollinearity.

Features	Univariate HR (95% CI)	p-value	Multivariable HR (95% CI)	p-value
Age (<50, 50-75, >75)	1.66 (0.58 - 4.69)	0.343		
Sex (Female/ Male)	0.47 (0.20 - 1.08)	0.075		
T stage (pT1/ pT2/ pT3 /pT4)	4.11 (2.21 - 7.64)	**<0.001**	3.56 (1.86 - 6.82)	**<0.001**
N stage (pN0/ pN1&pN2)	1.90 (0.95 - 3.80)	0.069		
GMS (0/ 1/ 2) *	2.05 (1.04 - 4.01)	**0.037**	1.33 (0.66 - 2.70)	0.431
KM (Low/ High)	0.22 (0.03 - 1.63)	0.139		
TSP (Low/ High)	1.81 (0.79 - 4.15)	0.163		
Tumour growth pattern method (pushing/ intermediate/ infiltrative)	1.91 (1.27 - 2.87)	**0.002**	1.54 (1.00 - 2.36)	**0.049**

The Kaplan-Meier curve with the log-rank test showed the infiltrative growth pattern had the worst survival outcomes when compared with intermediate and pushing growth patterns (p = 0.005) (Figure 3-9). In the rectal subgroup, the intermediate growth pattern curve lay closer to the infiltrative curve than in colon cancer. Because this subgroup is small, the curve is based on few events and its position is therefore unstable, so this most likely reflects sampling variation rather than a true biological difference.

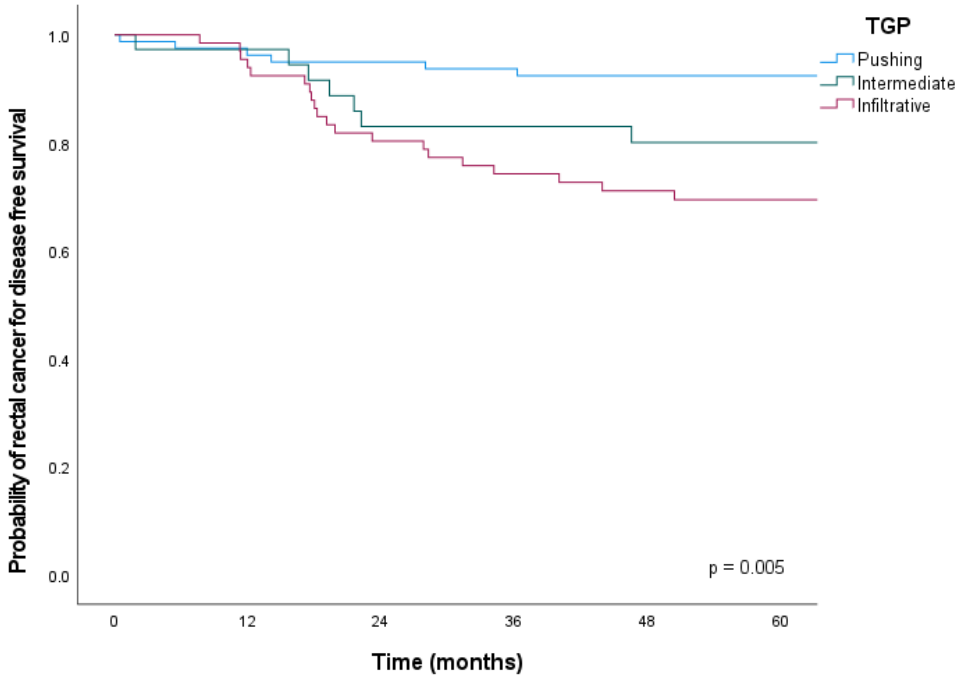


Figure 3-9: Kaplan–Meier curves for disease-free survival of 183 patients with rectal cancer in the TransSCOT cohort, stratified by the three TGP categories (pushing, intermediate, infiltrative). P-value calculated using the log-rank test.

In multivariable Cox regression, T stage (HR = 3.56 (1.86–6.82), p<0.001) and TGP (HR = 1.54 (1.00–2.36), p = 0.049) were retained as independent prognostic features for rectal cancer (Table 3-9). However, the TGP just crossed the conventional threshold of statistical significance with a p-value of 0.049. This finding contrasts with the GRI cohort that the TGP did not retain independence in multivariable analysis.

3.2.2.3 Summary of findings

In this chapter, the three invasive front scoring methods were evaluated, which are the Karamitopoulou, Taskin, and TGP methods, in two independent CRC

cohorts with different time periods of recruitment. The Glasgow Royal Infirmary (GRI) cohort served as the exploratory cohort, and the TransSCOT cohort as the independent validation cohort. There are three main important findings across these two cohorts.

First, all the three scoring methods were significantly associated with the cancer-specific survival in univariate Cox regression analysis in the all CRC and colon cancer subsets but were not significantly associated with overall survival or rectal cancer-specific survival. In multivariable analysis, the TGP retained independent prognostic value specifically for colon cancer, whereas the Karamitopoulou and Taskin methods did not.

Second, an independent prognostic value of TGP in the GRI colon cancer subgroup was reproduced in the TransSCOT validation cohort, providing a robust external validation of TGP as an independent prognostic feature for colon cancer. In rectal cancer, the prognostic value differed between the two cohorts: TGP was not significant in the GRI rectal cancer subgroup, but it reached borderline significance in multivariable analysis in the TransSCOT rectal cancer subgroup. Therefore, the prognostic value of the TGP for rectal cancer remains unclear.

Third, after stratifying for tumour budding, TGP can discriminate survival outcomes within the low tumour budding subgroup. Thus, the TGP and tumour budding can be applied as complementary assessments at the invasive front. Both features are evaluated on a single H&E-stained slide, the combination of TGP and tumour budding could identify a higher risk subgroup among the low tumour budding patients that would not be captured by using tumour budding alone.

3.3 Discussion

Comparison of the three invasive front scoring approaches

Across the three different invasive front scoring approaches, the TGP method presented the greatest prognostic value in this study. In more detail, although the Karamitopoulou method showed substantial interobserver agreement, it lost prognostic value in multivariable analysis. The Taskin method, originally proposed for pancreatic cancer, demonstrated a fair interobserver agreement, suggesting

that the five-point scale lacks consistency between observers and lacks prognostic value in multivariable analysis. The TGP method, a three-category classification, reached substantial interobserver agreement, indicating high reproducibility, and it was the only method to retain independent prognostic value in multivariable analysis.

Validation of the TGP method as a prognostic feature

The prognostic characteristic at the invasive front firstly proposed by Jass and colleagues that the infiltrative growth pattern was associated with worse survival outcomes in rectal cancer (Jass et al., 1987). Cianchi et al. supported this association in 238 stage IIA CRC patients, with a difference of more than 10% in 5-year survival between infiltrative and pushing growth patterns (81.8% and 92.9%, respectively).

The three-category classification adopted by Morikawa et al. was also evaluated with 1,139 colon and rectal cancer patients from two prospective cohorts, demonstrating that the infiltrative growth pattern was associated with worse prognosis independent of microsatellite instability, the CpG island methylation phenotype, LINE-1 methylation, and KRAS, BRAF, and PIK3CA mutations (Morikawa et al., 2012). More recent studies emphasised that the infiltrative growth pattern in stage II or III CRC was an independent prognostic factor (Tokodai et al., 2016).

To our knowledge, this study is the first to compare all three scoring methods within the same cohort and validate them in an independent cohort. The TGP method retained an independent prognostic feature in colon cancer in both cohorts. These findings suggest that across two independent cohorts, one of which is a single-centre unselected surgical cohort and the other is a multi-centre international clinical trial cohort, providing robust external validation that has not been studied for this TGP feature.

In addition to the Taskin method, this is the first adaptation of a method originally developed for pancreatic neuroendocrine tumours. Although it did not outperform existing methods, its evaluation extends the picture of available invasive front-scoring methods in primary CRC. The three methods compared in

this study were conducted by the same observers within the same cohort, removing the inter-study heterogeneity that previously made cross-method comparison difficult. Importantly, this study provides a more reliable basis for selecting between these scoring methods in future research and clinical routine pathology.

As an adjuvant chemotherapy trial population, the TransSCOT cohort represents the clinical setting in which a TGP-based prognostic tool would be most useful, since these patients are facing systemic treatment decisions and would benefit most from refined risk stratification. In contrast, the GRI cohort provided the histological, transcriptomic and spatial depth required for the test analyses in this thesis.

Site-specific findings between colon and rectal cancers

In colon cancer, the prognostic value of the TGP was reproducible across two cohorts. The TGP was retained as an independent prognostic feature for colon cancer-specific survival in the GRI cohort and for _DFS in the TransSCOT colon cancer subgroup. Although CSS and DFS represent different clinical endpoints, the association of an infiltrative growth pattern with poorer outcomes in both measures supports its adverse prognostic significance.

However, in rectal cancer, the prognostic value of the TGP method differed between cohorts. In the GRI rectal subgroup, TGP did not reach significance in both univariate and multivariable Cox regression analysis. In the TransSCOT rectal subgroup, TGP just reached the borderline of statistical significance (HR = 1.54 (1.00–2.36), p = 0.049). The lower bound of the 95% confidence interval at 1.00 indicates a fragile result. The absence of tumour budding as a covariate in this multivariable model means that whether TGP would retain independence after adjustment for tumour budding could not be assessed in the TransSCOT cohort.

The discrepancy in TGP prognostic value in rectal cancer between the two cohorts may reflect differences in patient population (GRI: an unselected single-centre surgical series between 1997-2013; TransSCOT: an international cohort recruited from 2008 onwards, selected for higher-risk pathological features) or the era of diagnosis. Therefore, the prognostic value of TGP in rectal cancer

remains uncertain and requires further validation in cohorts where tumour budding is available.

Despite this uncertainty in rectal cancer, these site-specific differences may be better understood in the context of the growing recognition that colon and rectal cancers are biologically distinct diseases. This biological divergence may help explain why histology-based prognostic features can behave differently between the two sites. Several studies have supported the biological differences between colon and rectal tumours. Salem et al. profiled over 10,000 colorectal tumours and reported that rectal cancers had higher rates of TOPO1 expression and HER2/neu amplification compared with left and right-sided colon cancers (Salem et al., 2017). This molecular heterogeneity might explain the differences between these two anatomical sites of the large bowel. Furthermore, Riihimäki et al. demonstrated that the pattern of distant metastasis differs between two sites in 49,096 CRC patients. Rectal cancers more frequently metastasised to the lungs, whereas right-sided colon cancers showed higher peritoneal metastasis (Riihimäki et al., 2016).

These molecular differences between colon and rectal cancers may provide a biological explanation for the effects of TGP, which may behave differently across the anatomical sites of the large bowel.

Combination of TGP and tumour budding prognostic features

Tumour budding is a high magnification feature that evaluates clusters of dedifferentiated tumour cells at the invasive front and is one of the most validated additional prognostic features in CRC (Lugli et al., 2017). Van Wyk et al. concluded in a systematic review that tumour budding is a robust prognostic feature in primary operable CRC across cohorts (van Wyk et al., 2015). More recently, tumour budding was confirmed as an independent prognostic factor in 1,048 stage III colon cancer patients in IDEA-France phase III randomised trial (Basile et al., 2022). In contrast to tumour budding, TGP is a low-magnification feature representing the overall invasive front architecture. Despite its long-recognised prognostic value, however, TGP has been less consistently included in routine reporting (Koelzer & Lugli, 2014).

In the GRI cohort, TGP and tumour budding showed a site-dependent dominance: tumour budding was an independent prognostic factor in all CRC and rectal cancers. While the TGP was an independent prognostic factor in colon cancer. These findings suggest that the two features provide complementary rather than redundant prognostic information at different anatomical sites. As tumour budding reflects focal tumour-cell dissociation, whereas infiltrative growth pattern reflects the broader architecture of tumour invasion, their differing prognostic effects in colon and rectal cancer may reflect site-specific differences in tumour biology and invasive behaviour. Notably, the absence of TGP's independent prognostic value in all CRC patients may reflect two factors: first, the dilution by rectal cancers, TGP did not reach statistical significance when both sites were analysed together as all CRC; second, the involvement of tumour budding as a covariate in multivariable analysis may absorb the TGP signal.

In the Kaplan-Meier analysis stratified by tumour budding, within the low tumour budding subgroup, the TGP retained the ability to discriminate survival outcomes, with the infiltrative growth pattern showing the worst prognosis compared with pushing and intermediate growth patterns. However, within the high tumour budding subgroup, TGP no longer discriminated. This pattern suggests that these two features assessed on the same H&E slide can be complementary. In practical terms, TGP identifies higher-risk patients who would be classified as low tumour budding but who in fact have worse outcomes because of an infiltrative growth pattern. Because TGP and tumour budding can be assessed from a standard H&E section, this complementarity can be captured within existing routine workflows without additional staining, or specialised methods. Furthermore, this complement may enhance prognostic stratification in CRC and colon cancer patients by identifying infiltrative patterns within the low budding patient’s subgroup who would not be detected by the tumour budding feature alone.

Limitations

A limitation in this study is that the tumour budding was not available as a covariate in the TransSCOT cohort Cox regression analysis, which may explain the different findings between the two cohorts in the all CRC analysis, showing that TGP was an independent prognostic factor in the TransSCOT cohort, but was not

in the GRI cohort, where tumour budding was included and may have absorbed the TGP signal.

Another limitation is that identical survival endpoints were not available across the two cohorts. CSS was the primary outcome in the GRI cohort, whereas DFS was the primary outcome in TransSCOT. Therefore, the TransSCOT findings should be interpreted as a reproduction of the prognostic association of TGP across different disease-related outcomes rather than direct validation using an identical endpoint.

Importantly, this limitation affects how the rectal cancer TransSCOT cohort result should be interpreted. In the TransSCOT cohort, TGP just reached significance in the multivariable model for disease-free survival (HR = 1.54 (1.00–2.36), p = 0.049), but the lower bound of the confidence interval at 1.00 indicates a fragile result. This result, therefore, should not be considered definitive evidence of TGP's independence in rectal cancer and requires confirmation in additional cohorts where tumour budding data is available as a covariate factor in multivariable Cox regression analysis.

On the basis of its reproducibility and prognostic performance in the GRI test cohort, together with external validation in TransSCOT, TGP was selected for molecular analyses. The GRI cohort had been classified into pushing, intermediate and infiltrative growth patterns. To investigate the biological differences associated with the most distinct tumour border morphologies, transcriptomic analyses focused on the two extreme groups, pushing and infiltrative, while intermediate cases were excluded. These analyses used existing bulk RNA-seq data generated from tumour centre tissue rather than the invasive front. As molecular profiles may differ between these regions, the next chapter analysis is framed as hypothesis-generating only.

Chapter 4 Transcriptomic characteristics of infiltrative and pushing growth patterns in primary CRC

4.1 Introduction

Previously, the infiltrative TGP was associated with a worse prognosis in two CRC cohorts (GRI and TransSCOT) compared with pushing and intermediate growth patterns. While histological identification of an infiltrative growth pattern provides prognostic information, deeper insight into the specific molecular biology at the invasive front is necessary to unravel the mechanisms driving this phenotype. A recent study investigated the tumour edge in CRC and found that CAFs and tumour-associated neutrophils (TANs) were enriched in this region, facilitating tumour progression. However, this study did not classify TGPs (TGPs) at the tumour edge (Tang et al., 2025). In addition, several studies utilised RNA-seq analysis in CRC metastasis in different histopathological growth patterns, providing invaluable insights into molecular mechanisms (Hu et al., 2022; Latacz et al., 2024), but no studies have focused on growth patterns in primary CRC. Unlike previous studies that focused on clinical information and TGP correlations, this chapter further investigated the molecular biology underlying TGP in primary stage I-III CRC with 154 FFPE samples stratified by infiltrative and pushing growth patterns from GRI cohort using TempO-Seq (Templated Oligo assay with Sequencing readout), a targeted bulk transcriptomic assay designed for FFPE material.

Bulk RNA sequencing (RNA-Seq) is widely used in cancer studies, uncovering gene expression profiles within tissue sections or biopsies. This technology offers average gene expression of pooled cell populations. However, conventional bulk RNA-Seq requires intact RNA and relies on reverse transcription, limiting its application to archival FFPE specimens, where RNA is often degraded. To overcome this limitation, we used TempO-Seq, an assay specifically designed for degraded and FFPE-derived RNA. It is a targeted ligation-based assay rather than an unbiased whole-transcriptome sequencing method: detector oligonucleotide pairs are designed against a pre-defined set of around 22,500 transcripts, so only those targets are quantified and novel transcripts or isoforms are not captured. This targeted, ligation-based design is what makes TempO-Seq tolerant of the

fragmented RNA typical of FFPE specimens and the reason it was selected for this archival cohort (Trejo et al., 2019).

Although the infiltrative growth pattern is associated with a worse prognosis in CRC, as shown in the previous chapter, the transcriptomic basis of this difference in primary tumours has not been characterised. Given that the infiltrative pattern is defined by diffuse, irregular invasion into the surrounding stroma, it was hypothesised that infiltrative pattern would show transcriptomic enrichment of EMT, alongside an altered TME, particularly its immune and stromal contexture, relative to pushing pattern. However, TempO-Seq signals reflect averaged expression across mixed cell populations and do not necessarily translate into protein-level differences. The most promising transcriptomic findings were further validated at the protein level by immunohistochemistry (IHC) in the same cohort. The specific aims are set out below.

This chapter aims to characterise the molecular signatures that distinguish infiltrative from pushing growth patterns in primary stage I–III CRC, and to validate the most promising findings at the protein level. The specific objectives were:

1. To identify differentially expressed genes between infiltrative and pushing TGPs in 154 FFPE primary CRC samples using TempO-Seq.

2. To identify the biological pathways enriched in each growth pattern using Gene Set Enrichment Analysis (GSEA) against the Hallmark and GO Biological Process (GO:BP) gene sets.

3. To estimate the composition of the tumour microenvironment in each TGP using MCPcounter deconvolution.

4. To compare the transcript-level expression of established markers for the pathways identified in objective 2 between TGPs.

5. To validate transcriptomic findings at the protein level by immunohistochemistry.

4.2 Results

4.2.1 Demographic characteristics of tumour growth patterns

The demographic of this study cohort is summarised in Table 4-1. A total of 154 primary CRC cases were included, comprising 80 infiltrative and 74 pushing tumours. T stage, N stage, age, and sex distributions were compared between the two TGP groups, with significant differences observed for T stage ($p = 0.005$), N stage ($p = 0.001$), and tumour budding ($p = 0.011$); age, sex, differentiation, and margin involvement did not differ significantly. Although tumour budding was more frequent in infiltrative tumours (46.3% versus 26.4%, $p = 0.011$), the two features were not coincident: over half of infiltrative tumours showed no budding, and budding was present in a subset of pushing tumours. This indicates that growth patterns and budding reflect related but distinct aspects of invasive front biology, consistent with the findings of Chapter 3. For the heatmap displaying the same demographic data as in Table 4-1, each column represents an individual patient, grouped by infiltrative and pushing growth patterns, while each row represents clinicopathological features (age, sex, tumour stage, differentiation, margin involvement, and tumour budding) (Figure 4-1).

Table 4-1: Demographic characteristics between infiltrative and pushing TGPs.

		Tumour Growth Pattern			
Characteristic	**N**	**Overall** N = 154	**Infiltrative** N = 80	**Pushing** N = 74	**p-value**
Age	154				0.2
<50		48 (31.2%)	29 (36.3%)	19 (25.7%)	
51-75		49 (31.8%)	27 (33.8%)	22 (29.7%)	
>75		57 (37.0%)	24 (30.0%)	33 (44.6%)	
Sex	154				0.7
Female		68 (44.2%)	34 (42.5%)	34 (45.9%)	
Male		86 (55.8%)	46 (57.5%)	40 (54.1%)	
T stage	154				**0.005**
pT1		3 (1.9%)	0 (0.0%)	3 (4.1%)	
pT2		18 (11.7%)	7 (8.8%)	11 (14.9%)	
pT3		83 (53.9%)	38 (47.5%)	45 (60.8%)	
pT4		50 (32.5%)	35 (43.8%)	15 (20.3%)	
Differentiation	153				0.5
Well or Moderate		138 (90.2%)	71 (88.8%)	67 (91.8%)	
Poor		15 (9.8%)	9 (11.3%)	6 (8.2%)	
N stage	154				**0.001**
pN0		89 (57.8%)	35 (43.8%)	54 (73.0%)	
pN1		43 (27.9%)	29 (36.3%)	14 (18.9%)	
pN2		22 (14.3%)	16 (20.0%)	6 (8.1%)	
Margin involvement	154				0.7
Absent		144 (93.5%)	74 (92.5%)	70 (94.6%)	
Involved		10 (6.5%)	6 (7.5%)	4 (5.4%)	
Tumour budding	152				**0.011**
Absent		96 (63.2%)	43 (53.8%)	53 (73.6%)	
Present		56 (36.8%)	37 (46.3%)	19 (26.4%)	

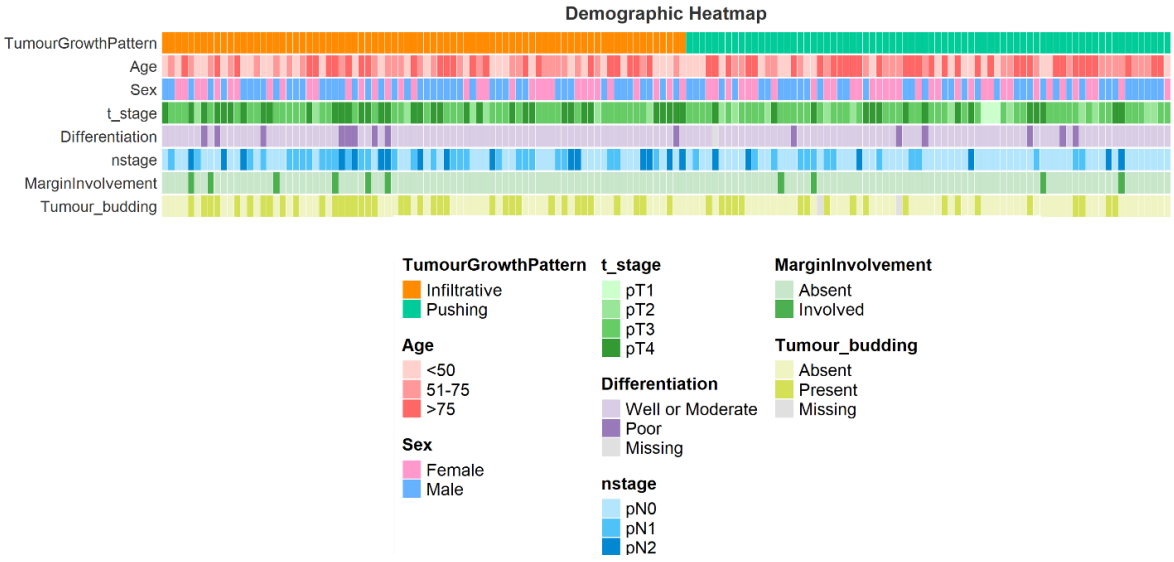


Figure 4-1: Heatmap of demographics for 154 patients
Infiltrative TGP patients are shown in orange and pushing TGP patients are shown in green at the top of the heatmap. Clinicopathological data are represented using different colours as shown in the legend.

4.2.2 The principal component analysis and differential gene expression

The principal component analysis of normalised gene expression data was applied and revealed that principal component 1 (PC1) explained 13.8 % variance and principal component 2 (PC2) explained 3.1 % variance across the whole dataset. The two TGP groups formed a single cluster, indicating a similar global expression profile (Figure 4-2).

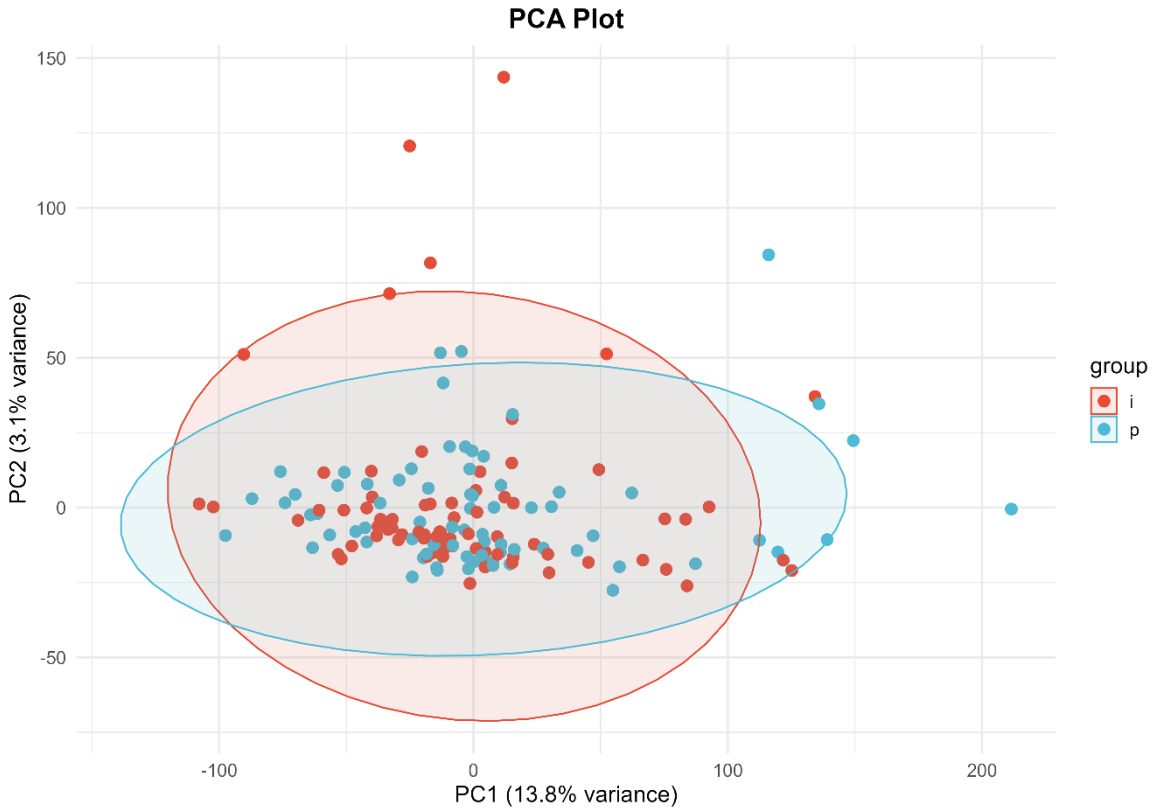


Figure 4-2: Principal Component Analysis (PCA) of normalised RNA-seq data
The first principal component (PC1) accounted for 13.8% and the second principal component (PC2) illustrated for 3.1%. Red dots represent infiltrative growth pattern patients, and blue dots represent pushing growth pattern patients.

To identify genes significantly associated with an infiltrative growth pattern compared to a pushing growth pattern, the differential gene expression analysis was performed. Overall, 400 genes were significantly differentially expressed between the infiltrative and pushing groups. The 323 up-regulated genes (orange dots as shown in Figure 4-3) were highly expressed in infiltrative growth pattern patients when compared to pushing growth pattern patients using log2FC > 1 and adjusted p-value < 0.05 as the cut-off. On the other hand, the 77 down-regulated genes (green dots as shown in Figure 4-3) were highly expressed in pushing growth pattern patients when compared to infiltrative growth pattern patients. The top ten up and down-regulated genes list were shown in Table 4-2 and Table 4-3, respectively.

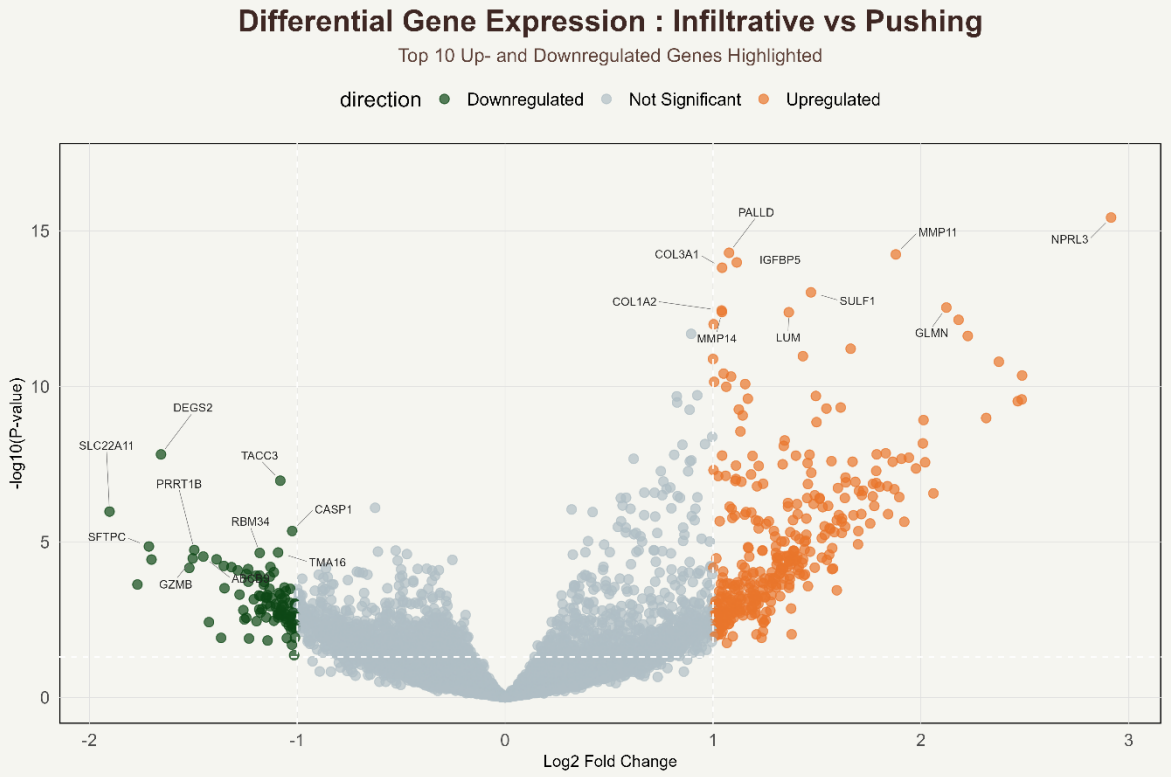


Figure 4-3: Volcano plot of differential gene expression between infiltrative and pushing growth patterns
Up and down-regulated genes of infiltrative growth pattern were highlighted in orange and green colours dots, respectively. The criteria for different gene expression are Ilog2FCI > 1 and adjusted p-value <0.05. Both top ten up- and down-regulated genes were pointed out and labelled in gene names.

Table 4-2: List of up-regulated genes in infiltrative growth pattern ranked by log2FC

Up-regulated gene	Log2FC	adj.P.Val
NPRL3	2.92	4.01×10^{-12}
LSM4	2.49	2.41×10^{-8}
PRRX1	2.49	9.78×10^{-8}
COL10A1	2.47	1.07×10^{-7}
THBS2	2.38	9.66×10^{-9}
SFRP2	2.31	3.03×10^{-7}
NTM	2.23	1.84×10^{-9}
ANTXR1	2.18	7.10×10^{-10}
GLMN	2.12	4.45×10^{-10}
COMP	2.06	2.96×10^{-5}

Table 4-3: List of down-regulated genes in infiltrative growth pattern (higher in pushing) ranked by log2FC

Down-regulated gene	Log2FC	adj.P.Val
SLC22A11	-1.90	8.78×10^{-5}
CLDN2	-1.77	6.98×10^{-3}
SFTPC	-1.71	7.60×10^{-4}
ADH6	-1.70	1.66×10^{-3}
DEGS2	-1.66	3.35×10^{-6}
SSTR1	-1.52	2.75×10^{-3}
GZMB	-1.50	1.54×10^{-3}
PRRT1B	-1.50	9.68×10^{-4}
ABCB9	-1.45	1.38×10^{-3}
MFSD3	-1.39	1.64×10^{-3}

4.2.3 Pathway enrichment analysis between infiltrative and pushing growth patterns

To investigate the biological pathways significantly associated with infiltrative compared to pushing growth patterns, Gene Set Enrichment Analysis (GSEA) was performed using Gene Ontology Biological Processes (GO:BP) and Hallmark gene sets. GSEA of GO:BP enrichment analyses revealed a clear biological contrast between TGPs (Figure 4-4). Infiltrative tumours were significantly enriched for multiple pathways. It showed biology programmes related to extracellular matrix organisation, stromal activation, cell migration, TGF-β signalling, and immune-regulatory activity. These included cell–substrate

junction assembly, positive regulation of cell migration, fibroblast activation, regulation of leukocyte activation, and regulation of cytokine production (Figure 4-4.)

In contrast, a pushing growth pattern was enriched in pathways associated with cellular metabolism and biosynthetic activity, including xenobiotic metabolism, amino acid metabolism, mitochondrial transport, cellular respiration, DNA repair, RNA processing, and ribosome biogenesis (Figure 4-4). Collectively, the pushing growth pattern correlates with energy metabolism, protein/RNA biosynthesis, and cellular maintenance programmes.

These GO:BP GSEA analysis provides distinguishing biological programmes between TGPs. The infiltrative growth reflects a tumour-stromal remodelling activity, while pushing growth is associated with metabolic and biosynthetic cellular processes.

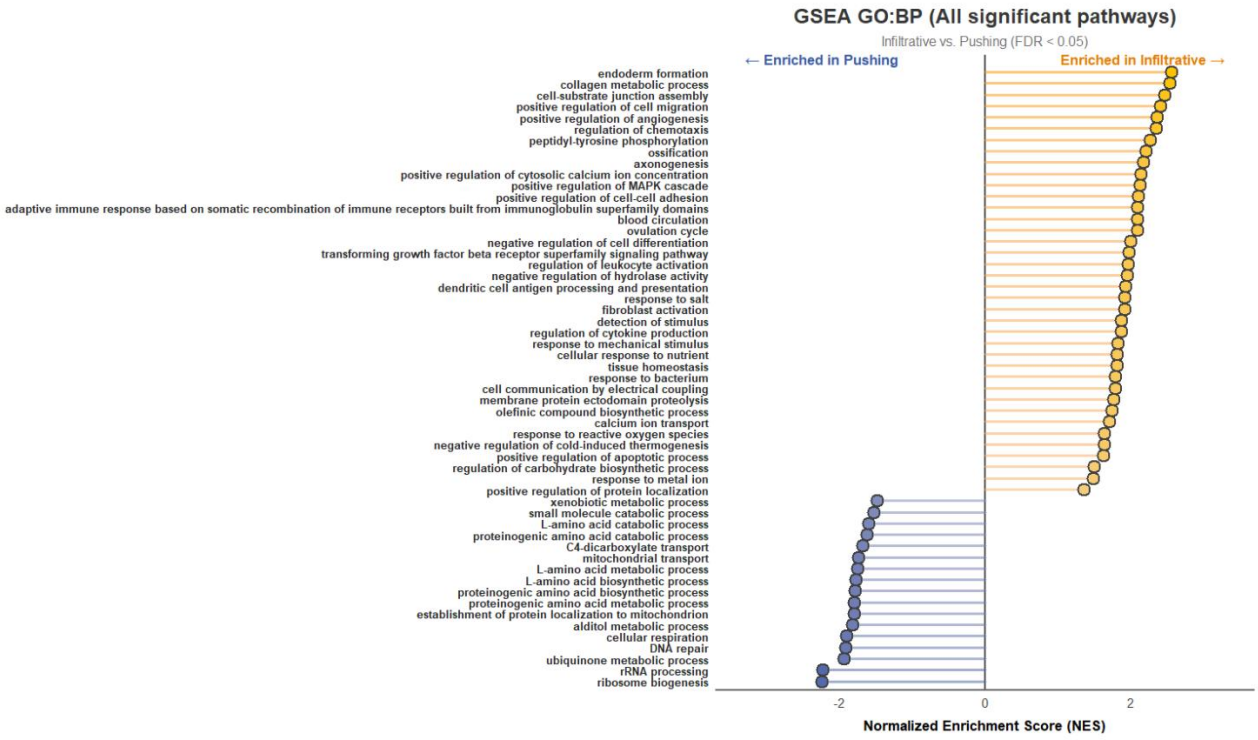


Figure 4-4: GSEA of Gene Ontology (GO) Biological processes (BP) enriched in infiltrative TGP
All GO:BP gene sets significantly enriched at FDR < 0.05 are shown, ranked by normalised enrichment score (NES). Pathways enriched in infiltrative tumours are shown on the right (positive NES, orange); pathways enriched in pushing tumours are shown on the left (negative NES, blue).

Hallmark GSEA analysis showed that the infiltrative growth was enriched in epithelial–mesenchymal transition (EMT), which is the strongest positive enrichment (Figure 4-5). Additionally, TNF-α signalling via NF-κB, the

inflammatory response, TGF-β signalling, IL6–JAK–STAT3 signalling, and hypoxia were enriched in infiltrative growth.

In contrast, pushing growth is associated with metabolic and proliferative cellular programmes (Figure 4-5), including fatty acid metabolism, MYC targets, G2M checkpoint, and E2F targets. This suggests that pushing growth presents intrinsic tumour proliferative and metabolic processes, while infiltrative shows tumour microenvironment remodelling and inflammatory responses.

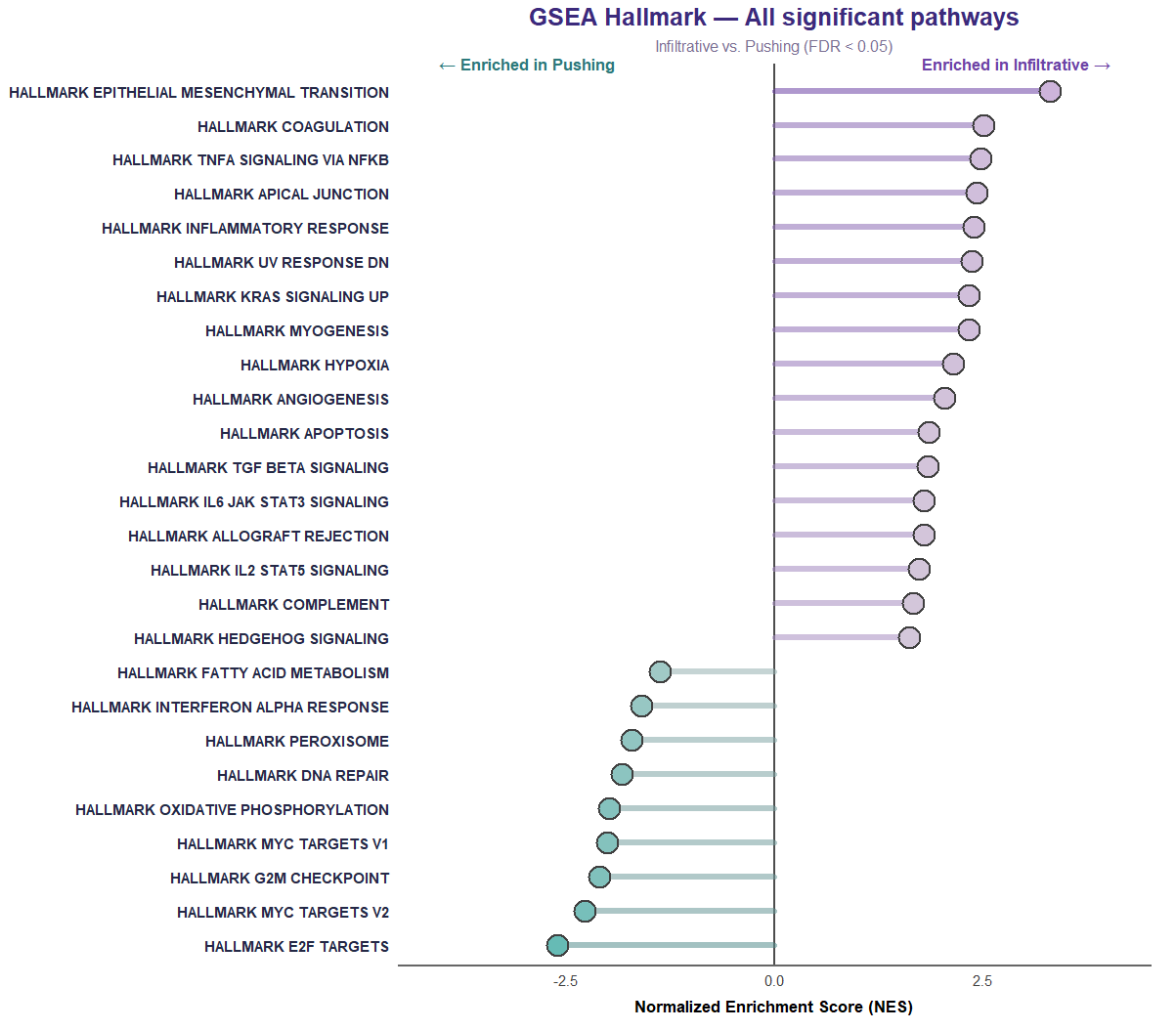


Figure 4-5: HALLMARK pathway enriched in infiltrative growth pattern
GSEA of the MSigDB hallmark gene sets that enriched in infiltrative when compared to pushing growth pattern ranked by NES.

4.2.4 Association between tumour growth pattern and consensus molecular subtype

CMS classification was available for 92 samples, comprising 40 infiltrative and 52 pushing. The remaining samples were reported as not classified by CMSClassifier and were excluded from this analysis. Among infiltrative tumours, CMS4 was the predominant subtype, accounting for 67.5%, followed by CMS2

(20.0%), CMS1 (10.0%), and CMS3 (2.5%). In contrast, pushing tumours were predominantly CMS2 (67.3%), while CMS4 accounted for only 9.6% (Figure 4-6). These findings show a strong association between infiltrative growth pattern and CMS4, while pushing growth pattern was mainly associated with CMS2.

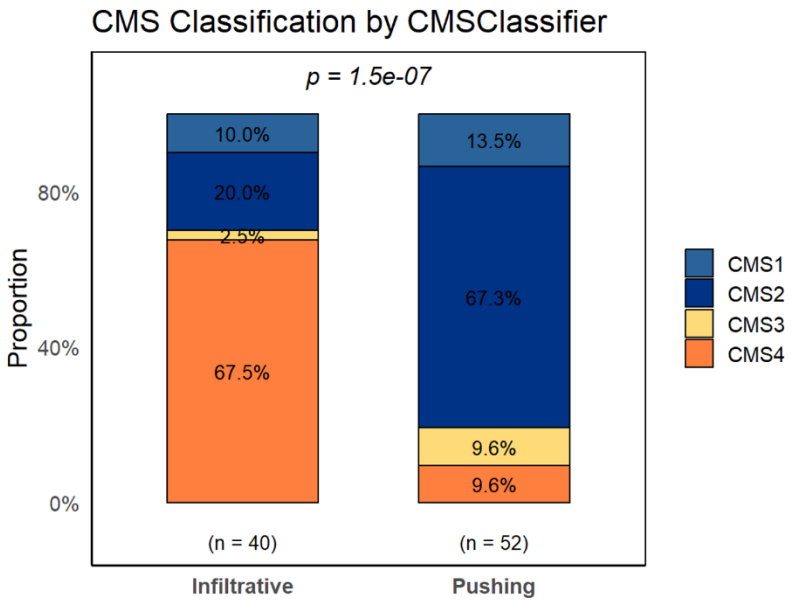


Figure 4-6: CMS proportion according to TGPs
Proportions of CMS1–4 among infiltrative (n = 40) and pushing (n = 52) growth patterns classified using CMSClassifier. Samples reported as not classified were excluded from the analysis. The distribution of CMS subtypes differed significantly between TGPs ($p = 1.5 \times 10^{-7}$)

4.2.5 The EMT-related genes comparison

Since the Hallmark GSEA results presented EMT as the strongest enriched pathway in infiltrative tumours, this finding suggests that infiltrative growth is associated with a more invasive phenotype and loss of epithelial features. To further explore the EMT pathway results by GSEA, individual EMT gene markers were compared between TGPs. The results of single EMT-associated gene comparison demonstrated that none of the selected EMT markers showed a statistically significant differential expression between TGPs (Figure 4-7). This suggests that the EMT enrichment pathway observed in GSEA analysis was not driven by changes in single EMT genes but reflects coordinated changes across a broader EMT-related gene programme. Moreover, the absence of single-EMT gene differences did not indicate that it would not be expressed at the protein level.

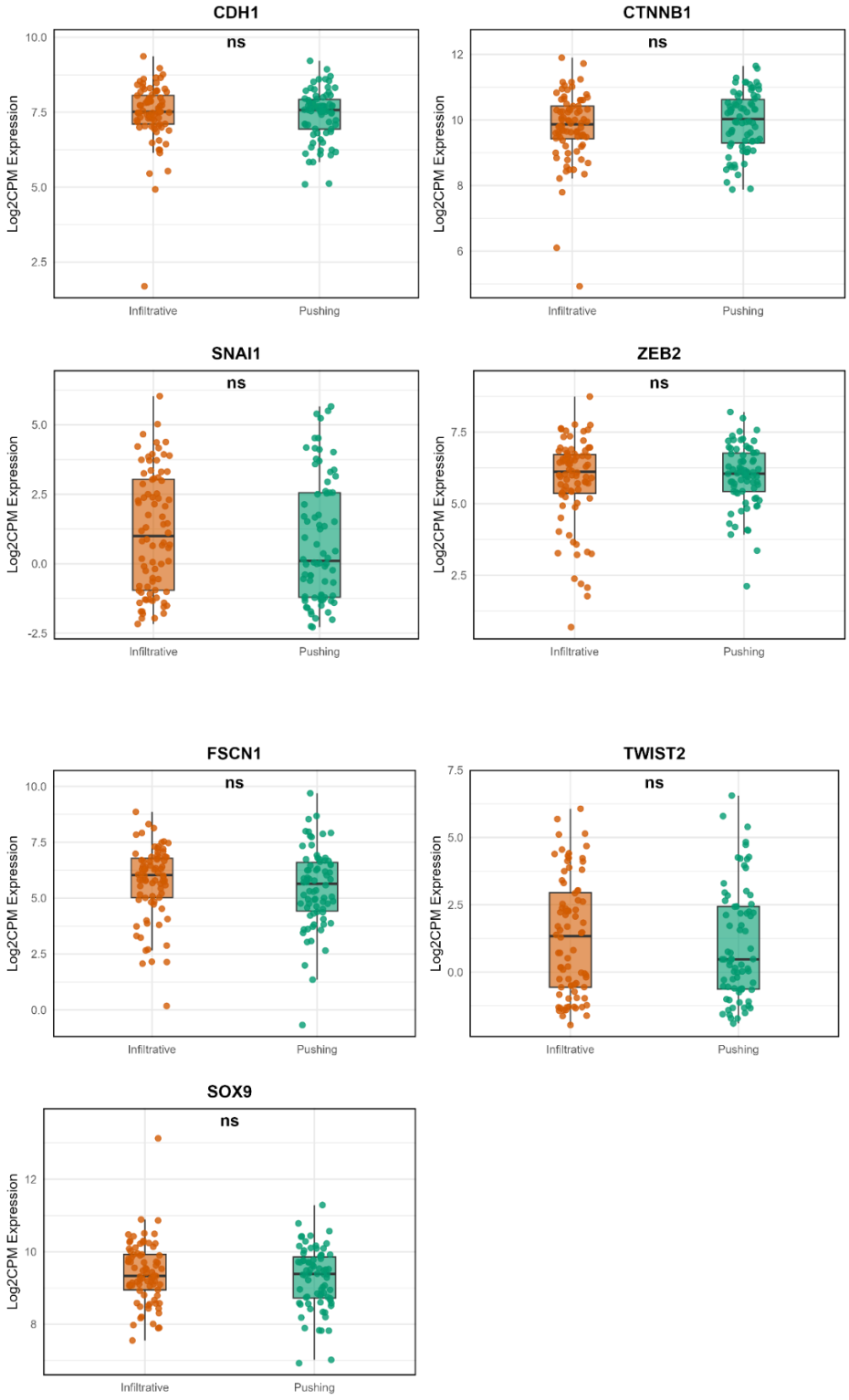


Figure 4-7: The gene expression of EMT markers
Box plots show TempO-Seq normalised expression (log2 counts per million). Orange boxes represent the infiltrative growth pattern; green boxes represent the pushing growth pattern. Group comparisons used the Wilcoxon rank-sum test with statistical significance defined as $p < 0.05$; ns indicates $p \geq 0.05$.

Therefore, the EMT-related protein level was then compared between TGPs using available IHC data at the tumour centre compartment as an additional analysis, rather than as direct validation of the invasive front.

4.2.6 EMT marker validation by IHC

The TMA cores from this cohort had previously been scored for a panel of EMT-related markers in earlier work, providing an opportunity to relate existing protein-level data to TGP. As these scores were generated on tumour-core rather than the invasive front, this analysis was undertaken opportunistically rather than as a planned validation of invasive front biology. Interestingly, infiltrative tumours expressed significantly higher nuclear staining for Snail and Twist markers ($p < 0.05$) as shown in Table 4-4. These results suggest that EMT transcription factors are linked to infiltrative tumour aggressiveness. In addition, cytoplasm staining of Fascin was significantly increased in infiltrative tumour ($p<0.05$) when compared to pushing tumour, indicating that Fascin might enhance cell motility of infiltrative tumour.

Table 4-4: EMT marker protein expression by tumour growth pattern

The proportion of cases with high versus low expression of each EMT marker is compared between infiltrative and pushing TGPs. Statistical comparison used the chi-squared test $p < 0.05$ was considered significant.

EMT Markers	Pushing growth pattern	Infiltrative growth pattern	p - value
E-cadherin (membrane)			0.175
Low expression	48 (66.7)	62 (76.5)	
High expression	24 (33.3)	19 (23.5)	
E-cadherin (cytoplasmic)			0.200
Low expression	7 (12.3)	15 (20.8)	
High expression	50 (87.7)	57 (79.2)	
Beta catenin (cytoplasmic)			0.148
Low expression	66 (50.4)	101 (58.7)	
High expression	65 (49.6)	71 (41.3)	
Beta catenin (membrane)			0.080
Low expression	11 (8.4)	26 (15.0)	
High expression	120 (91.6)	147 (85.0)	
Beta catenin (nuclear)			0.971
Low expression	111 (84.7)	146 (84.9)	
High expression	20 (15.3)	26 (15.1)	
Fascin (cytoplasm)			**0.001**
Low expression	41 (33.6)	27 (17.1)	
High expression	81 (66.4)	131 (82.9)	
Snail (cytoplasm)			0.299
Low expression	122 (88.4)	145 (84.3)	
High expression	16 (11.6)	27 (15.7)	
Snail (nuclear)			**0.023**
Low expression	35 (33.3)	31 (20.7)	
High expression	70 (66.7)	119 (79.3)	
Twist (cytoplasm)			0.238
Low expression	102 (77.9)	144 (83.2)	
High expression	29 (22.1)	29 (16.8)	
Twist (nuclear)			**0.047**
Low expression	110 (79.7)	120 (69.8)	
High expression	28 (20.3)	52 (30.2)	
SOX9 (cytoplasmic)			0.056
Low expression	21 (20.0)	17 (11.3)	
High expression	84 (80.0)	133 (88.7)	
SOX9 (nuclear)			0.773
Low expression	66 (62.9)	91 (61.1)	
High expression	39 (37.1)	58 (38.9)	
Zeb1 (cytoplasmic) at IE			0.660
Low expression	20 (71.4)	53 (75.7)	
High expression	8 (28.6)	17 (24.3)	
Zeb1 (nuclear) at IE			0.347
Low expression	3 (16.7)	1 (6.3)	
High expression	15 (83.3)	15 (93.8)	

4.2.7 Cell deconvolution using MCPcounter and CAF-related marker expression

Following pathway enrichment results, the MCPcounter deconvolution was applied to estimate the relative abundance of immune and stromal cell populations and subsequently were compared between TGPs. The analysis revealed that infiltrative tumours expressed significantly higher proportions of fibroblasts, endothelial cells, and monocytic lineage compared to pushing tumours (p.adj < 0.01) (Figure 4-8). These findings pointed out the significance of tumour microenvironment contributing to infiltrative behaviour. Of these three cell populations, fibroblasts were the most relevant to the stromal remodelling and EMT-related pathways identified by GSEA in infiltrative tumours. Thus, the five common CAF markers in CRC (ACTA2, FAP, PDPN, PDGFRA, and S100A4) were further compared at the gene expression level to determine whether they differ between TGPs.

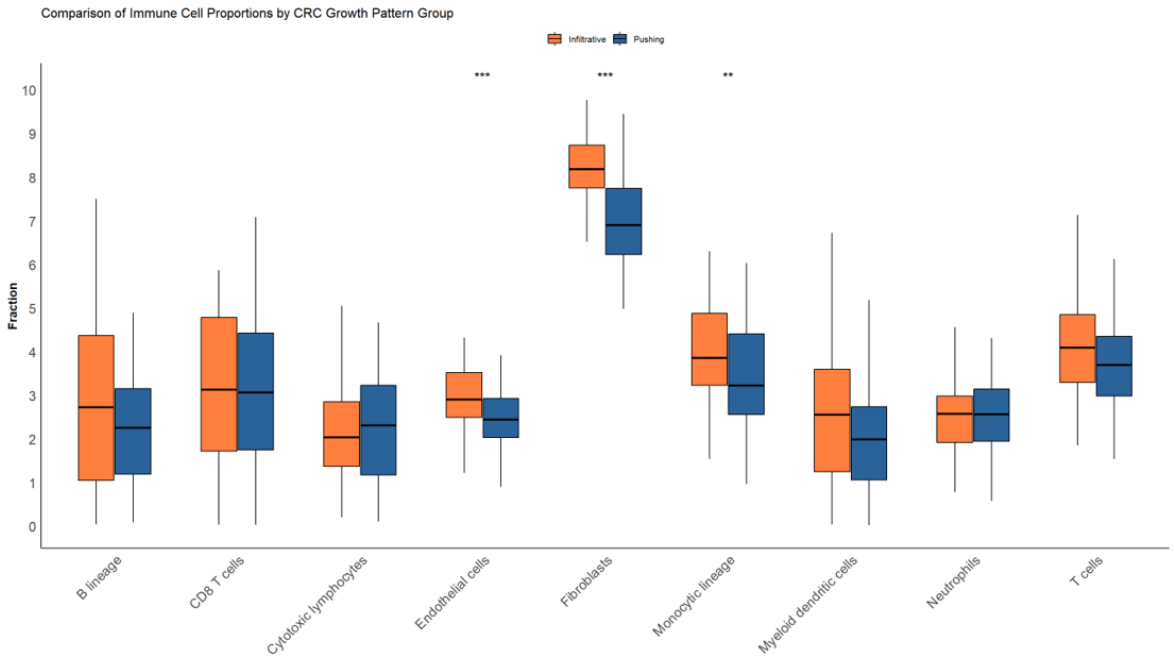


Figure 4-8: Cell populations between infiltrative and pushing growth patterns
The proportions of immune and stromal cell populations in TGPs were compared. An orange box plot represents an infiltrative growth pattern, while a blue box plot represents a pushing growth pattern. *** p adj <0.001, and ** p adj <0.01

All five common CAF marker gene expression comparisons showed clear results: infiltrative tumours expressed significantly higher levels of five common CAF markers than pushing tumours in CRC (p.adj < 0.001) (Figure 4-9). This supports the presence of CAF-related signals in infiltrative tumours.

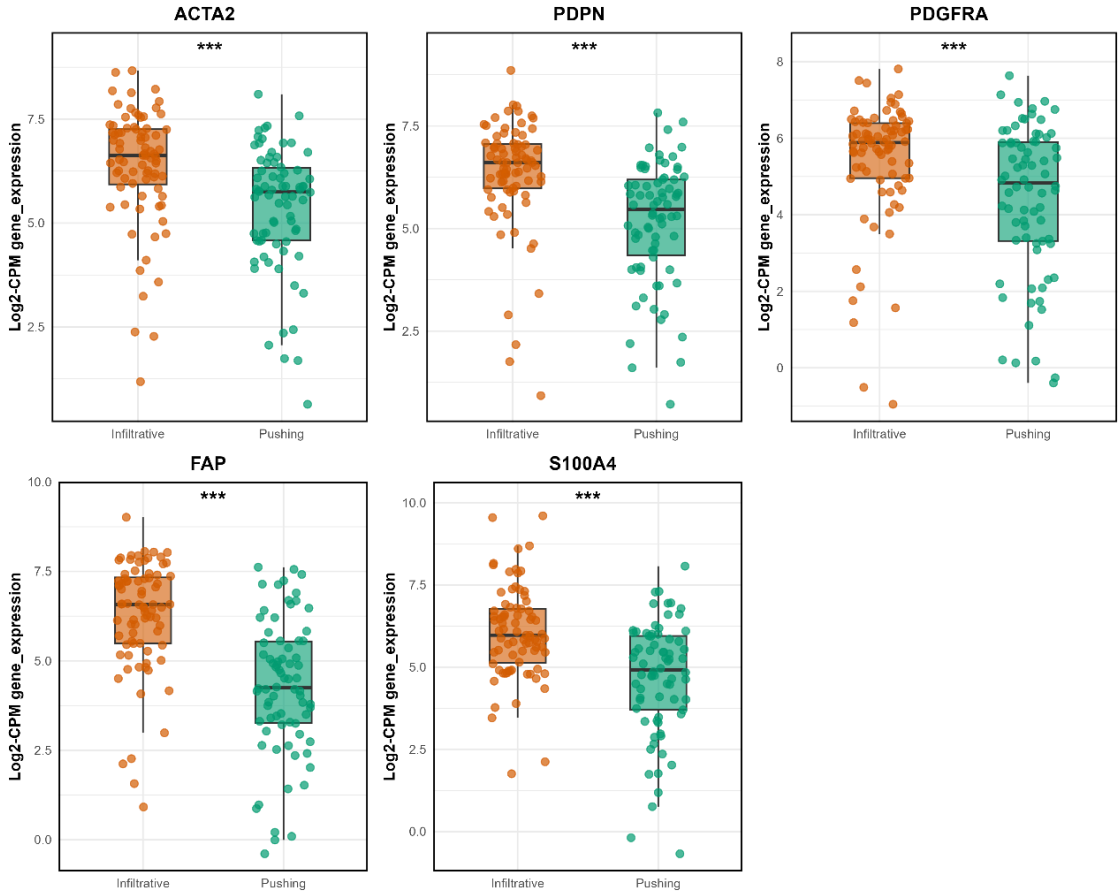


Figure 4-9: Common CAF markers expression between infiltrative and pushing growth patterns
Five common CAF marker expressions were compared between the two TGPs. An orange box plot represents the infiltrative growth pattern, and a green box plot represents the pushing growth pattern. *** p adj <0.001

4.2.8 Identification of CAF-associated candidate markers for mIF validation

After finding that common CAF markers in CRC demonstrated consistently higher expression in infiltrative tumours, the following step was to select candidate CAF markers to explore the spatial and the protein expression level between TGPs. Rather than referring to prior biological knowledge only, data-driven approach was applied by intersecting genes up-regulated in infiltrative tumours with common CAF markers in CRC. This approach provides both recognised biological knowledge and transcription signals. The intersection identified three overlapping genes: FAP, PDPN, and S100A4 (Figure 4-10). These results supported the selection of CAF-associated markers for protein-level and spatial assessment in the subsequent mIF chapter.

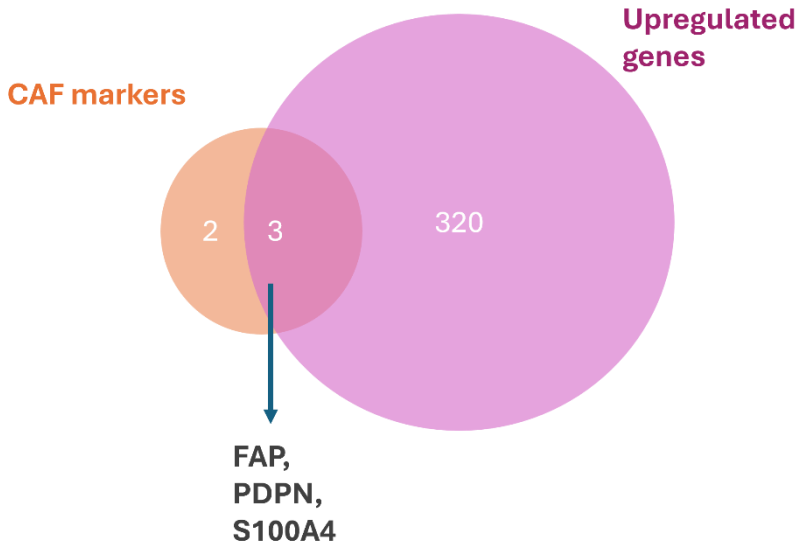


Figure 4-10: Venn diagram between Upregulated genes and CAF markers
Upregulated genes of infiltrative were intersected with five CAF markers. The arrow indicates the genes at the intersection area (3 genes).

4.3 Discussion

This chapter investigated the molecular features associated with TGPs in primary CRC. In the previous chapter, the infiltrative tumours were associated with poorer survival outcomes when compared with pushing growth. Thus, understanding the biological differences between TGPs can explain why infiltrative tumours are aggressive.

By performing TempO-Seq profiling of 154 primary CRC FFPE samples, 400 significantly differential gene expressions were identified between TGPs. Hallmark and GO:BP GSEA showed that infiltrative tumours were enriched in ECM, fibroblast activity, EMT, cell migration, inflammatory signaling, and immune-related pathways. In contrast, pushing tumours were enriched in metabolism, biosynthesis, cell-cycle regulation, and DNA repair. These findings suggest that infiltrative tumours are characterised by stromal remodelling and an inflammatory environment, while pushing tumours associated with tumour-cell-intrinsic metabolic and proliferative programmes.

The EMT-related pathways were the strongest finding from Hallmark GSEA analysis in infiltrative tumours and it is relevant to GO:BP pathways, including positive regulation of cell migration, cell-substrate junction assembly. EMT is a biological process which epithelial cells lose some epithelial features and acquire

more mesenchymal properties, leading to increased motility and invasive behaviours (Goossens et al., 2017). This is highly associated with histological growth pattern in infiltrative tumours which is defined by irregular borders and diffuse invasion into surrounding tissue (Karamitopoulou et al., 2010; Koelzer & Lugli, 2014).

However, none of the selected EMT-related genes showed statistically significant differential between TGPs, suggesting that the EMT pathway observed in GSEA analysis was not driven by a single EMT gene. Instead, it reflects coordinated changes from multiple EMT-related genes. Although EMT marker genes were not significantly different between TGPs, IHC was performed because transcript abundance does not always translate directly into protein expression. Therefore, EMT-related proteins were assessed to determine the difference between TGPs.

Nuclear Snail, nuclear Twist, and cytoplasmic Fascin were significantly higher in infiltrative tumours. Previous studies have shown that Snail and Twist can promote EMT-related phenotypes, including migration and invasion (Deng et al., 2016; Wang et al., 2023). Fascin is involved in cell motility (Tampakis et al., 2021). A higher cytoplasmic Fascin observed in infiltrative tumours supports that infiltrative growth pattern associated with a more motile phenotype. However, these IHC analysis was performed on tumour centre tissue. Thus, these results should be interpreted as supportive evidence that EMT-related protein expression differs by TGPs within the same patient cohort.

Following GSEA and MCPcounter results, infiltrative tumours are associated with CAF-associated biological processes, including ECM, fibroblast activation, and a higher abundance of fibroblasts compared to pushing tumours. These findings suggest that infiltrative tumours are associated with the tumour microenvironment, especially CAF. Common CAF markers in CRC were further determined to compare between TGPs, and the results were consistent with MCPcounter results that all common CAF markers gene expressions were higher in infiltrative tumours. For further CAF validation using mIF in the next chapter, CAF candidates were selected by intersections, which were FAP, PDPN, and S100A4. Although S100A4 was identified, αSMA was incorporated because it is an established marker of myofibroblastic CAFs and is closely linked to stromal

activation, extracellular matrix remodelling, collagen organisation, and tissue contractility (Nurmik et al., 2020). This was correlated to the transcriptomic findings in infiltrative tumours, which showed enrichment of fibroblast activation, collagen metabolic process, ECM remodelling, and TGF-β signalling. Therefore, combining FAP, PDPN, and αSMA allowed the mIF analysis to assess CAF expressions at the protein and spatial levels.

The CMS classification was consistent with the pathway-level differences observed between the two growth patterns. Most infiltrative tumours were classified as CMS4, whereas pushing tumours were predominantly CMS2. The predominance of CMS4 among infiltrative tumours is consistent with the enrichment of EMT, TGF-β signalling, extracellular matrix remodelling and fibroblast-related pathways observed by GSEA, as CMS4 is characterised by a mesenchymal and stromal-rich phenotype. In contrast, the predominance of CMS2 among pushing tumours is consistent with the enrichment of MYC-associated and proliferative programmes, including E2F targets and G2M checkpoint pathways. However, TGP and CMS describe different aspects of the tumour: TGP describes the low-magnification morphology of the invasive border, whereas CMS reflects the bulk transcriptional profile of the tumour. Relating TGP to CMS provides a link between histological growth pattern and an established molecular classification of CRC, allowing the biological characteristics associated with TGP to be interpreted within a broader molecular context.

From the multiple biological pathways enrichment in infiltrative results, it suggests that the infiltrative tumours cannot be explained by one pathway alone. In this study, EMT, TGF-β signalling, TNF-α/NF-κB signalling, hypoxia, angiogenesis, inflammatory response, apical junction, coagulation, and several ECM- and fibroblast-related GO:BP pathways were enriched in infiltrative tumours. These pathways are connected in the TME. For example, TGF-β signalling can promote both EMT-related changes in tumour cells and fibroblast activation (Lee & Massagué, 2022), while activated fibroblasts can remodel the extracellular matrix and contribute to stromal stiffness (Belhabib et al., 2021). Inflammatory and hypoxia pathways enhance EMT, angiogenesis, and immune-stromal interactions (Wei et al., 2021; Yeh et al., 2018). Therefore, the infiltrative tumours can be caused by crosstalk between tumour cells, CAFs, ECM remodelling,

inflammatory signalling, and immune pathways, instead of a single pathway activation.

The immune-related pathways were enriched in infiltrative tumours. For GO:BP terms, leukocyte activation, cytokine production, and adaptive immune response were enriched. Whereas Hallmark GSEA analysis showed enrichment for IL6/JAK/STAT3 signalling and TNF-α/NF-κB signalling. These findings suggest that infiltrative tumours have a more inflamed microenvironment. However, TempO-Seq data identifies immune-related pathways but cannot determine whether these pathways are driven by increased immune cells, changes in activation state, or spatial localisation. In addition, it cannot identify the specific immune cell type. Therefore, the mIF approach was performed in the next chapter to assess CD3-T-cell abundance and spatial organisation in relation to tumour and CAF cells.

Taken together, this chapter suggests that infiltrative tumours are characterised by enrichment of EMT, ECM, CAF-associated, inflammatory, and immune-related pathways. While pushing shows different biological pathways, including metabolic, biosynthesis, DNA repair, and proliferative pathways. These findings indicate that infiltrative tumours are associated with tumour-stroma interactions, while pushing tumours are correlated with tumour-intrinsic metabolic processes.

One limitation is that the infiltrative and pushing growth patterns were not matched for clinicopathological characteristics, with differences observed in T stage, N stage and tumour budding. Therefore, some transcriptomic differences between TGPs may reflect the associated clinicopathological features. A stage-matched analysis, adjusted for other variables, would help better define the biological differences associated with TGP independently of other clinicopathological features. Another limitation of this chapter is that TempO-Seq was performed on bulk tumour tissue, so the gene expression represents an average across mixed cell populations. In addition, it cannot provide spatial information; thus, it cannot determine how CAFs, immune cells, or tumours are organised at the invasive front. Therefore, the following mIF chapter was designed to address this limitation. The exclusion of intermediate growth pattern cases also limits the molecular findings across the full spectrum of TGPs. This extreme-group approach was used to maximise biological contrast and was not intended to

redefine TGP as a two-group prognostic classification. Another limitation is that IHC was performed in the tumour centre compartment. These IHC findings should be interpreted as complementary evidence rather than direct validation of the invasive margin biology. In addition, they may not capture the association between marker expression and growth pattern at the invasive front itself, where the growth pattern is defined. However, an invasive edge TMA for this cohort is now available and staining it in the future would allow these markers to be assessed directly at the invasive front, where this association may differ from that observed in the tumour centre.

Chapter 5 Spatial proteomics at the invasive front of primary colorectal cancer

5.1 Introduction

In the previous chapter, bulk RNA-seq analysis revealed that fibroblast populations were significantly more abundant in infiltrative compared to pushing TGPs. However, bulk RNA-seq provides an average gene expression profile across all cells, lacking both single-cell resolution and spatial or architectural information. This limitation means that while fibroblast enrichment could be detected, it was not possible to determine which fibroblast subtypes were responsible, where they were located relative to tumour and immune cells, or how their spatial arrangement might differ between growth patterns.

To address these questions, mIF was selected as the next investigative approach. Unlike conventional immunohistochemistry (IHC), which is limited to the assessment of a single protein per tissue section, mIF allows simultaneous detection of multiple biomarkers on the same section, enabling co-expression analysis and spatial mapping of multiple cell populations within the tumour microenvironment (TME) (Hernandez et al., 2021). Importantly, mIF preserves the spatial coordinates of every cell, it enables quantification of how cells are arranged relative to one another, not only how many are present.

The TME plays a crucial role in tumour progression. CAFs are the most abundant stromal component in the TME, and several studies have shown that they correlate with poor prognosis (Liu et al., 2025; Min et al., 2021), although some CAF subtypes are associated with favourable outcomes in CRC (Cords et al., 2024). Given these discrepancies, spatial analysis is essential for understanding how the TME is organised across different TGPs (TGPs) and how this organisation may influence disease progression.

In the previous chapter, five commonly used CRC CAF markers were compared between growth patterns; all were expressed more highly in infiltrative tumours, and a data-driven intersection of the significantly up-regulated genes with these markers returned FAP, PDPN, and S100A4. FAP and PDPN were supported by both the literature and the transcriptomic data. S100A4 also

emerged from the intersection but was not carried forward, as it is not restricted to fibroblasts, being expressed by endothelial and epithelial cells (Zhao et al., 2023), which would confound its interpretation at the single-cell level; αSMA was used instead to represent the myofibroblastic CAF subset. Among the markers not taken forward, PDGFRA was elevated in infiltrative tumours but did not emerge from the data-driven intersection. Other commonly used markers, although also expressed by CAFs, were not included because they are not specific to fibroblasts and would be difficult to interpret at the single-cell level: vimentin is expressed by endothelial cells and by tumour cells undergoing EMT, whereas PDGFRB is also present on pericytes and perivascular cells (Zhao et al., 2023).

FAP expression has been shown to be enriched at the CRC invasive front and associated with poor prognosis (Sandberg et al., 2019; Wikberg et al., 2013). αSMA was included as a marker of myofibroblastic CAFs, which represent a functionally distinct subtype associated with contractile properties and wound healing responses (Sahai et al., 2020). PDPN was selected because, although widely used as a lymphatic vessel marker, it is also expressed by CAF subpopulations and has been associated with both favourable and unfavourable outcomes depending on cancer type (Pula et al., 2013; Yamanashi et al., 2009). The combination of FAP, αSMA, and PDPN enables the identification of multiple CAF subtypes based on marker co-expression patterns, including single-positive, double-positive, and triple-positive populations. Pan-cytokeratin (Pan-CK) was included to identify tumour epithelial cells. CD3 was included as a marker of T lymphocytes, given the well-established prognostic significance of T cell infiltration in CRC (Williams et al., 2024), though previous studies have largely focused on cell density without spatial context. In this exploratory study, CD3 was prioritised to capture the broader spatial context of T cells in relation to tumour and CAF populations rather than to resolve specific T-cell subsets. CD8 was therefore not included because of the limited number of fluorescence channels available in the multiplex panel, and more detailed T-cell phenotyping was beyond the scope of this analysis.

To facilitate spatial analysis at the invasive margin, tissue microarrays (TMAs) were constructed with cores targeted to the invasive margin of primary CRC specimens. Each patient was represented by 2–3 cores, enabling assessment of spatial relationships between CAF subtypes, tumour cells, and immune cells.

Two hypotheses were tested. First, the spatial organisation of CAF subtypes relative to tumour cells and immune cells differs between infiltrative and pushing TGPs. Second, these spatial relationships carry independent prognostic significance for cancer-specific survival beyond conventional clinicopathological features. To address these hypotheses, this chapter has three specific aims:

1. To characterise CAF subtype composition at the invasive margin across the two growth patterns

2. To quantify the spatial relationships between CAF subtypes, tumour cells, and T cells using the NDI

3. To determine whether these spatial metrics are independently associated with cancer-specific survival.

5.2 Results

5.2.1 Cohort characteristics

A total of 101 patients with primary CRC were included in this study. Tissue microarrays were constructed with cores targeted to the invasive margin, with each patient consisting of 2–3 cores.

To determine whether the cellular composition at the invasive margin differed between TGPs, the relative abundance and density of each cell phenotype were compared.

5.2.2 Cell composition and density analysis

Investigating the cellular composition and density of cell phenotypes, the infiltrative and pushing growth patterns were compared. Full methods are described in Chapter 2 (Methods).

For the proportion of cell phenotypes, as shown in Figure 5-1, there were no statistically significant differences in the proportional abundance of any cell phenotype between the infiltrative and pushing margins (all BH-adjusted p-values > 0.05).

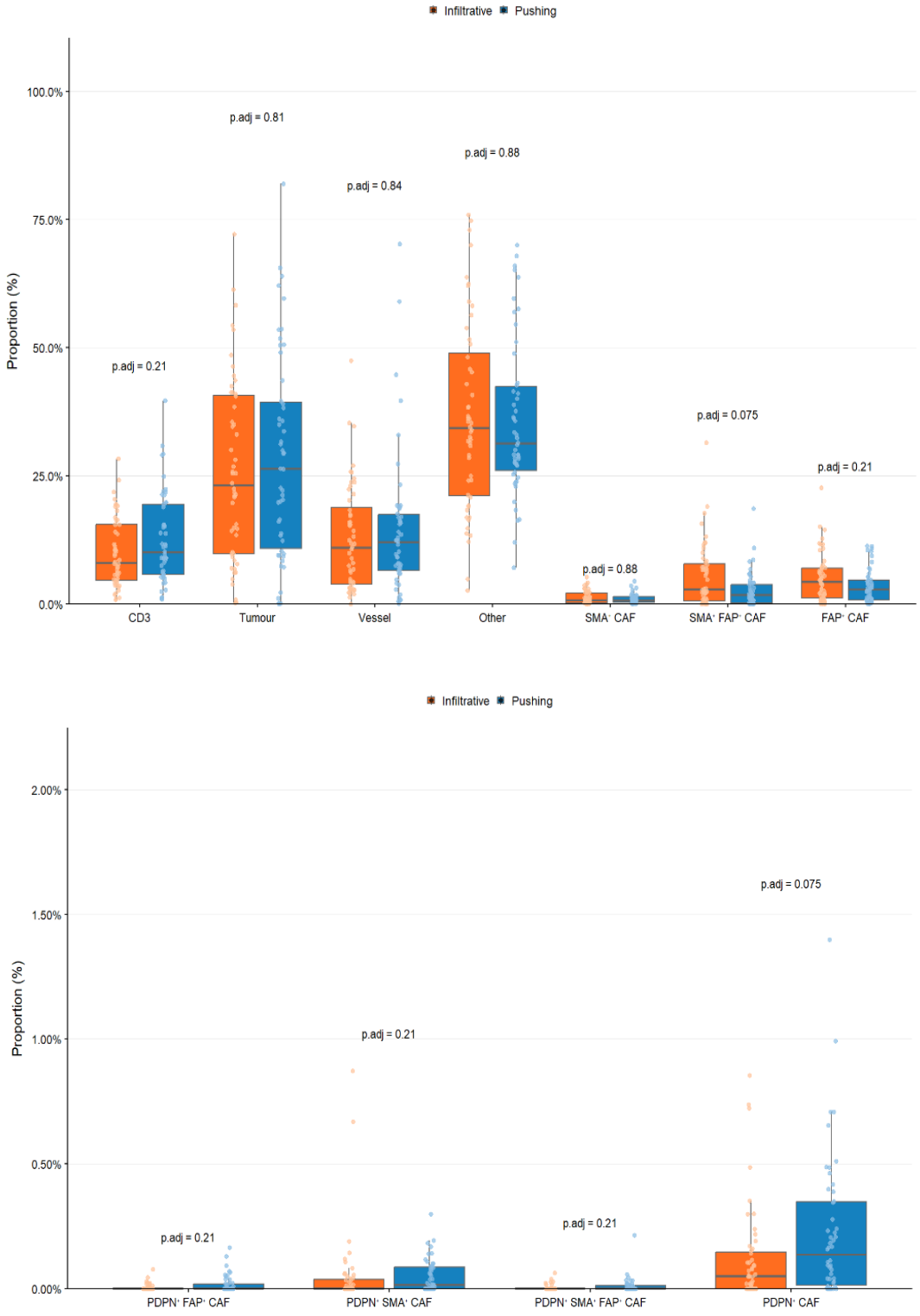


Figure 5-1: Cell proportion between tumour growth patterns. Boxplots comparing cell proportion (%) between TGPs at the invasive margin. The infiltrative margin is represented in orange, and the pushing margin in blue. Each dot represents an individual sample.

Consistent with the abundance results, no significant differences in cell density across all cell phenotypes were observed between TGP groups (Figure 5-2; all p.adj > 0.05).

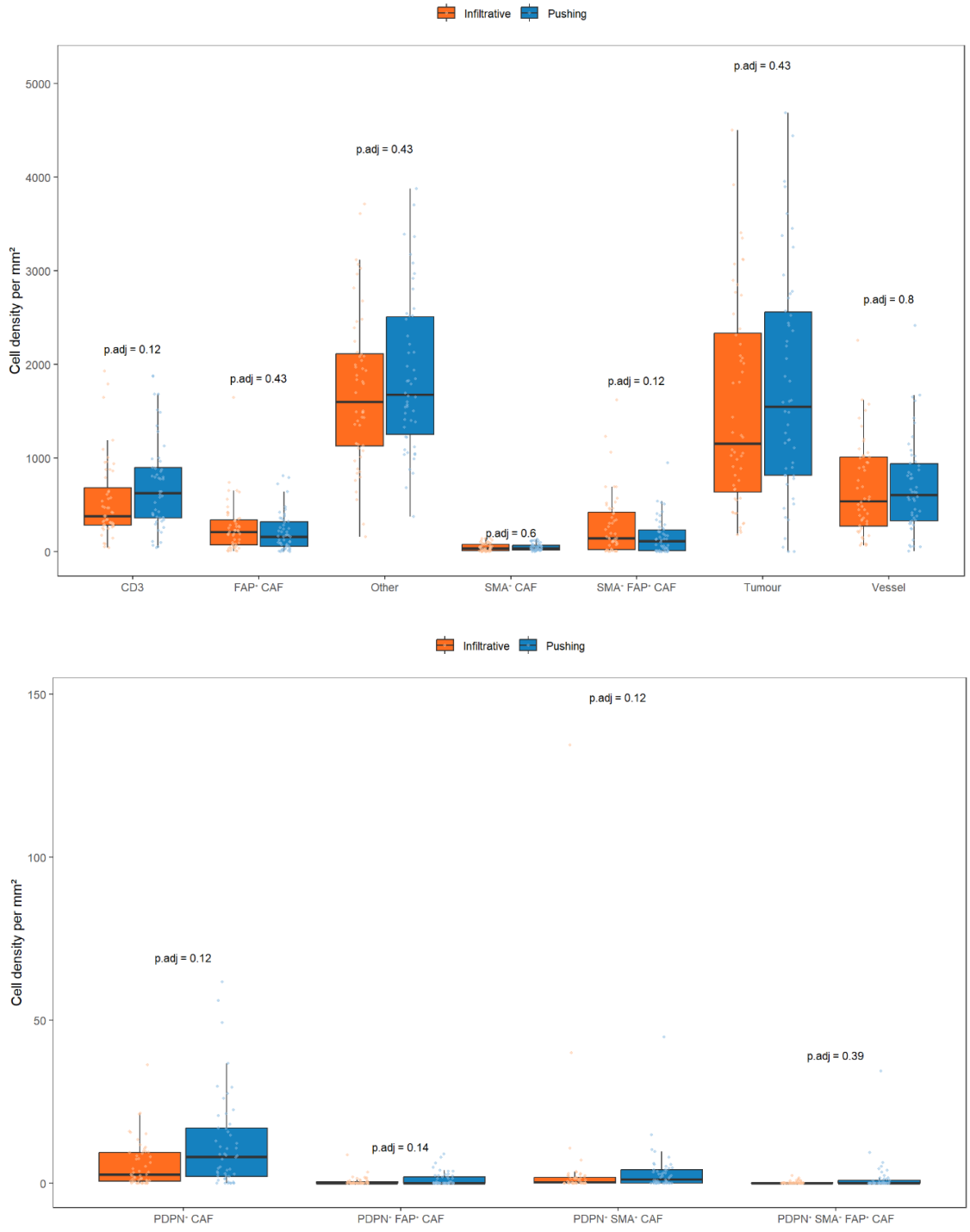


Figure 5-2: Cell densities between tumour growth patterns

Boxplots displaying the absolute cell density (cells per mm²) of cell phenotypes. The infiltrative margin is represented in orange, and the pushing margin in blue. Each dot represents an individual sample.

Collectively, these findings provide a crucial biological insight that the difference between infiltrative and pushing growth patterns cannot be explained by the proportion or density of cells. Therefore, employing spatial analysis was required to investigate the actual factors associated with the two different TGP.

5.2.3 Spatial proximity analysis: Nearest Distance Index (NDI)

To examine whether the spatial organisation of cells at the invasive front differ between TGP, the NDI was applied to quantify the spatial proximity between cells.

The NDI is the $\log_2$-transformed ratio of the observed median nearest-neighbour distance to the expected distance under complete spatial randomness (CSR), providing a density-normalized measure of spatial proximity. Full methods are described in Chapter 2 (Methods).

Figure 5-3 presents the NDI comparison between infiltrative (orange) and pushing (blue) for nine NDI pairwise cell-type proximities (NDI metrics). calculated. A minimum of five patients per group was required for an NDI metric to be calculated. As a result, some NDI metrics did not reach this threshold, and therefore, NDI metrics were not calculated. Five endpoints showed statistically significant differences (BH-adjusted $p < 0.05$). The pushing group had significantly higher NDI than the infiltrative group for three endpoints: CD3 to tumour, FAP+ to tumour, and αSMA+FAP+ to tumour, indicating greater spatial exclusion between these cell populations in pushing tumours and equivalently, lower NDI (closer proximity) in infiltrative tumours. Conversely, FAP+ to CD3 and αSMA+FAP+ to CD3 NDI values were significantly higher in the infiltrative group, indicating that FAP-expressing CAFs were further from T cells in infiltrative tumours (closer in pushing tumours).

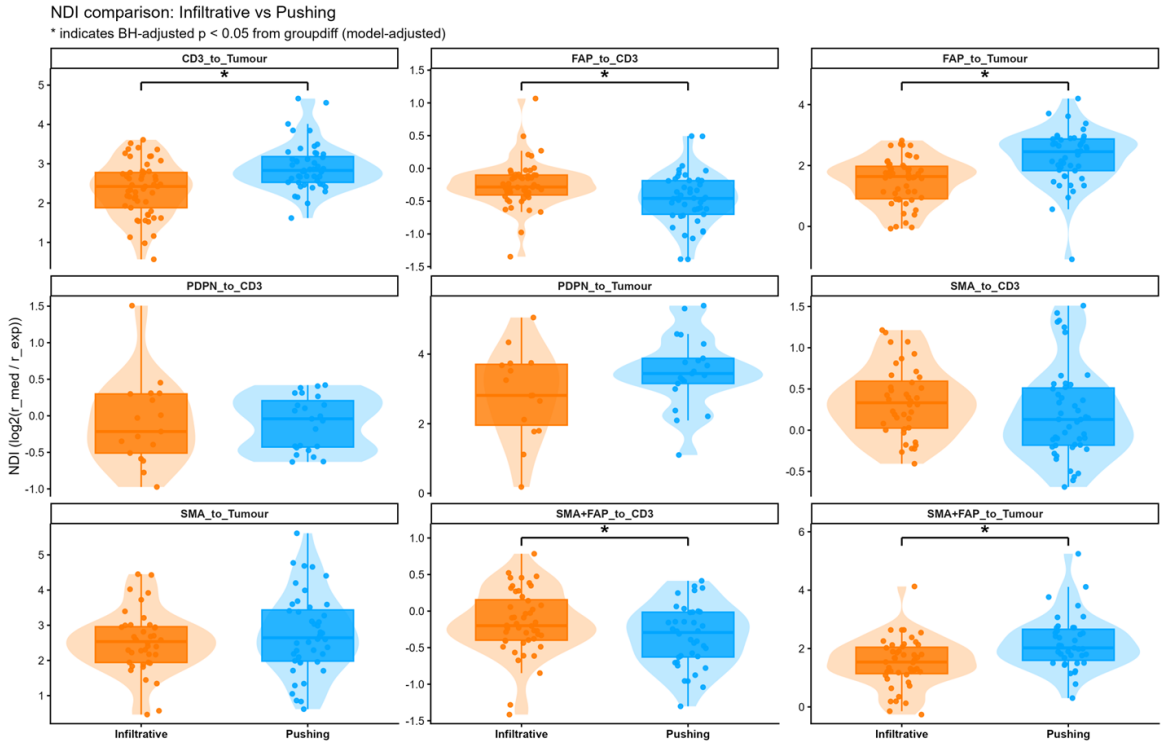


Figure 5-3: Nearest Distance Index comparison between tumour growth patterns

The infiltrative margin is represented in orange, and the pushing margin in blue. Each dot represents an individual sample.

Representative cores are shown in Figure 5-4 and Figure 5-5, which display the mIF staining and the corresponding single-cell maps from a pushing and an infiltrative tumour.

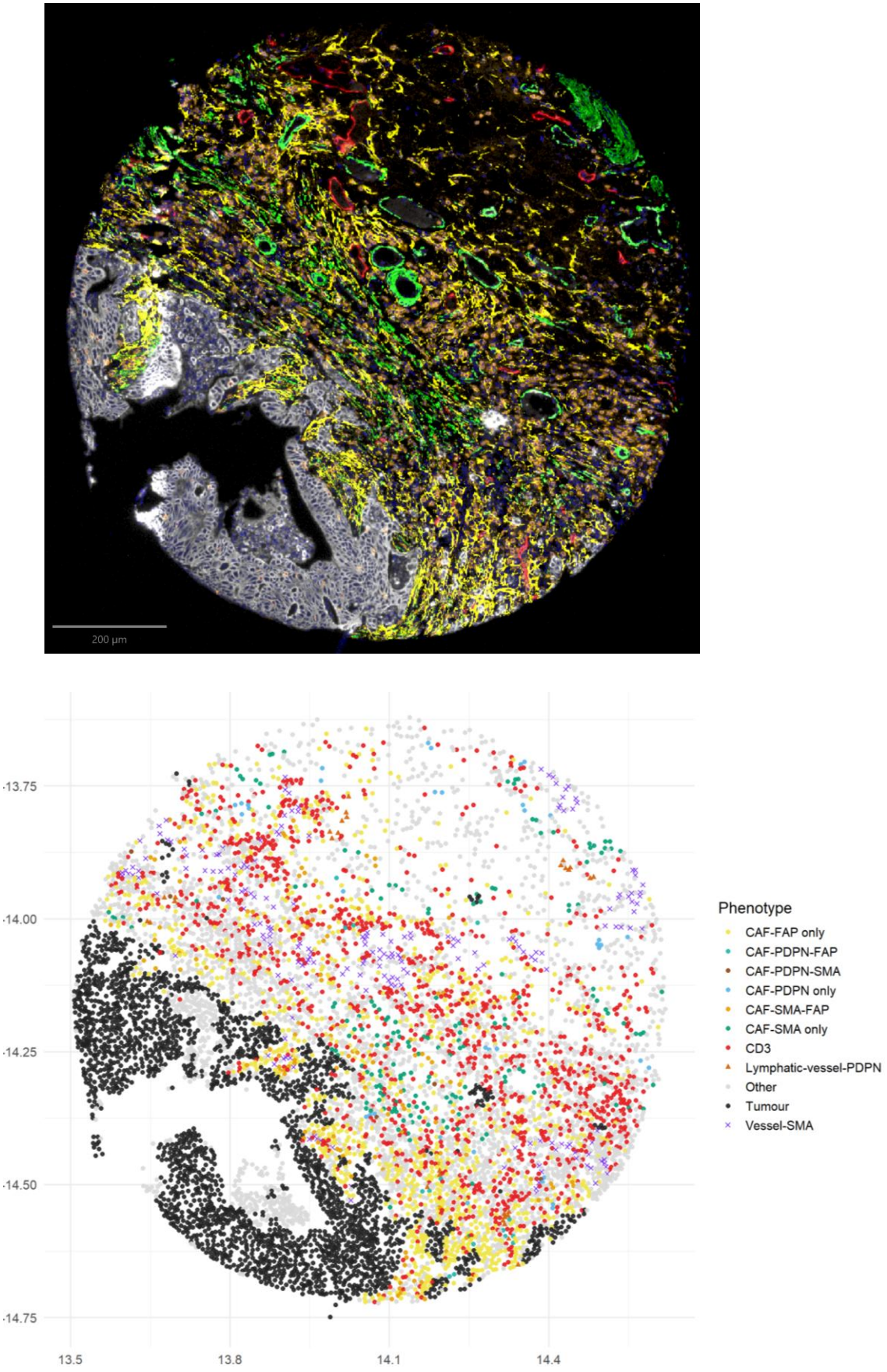
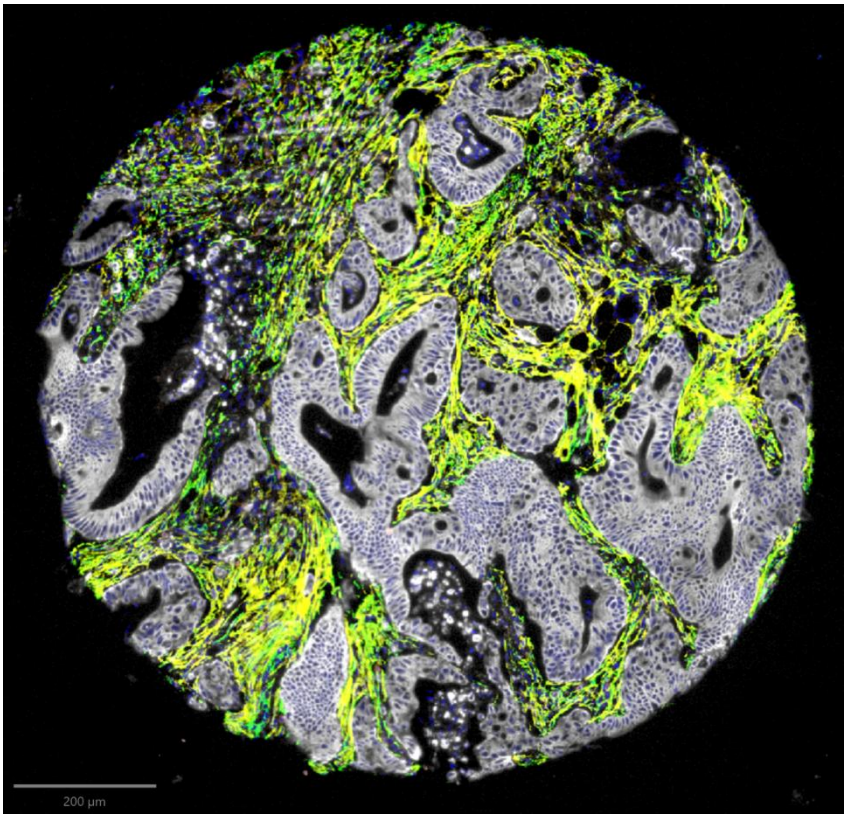


Figure 5-4: Representative cores from a pushing tumour. The mIF stained, shown in 200 µm scale bar (Top). Single-cell phenotype maps of the same core and coloured by phenotype (Bottom). As the NDI is density-corrected, these proximity relationships cannot be read directly from the images.



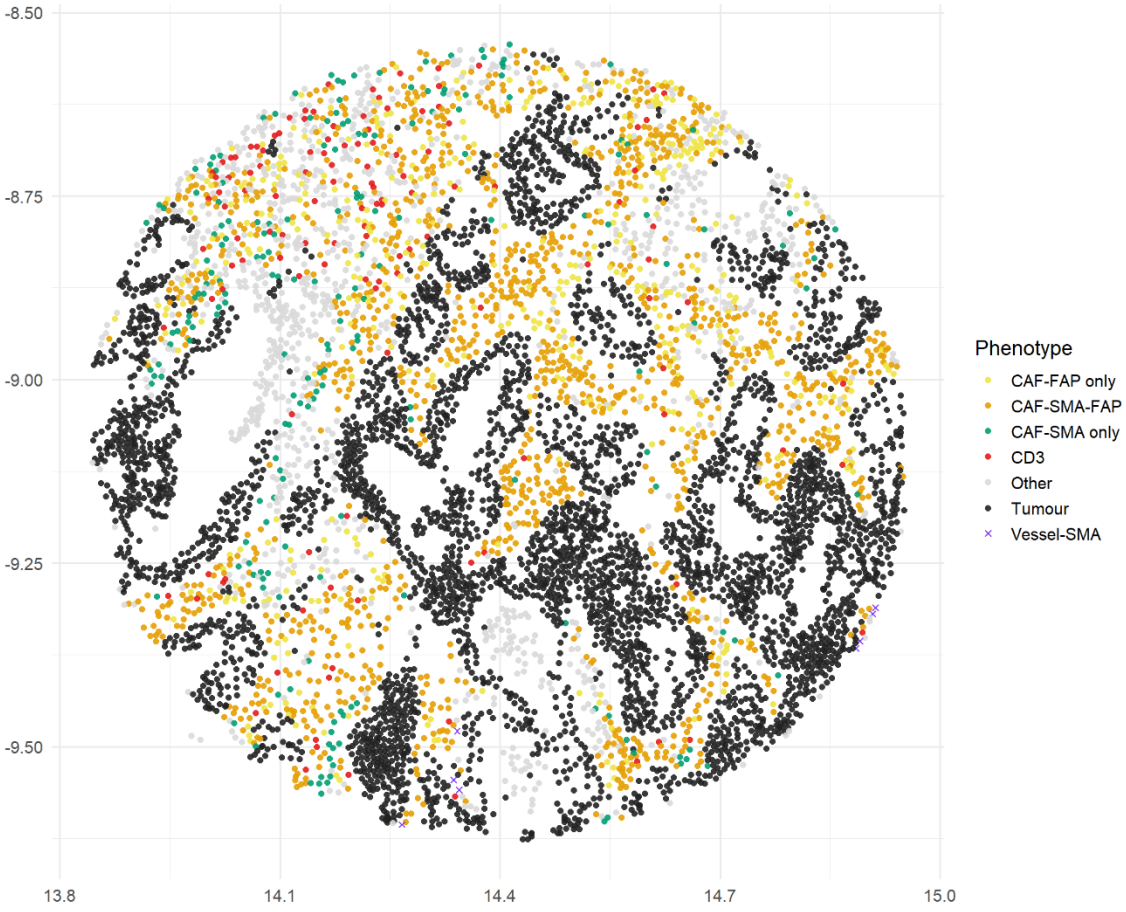


Figure 5-5: Representative cores from an infiltrative tumour. The mIF stained, shown in 200 µm scale bar (Top). Single-cell phenotype maps of the same core and coloured by phenotype (Bottom). As the NDI is density-corrected, these proximity relationships cannot be read directly from the images.

These spatial differences raised the question of whether the observed proximity patterns carried prognostic significance. To address this, survival analysis was performed to evaluate the relationship between NDI and cancer-specific survival.

The univariable and multivariable Cox regression analyses were performed to assess whether NDI metrics were independent predictors of CRC-specific survival. To categorise NDI data into low and high groups, the Maximally Selected Rank Statistics (MaxStat) was utilised to get the optimal cutoff. This provided the most significant separation between groups.

Table 5-1 showed that tumour budding, N stage, NDI FAP+ to tumour, and NDI αSMA+FAP+ to tumour were significantly associated with cancer-specific survival ($p < 0.05$) in univariable analysis. To determine whether the significant NDI metrics remained prognostic after adjusting for clinicopathological factors, multivariable Cox regression was performed.

Table 5-1: Univariable Cox regression analysis of spatial proximity indices (NDI) and clinicopathological characteristics for cancer-specific survival in CRC. Bold p-values are statistically significant. Each NDI metric was dichotomised with the high-NDI group as the reference category; a hazard ratio above one, therefore, denotes higher cancer-specific mortality in the low-NDI group. Values in parentheses denote the growth pattern to which the low-NDI group corresponds.

Variable	HR 95CI	P-Value
Tumour_budding (present vs absent)	2.08 (1.01-4.28)	**0.047**
Age (>75 yr vs ≤75 yr)	3.48 (0.42-28.76)	0.247
Tstage (T3-T4 vs T1-T2)	2.01 (0.48-8.43)	0.341
Nstage (N1-N2 vs N0)	3.04 (1.36-6.82)	**0.007**
TGP (pushing vs infiltrative)	0.55 (0.27-1.13)	0.105
NDI FAP+ to Tumour (low vs high)	2.45 (1.16-5.15)	**0.019**
NDI αSMA+FAP+ to Tumour (low vs high)	2.72 (1.14-6.48)	**0.024**
NDI FAP+ to CD3 (low vs high)	0.50 (0.24-1.04)	0.064
NDI αSMA+FAP+ to CD3 (low vs high)	0.47 (0.16-1.38)	0.169
NDI CD3 to Tumour (low vs high)	1.68 (0.77-3.64)	0.193

For the multivariable Cox regression analysis, backward selection was performed separately for each NDI metric to avoid collinearity, and only variables significant in the univariable analysis were retained (Table 5-2).

Table 5-2: Multivariable Cox regression analysis of FAP+ to tumour NDI and clinicopathological characteristics for cancer-specific survival in CRC. Bold p-values are statistically significant. Each NDI metric was dichotomised with the high-NDI group as the reference category; a hazard ratio above one, therefore, denotes higher cancer-specific mortality in the low-NDI group. Values in parentheses denote the growth pattern to which the low-NDI group corresponds.

Variable	HR (95% CI)	P-value
Tumour budding (present vs absent)	1.78 (0.83-3.82)	0.136
Nstage (N1-N2 vs N0)	2.76 (1.16-6.55)	**0.022**
NDI FAP+ to Tumour (low vs high)	2.19 (0.99-4.86)	0.053

As shown in Table 5-2, N stage remained an independent predictor of cancer-specific mortality. A lower FAP+ to tumour NDI (indicating closer spatial proximity between FAP+ CAFs and tumour cells) showed a strong trend towards a 2.19-fold increased risk of cancer-specific mortality compared to those with high NDI (HR = 2.19, 95% CI 0.99–4.86, p = 0.053).

A low αSMA+FAP+ to tumour NDI was significantly associated with cancer-specific survival in multivariable analysis after adjusting for tumour budding and N stage (HR = 2.46, 95% CI 1.00–6.05, p = 0.05), indicating that closer spatial proximity between αSMA+FAP+ CAFs and tumour cells was independently associated with a 2.46-fold higher risk of cancer-specific mortality in patients with a low NDI than in those with a high NDI (Table 5-3).

Table 5-3: Multivariable Cox regression analysis of αSMA+FAP+ to tumour NDI and clinicopathological characteristics for cancer-specific survival in CRC. Bold p-values are statistically significant. Each NDI metric was dichotomised with the high-NDI group as the reference category; a hazard ratio above one, therefore, denotes higher cancer-specific mortality in the low-NDI group. Values in parentheses denote the growth pattern to which the low-NDI group corresponds.

Variable	HR (95% CI)	P-value
Tumour_budding (present vs absent)	2.00 (0.88-4.54)	0.097
nstage (N1-N2 vs N0)	3.15 (1.21-8.23)	0.019
NDI αSMA+FAP+ to Tumour (low vs high)	2.46 (1.00-6.05)	**0.050**

5.3 Discussion

This chapter presents four important findings. First, the cellular composition at the invasive margin, measured by both relative abundance and density, did not differ significantly between infiltrative and pushing TGPs. Second, NDI analysis revealed that FAP-expressing CAFs were closer to tumour cells in infiltrative tumours than in pushing tumours and were also further from T cells. Third, T cells were spatially excluded from tumour cells in pushing tumours. Fourth, the spatial proximity of αSMA+FAP+ CAFs to tumour cells was independently associated with cancer-specific survival after adjustment for tumour budding and N stage, with the FAP+ single-positive endpoint showing a borderline result. Together, these findings support the view that cell arrangement, rather than cell abundance, distinguishes the two growth patterns and carries prognostic information.

Cell composition does not classify TGPs

The cell proportion and density comparison between TGPs was not different, demonstrating that the distinction between the two growth patterns cannot be explained by the quantity or density of immune cells, or CAF subtypes at the invasive margin. These findings provided the rationale for employing spatial analysis using the NDI to investigate whether the arrangement, rather than the abundance, of cells distinguishes TGPs.

Spatial organisation of FAP-expressing CAFs differs between TGPs

The NDI analysis revealed significantly lower FAP+ to tumour and αSMA+FAP+ to tumour NDI values in the infiltrative group compared to the pushing group, indicating that FAP-expressing CAFs lie relatively closer to tumour cells in infiltrative tumours than in pushing tumours. The close spatial association between FAP+ CAFs and tumour cells in infiltrative tumours is consistent with the known biology of this CAF population. FAP is a type II transmembrane serine protease highly expressed in tumour stroma that promotes extracellular matrix remodelling, invasion, and epithelial-to-mesenchymal transition (Xin et al., 2021). The juxtatumoural positioning of FAP+ CAFs observed here places them in contact with tumour cells, providing a spatial substrate for direct tumour–stroma crosstalk

through localised ECM remodelling and paracrine signalling that may facilitate local invasion (Qi et al., 2022).

Notably, SMA+ and PDPN+ CAFs without FAP co-expression did not show significant differences in NDI between the two growth patterns. This suggests that the spatial distinction between TGPs is specific to FAP-expressing CAF populations.

FAP-expressing CAFs are spatially further from T cells in infiltrative tumours

The finding that FAP+ to CD3 and αSMA+FAP+ to CD3 NDI values were significantly higher in the infiltrative group indicates that, in infiltrative tumours, FAP-expressing CAFs are located further from CD3+ T cells than in pushing tumours. A higher FAP+ to CD3 NDI together with a lower FAP+ to tumour NDI in infiltrative tumours explains the spatial configuration that FAP+ CAFs lie close to tumour cells, while T cells remain distant. This configuration is biologically significant because FAP+ CAF are known to exclude T cells through CXCL12-CXCR4 signaling (Feig et al., 2013).

A key finding in this study is that the spatial proximity of αSMA+FAP+ CAFs to tumour cells was an independent factor of cancer-specific survival in multivariable analysis, after adjusting for tumour budding and N stage (HR = 2.46, p = 0.05). Similarly, a low FAP+ to tumour NDI showed a closely parallel result (HR = 2.19, p = 0.053). In both models, a lower NDI indicating closer proximity between FAP+ / αSMA+FAP+ CAFs and tumour cells was associated with worse survival. These results suggest that the NDI captures biologically relevant information about the tumour-stroma interaction that conventional staging does not.

This study suggests that it is not simply the presence of FAP+/ αSMA+FAP+ CAFs, but specifically their spatial proximity to tumour cells, that carries prognostic information. This distinction is important because cell density and spatial proximity need not align: two patients may have similar FAP+ / αSMA+FAP+ CAFs densities but different spatial arrangements, with the patient whose CAFs are in closer proximity to tumour cells having worse outcomes.

CD3 T cells are spatially excluded from tumour cells in pushing tumours

The pushing group showed significantly higher NDI for CD3 to tumour, indicating that T cells are further from tumour cells in pushing tumours. This likely reflects the distinct tissue architecture of pushing tumours, which have a well-defined, smooth tumour border that creates a clearer physical separation between the tumour epithelium and the surrounding stroma where immune cells reside (Koelzer & Lugli, 2014). In contrast, infiltrative tumours have irregular, finger-like projections that interdigitate with the stroma, placing tumour cells closer to stromal cells. However, a closer proximity of CD3 T-cells to tumour cells in infiltrative tumours should not necessarily be interpreted as a more effective anti-tumour immune response. T cells within the tumour microenvironment may become dysfunctional or exhausted, and spatial proximity to tumour cells may coexist with impaired anti-tumour activity. As this exploratory mIF panel used CD3 to capture the broader T-cell population and did not include markers of T-cell subsets, the functional state of these T cells could not be determined. Further immune phenotyping would be required to distinguish active from exhausted T-cell populations.

Comparison with existing spatial studies in CRC

Several studies have shown the prognostic value of immune cells in CRC, but few studies have focused on the spatial arrangement of CAF subtypes. The Immunoscore, CD3 and CD8 densities were found to be prognostic parameters at the tumour centre and invasive margin (Galon et al., 2006; Pagès et al., 2018). However, it is a density-based metric and does not capture the spatial proximity between immune cells and tumour or CAF population.

Nearchou et al. demonstrated that a low number of CD3+CD8+ T cells within 50 μm of tumour buds was associated with reduced disease-specific survival in stage II CRC (HR = 8.91, 95% CI 2.83–28.0) (Nearchou et al., 2019). The present study extends this spatial analysis in two directions. First, rather than density analysis, the similar spatial proximity logic was applied. Second, Nearchou et al. reported that closer lymphocyte–tumour bud proximity predicted better outcome; the present data showed that closer CAF–tumour proximity predicts worse outcome. These results are not in conflict because lymphocyte proximity reflects active antitumour response, whereas CAF proximity could reflect stromal support

of invasion. Together, they support the broader view that the prognostic meaning of spatial proximity depends on the identities of the cell types involved.

To our knowledge, this is the first study to use the NDI to quantify the spatial relationships between CAF subtypes and tumour cells in CRC stratified by TGP. While the NDI has been applied in other contexts, its application to the specific question of CAF–tumour spatial proximity in TGP is novel. This approach reveals a layer of spatial information that is missed by both density-based analyses and immune-cell-focused spatial metrics.

Limitations

Several limitations should be acknowledged. First, the less common CAF subtypes did not meet the minimum cell threshold, so that the NDI calculation could not be performed. Second, the mIF panel included five markers and therefore could not capture several other cell subtypes. Finally, the cross-sectional design cannot distinguish whether close CAF–tumour proximity is a cause of aggressive disease behaviour or a downstream marker of it; this limits causal inference and should be addressed in future functional and longitudinal studies.

While mIF has resolved how CAFs and T cells are spatially arranged relative to tumour cells, it cannot identify the molecular signals within these spatial arrangements. To begin to address this, the following chapter applies a spatial transcriptomics approach (GeoMx) at the invasive front, with the aim of resolving the signalling pathways that may underlie the CAF-tumour and CAF-T cell spatial relationships identified here.

Chapter 6 Spatial biology of the tumour growth patterns in primary CRC

6.1 Introduction

In the previous chapter, mIF analysis at the invasive front of primary CRC identified five NDI endpoints that differed significantly between infiltrative and pushing growth patterns: FAP+ to Tumour, αSMA+FAP+ to Tumour, CD3 to Tumour, FAP+ to CD3, and αSMA+FAP+ to CD3. Critically, a greater αSMA+FAP+ to Tumour distance emerged as an independent favourable prognostic factor for CRC-specific survival, while a greater FAP+ to Tumour distance showed a strong favourable prognostic trend ($p = 0.053$). These findings established that the spatial arrangement of CAF subtypes and T cells relative to the tumour is different between the two growth patterns and carries clinical significance. However, identifying these spatial distances alone could not answer the questions of which specific cell populations correlate with these spatial arrangements and what molecular mechanisms underlie them. While mIF provides single-cell phenotypic resolution and spatial coordinates, it is inherently limited in the number of markers that can be assessed simultaneously. To bridge this gap, an integrative approach combining comprehensive spatial transcriptomic profiling with computational inference of cell states and ligand-receptor interactions was required.

The aim of this chapter was to investigate the cellular and molecular biology underlying the five spatial endpoints identified in the previous chapter. The objectives of this study were:

1. To identify the transcriptional programmes whose activity correlates with each spatial distance, using single-sample gene set enrichment analysis (ssGSEA) of Hallmark pathways.

2. To identify the specific cell subpopulations whose abundance correlates with each spatial endpoint within each tissue compartment, using EcoTyper cell-state deconvolution.

3. To identify candidate ligand-receptor signalling axes that may underlie the observed spatial relationships, using LIANA ligand-receptor inference.

6.1.1 GeoMx cohort selection

The GeoMx cohort was derived from the existing invasive margin TMA of GRI cohort. Owing to the cost of GeoMx WTA profiling, the analysis was limited to two TMA slides. The two slides containing the greatest number of infiltrative and pushing growth patterns cases were selected to maximise the number of informative cases available for comparison. Individual patients were not selected on the basis of clinicopathological characteristics or clinical outcome. Intermediate growth pattern cases were excluded from the comparison. The final cohort comprised 51 patients, including 31 infiltrative and 20 pushing tumours.

6.1.2 GeoMx digital spatial profiling: whole-transcriptome interrogation of the invasive front in primary CRC

To capture the comprehensive molecular landscape of the tumour microenvironment (TME) at the invasive front, GeoMx Whole Transcriptome Atlas (WTA) spatial transcriptomic profiling was performed. The GeoMx Digital Spatial Profiler (DSP) enables simultaneous and guided detection of thousands of RNA targets from user-defined regions of interest (ROIs) on FFPE tissue, using barcoded in situ hybridisation probes conjugated to UV-photocleavable oligonucleotides (Van & Blank, 2019). The GeoMx WTA covers over 18,000 protein-coding genes with high sensitivity and specificity, demonstrating strong concordance with RNA-seq and RNAscope even in small ROI sizes. Previous studies have demonstrated that the integration of multiplex imaging with spatial transcriptomic profiling provides a synergistic strategy: mIF delivers single-cell spatial resolution and phenotypic characterization, while GeoMx delivers comprehensive whole-transcriptome profiling within defined tissue compartments, effectively linking cellular phenotypes with molecular signatures (Bungaro et al., 2025).

6.1.3 Three-mask segmentation selection: CK+, αSMA+, and Rest

A critical step in the GeoMx workflow is the selection and segmentation of ROIs using fluorescently labelled visualisation markers, which enable the demarcation of distinct tissue compartments based on morphology (Van & Blank, 2019). The DSP platform supports three fluorescent visualisation markers plus a

nuclear stain, and previous studies have demonstrated that Pan-CK-based segmentation into tumour and TME segments effectively distinguishes epithelial from stromal biology. In this study, three tissue compartments were defined at the invasive front: the CK-positive (CK+) compartment to identify tumour-rich epithelial regions, the αSMA-positive (αSMA+) compartment to identify αSMA-expressing stromal cells, which include myofibroblast-like cancer-associated fibroblasts (myCAFs), vascular smooth muscle cells, and pericytes, and the Rest compartment (CK−αSMA−) to capture the remaining tissue microenvironment, including immune infiltrates and other unclassified stromal populations. This three-mask approach reflects the cellular compartments most relevant to the biology of TGPs: the tumour epithelium that defines the growth front, the αSMA-enriched stroma where myofibroblastic remodelling, vascular regulation, and potential immune exclusion are expected to occur, and the broader TME where immune and other non-epithelial populations reside.

6.1.4 EcoTyper: Cell-state and ecosystem inference from bulk-like spatial data

Since GeoMx data are acquired at the ROI level rather than at single-cell resolution, each mask captures a mixture of cell types within a defined tissue region. To identify which specific cell populations within each compartment are associated with TGP-specific spatial distances, computational deconvolution of cell states from these bulk-like expression profiles was required. For this purpose, EcoTyper was used, a machine learning framework designed for large-scale identification of transcriptionally defined cell states and multicellular communities from bulk, single-cell, and spatially resolved gene expression data (Luca et al., 2021). EcoTyper differs from related approaches in several key ways: it imputes cellular heterogeneity directly from RNA profiles of intact tissue, avoids distortions caused by physical cell isolation, does not require antibodies or preselection of phenotypic markers, and can be applied to fresh, frozen, and fixed samples (Luca et al., 2021). Additionally, EcoTyper goes beyond simple cell-type abundance estimation by identifying transcriptional states within each cell type and discovering co-occurrence patterns that define multicellular communities called “ecotypes”.

The choice of EcoTyper over alternative approaches is driven by the nature of GeoMx data. Spatial domain identification tools such as BayesSpace or SpaGCN

are designed for spot-level spatial transcriptomics platforms with regular spatial grids and are not directly applicable to GeoMx's irregularly shaped, user-defined ROIs (Liu et al., 2026). Similarly, single-cell deconvolution methods such as CIBERSORTx or MuSiC estimate cell-type proportions but do not resolve transcriptional states within each cell type, whereas EcoTyper identifies distinct cell states and their co-occurrence patterns, providing a finer-grained view of the cellular composition within each mask. In the original EcoTyper framework, 69 transcriptionally defined cell states were identified across 12 major cell lineages in 16 types of human carcinoma, and these cell states were shown to be recoverable across independent datasets and platforms (Luca et al., 2021). By applying EcoTyper in recovery mode to GeoMx WTA data from the three tissue compartments, this study aimed to estimate cell-state abundances within each mask and correlate them with the mIF-derived spatial distance metrics, thereby linking TGP-specific spatial arrangements to specific cell populations at the invasive front.

6.1.5 LIANA: Cell-Cell communication inference

In addition to identifying cell states, understanding the ligand-receptor (LR) communication that mediates crosstalk between tissue compartments is essential for decoding the molecular mechanisms underlying TGP-associated spatial differences. To obtain a curated set of LR interaction pairs, the LIANA Consensus resource was used (Dimitrov et al., 2022). LIANA systematically compared 16 cell-cell communication resources and demonstrated that individual resources vary considerably in their coverage of specific pathways and tissue-enriched proteins, with limited overlap between them (Dimitrov et al., 2022). The LIANA Consensus resource addresses this by integrating interactions from multiple curated databases, including CellPhoneDB, CellChatDB, and OmniPath, thereby providing a more comprehensive and less biased catalogue of known LR pairs than any single resource alone. In this study, LR pairs extracted from the LIANA Consensus resource were combined with EcoTyper-inferred cell-state abundances and GeoMx mask-level expression data to compute weighted interaction scores, which were then correlated with the mIF-derived spatial distance metrics. This approach provided a molecular signalling layer to the spatial associations, linking specific LR interactions from defined sender cell states to the TGP-specific spatial arrangements observed at the invasive front in primary CRC.

6.2 Results

6.2.1 Cohort characteristics

Table 6-1 Clinicopathological characteristics of the 51 patients stratified by TGP

Characteristic	N	Overall N = 51[1]	Infiltrative N = 31[1]	Pushing N = 20[1]	p-value[2]
Age	51				0.617
<50		1 (2.0%)	0 (0%)	1 (5.0%)	
51-75		32 (63%)	20 (65%)	12 (60%)	
>75		18 (35%)	11 (35%)	7 (35%)	
Sex	51				0.778
Female		22 (43%)	14 (45%)	8 (40%)	
Male		29 (57%)	17 (55%)	12 (60%)	
T stage	51				**0.003**
pT1		2 (3.9%)	0 (0%)	2 (10%)	
pT2		4 (7.8%)	0 (0%)	4 (20%)	
pT3		28 (55%)	17 (55%)	11 (55%)	
pT4		17 (33%)	14 (45%)	3 (15%)	
Differentiation	50				0.123
Well or Moderate		42 (84%)	23 (77%)	19 (95%)	
Poor		8 (16%)	7 (23%)	1 (5.0%)	
N stage	51				**0.005**
pN0		28 (55%)	12 (39%)	16 (80%)	
pN1		16 (31%)	12 (39%)	4 (20%)	
pN2		7 (14%)	7 (23%)	0 (0%)	
Margin involvement	51				-
Absent		51 (100%)	31 (100%)	20 (100%)	
Present		0 (0%)	0 (0%)	0 (0%)	
Tumour budding	50				0.079
Absent		29 (58%)	14 (47%)	15 (75%)	
Present		21 (42%)	16 (53%)	5 (25%)	

[1]n (%)

[2]Fisher's exact test

Note: Fisher’s exact test was used for comparisons. Margin involvement was not tested because all 51 patients had absent margin involvement.

The GeoMx analysis comprised 51 patients, including 31 infiltrative and 20 pushing tumours. The clinicopathological characteristics are summarised in Table 6-1. T stage and N stage distributions differed significantly between the TGPs ($p = 0.003$ and $p = 0.005$, respectively). Infiltrative tumours had a higher proportion of pT4 than pushing tumours (45% and 15%, respectively) and were more frequently node-positive. In contrast, age, sex, tumour differentiation and tumour budding did not differ significantly between the two TGPs. All patients had absent margin involvement, therefore, margin status was not statistically compared.

6.2.2 GeoMx mask-based segmentation and quality control

GeoMx Digital Spatial Profiling was performed on tissue sections from the invasive margin of colorectal cancer specimens, using three fluorescence-guided masks to capture distinct tissue compartments: Tumour (Pan-CK-positive) identifies epithelial tumour cells, αSMA (αSMA-positive) identifies stromal cells, and Rest (remaining tissue which is Pan-CK-negative and αSMA-negative). Representative fluorescence images of a sample with pushing and infiltrative growth patterns showing the three GeoMx segmentation masks at the invasive margin. Tumour mask (Pan-CK-positive, green), αSMA mask (αSMA-positive, red), and Rest mask (Pan-CK-negative and αSMA-negative, blue) captures the remaining tissue, including immune cell populations (Figure 6-1 and Figure 6-2)

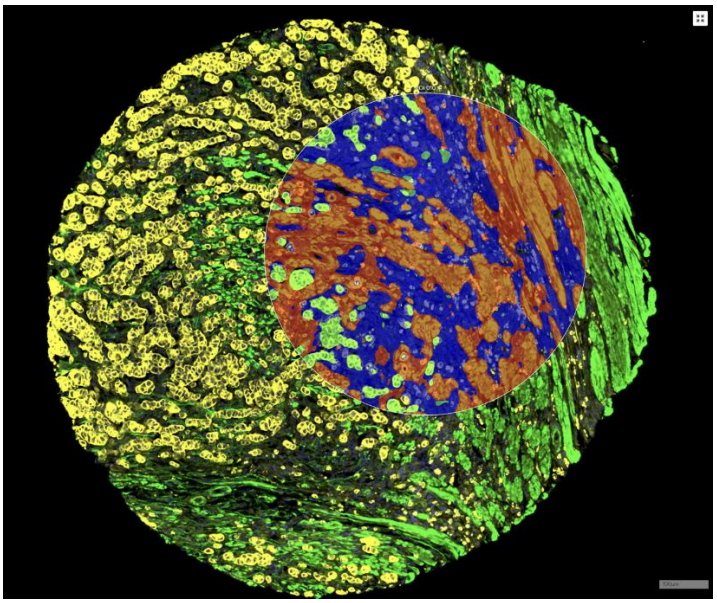


Figure 6-1: Representative infiltrative growth pattern image of a GeoMx region of interest (ROI) at the invasive margin of colorectal cancer. Three masks were applied: Tumour (Pan-CK+, green), αSMA (αSMA+, red), and Rest (double-negative, blue). Scale bar = 100 µm.

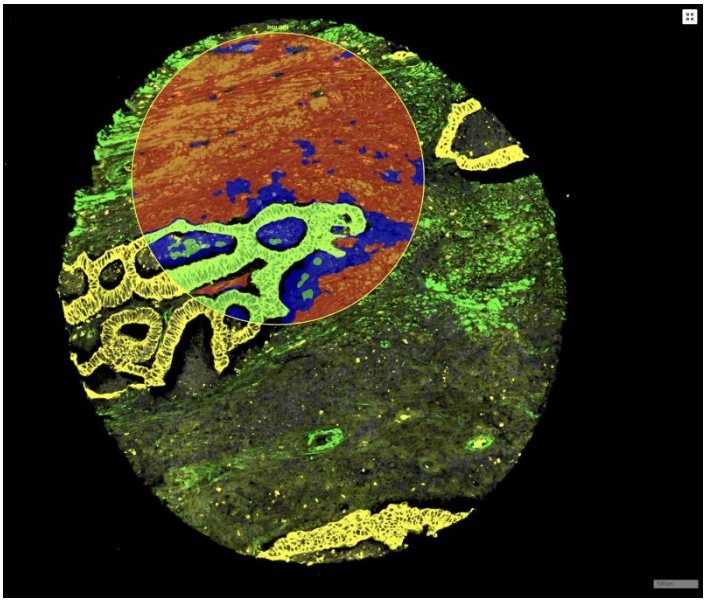


Figure 6-2 Representative pushing growth pattern image of a GeoMx region of interest (ROI) at the invasive margin of colorectal cancer. Three masks were applied: Tumour (Pan-CK+, green), αSMA (αSMA+, red), and Rest (double-negative, blue). Scale bar = 100 μm.

Following data acquisition, quality control was performed as described in method section. Nuclei counts and raw sequencing reads were evaluated across all three masks (Figure 6-3). Tumour mask showed the highest in both nuclei counts and raw reads, while Rest mask showed the lowest. Grey lines connecting matched AOIs from the same specimen across masks. Following quality control, the LOQ+Q3-normalised expression matrix was further processed for downstream analysis.

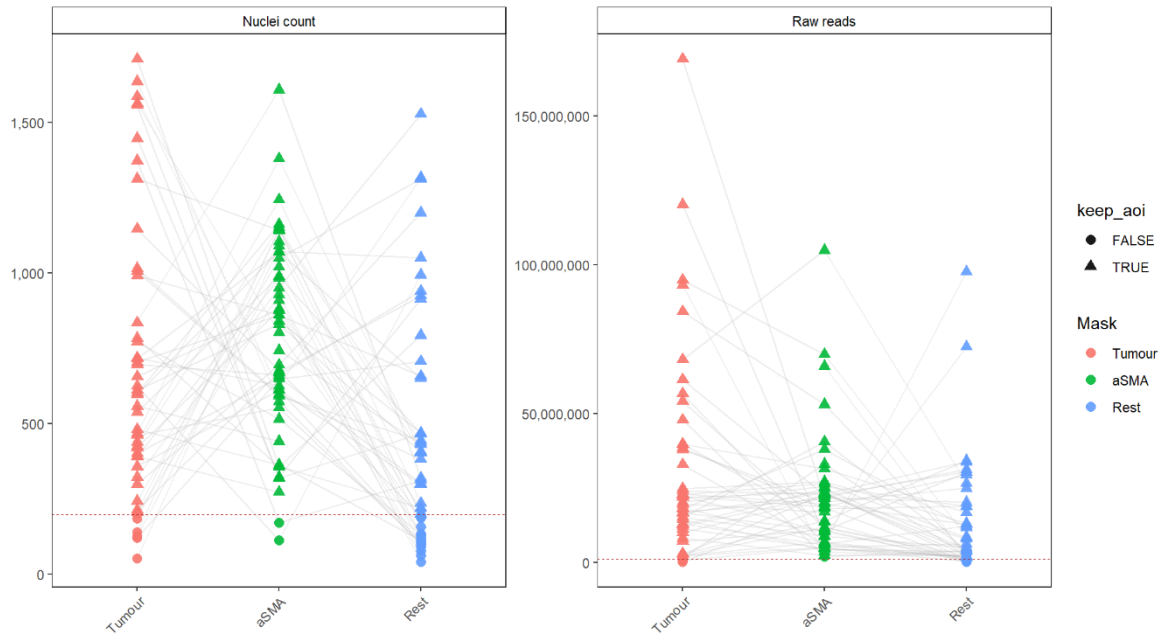


Figure 6-3 Quality control metrics separated by masks. Nuclei counts (left) and raw sequencing reads (right) for each AOI, grouped by mask: Tumour (red), αSMA (green), and Rest (blue). Triangles indicate AOIs passing QC; circles indicate excluded AOIs. Grey lines connect matched AOIs from the same specimen across masks. Dashed red lines indicate minimum QC thresholds.

6.2.3 Transcriptomic separation of tissue compartments

A principal component analysis was performed on the retained AOIs to gain an overview of the transcriptomics variation across compartments and TGPs (Figure 6-4). The top 2,000 most variable genes, ranked by row variance, were selected and PCA was computed. AOIs were plotted by the first two principal components and annotated by compartments and TGPs. PC1 showed 36.2% of the total variance and clearly separated tumour mask from αSMA and Rest masks, supporting fluorescence-guided captured different biological tissue areas. PC2 explained 9.1% of the variance but did not clearly separate the αSMA and Rest masks. No clear separation between infiltrative and pushing growth patterns was observed, suggesting that growth pattern-associated differences are not driven by large-scale transcriptomic shifts.

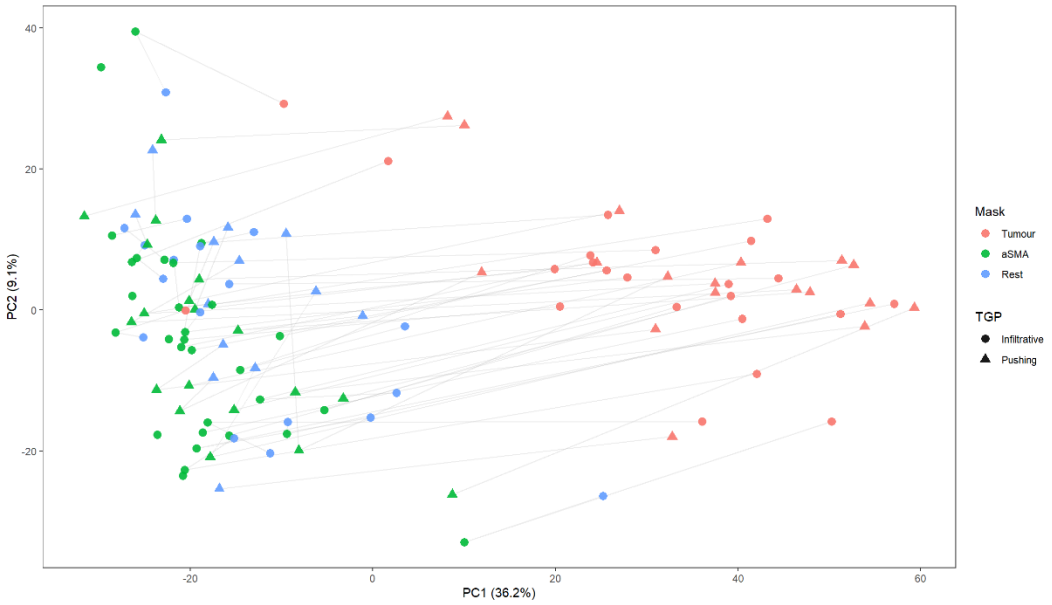


Figure 6-4 Principal compartment analysis of GeoMx AOIs. AOIs are coloured by mask: Tumour (red), αSMA (green), and Rest (blue). Shapes indicate tumour growth pattern: circles = infiltrative, triangles = pushing. Grey lines connect AOIs from the same specimen.

To confirm that each mask captured biologically distinct tissue compartments, the top enriched genes per mask were identified by comparing in-segment versus out-of-segment mean expression and visualised as a row-scaled z-score heatmap (Figure 6-5).

Genes enriched in the Tumour mask included canonical epithelial markers such as KRT8, EPCAM, CDH1, CEACAM5, and CEACAM1, as well as claudin family members CLDN4 and CLDN7. The presence of intestinal differentiation markers TFF3, KLF5, and PHGR1 further supported that the Tumour mask specifically captured colorectal epithelial cells rather than epithelial contamination from other tissue types. Genes enriched in the αSMA mask included smooth muscle and myofibroblast markers ACTA2, MYL9, and TAGLN, alongside extracellular matrix components COL1A1, COL1A2, COL3A1, COL6A2, BGN, POSTN, and SPARC. This expression profile is consistent with a myofibroblast-like cancer-associated fibroblast (myCAF) identity. Genes enriched in the Rest mask were dominated by immunoglobulin genes, including IGHG1–4, IGHA1, IGKC, IGLL5, and JCHAIN, indicating a strong B cell and plasma cell presence, alongside leukocyte-associated genes LCP1, RGS1, IFI30, and APOE. This heterogeneous profile reflects the broader tumour microenvironment beyond the epithelial and stromal compartments, including both adaptive and innate immune populations not

captured by either the Tumour or αSMA masks. Notably, no clear expression differences between infiltrative and pushing growth patterns were observed within any mask, consistent with the PCA findings.

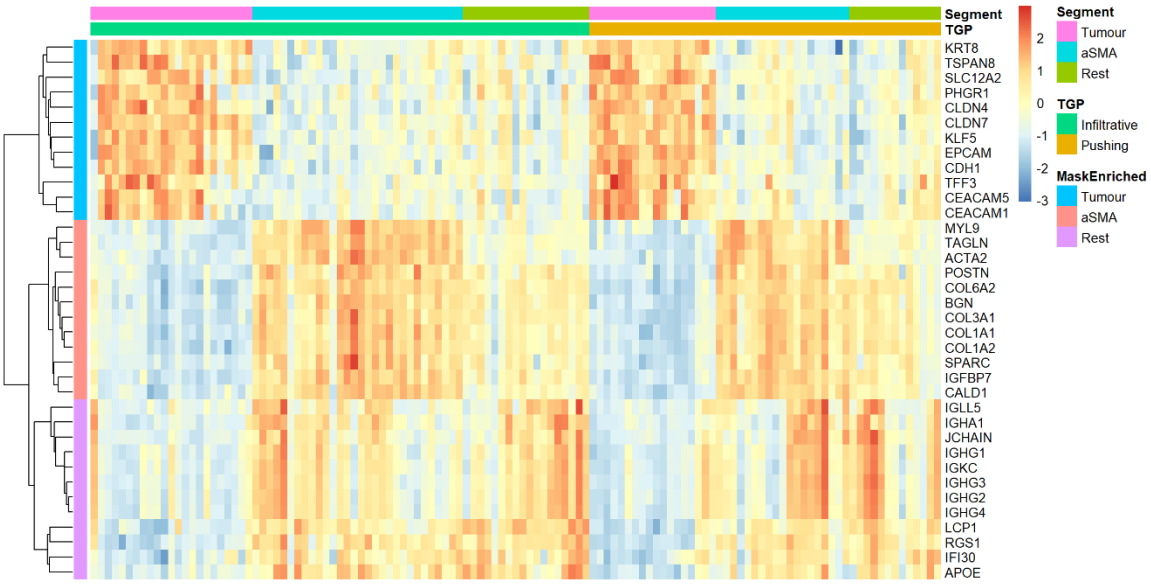


Figure 6-5 Heatmap of top enriched genes separated by masks and tumour growth patterns. Rows (genes) are grouped by the mask in which they are enriched: Tumour (cyan), αSMA (pink), and Rest (purple), as indicated by the colour bar on the left. Columns (AOIs) are ordered first by segment (Tumour, αSMA, Rest; top colour bar) and then by tumour growth pattern (infiltrative, green; pushing, orange; second colour bar). Red indicating higher and blue indicating lower expression.

6.2.4 Differential gene expression between spatial compartments

Differential gene expression analysis was performed to assess the transcriptomic differences. All pairwise comparisons between three compartments were assessed. Infiltrative and pushing growth patterns were compared. Results were visualised in volcano plots (Figure 6-6, Figure 6-7, and Figure 6-8)

The comparison between Tumour and αSMA masks revealed the largest number of differentially expressed genes (Figure 6-6). Genes upregulated in the Tumour mask included epithelial markers KRT8, EPCAM, CEACAM5, CLDN7, SLC12A2, and ELF3, while genes upregulated in the αSMA mask included stromal and myofibroblast markers TAGLN, ACTA2, IGFBP7, COL1A1, VIM, CALD1, and AEBP1. This strong separation confirms that the Tumour and αSMA masks captured different tissue biology.

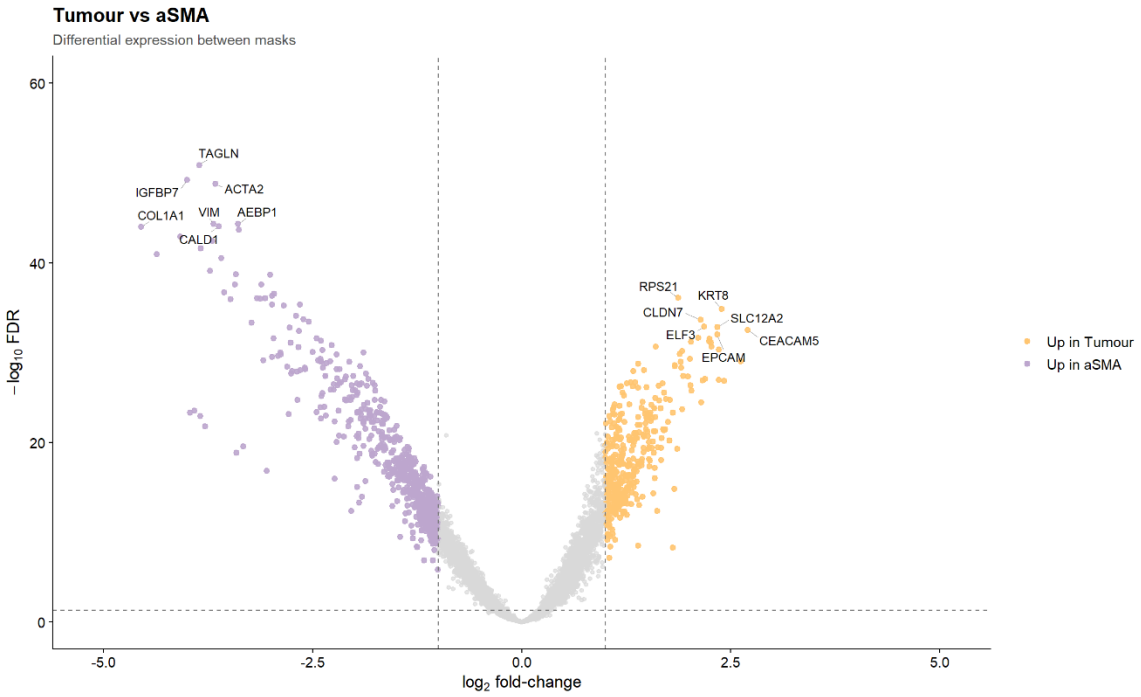


Figure 6-6 Differential gene expression between Tumour and αSMA masks. The X-axis shows log2fold-change and the Y-axis shows -log10FDR. Yellow points show genes significantly upregulated in Tumour mask, and purple points show genes significantly upregulated in αSMA mask. Dash lines indicate significant thresholds (FDR < 0.05 and |log2fold-change| ≥ 1). The top significant genes are labelled.

The comparison between Tumour and Rest masks showed a similar pattern, with the same epithelial markers upregulated in the Tumour mask (KRT8, CLDN7, SLC12A2, SOX9) (Figure 6-7). Genes upregulated in the Rest mask included VIM, RGS1, IGFBP7, COL3A1, COL6A2, LAPTM5, and A2M, reflecting a mixture of stromal and immune-related gene expression consistent with the heterogeneous composition of the Rest compartment.

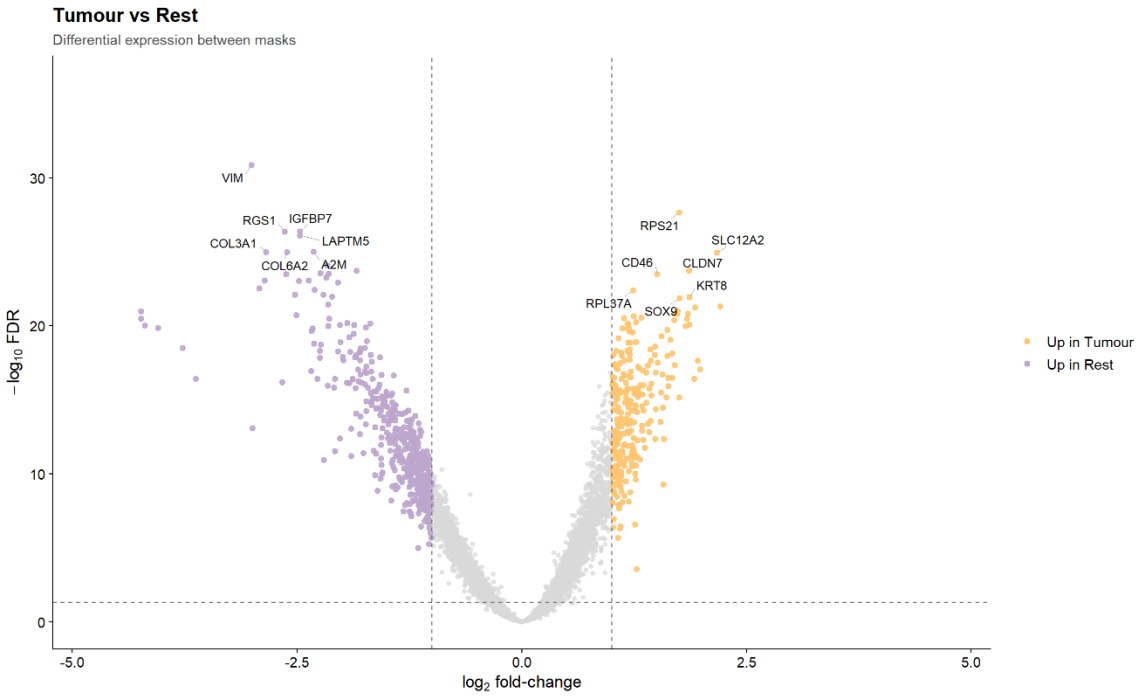


Figure 6-7 Differential gene expression between Tumour and Rest masks. The X-axis shows log2fold-change, and the Y-axis shows -log10FDR. Yellow points show genes significantly upregulated in Tumour mask, and purple points show genes significantly upregulated in Rest mask. Dash lines indicate significant thresholds (FDR < 0.05 and |log2fold-change| ≥ 1). The top significant genes are labeled.

Finally, the comparison between αSMA and Rest masks showed that all significant hits were upregulated in the αSMA mask (Figure 6-8). These included smooth muscle contractile genes ACTA2, TAGLN, MYL9, ACTG2, TPM2, MYLK, and CALD1. No genes were significantly upregulated in the Rest mask. This is consistent with the PCA results, where the αSMA and Rest masks were not clearly separated along PC2, and reflects the expected transcriptomic overlap between stromal and immune compartments relative to the more distinct epithelial tumour compartment.

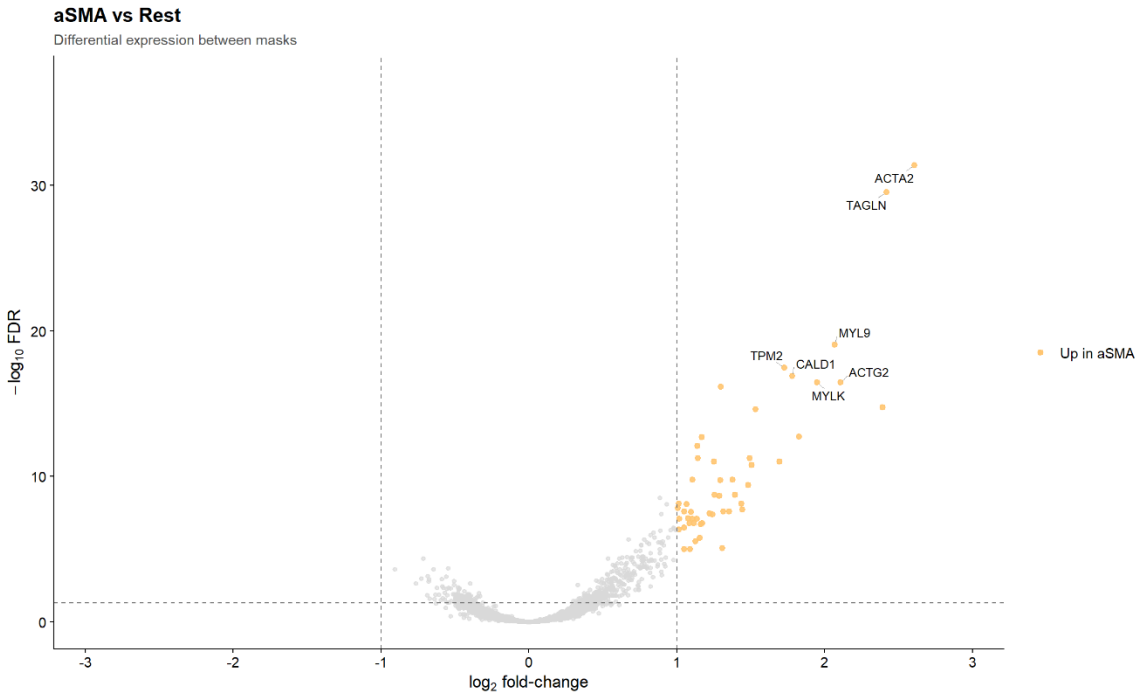


Figure 6-8 Differential gene expression between αSMA and Rest masks. The X-axis shows log2fold-change, and the Y-axis shows -log10FDR. Yellow points show genes significantly upregulated in αSMA mask. Dash lines indicate significant thresholds (FDR < 0.05 and |log2fold-change| ≥ 1). The top significant genes are labelled.

To assess differential gene expressions between infiltrative and pushing growth patterns, analyses were performed within each mask. The results revealed that no genes reached significance in any of the three masks, indicating that TGP differences are not detectable from the gene expression level within individual compartments. This finding is consistent with PCA and heatmap results.

6.2.5 HALLMARK Pathway-level characterisation of spatial compartments

To characterise the biological programs distinguishing each spatial compartment at the pathway level, single-sample gene set enrichment analysis (ssGSEA) was performed using the GSVA package with MSigDB Hallmark gene sets. Only pathways with FDR < 0.05 are shown (Figure 6-9, Figure 6-10, and Figure 6-11).

The Tumour mask was consistently enriched for proliferation and metabolic pathways across both comparisons. MYC targets (V1 and V2), E2F targets, G2M checkpoint, oxidative phosphorylation, and cholesterol homeostasis were among

the top upregulated pathways in the Tumour mask relative to both αSMA and Rest, reflecting the high proliferative activity and metabolic demand of tumour epithelial cells (Figure 6-9 and Figure 6-10).

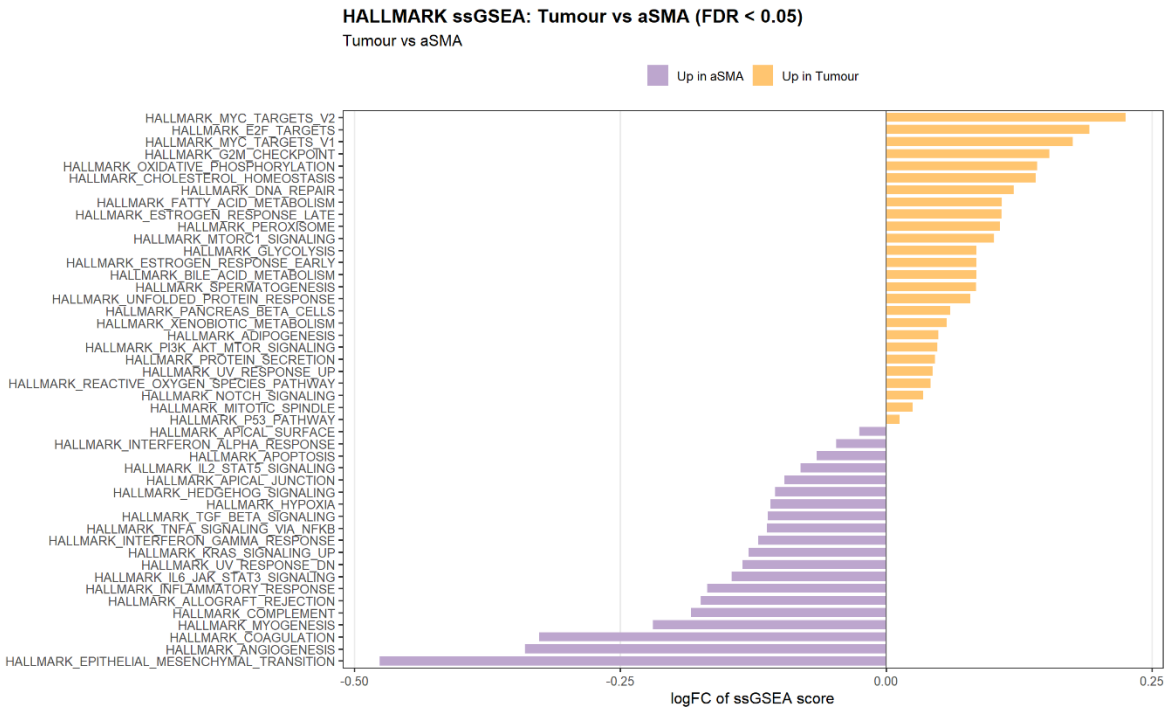


Figure 6-9 ssGSEA HALLMARK pathway between Tumour and αSMA masks. Bar plots showing log2 fold-change of ssGSEA enrichment scores for Tumour vs αSMA. Yellow bars indicate pathways enriched in the Tumour mask; purple bars indicate pathways enriched in the αSMA mask. Only pathways with FDR < 0.05 are shown.

The αSMA mask showed strong enrichment for stromal remodelling programs (Figure 6-9 and Figure 6-10). Epithelial-mesenchymal transition was the most enriched pathway in the αSMA mask in both comparisons, consistent with the myofibroblast-like identity observed in the DGE and heatmap analyses. Myogenesis, coagulation, angiogenesis, and TGF-β signalling were also upregulated, reflecting active stromal remodelling and fibrotic signalling within the αSMA-positive compartment.

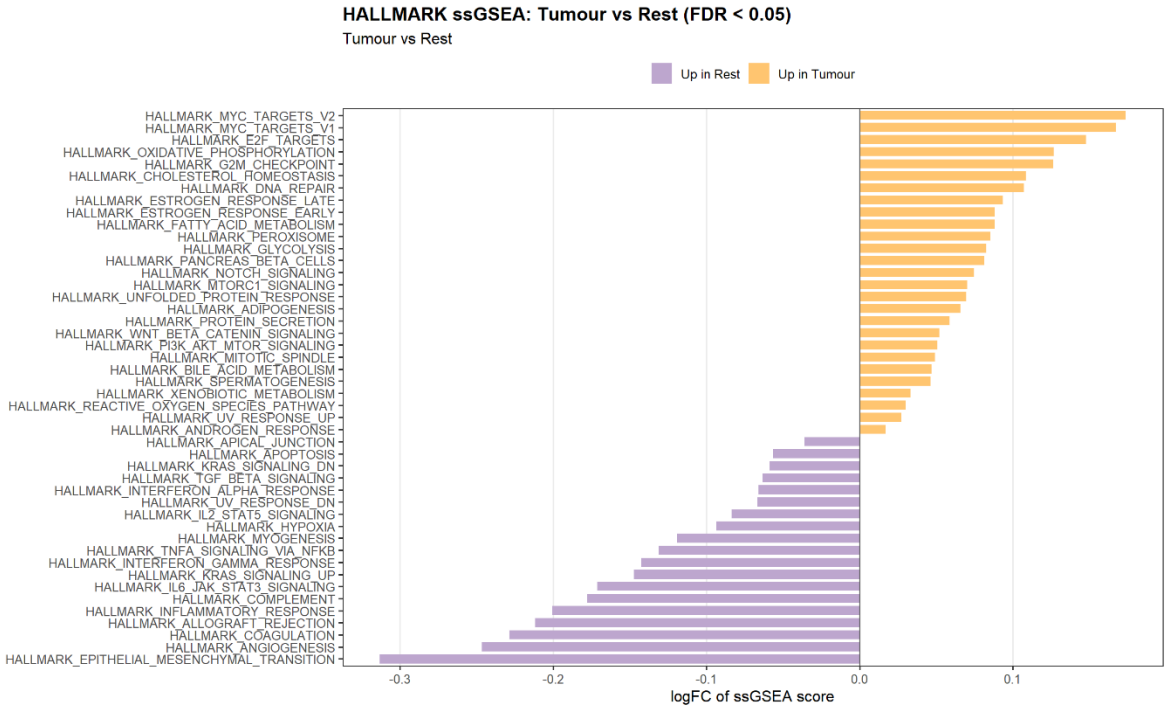


Figure 6-10 ssGSEA HALLMARK pathway between Tumour and Rest masks. Bar plots showing log2 fold-change of ssGSEA enrichment scores for Tumour vs Rest. Yellow bars indicate pathways enriched in the Tumour mask; purple bars indicate pathways enriched in the Rest mask. Only pathways with FDR < 0.05 are shown.

The Rest mask showed enrichment for immune and inflammatory pathways relative to the Tumour mask, including interferon gamma response, IL6-JAK-STAT3 signalling, inflammatory response, complement, allograft rejection, and TNF-α signalling via NF-κB, consistent with the immune signatures identified in the mask-enriched gene analysis (Figure 6-10). Relative to the αSMA mask, the Rest compartment showed higher enrichment for proliferation-related pathways (MYC targets, E2F targets) and metabolic programs, likely reflecting the proliferative activity of immune cell populations within this compartment (Figure 6-11).

The comparison between αSMA and Rest revealed that the αSMA mask was distinguished by EMT, myogenesis, angiogenesis, hedgehog signalling, and WNT-beta-catenin signalling, whereas the Rest mask was characterised by cell cycle and immune-related pathways (Figure 6-11). This supports the biological rationale for separating these two non-epithelial compartments into distinct masks.

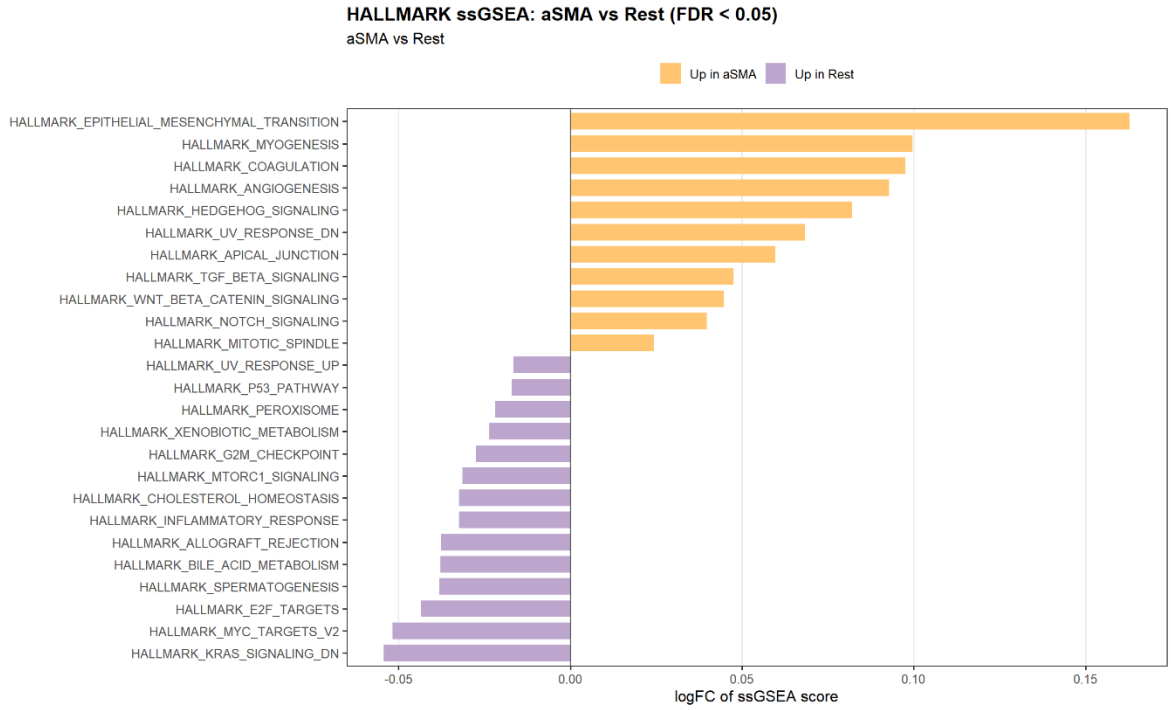


Figure 6-11 ssGSEA HALLMARK pathway between αSMA and Rest masks. Bar plots showing log2 fold-change of ssGSEA enrichment scores for αSMA vs Rest. Yellow bars indicate pathways enriched in the αSMA mask; purple bars indicate pathways enriched in the Rest mask. Only pathways with FDR < 0.05 are shown.

For the ssGSEA scores comparing TGPs, similar to the differential gene expression results, no pathways reached significance (FDR < 0.05) in any of the three masks, emphasising that TGP differences are not detectable at the pathway level.

6.2.6 Association between HALLMARK pathway enrichment and spatial distance metrics

In the previous chapter, multiplex immunofluorescence analysis identified five NDI endpoints that differed significantly between infiltrative and pushing growth patterns: αSMA+FAP+ to CD3, αSMA+FAP+ to Tumour, CD3 to Tumour, FAP+ to Tumour, and FAP+ to CD3. To explore whether these spatial differences are associated with specific biological programs captured by spatial transcriptomics, a pathway-NDI correlation analysis was performed using Hallmark gene sets.

For all spatial metrics NDI, associations between HALLMARK ssGSEA pathway scores and NDI were assessed across Tumour, Rest, and αSMA masks.

Higher NDI values indicate greater distance, whereas lower NDI values indicate closer proximity.

Figure 6-12 presents pathways associated with CD3 to Tumour NDI. The strongest association was in the tumour mask, where several pathways correlated with low NDI (FDR < 0.05), whereas the αSMA and Rest masks showed no significant correlations between CD3 to tumour NDI and HALLMARK ssGSEA pathway scores, suggesting that the molecular mechanisms correlating with the short NDI are primarily localised within the tumour nest itself, rather than the surrounding stroma (αSMA or Rest masks). Notably, the immune and stromal-related pathways, TNFA SIGNALING VIA NFKB, KRAS SIGNALING UP, INFLAMMATORY RESPONSE, ALLOGRAFT REJECTION, HYPOXIA, and EPITHELIAL MESENCHYMAL TRANSITION (EMT) were significantly negatively correlated with NDI (Spearman R < 0, FDR < 0.05), indicating higher pathway activity in samples with shorter CD3 to tumour distances. Conversely, the proliferation-related pathway, MYC TARGETS V1, showed a significant positive correlation with NDI (Spearman R > 0, FDR < 0.05), suggesting that high NDI is associated with cellular proliferation and tumour bulk expansion.

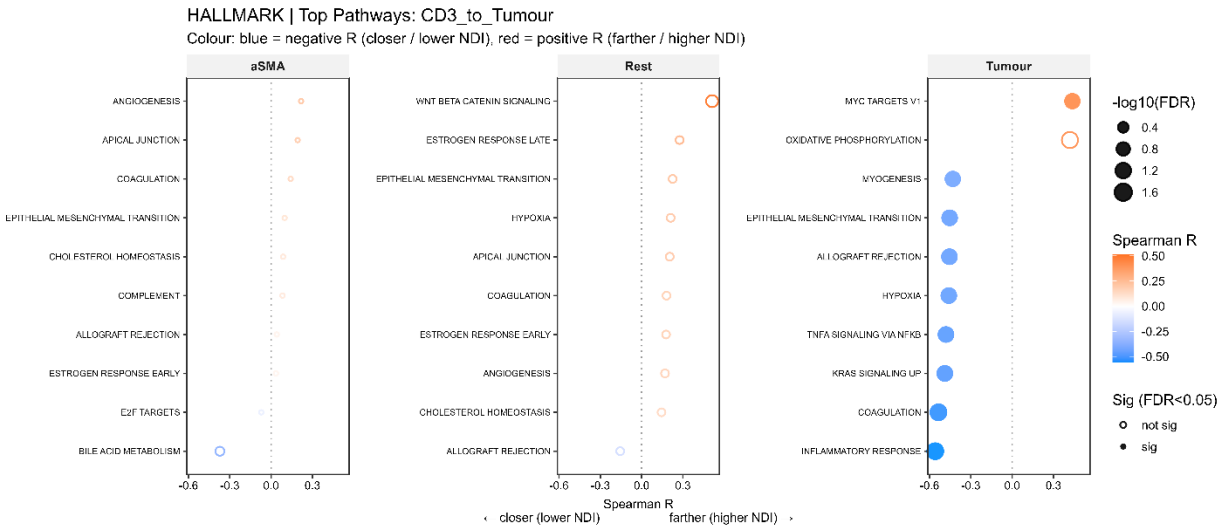


Figure 6-12 HALLMARK pathway correlations with CD3 to Tumour spatial proximity across masks. Dot plots showing Spearman correlation between ssGSEA Hallmark pathway scores and CD3 to Tumour NDI, separated by mask (aSMA, Rest, Tumour). Colour indicates direction and strength of correlation (red = positive, blue = negative). Dot size represents −log10 FDR. Filled dots indicate FDR < 0.05; open dots indicate non-significant associations. Top 10 pathways per mask are shown, ranked by FDR.

As illustrated in Figure 6-13, the analysis of the HALLMARK pathway and FAP+ to tumour NDI correlation demonstrated that a higher NDI was associated with the proliferative pathways, including E2F TARGETS, MYC TARGETS V2, and G2M CHECKPOINT, in the tumour compartment (Spearman R > 0, FDR < 0.05). This evidence is consistent with the pushing growth pattern, in which the tumour primarily grows locally. Interestingly, the shorter NDI was associated with EMT and MYOGENESIS pathways related to cellular plasticity and cell motility (Spearman R < 0, FDR < 0.05), supporting the infiltrative growth pattern in which FAP fibroblasts close to cancer cells are associated with EMT activation, thereby enhancing cancer cell invasion.

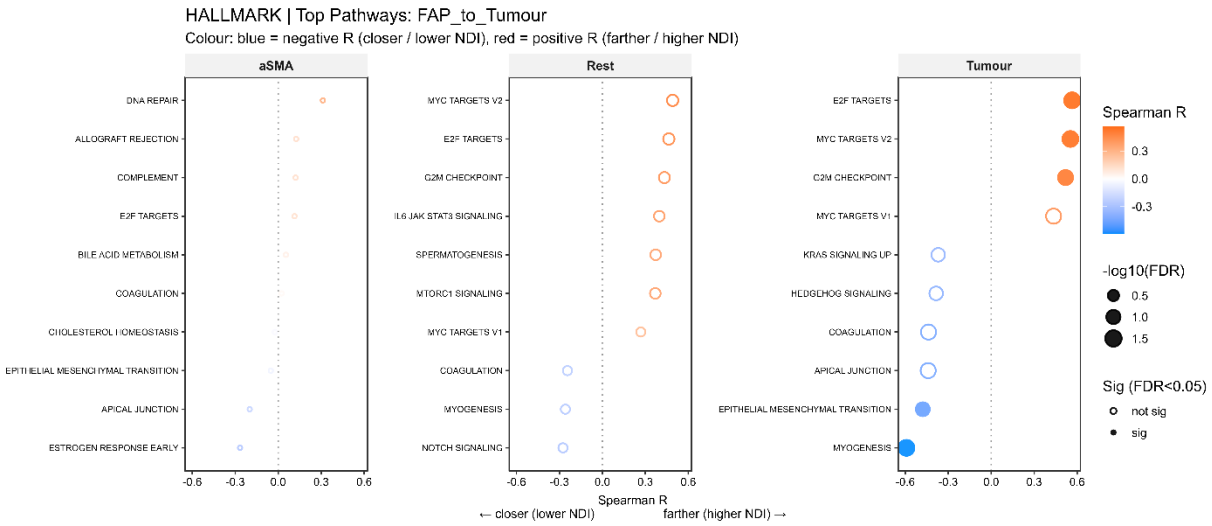


Figure 6-13 HALLMARK pathway correlations with FAP to Tumour spatial proximity across masks. Dot plots showing Spearman correlation between ssGSEA Hallmark pathway scores and FAP to Tumour NDI, separated by mask (aSMA, Rest, Tumour). Colour indicates direction and strength of correlation (red = positive, blue = negative). Dot size represents −log10 FDR. Filled dots indicate FDR < 0.05; open dots indicate non-significant associations. Top 10 pathways per mask are shown, ranked by FDR.

Similar to FAP+ to tumour NDI, the correlation between closer αSMA+FAP+ to tumour NDI and HALLMARK pathways demonstrated statistically significant pathways in the tumour compartment, which are HEDGEHOG SIGNALING, and APICAL JUNCTION (Figure 6-14) (Spearman R < 0, FDR < 0.05). Given that a shorter NDI is characteristic of an infiltrative phenotype, these findings suggest a link between these molecular signatures and tumour invasion. Notably, no statistically significant Hallmark pathways were identified in the αSMA and Rest compartments. This indicates that the molecular activities associated with this

spatial organisation are likely intrinsic to cancer cells rather than to the stromal microenvironment.

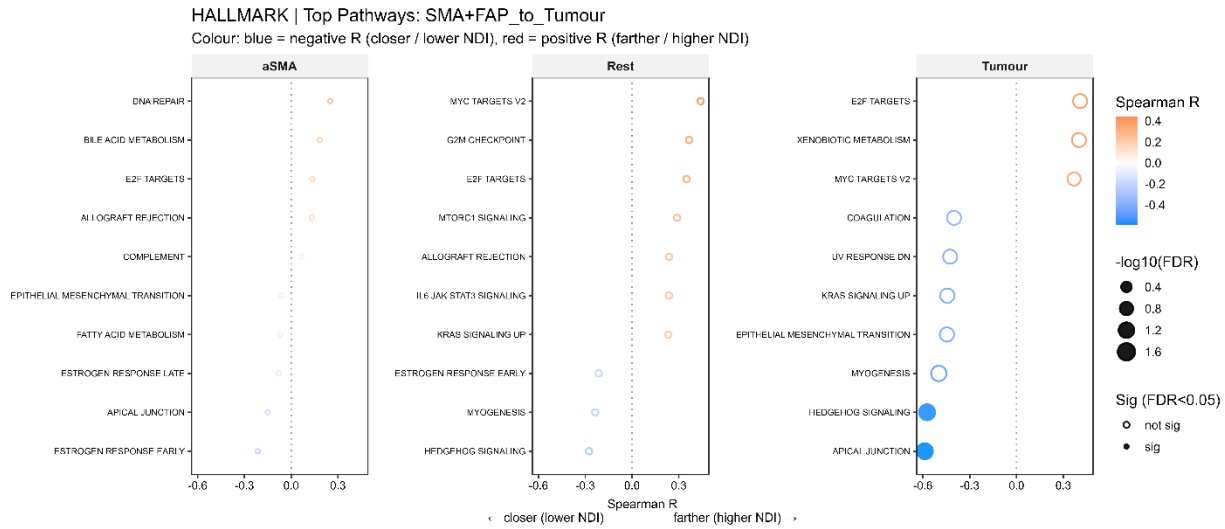


Figure 6-14 HALLMARK pathway correlations with αSMA+FAP+ to Tumour spatial proximity across masks. Dot plots showing Spearman correlation between ssGSEA Hallmark pathway scores and αSMA+FAP+ to Tumour NDI, separated by mask (αSMA, Rest, Tumour). Colour indicates direction and strength of correlation (red = positive, blue = negative). Dot size represents −log10 FDR. Filled dots indicate FDR < 0.05; open dots indicate non-significant associations. Top 10 pathways per mask are shown, ranked by FDR.

Regarding the correlation between FAP+ to CD3 NDI and Hallmark pathways, there were no statistically significant correlations in stromal compartments (αSMA and Rest), while the tumour compartment exhibited the correlation enrichment of immune and stress-related pathways with spatial NDI (Figure 6-15). Specifically, positive statistically significant correlations were found in inflammatory signalling pathways, including TNFA SIGNALING VIA NFKB, INFLAMMATORY RESPONSE, IL6 JAK STAT3 SIGNALING, and COMPLEMENT (Spearman R > 0, FDR < 0.05), indicating these pathways were associated with a higher NDI. Moreover, a higher FAP+ to CD3 NDI is associated with HYPOXIA, COAGULATION, and EPITHELIAL MESENCHYMAL TRANSITION (EMT). However, SPERMATOGENESIS was the only pathway showing a significant negative correlation (Spearman R < 0, FDR < 0.05).

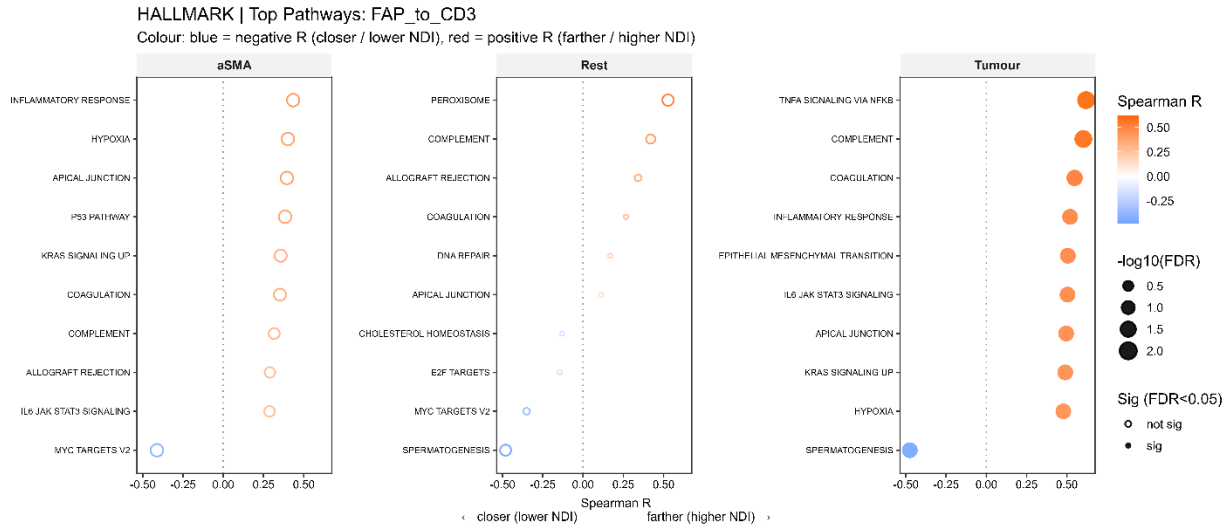


Figure 6-15 HALLMARK pathway correlations with FAP+ to CD3 spatial proximity across masks. Dot plots showing Spearman correlation between ssGSEA Hallmark pathway scores and αSMA+FAP+ to Tumour NDI, separated by mask (αSMA, Rest, Tumour). Colour indicates direction and strength of correlation (red = positive, blue = negative). Dot size represents −log10 FDR. Filled dots indicate FDR < 0.05; open dots indicate non-significant associations. Top 10 pathways per mask are shown, ranked by FDR.

Further analysis was performed to determine the correlation of Hallmark pathway and αSMA+FAP+ to CD3 NDI (Figure 6-16). In contrast to FAP+ to CD3 NDI, αSMA+FAP+ CAF subtype did not exhibit any statistically significant correlations with Hallmark pathways across the αSMA, rest, or tumour compartments (FDR > 0.05). However, a higher NDI of this endpoint, the tumour compartment displayed similar positive trends to those of FAP+ to CD3 NDI towards inflammatory signatures (COMPLEMENT, INFLAMMATORY RESPONSE, TNFA SIGNALING).

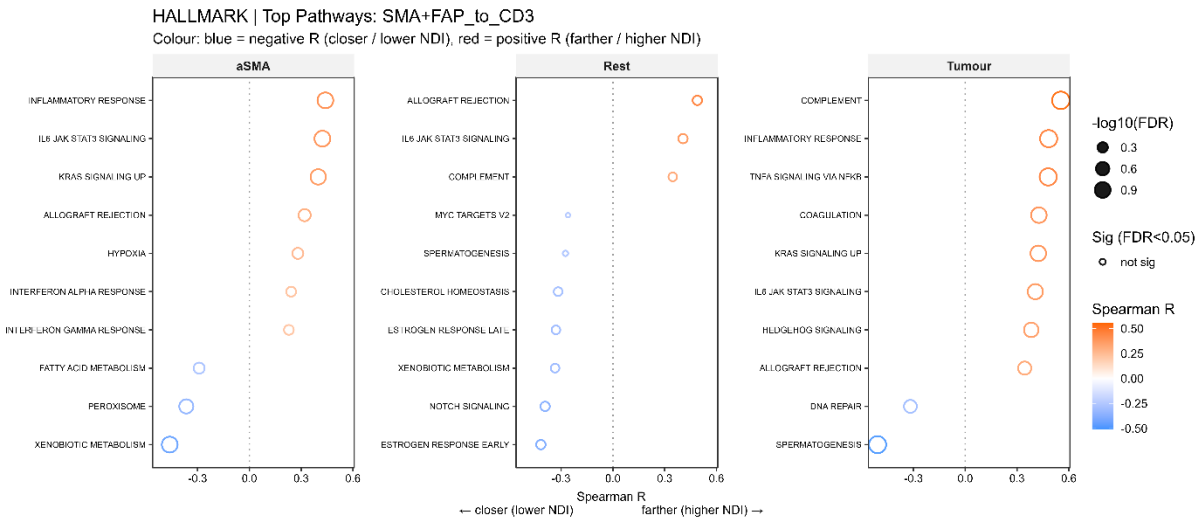


Figure 6-16 HALLMARK pathway correlations with SMA+FAP+ to CD3 spatial proximity across masks. Dot plots showing Spearman correlation between ssGSEA Hallmark pathway scores and SMA+FAP+ to Tumour NDI, separated by mask (aSMA, Rest, Tumour). Colour indicates direction and strength of correlation (red = positive, blue = negative). Dot size represents −log10 FDR. Filled dots indicate FDR < 0.05; open dots indicate non-significant associations. Top 10 pathways per mask are shown, ranked by FDR.

The pathway and NDI correlation analysis revealed biological programs associated with spatial distance but did not identify which specific cell populations within each mask correlate with spatial distance. Since each GeoMx mask captures a mixture of cell types, the observed pathway and NDI correlations might arise from different cell subpopulations. To address this, EcoTyper cell-state deconvolution was applied to estimate specific cell-state abundances within each mask, and the correlation between cell-state abundances and NDI was assessed using Spearman's rank correlation.

6.2.7 EcoTyper cell-state associations with spatial distance metrics

The results were shown using volcano plots, defined by two axes: X-axis (Spearman's Rho), which represents the direction and strength of the correlation. A positive value indicates that as the NDI increases (reflecting a greater distance), the proportional abundance of the cell state concurrently increases. Y-axis (-log10 FDR): Represents statistical significance, adjusted via the BH procedure to control the False Discovery Rate (FDR). Two significance thresholds were established: a blue dashed line indicating FDR < 0.05 and a blue dotted line indicating FDR < 0.10.

The results were stratified into three compartments: Tumour mask (representing cancer epithelial cells), αSMA mask (representing the stromal area characterised by high αSMA expression), and Rest mask (representing the remaining microenvironment, including diverse elements such as immune infiltrates, adjacent normal tissue, and unclassified stromal components).

To interpret the EcoTyper-inferred analysis that showed significant cell states results, the key gene markers and clinical implications were referred to Table 6-2 from the EcoTyper framework (Luca et al., 2021).

Table 6-2: Cell states annotation from EcoTyper framework

Cell Type	State	Annotation
B cells	S04	Activated
CD4 T cells	S03	Unknown/ Dysfunctional
CD8 T cells	S02	Late-stage differentiated effector
Endothelial cells	S01	Normal-enriched/ Angiogenic
Fibroblasts	S02	CAF2 (normal-enriched)
Monocytes/ Macrophages	S09	Unknown/ Regulatory

Firstly, the correlation between NDI metrics and enriched cell states in αSMA compartment was shown in Figure 6-17. Interestingly, the longer αSMA+FAP+ to CD3 NDI is significantly positively correlated with the abundance of Fibroblast S02 (annotated by EcoTyper as CAF2/ normal-enriched) and Endothelial cells S01 (Normal-enriched/ Angiogenic). Additionally, a significant positive correlation was observed in FAP+ to CD3 NDI with CD4 T cells S03 (Unknown/ Dysfunctional). Conversely, the wider distance between CD3 to Tumour NDI is significantly positive correlated with the abundance of CD8 T cells S02 (Late-stage differentiated effector).

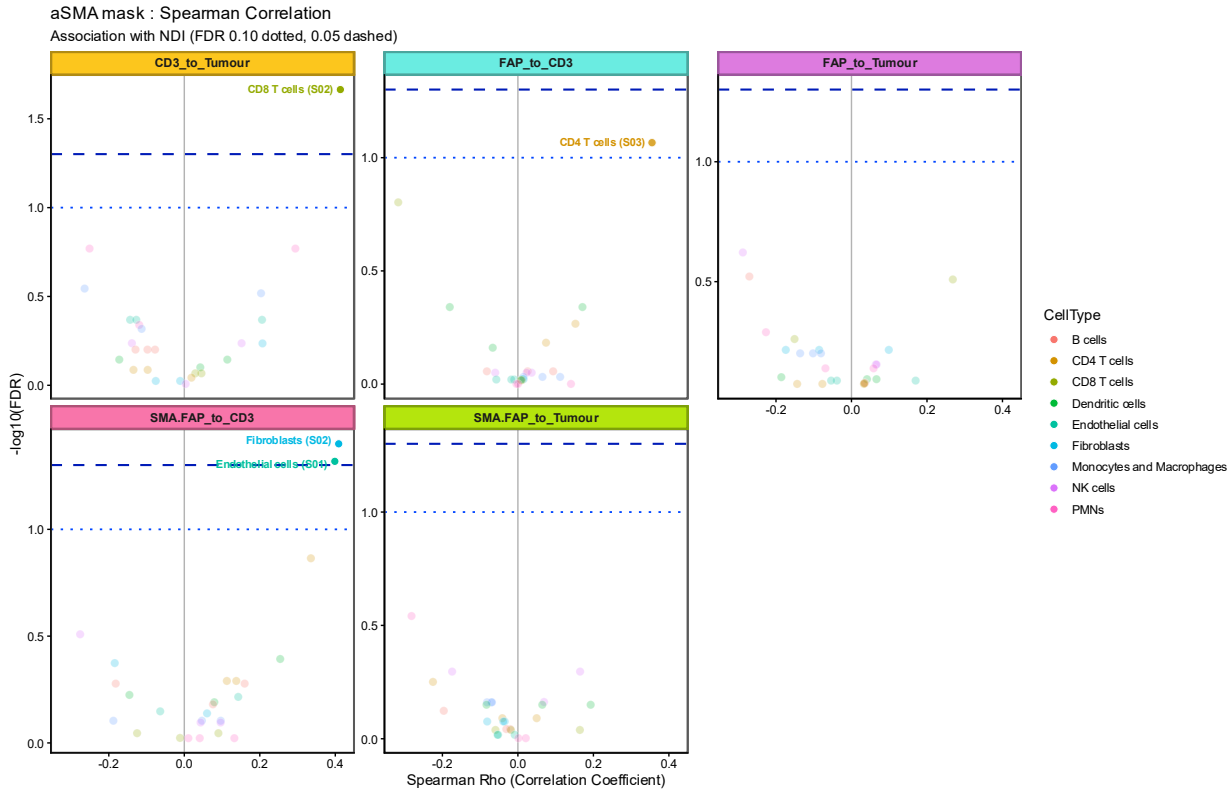


Figure 6-17 Cell state and spatial NDI correlation in aSMA compartment. Volcano plots showing the Spearman correlation between EcoTyper-inferred cell state abundance and NDI and statistical values, -log10(FDR), within the aSMA mask. Statistical thresholds are marked by horizontal lines: the dotted line indicates an FDR of 0.10, and the dashed line indicates a significance cutoff of FDR < 0.05

Within the Rest mask (Figure 6-18), different immune cell states demonstrated a significant association with specific spatial NDI. An increase of FAP+ to Tumour NDI showed a significantly positive correlation (Spearman > 0, FDR < 0.05) with the abundance of B cells S04 (Activated B cell). An increase in the αSMA+FAP+ to CD3 NDI presented a positive correlation (Spearman > 0, FDR < 0.10) with the abundance of Monocytes and Macrophages S09 (Unknown/ Regulatory).

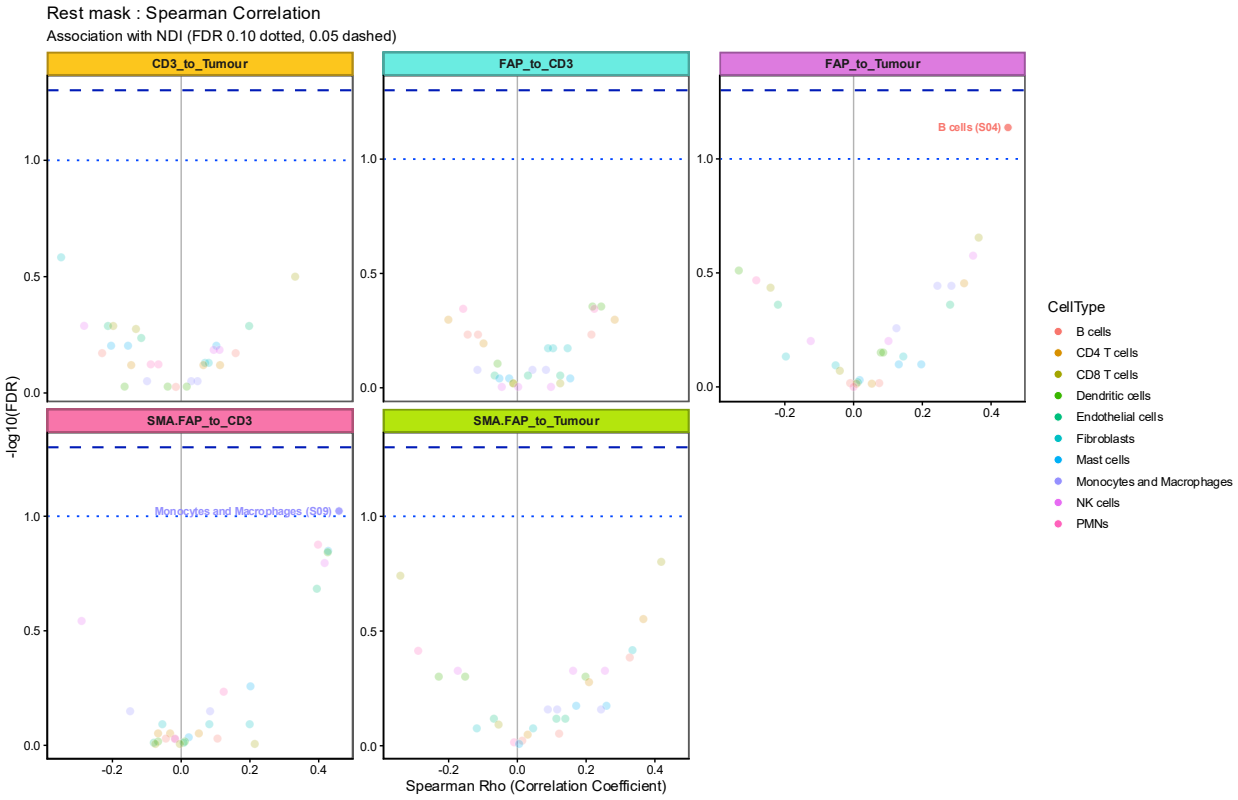


Figure 6-18 Cell state and spatial NDI correlation in the Rest compartment. Volcano plots showing the Spearman correlation between EcoTyper-inferred cell state abundance and NDI and statistical values, -log10(FDR), within the Rest mask. Statistical thresholds are marked by horizontal lines: the dotted line indicates an FDR of 0.10, and the dashed line indicates a significance cutoff of FDR < 0.05

Within the cancer epithelial compartment (Figure 6-19), the Spearman correlation analysis showed no statistically significant associations. Across all NDI metrics, no EcoTyper-inferred cell states were significant at the FDR < 0.10 threshold. This suggests that the spatial pathobiology at the invasive front of primary CRC is predominantly associated with the extrinsic microenvironmental (aMSA and Rest compartments)

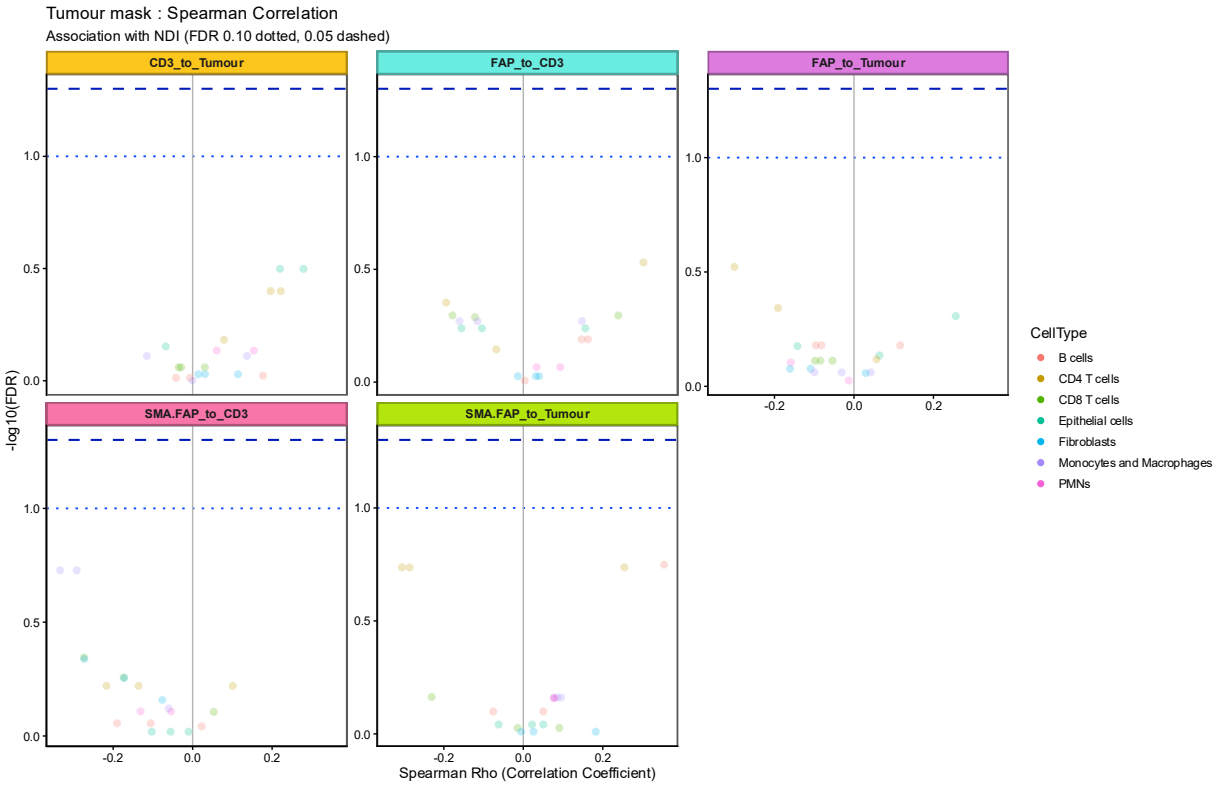


Figure 6-19 Cell state and spatial NDI correlation in the Tumour compartment. Volcano plots showing the Spearman correlation between EcoTyper-inferred cell state abundance and NDI and statistical values, -log10(FDR), within the Tumour mask. Statistical thresholds are marked by horizontal lines: the dotted line indicates an FDR of 0.10, and the dashed line indicates a significance cutoff of FDR < 0.05

6.2.8 Ligand-receptor interactions linking cell states to spatial organisation

To investigate which molecular signals from cell state abundances may be involved, a ligand-receptor association analysis was performed using the LR pairs resource from the LIANA Consensus resource (Dimitrov et al., 2022). This approach links the cell states from EcoTyper with the specific ligand-receptor interactions, providing a molecular layer to the spatial associations observed in the previous analysis.

For each spatial endpoint, sender cell states were identified from the EcoTyper analysis as those with significant positive correlations between state abundance and NDI (adjusted $p < 0.10$). The GeoMx mask, which identified each significant cell state, served as the source mask for ligand expression. The target mask was assigned based on the biological identity of the target cell population represented by the NDI. For example, for the αSMA+FAP+ to CD3 endpoint, Fibroblasts (S02) and Endothelial cells (S01) were significant in the αSMA mask, so the αSMA mask served as the source and the Rest mask as the target. Monocytes

and Macrophages (S09) were significant in the Rest mask for the same endpoint, so an additional mapping using the Rest mask as both source and target was included to capture signalling within the immune compartment. It should be noted that receptor expression was measured at the mask level, not at single-cell resolution, so that the specific cell type expressing each receptor within a given mask cannot be determined from these analyses.

Figure 6-20 illustrates the direction of signalling flow from the sender cell states to spatial NDI, and a detailed summary of all 20 LR pairs, including source and target masks, sender cell states, correlation coefficients, and FDR value, is provided in Table 6-3. Due to the filter threshold, not all significant EcoTyper cell states or NDI metrics are presented in the final LIANA network results

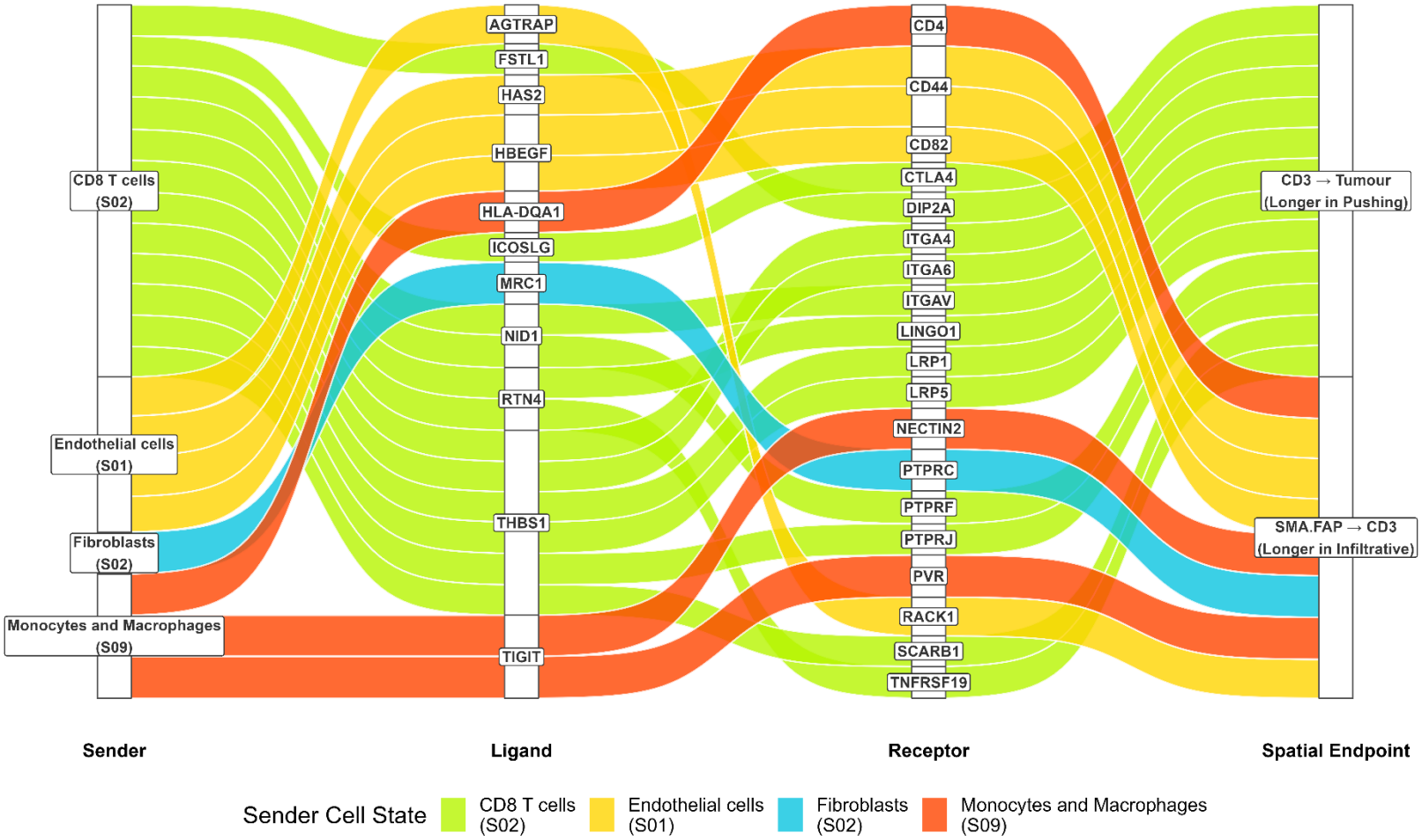


Figure 6-20: Ligand-Receptor signalling network between cell states and spatial metrics

Sankey diagram illustrating the communication between sender cells (significant cell states from EcoTyper) and their corresponding receptors associated with spatial endpoints.

Table 6-3: All 20 ligand-receptor significant pairs results

Endpoint	Source mask	Target mask	Sender cell type	Sender state	Ligand	Receptor	Spearman ρ	Linear FDR
αSMA+FAP+ to CD3	Rest	Rest	Monocytes and Macrophages	S09	TIGIT	PVR	0.60	0.048
αSMA+FAP+ to CD3	αSMA	Rest	Fibroblasts	S02	MRC1	PTPRC	0.60	0.048
αSMA+FAP+ to CD3	Rest	Rest	Monocytes and Macrophages	S09	HLA-DQA1	CD4	0.59	0.048
αSMA+FAP+ to CD3	Rest	Rest	Monocytes and Macrophages	S09	TIGIT	NECTIN2	0.59	0.048
αSMA+FAP+ to CD3	αSMA	Rest	Endothelial cells	S01	HBEGF	CD44	0.58	0.047
αSMA+FAP+ to CD3	αSMA	Rest	Endothelial cells	S01	HAS2	CD44	0.58	0.048
αSMA+FAP+ to CD3	αSMA	Rest	Endothelial cells	S01	AGTRAP	RACK1	0.56	0.048
αSMA+FAP+ to CD3	αSMA	Rest	Endothelial cells	S01	HBEGF	CD82	0.51	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	NID1	PTPRF	0.47	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	RTN4	TNFRSF19	0.46	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	THBS1	LRP5	0.45	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	THBS1	PTPRJ	0.45	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	FSTL1	DIP2A	0.45	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	THBS1	ITGA4	0.45	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	NID1	ITGAV	0.44	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	RTN4	LINGO1	0.44	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	THBS1	SCARB1	0.44	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	THBS1	ITGA6	0.43	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	THBS1	LRP1	0.43	0.048
CD3 to Tumour	αSMA	Tumour	CD8 T cells	S02	ICOSLG	CTLA4	0.43	0.039

For CD3 to Tumour NDI, which was higher in pushing growth pattern, 12 LR pairs were identified, all from CD8 T cells (S02). The most frequently occurring ligand was THBS1, engaging six receptors: LRP5 (ρ = 0.45), PTPRJ (ρ = 0.45), ITGA4 (ρ = 0.45), SCARB1 (ρ = 0.44), ITGA6 (ρ = 0.43), and LRP1 (ρ = 0.43). Additional pairs included NID1–PTPRF (ρ = 0.47), RTN4–TNFRSF19 (ρ = 0.46), NID1–ITGAV (ρ = 0.44), RTN4–LINGO1 (ρ = 0.44), FSTL1–DIP2A (ρ = 0.45), and ICOSLG–CTLA4 (ρ = 0.43). All correlations were positive, indicating that higher LR expression scores were associated with greater CD3 to Tumour spatial distance.

For αSMA+FAP+ to CD3 NDI, which was higher in infiltrative growth patterns, Eight LR pairs were identified from three sender cell states. Monocytes and Macrophages (S09) contributed TIGIT–PVR (ρ = 0.60), HLA-DQA1–CD4 (ρ = 0.59), and TIGIT–NECTIN2 (ρ = 0.59). Fibroblasts (S02) contributed MRC1–PTPRC (ρ = 0.60). Endothelial cells (S01) contributed HBEGF–CD44 (ρ = 0.58), HAS2–CD44 (ρ = 0.58), AGTRAP–RACK1 (ρ = 0.56), and HBEGF–CD82 (ρ = 0.51). All correlations were positive.

The remaining three endpoints, FAP+ to Tumour, FAP+ to CD3, and αSMA+FAP+ to Tumour, did not show significant LR pairs after filtering, despite showing significant cell state associations in the EcoTyper analysis. Importantly, it should be noted that these associations between LR expression scores and spatial distances are not direct evidence of causal signalling.

6.3 Discussion

This chapter addresses the questions left open by the previous chapter. The mIF method identified five spatial endpoints, measured as NDI values, that significantly differed between infiltrative and pushing growth patterns at the invasive front of primary CRC: FAP+ to Tumour, αSMA+FAP+ to Tumour, CD3 to Tumour, FAP+ to CD3, and αSMA+FAP+ to CD3. A greater αSMA+FAP+ to Tumour NDI emerged as an independent prognostic factor for CRC-specific survival, and a greater FAP+ to Tumour NDI showed a strong prognostic trend ($p = 0.053$). However, the mIF approach alone could not identify which cell populations were associated with these spatial endpoints or which molecular signals supported them. To address these, a combination of whole-transcriptome GeoMx spatial profiling, EcoTyper cell-state deconvolution, and LIANA ligand-receptor inference was performed.

6.3.1 Overview

To our knowledge, this is among the first studies to integrate the spatial architecture of the invasive front with its underlying transcriptomic programmes, characterising how CAFs, immune cells and tumour cells are spatially organised and communicate according to their configuration rather than composition alone. These findings develop our understanding of how stromal and immune populations engage tumour cells at the invasive front and carry implications for the wider field.

The three masks GeoMx segmentation approach were validated by exploratory analyses. PCA analysis showed clear separation of the Tumour mask from the αSMA and Rest masks, with 36.2% of the variance explained by PC1. Moreover, the differential gene expression confirmed mask identity, canonical epithelial markers (KRT8, EPCAM, CDH1) enriched in the Tumour mask, myofibroblast and matrix genes (ACTA2, TAGLN, COL1A1) enriched in the αSMA mask, and immunoglobulin and immune genes (IGHG1-4, JCHAIN) enriched in the Rest mask. Pathway analysis further supported the mask characteristics: the Tumour mask enriched for proliferative and metabolic pathways, the αSMA mask for stromal remodelling programmes such as EMT and myogenesis, and the Rest mask for immune and inflammatory pathways. Altogether, these findings

supported that the three masks captured biologically distinct characteristics in these three compartments.

A consistent finding across all bulk-level analyses (PCA, differential expression, and ssGSEA) was the absence of significant differences in results between infiltrative and pushing growth patterns within all three different masks. Therefore, the two growth patterns did not arise from large-scale transcriptomics differences. However, ssGSEA scores correlated with all spatial endpoints, in different pathways, particularly in the Tumour mask. This indicates that tumour cells' transcriptional programmes were revealed along the spatial gradient rather than through growth pattern comparisons. Because bulk-level analysis cannot determine which cell states or ligand-receptor signals within each mask correlate with spatial endpoints, EcoTyper cell-state deconvolution and LIANA ligand-receptor inference were applied to reveal them.

Interestingly, across all endpoints, a clear mask-specific division was apparent. Significant pathway correlations were almost exclusively detected within the tumour mask, while significant cell-state correlations were detected within the αSMA and Rest masks. These observations indicate that cancer cells at the invasive front expressed their biology through transcriptational programmes, whereas stromal and immune masks respond through cell-state composition. These patterns support the rationale for combining the three analytic approaches, pathway-NDI correlation (ssGSEA), cell-state deconvolution (EcoTyper), and ligand-receptor inference (LIANA).

6.3.2 Spatial endpoint findings

Each spatial endpoint is discussed by integrating the pathway-level (ssGSEA), cell-state (EcoTyper), and ligand-receptor (LIANA) analysis findings in this section. Table 6-4 summarises the significant findings across the three analytical layers (ssGSEA, EcoTyper, and LIANA) for all five endpoints, together with their prognostic status from the previous chapter. This serves as a reference for the endpoint-by-endpoint discussion.

Table 6-4: Summary of significant findings across the five spatial endpoints from integrated ssGSEA, EcoTyper, and LIANA analyses.

Endpoint NDI	Shorter NDI associated with	Pathway (ssGSEA)	Cell state (EcoTyper)	LR pairs (LIANA)	Prognostic
CD3 to Tumour	Infiltrative	shorter NDI: Inflammatory Hypoxia EMT longer NDI: MYC TARGETS V1	CD8 T cells S02 (αSMA mask)	12 pairs from CD8 T S02 (THBS1, NID1, RTN4, FSTL1) targeting integrins and LRPs in Tumour mask	No
FAP+ to Tumour	Infiltrative	shorter NDI: EMT MYOGENESIS longer NDI: E2F TARGETS MYC TARGETS V2 G2M CHECKPOINT	Activated B cells S04 (Rest mask)	Not significant	Trend (p = 0.053)
αSMA+FAP+ to Tumour	Infiltrative	shorter NDI: HEDGEHOG SIGNALING APICAL JUNCTION	Not significant	Not significant	**Independent prognostic factor**
FAP+ to CD3	Pushing	longer NDI: Inflammatory Hypoxia EMT COAGULATION	CD4 T cells S03 (αSMA mask)	Not significant	No
αSMA+FAP+ to CD3	Pushing	Not significant	Fibroblast S02 (αSMA mask) Endothelial S01 (αSMA mask) Mono/Mac S09 (Rest mask)	8 pairs (TIGIT–PVR, TIGIT–NECTIN2, HLA-DQA1–CD4, MRC1–PTPRC, HAS2–CD44, HBEGF–CD44, HBEGF–CD82, AGTRAP–RACK1)	No

6.3.2.1 CD3 to Tumour

The CD3 to Tumour spatial arrangement showed all pathway associations only in the Tumour mask. At the pathway level, a longer CD3 to Tumour distance correlated with MYC TARGETS V1, a proliferation pathway, consistent with the architecture of the pushing growth pattern in which tumour expands as a coherent mass rather than as infiltrated cells. In contrast, shorter CD3 to Tumour distances were associated with inflammatory programmes (TNFA SIGNALING VIA NFKB,

INFLAMMATORY RESPONSE, COMPLEMENT, ALLOGRAFT REJECTION) and stress responses (HYPOXIA, KRAS SIGNALING UP, EMT) in the Tumour mask.

The co-occurrence of the inflammatory, hypoxic, and EMT pathway is associated with shorter distances, suggesting that T cells simultaneously produce inflammatory cytokines and occur under hypoxic conditions. Hypoxia is known to suppress T cell function through several mechanisms, which explains why tumours appear inflamed but are not functionally effective. This is consistent with the prognostic value of CD3 to Tumour not being an independent factor: closeness between T cells and tumour cells alone does not mean the T cells are functionally active.

At the cell-state level, wider CD3 to Tumour distances correlated with enrichment of CD8 T cells in the late-stage differentiated effector state (S02) within the αSMA mask. This indicates that a wider CD3-tumour separation found in pushing tumours coincides with a more differentiated CD8 T cell phenotype at the stromal interface. Ligand-receptor from LIANA inference identified 12 significant ligand-receptor pairs at this endpoint. The most frequent ligand was THBS1, a matricellular glycoprotein linked to stromal remodelling and immune modulation (Omatsu et al., 2023). Its corresponding receptors included the integrins ITGA4 and ITGA6, which mediate cell adhesion to extracellular matrix, and LRP family receptors (LRP1, LRP5) involved in matrix-related signalling. The expression of integrins and LRP receptors on tumour cells indicates that the cancer cells of the pushing growth pattern are biologically related to the cohesive, non-invasive nature of the pushing growth pattern. In this configuration, the tumour cells remain attached to their matrix environment rather than reorganising their adhesion for migration as occurs during the EMT pathway observed at shorter CD3 to Tumour distances within this same endpoint.

In multivariable Cox regression, the CD3 to Tumour endpoint was not an independent prognostic factor for CRC-specific survival analysis. These findings should be interpreted as reflecting a tumour-immune architecture in pushing tumours rather than a primary driver of CRC clinical outcomes.

6.3.2.2 FAP+ to Tumour

An infiltrative tumour showed a shorter distance of FAP+ to Tumour and presented a strong prognostic trend towards poorer CRC-specific survival ($p = 0.053$). At the pathway level, shorter FAP+ to Tumour distances correlated with the EMT and MYOGENESIS pathways in the tumour mask, suggesting that shorter distances are linked to the epithelial plasticity pathway, consistent with the infiltrative phenotype. Longer FAP+ to Tumour distances correlated with proliferative programmes, including E2F TARGETS, MYC TARGETS, G2M CHECKPOINT, consistent with the locally expanding tumour bulk characteristic of the pushing tumours.

At the cell-state level, longer FAP+ to Tumour distances were associated with enrichment of activated B cells (S04) in the rest mask. The accumulation of activated B cells raises the possibility of tertiary lymphoid structure (TLS) formation, though this interpretation remains tentative. TLS are organised lymphoid aggregates that, when present and functional, can support effector T cell priming and humoral responses (Sautès-Fridman et al., 2019). If TLS formation underlies this association, it could contribute to better survival outcomes in patients with a pushing growth pattern, despite T cells not infiltrating the tumour mass. However, this analysis did not include a direct histological confirmation of TLS and did not assess the spatial arrangement of these B cells. So, whether they form organised structures or accumulate diffusely remains undetermined.

Ligand-receptor LIANA analysis did not identify significant correlations for this endpoint, despite the cell-state significance correlation. It is important to explain why this endpoint showed significant cell-state associations in the EcoTyper deconvolution analysis but did not yield significant LR pairs in the LIANA inference. Rather than being a limitation, this divergence provides important biological insights.

First, FAP functions primarily as a cell-surface serine protease that remodels the extracellular matrix through enzymatic cleavage rather than through classical ligand-receptor signalling (Puré & Blomberg, 2018). Because enzymatic cleavage products and structural matrix alterations lack mRNA transcripts, the significant

spatial influence of FAP-positive fibroblasts is inherently difficult for transcriptomic-based LR inference to detect.

Second, the accumulation of activated B cells (S04) at greater FAP to Tumour distances is consistent with TLS formation, where B cell recruitment is primarily driven by long-range chemokines such as CXCL13 rather than direct cell-cell contact (Sautès-Fridman et al., 2019). Static snapshot transcriptomics may fail to capture these early diffusible signals once the TLS is fully formed. Consequently, while the EcoTyper can detect the B cells that remain in the tissue, LIANA misses the initial signals that recruited them.

Taken together, the FAP+ to Tumour spatial arrangement captures the two distinct biological correlates at the opposite ends of its spatial gradient: stromal-associated epithelial plasticity when FAP+ fibroblasts are in closer proximity to tumour cells, and peripheral activated B cell accumulation at greater distance.

6.3.2.3 αSMA+FAP+ to Tumour

The αSMA+FAP+ to Tumour spatial endpoint was the only spatial NDI that served as an independent prognostic factor for CRC-specific survival from the previous chapter, with a wider distance correlated with better outcomes. However, it showed limited pathway-level associations in ssGSEA and no significant cell state or ligand-receptor associations in the EcoTyper and LIANA analyses. This apparent disconnect between prognostic significance and molecular signal provides important insight.

First, this finding highlights the key difference between physical tissue displacement and active molecular reprogramming. As solid tumours grow, they generate mechanical pressure known as solid stress that physically displaces the surrounding stroma outward (Nia et al., 2020). The widened gap between CAFs and tumour cells is likely maintained by acellular extracellular matrix components, particularly dense fibrillar collagens, which represent a well-recognised stromal response to cancer (Pearce et al., 2018). Because EcoTyper measures relative proportions of active cell states based on RNA expression, an expansion of the stroma driven mainly by acellular ECM deposition and physical

compression would alter the spatial distance without necessarily changing the transcriptomic cell state composition.

Second, this observation reflects the temporal dynamics of tissue remodelling. Current spatial transcriptomic technologies capture a single snapshot at the time of surgical resection. The signalling required to initially activate fibroblasts and establish the stromal barrier most likely occurred during earlier stages of tumour development. Once this barrier reached a stable state, the transcriptomic signals that drove its formation would have subsided, leaving behind a structurally expanded but transcriptionally quiescent stromal zone.

Importantly, this spatial endpoint demonstrated two significant pathway associations within the tumour mask, with shorter distances correlated with increased HEDGEHOG SIGNALING and APICAL JUNCTION pathway activities. These signals indicate tumour-intrinsic activity because there were no significant enrichment pathways in the rest or αSMA masks. The shorter αSMA+FAP+ to Tumour distances, characteristic of the infiltrative growth pattern, are consistent with tumour cells engaged in paracrine signalling and junctional remodelling.

Hedgehog signalling acts as short-range paracrine signal from tumour cells, driving the differentiation of nearby fibroblasts into myCAFs (αSMA+FAP+ CAF) and leading to a plausible mechanism for their close position to tumour cells (Gerling et al., 2016). The apical junction activity reflects junctional remodelling and is associated with partial EMT states underlying tumour budding (Grigore et al., 2016). Tumour budding, defined as small clusters of fewer than five cancer cells dissociating at the invasive front, is a histological hallmark of the infiltrative growth pattern and a recognised independent prognostic marker in CRC (Lugli et al., 2017). Together, these molecular signatures align with the aggressive, invasive phenotype that characterises infiltrative tumours at the histological level.

At wider distances, characteristic of pushing tumours, no signalling was detected, and the molecular profile is consistent with the physical-displacement framing described above. Therefore, these findings explain two contrasting biological states of different distances of the αSMA+FAP+ to the tumour endpoint:

an actively signalling, invasion-permissive interface in infiltrative tumours and a stable, physically displaced interface in pushing tumours.

6.3.2.4 FAP+ to CD3

The longer FAP+ to CD3 distances, correlated with the infiltrative tumours, were positively correlated with inflammatory pathways (TNFA SIGNALING VIA NFKB, INFLAMMATORY RESPONSE, IL6 JAK STAT3 SIGNALING, COMPLEMENT), HYPOXIA, COAGULATION, and EMT within the tumour mask. One possible explanation for these findings is that when FAP+ is located near T cells, it normally produces immunoregulatory signals that suppress T cell activity. In contrast, when FAP+ cells are far from T cells, this suppressive effect is reduced, allowing T cells to release more inflammatory cytokines.

At the cell-level state, longer FAP+ to CD3 distances correlated with enrichment of an Unknown/Dysfunctional CD4 T cell state (S03) in the αSMA mask. This annotation does not specify which type of dysfunctional T cells is involved. Future immunophenotyping will be needed to determine which type of dysfunctional CD4 T cell state (regulatory T cells, exhausted T cells, anergic effector cells, or dysfunctional follicular helper T cells).

It is important to explain why FAP+ to CD3 distances showed significant cell state associations in EcoTyper analysis but did not yield significant LR pairs in the LIANA analysis. Rather than being a limitation, this divergence provides important biological insight. One plausible mechanism is metabolic competition: fibroblasts in dense fibrotic environments consume essential nutrients such as glucose and release suppressive byproducts such as lactate (Monteran & Erez, 2019). Because LIANA relies on protein-coding RNA, it cannot detect these non-protein metabolic factors.

6.3.2.5 αSMA+FAP+ to CD3

The αSMA+FAP+ to CD3 spatial endpoint showed several significant cell-state and ligand-receptor associations yet yielded no significant pathway correlations across the three masks. These dissociations support the interpretation that the molecular activity underlying myCAF subtype-driven immune exclusion, observed at longer distances, is characteristic of infiltrative tumours, operates at the level

of cell-state composition (EcoTyper) and ligand-receptor signalling (LIANA) rather than at the transcriptional programme level (ssGSEA). This pattern reinforces the methodological argument that ssGSEA, EcoTyper, and LIANA capture complementary rather than redundant aspects of the spatial biology.

At the cell-state level, longer αSMA+FAP+ to CD_3 distances were associated with enrichment of Fibroblast S02 and Endothelial cells S01 (normal-enriched/angiogenic) in the αSMA mask, and Monocytes/Macrophages S09 (regulatory) in the Rest mask. The convergence of multiple stromal and immune cell states on this single endpoint is consistent with a coordinated multicellular programme leading to the spatial immune exclusion.

LIANA identified eight ligand-receptor pairs associated with this endpoint, which suggest a dual-barrier mechanism: a biochemical layer that suppresses T cell function through signalling, and a physical layer that blocks T cell movement through extracellular matrix accumulation.

The biochemical layer was supported by two cell types located in different compartments. From the immune compartment (Rest mask), Monocytes/Macrophages (S09) contributed TIGIT–PVR (ρ = 0.60), TIGIT–NECTIN2 (ρ = 0.59), and HLA-DQA1–CD4 (ρ = 0.59). TIGIT is a well-established co-inhibitory receptor that suppresses T cell and NK cell cytotoxicity when engaged by PVR or NECTIN2 (Chauvin & Zarour, 2020). The HLA-DQA1–CD4 pair reflects retained MHC class II expression on these macrophages but, in the context of a regulatory phenotype, may support tolerogenic rather than activating engagement of CD4 T cells. Although the EcoTyper annotation for the S09 state is Unknown/Regulatory, the combined ligand profile observed in this analysis (TIGIT, NECTIN2, MRC1, HLA-DQA1) is consistent with an M2-like, immunosuppressive tumour-associated macrophage phenotype, providing an evidence-based interpretation of this otherwise unannotated state.

From the stromal zone (αSMA mask), Fibroblasts (S02) contributed MRC1–PTPRC (ρ = 0.60). MRC1 (CD206) is recognised as a marker of immunosuppressive, M2-like macrophage polarisation (Murray et al., 2014), though its expression has also been reported in stromal cells; its interaction with PTPRC (CD45) on

leukocytes suggests an immunomodulatory axis at the stromal–immune interface that may reduce TCR signalling through phosphatase activity.

On the other hand, the physical layer was supported by Endothelial cells (S01) in the stromal zone, which contributed HAS2–CD44 ($\rho = 0.58$). HAS2 is the enzyme responsible for producing hyaluronan, a large extracellular matrix polysaccharide (McCarthy et al., 2018). Accumulation of hyaluronan in the tumour stroma has been shown to create physical barriers that exclude CD8+ T cells in mesenchymal CRC, and its degradation by hyaluronidase restores T cell infiltration and sensitivity to immune checkpoint blockade (Martinez-Ordoñez et al., 2023).

Taken together, these signals indicate that mature myCAFs (αSMA+FAP+ CAF) at the infiltrative front coordinate a multicellular programme of T cell exclusion that operates at two complementary levels: M2-like macrophages and Fibroblast S02 cells reduce T cell signalling biochemically, while Endothelial S01 cells contribute a physical hyaluronan barrier.

6.3.3 Distinct biology of pushing and infiltrative growth patterns

The five spatial endpoint findings reveal two distinct cellular and molecular mechanisms underlying the two TGPs at the invasive front of the primary CRC, with the αSMA+FAP+ to Tumour endpoint representing an independent prognostic factor.

The pushing growth pattern showed proliferative activity and peripheral immune accumulation, including activated B cells, consistent with TLS-like organisation. This combination of structural containment and proximal immune surveillance may contribute to the better survival observed in pushing margins, even when T cells do not deeply infiltrate the tumour mass, although direct functional validation was not performed.

The infiltrative growth pattern demonstrated that the two different fibroblast subtypes play complementary but distinct roles, together driving an aggressive phenotype. FAP+ fibroblasts close to the tumour are linked to the EMT activity. While the αSMA+FAP+ fibroblast subtype is located farther from T cells, it orchestrates a dual barrier of immune exclusion through biochemical (TIGIT,

MRC1) and physical (HAS2–hyaluronan) mechanisms, while the few CD4 T cells that reach the stromal zone are dysfunctional.

Although the two fibroblast subtypes occupy different positions and perform different functions, they act together to make the infiltrative phenotype aggressive. FAP-only fibroblasts at close range push the cancer cells towards a motile, invasive phenotype, while αSMA+FAP+ myCAFs at greater distance from T cells shield the resulting invasive cells from immune attack. These may explain the aggressive nature of the infiltrative phenotype arising from this combination: cancer cells become motile and invade the stroma (via FAP-only signalling), and once they invade, they are protected from immune clearance by the dual barrier formed by αSMA+FAP+ myCAFs. Therefore, the infiltrative phenotype is associated with poor prognosis because both mechanisms operate simultaneously through complementary fibroblast subsets at different spatial scales.

6.3.4 Therapeutic implications

These findings suggest that immune exclusion in infiltrative CRC is associated with a coordinated spatial architecture and cellular communication network that may explain why standard immunotherapies such as anti-PD-1 monotherapy are often insufficient in these patients. The associations identified here suggest that T cells in infiltrative tumours are not simply inactive but are simultaneously physically obstructed by a hyaluronan-rich matrix barrier and biochemically suppressed by alternative checkpoint signalling through TIGIT before they can reach the tumour. In parallel, the FAP-only fibroblast subset appears to drive cancer cell plasticity and invasion in close apposition to the tumour, a process that operates independently of T cell exclusion.

These observations support the rational design of combination therapies for patients with infiltrative growth patterns. Stromal-depleting agents targeting HAS2 to break down the physical barrier could be combined with alternative immune checkpoint inhibitors, such as anti-TIGIT, to overcome the biochemical suppression. Agents targeting FAP-positive fibroblasts or their downstream effects on cancer cell EMT could be considered in parallel to limit the invasive component driven by the less mature fibroblast subset. Growth pattern, which is assessable

in routine H&E sections, could serve as a simple biomarker to identify patients who may benefit from such combination strategies.

6.3.5 Strengths and limitations

This study has several strengths. The analyses were performed on a comparatively large cohort for spatial transcriptomic profiling, comprising 51 cases (31 infiltrative and 20 pushing growth patterns), which _improves the statistical power to the detect associations against the substantial inter-patient and spatial variability inherent to expression measurements at the invasive front. Sampling was directed specifically at the invasive front, material that is not widely accessible and that is essential for interrogating growth pattern biology at its functional interface. The study also integrates spatial single-cell proteomics (mIF) with whole-transcriptome spatial profiling (GeoMx WTA), coupled with EcoTyper cell-state deconvolution and LIANA ligand-receptor inference, providing a multimodal framework well suited to dissecting the cellular and molecular basis of growth pattern architecture. Such an integrated dissection of growth pattern biology, combining spatial proteomics and spatial transcriptomics at the invasive front, has few _comparable studies that exist in the CRC literature.

Several limitations should be acknowledged. First, GeoMx spatial transcriptomics provides expression from mask-defined areas, not single-cell resolution. EcoTyper deconvolution provides estimates of cell state abundance but does not confirm the spatial positioning of individual cells relative to each other; the relative ordering of cell types cannot be determined from NDI measurements alone. Second, all findings represent statistical associations between independently measured modalities (GeoMx expression and mIF-derived NDI), not causal evidence. Third, the LIANA Consensus resource represents a curated subset of known LR interactions and may miss tissue-specific or non-canonical interactions. Fourth, the tertiary lymphoid structure interpretation at the FAP+ to Tumour endpoint is based on activated B cell abundance and was not confirmed by histological identification of organised lymphoid aggregates; diffuse, non-organised B cell accumulation cannot be excluded. Finally, receptor expression was measured at the mask level; the specific cell type expressing each receptor within a given mask cannot be determined.

Chapter 7 General Discussion

7.1 Overview

CRC remains one of the most common malignancies worldwide and a leading cause of cancer-related mortality. Although the TNM staging classification guides clinical decision-making, patients with the same stage exhibit different survival outcomes (Compton, 2003). At the invasive front, tumours meet the stroma and immune cells, and actively reshape that environment as they spread (Brabletz et al., 2005). This may be the reason why several of the strongest prognostic features in CRC are examined at the invasive front rather than at the tumour centre, such as tumour budding (Lugli et al., 2017) and growth patterns (Morikawa et al., 2012) which are both assessed on routine H&E-stained sections. Their established prognostic value reveals important biological processes concentrated at this interface. However, how the stromal and immune cells at the invasive front are organised in space remains poorly understood.

This thesis focused on the invasive front of primary CRC using an integrated multimodal strategy to reveal both cellular organisation and underlying molecular biology. The TGP was established as a reproducible prognostic feature on routine H&E sections across the two independent cohorts: the Glasgow Royal Infirmary (GRI) and the TransSCOT cohorts. The molecular biology and cellular spatial arrangement underpinning the TGP were then characterised by the multi-omics technologies: the transcriptomic profiling (TempO-Seq), the mIF with spatial proximity analysis using the NDI, and the compartment-resolved spatial transcriptomic profiling (GeoMx Whole Transcriptome Atlas), complemented by cell-state inference (EcoTyper) and ligand-receptor analysis (LIANA).

This thesis proposes that the TGP is the morphological projection of the two biologically distinct stromal-immune ecosystems at the invasive front. Across multiple modalities, the biological basis of this distinction emerged as architectural rather than transcriptional. Cells within tissue compartments expressed similar gene programmes across TGP, but their proportions, spatial proximity, and inter-cell communication networks differed in clinically meaningful ways. Importantly, the spatial proximity of αSMA+FAP+ CAF to the tumour endpoint, captured by the NDI metric, emerged as an independent

prognostic factor beyond the conventional TNM staging system, repositioning the existing prognostic dimension from cell abundance to cell geometry.

7.2 Tumour growth pattern is more than morphology

Chapter 3 established the three-tier TGP classification as a reproducible histological feature and an independent prognostic factor, after adjusting for TNM staging, across two large cohorts. Therefore, TGP captures prognostic information that conventional TNM staging cannot, raising the question of what biology underlies it. The remainder of this discussion chapter sets out to address this question, focusing on the molecular and spatial findings from Chapters 4, 5, and 6.

7.3 Compositional versus architectural biology of TGPs

Across the histological, transcriptomic, and spatial analyses examined in this thesis, the differences between growth patterns concentrate in the stromal compartment: tumour stroma percentage (TSP) was higher in the infiltrative growth pattern (Chapter 3), bulk transcriptomics showed stromal and EMT programmes enriched in infiltrative growth pattern (Chapter 4), and compartment-resolved spatial transcriptomics further confirmed that αSMA compartment expressed the expected stromal programmes (Chapter 6). The character of this stromal involvement, however, is more complicated than bulk transcriptomics findings alone might suggest.

Bulk RNA-seq (Chapter 4) identified several hundred genes differentially expressed between infiltrative and pushing growth patterns, with stromal and EMT-related programmes in the infiltrative pattern and proliferative-metabolic programmes in the pushing. However, when compartment-resolved spatial transcriptomics were performed (Chapter 6), a paradoxical finding emerged. Between compartments (CK+, αSMA+, Rest), as expected, there was strong transcriptomic separation. But when the same compartment was compared between TGPs, αSMA+ in infiltrative vs αSMA+ in pushing, no differential genes reached significance. This suggests that cells within each compartment expressed similar programmes regardless of TGP.

These findings reconcile the two observations. If cells within each compartment do not differ transcriptionally between TGPs, then the bulk signal cannot be interpreted as cell-intrinsic reprogramming. Instead, the observed differences may not reflect altered biological activity, but rather differences in the relative abundance of cellular compartments: the infiltrative pattern contains a greater proportion of αSMA+ stromal compartments, resulting in a bulk transcriptomic signal enriched for stromal genes even though the underlying transcriptional programmes of αSMA+ regions remain similar across TGPs.

The mIF findings (Chapter 5) show that the density of αSMA+FAP+ CAFs does not distinguish infiltrative from pushing patterns. But when αSMA+FAP+ CAF were assessed for spatial proximity to Tumour cells using the density-corrected distance metric, the NDI, a prognostic signal emerged as an independent factor for CRC-specific survival outcomes after adjustment for conventional TNM staging. This indicates that NDI captures information beyond cell abundance and density, highlighting the prognostic importance of spatial organisation and supporting the idea that TGP may reflect not only how tumours invade but also how stromal cells are positioned at the invasive front.

7.4 Immune presence versus immune function in the infiltrative growth pattern

Bulk RNA-seq analysis showed enrichment of immune-related programmes in infiltrative pattern, including IL6-JAK-STAT3 signalling, TNF-α signalling via NF-κB, and inflammatory response (Chapter 4). At the protein and spatial levels, CD3+ T cell density did not differ between TGPs, but spatial analysis showed that T cells were closer to tumour cells in infiltrative tumours than in pushing tumours (lower CD3 to Tumour NDI in infiltrative tumours; Chapter 5). With the T cell abundance, transcriptomic inflammation, and closer T cell proximity to tumour cells, these characteristics are used to classify a tumour as an immune-inflamed phenotype based on Hedge et al. criteria (Gerling et al., 2016).

However, the infiltrative tumours are associated with worse survival outcomes (Chapter 3), and they meet the criteria of an immune-inflamed phenotype but do not deliver favourable survival outcomes. That T cell-inflamed

tumours that are dysfunctional are well established (Trujillo et al., 2018). What the present findings add is a specific spatial and signalling pattern in primary CRC that corresponds to this dysfunctional T cell in the TGPs context.

The GeoMx-LIANA inference approach (Chapter 6) links this CAF-T cell separation to several ligand-receptor signalling axes inferred from ligand-receptor correlation analysis: TIGIT-PVR and TIGIT-NECTIN2 pairs associated with regulatory monocyte/macrophage cell states (S09) in the Rest compartment, HAS2-CD44 axis associated with endothelial cell states (S01), and an MRC1-PTPRC axis associated with fibroblast cell states (S02). The individual components of this picture are not new: TIGIT-mediated immunosuppression is a recognised immune evasion mechanism currently being targeted in clinical trials (Chauvin & Zarour, 2020), and hyaluronic acid-mediated stromal organisation has been studied as a barrier to immune access (Provenzano et al., 2012). What is specific to study is the integration of these spatial arrangements (CAF-T cell separation) and signalling elements (inhibitory signalling/ stromal organisation) in the context of TGPs in primary CRC.

Taken together, the infiltrative growth pattern can therefore be considered a dysfunctional subset of the T cell-inflamed category, in which T cells reach the tumour interface but are functionally held back by the structural and signalling niches they occupy. The bulk inflammation signal reflects T cells having access to tumour cells but appearing unable to function. This may explain the mismatch between the inflammation signal from bulk RNA-seq and the poor clinical outcomes. The therapeutic implications of this positioning are discussed in Section 7.7.

7.5 An integrated model of the invasive front

Drawing the histological, transcriptomics, and spatial arrangement findings, the contrasting biology of the two TGPs can be summarised schematically in Figure 7-1, with the pushing pattern on the left and the infiltrative pattern on the right. The figure is offered only as a visual aid to make the two configurations easier to picture. It is drawn from the conclusions of Chapters 3 to 6. Because the elements rest on inference and cell-state deconvolution, the molecular axes it depicts are candidate associations generated as hypotheses, not causal relationships, and they

remain to be confirmed. The spatial arrangement is also schematic, as the underlying measurements are pairwise distances and do not fix the precise ordering of the cell populations. The figure should therefore be read as an aid to picturing the integrated findings.

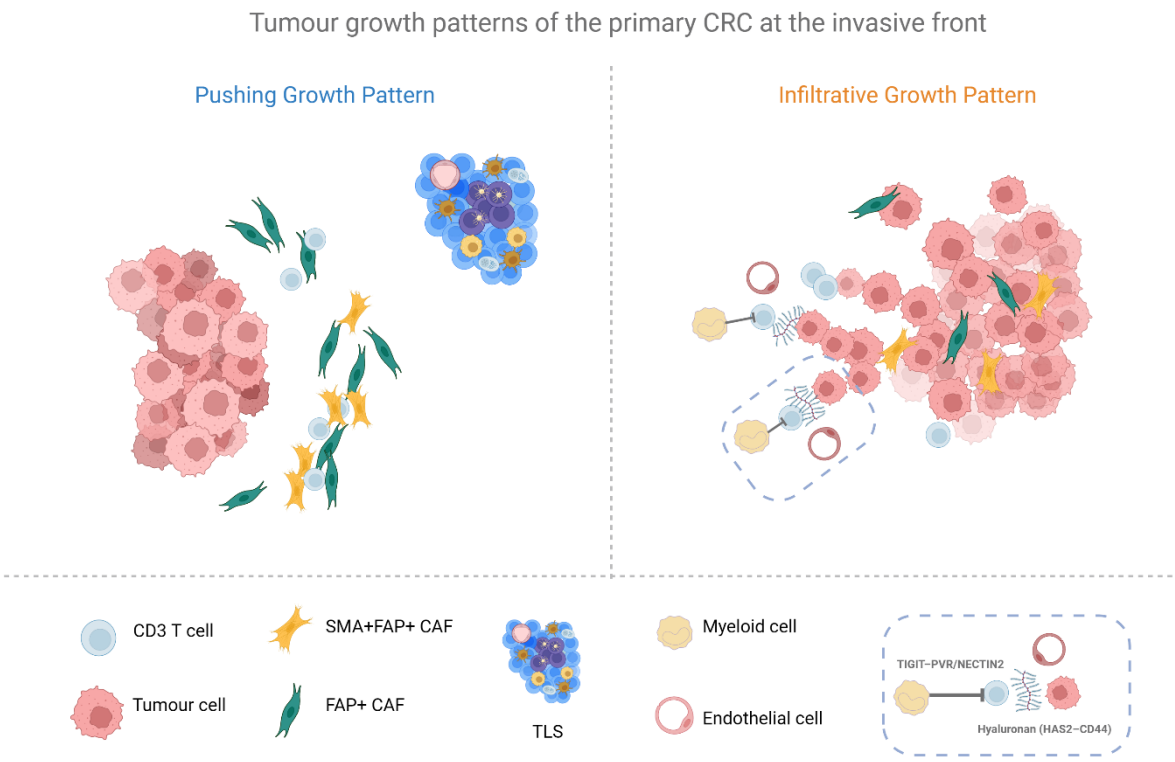


Figure 7-1: Schematic of the two growth patterns at the CRC invasive front.

Conceptual summary drawn from Chapters 3–6. Pushing (left): cohesive tumour, with CAFs and CD3+ T cells in stromal regions separated from tumour cells, and abundant activated B cells shown as a putative TLS (not histologically confirmed). Infiltrative (right): CAFs and T cells closer to and intermixed with dissociating tumour cells, associated with two candidate inhibitory mechanisms; a TIGIT–PVR/NECTIN2 axis from regulatory myeloid cells and hyaluronan (HAS2–CD44) from angiogenic endothelial cells (inset). The signalling axes were inferred, not functionally validated. As the NDI measures pairwise distances and cannot fix the position of three or more cell types, the figure is an aid to interpretation, not a literal map of tissue architecture. The figure was created by BioRender.

In the pushing growth pattern, αSMA+FAP+ CAFs, and FAP+ CAFs are located at a greater distance from tumour cells, and the T cells lie more peripherally relative to tumour cells (Chapter 5). At a greater distance from the FAP+ to Tumour relationship in pushing tumours, the deconvolution analysis showed a higher abundance of activated B cells (Chapter 6). This is an abundance-level association only, and it does not establish where these cells sit or whether they form organised structures, but it raises the possibility of tertiary lymphoid

structures. For the tumour itself, the bulk transcriptomic programme of the pushing tumours was proliferative and enriched for MYC, E2F, and G2M pathways (Chapter 4), and the proliferative signatures reappeared at the wider distance of the spatial endpoint in Chapter 6, fitting a tumour that expands as a coherent mass rather than dissociating into stroma.

In the infiltrative growth pattern, αSMA+FAP+ CAFs and FAP+ CAFs lie close to tumour cells, but both CAF subtypes lie far from T cells. In the spatial transcriptomic analysis, shorter FAP+ to Tumour distances were associated with EMT and tumour cell motility and showed a strong adverse prognostic trend ($p = 0.053$). The αSMA+FAP+ to Tumour NDI, an independent prognostic endpoint, carried no candidate signalling, behaving as physical stromal displacement rather than active signalling. By contrast, the αSMA+FAP+ CAFs to T cells showed a candidate dual signal, a biochemical arm, TIGIT-PVR / TIGIT-NECTIN2 from regulatory macrophages, and a physical arm, hyaluronan (HAS2-CD44) from endothelial cells. This configuration, a dispersed tumour front in which the T-cell presence does not translate into tumour control, is the one associated with an aggressive phenotype.

One thing that this study cannot answer is what makes a tumour become one type rather than the other type of TGPs. This could be set by the cell of origin, particular driver mutations, the physical forces at the interface area, or some combination of these.

7.6 Prognostic is geometric, and a predictive marker is not the therapeutic target

Across spatial chapters, the findings point to two main things that no single chapter alone can explain. The first is that what predicts survival outcome here is where cells sit, not how many there are. Second, the relationship which predicts survival outcomes is not the one a drug would target.

The first point is that, among all NDI proximities measured in Chapter 5, the distance between αSMA+FAP+CAF and tumour cells predicted cancer-specific survival after adjustment for tumour budding and nodal stage, with the FAP+ to

Tumour distance showing borderline significance; no other proximity or density did.

On the second point, although the αSMA+FAP+ to Tumour distance predicted survival outcomes, no candidate ligand-receptor signalling was detected on it. As discussed in Chapter 6, this distance reflects physical stromal displacement rather than active signalling and therefore cannot itself be a drug target. It carries no molecular signal to act on, reflecting a barrier that has already formed and become quiescent. By the time that the tumour is resected, the signalling that formed it has likely subsided, leaving nothing active for a drug to engage. Moreover, the direction of αSMA+FAP+ to Tumour NDI endpoint runs the reversed way, since the wider distance is a favourable survival outcome and acting on it would mean pushing the stroma further from the tumour, which is not a straightforward therapeutic aim.

The signalling that could be targeted lay instead on a different, non-prognostic spatial NDI, the separation of αSMA+FAP+ CAFs from CD3 T-cells. The mechanistic basis was set out in Chapter 6. What matters here is the dissociation: the prognostic marker and the point of intervention are not on the same axis. This is helpful rather than a contradiction, but it means the marker and the therapy must be developed as two distinct things. The therapeutic consequences are taken up in Section 7.8.

7.7 Clinical translation and therapeutic implications

7.7.1 Translational potential of the integrated findings

The translational pipeline in this thesis is unusually accessible for precision medicine. The present work applies high-cost multi-modal research to validate a low-cost, first-line clinical readout obtained from standard histology, requiring no new technology for initial stratification. The TGP scoring method can be assessed on an H&E-stained section and requires no additional staining beyond routine reporting. The findings from this study position TGP not as a standalone marker but as one component of an integrated risk-stratification framework with selectively applied molecular and spatial confirmation.

A practical way suggested by these findings begins with TGP scoring at histopathological reporting. Patients with an infiltrative growth pattern could undergo a confirmatory spatial-biomarker assessment using multiplex panels appropriate to clinical pathology infrastructure. The αSMA+FAP+CAF to Tumour NDI, an independent prognostic factor introduced here, is feasible and requires a panel encompassing αSMA, FAP, and panCK, together with automated cell segmentation and proximity computation.

Notably, the αSMA+FAP+ to Tumour NDI identifies risk, but it does not establish that an individual tumour carries the signalling to be targeted, as set out in Section 7.6, the prognostic distance and actionable signalling lie on different axes. Selecting patients for TIGIT-axis or hyaluronan-targeted therapy would require direct measurement of those targets. Both TGP and the NDI metric are, in this sense, prognostic rather than predictive features that stratify risk but were not tested against treatment response, and neither identifies which tumours would respond to a given therapy.

The TGP scoring and the αSMA+FAP+ to Tumour NDI are both read from the resected specimen; their role is in post-operative prognostication, refining decisions about adjuvant treatment and the intensity of follow-up in patients who have undergone surgery. A separate question is how biology might inform treatment beyond the resection setting. As set out in Section 7.6, the prognostic geometry and the actionable signalling lie on different axes: selecting patients for TIGIT-axis or hyaluronan-targeted therapy would require direct measurement of those molecular targets, not the spatial distance.

Because these targets are molecular rather than architectural, they could be assessed on a diagnostic biopsy. For example, TIGIT, PVR, and NECTIN2 could be assessed by immunohistochemistry, while hyaluronan could be evaluated by tissue histochemistry, as mismatch-repair deficiency, identified by immunohistochemistry, is already used to select CRC patients for checkpoint inhibition (André et al., 2020). In contrast to the resection-based prognostic metric, this biopsy-level approach would not be limited to patients undergoing surgery and could therefore support treatment selection in patients with unresectable or metastatic disease.

7.7.2 Tumour growth pattern and tumour budding as complementary features of assessment

Tumour budding is routinely reported in many departments, including Glasgow, in line with the recommendations of the International Tumour Budding Consensus Conference (Lugli et al., 2017). As discussed in Chapter 3, TGP and tumour budding act as complementary features, with TGP discriminating survival outcomes when tumour budding is low. The translational implication is that the two features could be combined into an H&E scoring framework, which is read from the same section, requires no additional reagents, and together, may capture the biology of the invasive front more completely than either alone.

The clinical viability of spatial metrics in CRC is not hypothetical. The Immunoscore, which quantifies CD3+ and CD8+ T-cell densities in the tumour centre and invasive margin, is the first immune-based biomarker recommended by international guidelines for prognostication beyond TNM staging in colon cancer (Pagès et al., 2018). It was further developed to Immunoscore-IC, incorporating the spatial proximity between CD8+ and PD-L1+ cells, reflecting a field-level shift from cell density to cell geometry and has shown candidate predictive value for checkpoint inhibition in pMMR metastatic disease (Hsu & Leung, 2026).

The CAF-tumour NDI introduced in this thesis follows the same trajectory but extends it to the stromal compartment, which current Immunoscore implementations omit. As in the CRC invasive front, where cells are non-uniformly distributed, a density value averages over their spatial arrangement. For instance, two tumours can have the same density yet differ in how those cells are organised, and it is that organisation, not density, that carries the prognostic signal. However, the computational obstacle of this NDI approach is that proximity cannot be scored by eye, requiring imaging, automated segmentation, and NDI calculation. Even the guideline-backed assay, Immunoscore, has seen limited uptake owing to cost and training requirements. Fortunately, AI-based digital pathology is diminishing this barrier. Paired with TGP as a low-cost first-line stratifier, the NDI offers a pragmatic, if still prospective, route toward clinical use in primary CRC.

7.8 Strengths and Limitations

The limitations of the individual analyses have been discussed in the relevant chapters and are not repeated here. This section instead focuses on broader limitations of the thesis and those that emerge from considering the findings together.

First, the NDI was modelled as a continuous variable, and it lacks an established cut-off as a prognostic feature; a validated threshold would be required and validated in an independent cohort. Furthermore, the NDI measures distances between pairs of cell populations and does not capture the full spatial arrangement of multiple cell types. For this reason, the TGP model presented in Figure 7-1 should be viewed as a conceptual framework rather than a direct representation of tumour architecture.

In addition, tumours may display combinations of spatial features that are not represented by this model since the spatial endpoints were analysed independently. For instance, a tumour could show a short αSMA+FAP+ to Tumour distance together with a short αSMA+FAP+ to T cell distance, whereas the model would predict a longer αSMA+FAP+ to T cell distance. The biological significance of such patterns remains unclear. Further work is therefore required to determine how these spatial endpoints interact before the model can be used for tumour classification.

Second, the prognostic and therapeutic implications of these findings should be considered separately. The TGP and CAF-tumour NDI were associated with clinical outcomes and represent prognostic, not predictive, markers, as they were not evaluated in relation to treatment response. Similarly, the proposed combination of the TIGIT axis and a hyaluronan-targeted intervention remains a hypothesis derived from the integrated analyses rather than a validated therapeutic strategy. Because the prognostic distance and the actionable signalling lie on different axes, the NDI identifies risk but cannot indicate which tumours carry the target; selecting patients for therapy would require direct measurement of the targets themselves.

Third, both the GRI and TransSCOT cohorts comprised patients with resectable, non-metastatic disease (stage I-III). The prognostic value of the TGP and the spatial findings therefore applies to this group, and it is not known how they would behave in locally advanced unresectable or metastatic CRC, where the biology of the invasive front and treatment context differ. Extending these findings to later-stage disease would require a separate study.

Despite these limitations, the work has notable strengths. The findings were derived from two large, well-annotated clinical cohorts with mature survival follow-up and matched multimodal data, spanning histopathology, bulk transcriptomics, and spatial profiling. Few spatial biomarkers in the published literature have shown prognostic value independent of TNM stage, as the αSMA+FAP+CAF to Tumour NDI does here. Taken together, integrating these modalities with cell-state deconvolution and ligand-receptor inference provides a powerful model for studying the biology of the invasive front in CRC.

7.9 Future directions

The findings of this thesis suggest several directions for future research. Some arise from the limitations discussed above, while others are new opportunities generated by the findings themselves. Together, they point to four broad areas for future work: external validation, mechanistic confirmation, investigation of the factors that drive TGPs, and clinical translation.

The highest priority is external validation of the spatial findings. While the prognostic value of the TGP was confirmed in the independent TransSCOT cohort, the αSMA+FAP+ CAFs to Tumour NDI and the associated cell states and signalling pathways were identified in a single cohort and required validation in independent datasets. These datasets should include matched H&E assessment, multiplex imaging, and spatial transcriptomic profiling of the invasive front in primary CRC. In addition, the prognostic significance of TGP in rectal cancer remains uncertain and should be investigated in a larger rectal cancer cohort with tumour budding data available for multivariable Cox regression analysis.

As the mIF approach is not routinely available in diagnostic pathology, conventional IHC could provide a practical initial approach for validating selected

cell populations identified in the spatial analyses. For instance, CD8 IHC could be used to assess the presence and distribution of CD8 T cells associated with the spatial findings. However, this would validate the broader cell population rather than the specific transcriptional cell state by EcoTyper, and it would not reproduce the mIF-derived NDI spatial metric.

A second area for future research is the validation of the immune-barrier mechanisms proposed in this thesis. The candidate signalling pathways were identified from compartment-level ligand–receptor analyses, and it was not possible to determine which specific cells within each compartment expressed the relevant receptors. For example, TIGIT–PVR and TIGIT–NECTIN2 signalling was associated with the regulatory monocyte/macrophage state (S09), while HAS2–CD44 signalling was linked to the endothelial state (S01). However, confirmation of these cellular interactions will require higher-resolution approaches, such as single-cell spatial technologies.

Functional studies are also needed to determine whether these signalling pathways contribute directly to immune exclusion. Co-culture or organoid models incorporating tumour cells, myeloid cells, endothelial cells, and T cells could be used to test the effects of disrupting PVR/NECTIN2 signalling or hyaluronan production on T-cell access to tumour cells.

Furthermore, the present study was based on tissue collected at a single time point. Longitudinal sampling would help determine whether the spatial arrangements observed at the invasive front contribute to tumour progression or arise as a consequence of tumour evolution.

A third area for future research is understanding what determines TGP itself. While this thesis characterises the biological and spatial features associated with pushing and infiltrative tumours, it does not explain why a tumour adopts one pattern rather than the other. The factors that drive TGPs remain unclear and may include the cell origin, specific genomic alterations, or interactions between these factors.

In addition, the NDI quantifies pairwise distance between cell populations in two-dimensional tissue sections, whereas the invasive front is inherently a three-

dimensional structure. Therefore, the full spatial relationships among multiple cell populations cannot be captured using the current approach. Reconstructing the invasive front in three dimensions may reveal organisational features that are not apparent in two-dimensional sections.

A fourth area for future research is clinical translation. The candidate combination of TIGIT and hyaluronan-targeted therapy remains hypothetical and requires further validation before clinical application. Importantly, the spatial features associated with prognosis are distinct from the signalling pathways that may represent therapeutic targets. As a result, the distance-based metrics developed in this study could not be used to select patients for these therapies. Instead, patient selection would require direct assessment of the relevant molecular targets, followed by prospective clinical evaluation.

More broadly, the integrative framework developed in this thesis, combining spatial proteomics, spatial transcriptomics, cell-state deconvolution, and ligand–receptor analysis, may provide a useful approach for studying the invasive front of other solid tumours. Applying similar strategies across tumour types may help identify common and tumour-specific principles of spatial organisation within the tumour microenvironment. This applies to the analytical framework rather than to the scoring system, since pattern distribution and prognostic value vary between tumour types (Chapter 1) and would require tumour-type-specific validation.

7.10 Concluding remarks

This thesis investigated whether TGPs at the invasive front of primary CRC represent a clinically meaningful and biologically interpretable phenotype and explored what may underlie the poorer prognosis associated with the infiltrative growth pattern. Across histopathology, bulk transcriptomics, multiplex immunofluorescence, and spatial transcriptomics, these analyses suggest that the adverse biology of the infiltrative pattern cannot be explained by a single cell population, marker, or pathway alone, but is better understood through the spatial organisation of tumour, stromal, and immune compartments at the tumour-stromal interface.

This work reframes TGPs not simply as a descriptive histopathological category, but as a spatial phenotype with prognostic value. Across two cohorts, the infiltrative growth pattern remained an independent prognostic factor in colon cancer, while spatial analyses indicated that infiltrative and pushing tumours differed in the arrangement of cell populations at the invasive front, rather than in their abundance. This led to the identification of a spatial feature with independent prognostic value: the distance between αSMA+FAP+ CAFs and tumour cells. It also showed that this prognostic geometry is not the same as the inhibitory signalling axes that may be therapeutically actionable.

Given that these findings were derived from a single cohort and are based on inferred rather than experimentally validated mechanisms, they should be viewed as hypothesis-generating and require further validation.

This thesis offers a framework for understanding the invasive front as a spatially organised ecosystem, in which tissue architecture carries prognostic and therapeutic information that may be missed by abundance-based measurements alone. TGPs, assessable on routine H&E sections, provide a practical entry point into this architecture, while the integrative strategy developed here offers a route to interrogate it in greater biological detail in CRC and may be adapted to the invasive front of other tumour types, although pattern definitions and their prognostic value would need to be established separately in each setting.

List of References

Alexander, P. G., Roseweir, A. K., Pennel, K. A. F., van Wyk, H. C., Powell, A., McMillan, D. C., Horgan, P. G., Kelly, C., Hay, J., Sansom, O., Harkin, A., Roxburgh, C. S. D., Graham, J., Church, D. N., Tomlinson, I., Saunders, M., Iveson, T. J., Edwards, J., & Park, J. H. (2021). The Glasgow Microenvironment Score associates with prognosis and adjuvant chemotherapy response in colorectal cancer. *Br J Cancer*, *124*(4), 786-796. https://doi.org/10.1038/s41416-020-01168-x

Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., Gershenwald, J. E., Brookland, R. K., Meyer, L., Gress, D. M., Byrd, D. R., & Winchester, D. P. (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*, *67*(2), 93-99. https://doi.org/10.3322/caac.21388

André, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Zaninelli, M., Clingan, P., Bridgewater, J., Tabah-Fisch, I., & Gramont, A. d. (2004). Oxaliplatin, Fluorouracil, and Leucovorin as Adjuvant Treatment for Colon Cancer. *New England Journal of Medicine*, *350*(23), 2343-2351. https://doi.org/doi:10.1056/NEJMoa032709

André, T., Shiu, K.-K., Kim, T. W., Jensen, B. V., Jensen, L. H., Punt, C., Smith, D., Garcia-Carbonero, R., Benavides, M., Gibbs, P., Fouchardiere, C. d. l., Rivera, F., Elez, E., Bendell, J., Le, D. T., Yoshino, T., Cutsem, E. V., Yang, P., Farooqui, M. Z. H.,…Diaz, L. A. (2020). Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer. *New England Journal of Medicine*, *383*(23), 2207-2218. https://doi.org/doi:10.1056/NEJMoa2017699

Antoniotti, C., Rossini, D., Pietrantonio, F., Catteau, A., Salvatore, L., Lonardi, S., Boquet, I., Tamberi, S., Marmorino, F., Moretto, R., Ambrosini, M., Tamburini, E., Tortora, G., Passardi, A., Bergamo, F., Kassambara, A., Sbarrato, T., Morano, F., Ritorto, G.,…Cremolini, C. (2022). Upfront FOLFOXIRI plus bevacizumab with or without atezolizumab in the treatment of patients with metastatic colorectal cancer (AtezoTRIBE): a multicentre, open-label, randomised, controlled, phase 2 trial. *Lancet Oncol*, *23*(7), 876-887. https://doi.org/10.1016/s1470-2045(22)00274-1

Basile, D., Broudin, C., Emile, J. F., Falcoz, A., Pagès, F., Mineur, L., Bennouna, J., Louvet, C., Artru, P., Fratte, S., Ghiringhelli, F., André, T., Derangère, V., Vernerey, D., Taieb, J., & Svrcek, M. (2022). Tumor budding is an independent prognostic factor in stage III colon cancer patients: a post-hoc analysis of the IDEA-France phase III trial (PRODIGE-GERCOR)☆. *Annals of Oncology*, *33*(6), 628-637. https://doi.org/https://doi.org/10.1016/j.annonc.2022.03.002

Belhabib, I., Zaghdoudi, S., Lac, C., Bousquet, C., & Jean, C. (2021). Extracellular Matrices and Cancer-Associated Fibroblasts: Targets for Cancer Diagnosis and Therapy? *Cancers (Basel)*, *13*(14). https://doi.org/10.3390/cancers13143466

Brabletz, T., Jung, A., Spaderna, S., Hlubek, F., & Kirchner, T. (2005). Migrating cancer stem cells — an integrated concept of malignant tumour progression. *Nature Reviews Cancer*, *5*(9), 744-749. https://doi.org/10.1038/nrc1694

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & 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), 229-263. https://doi.org/https://doi.org/10.3322/caac.21834

Bungaro, C., Guida, M., & Apollonio, B. (2025). Spatial proteomics of the tumor microenvironment in melanoma: current insights and future directions. *Front Immunol*, *16*, 1568456. https://doi.org/10.3389/fimmu.2025.1568456

Cercek, A., Lumish, M., Sinopoli, J., Weiss, J., Shia, J., Lamendola-Essel, M., Dika, I. H. E., Segal, N., Shcherba, M., Sugarman, R., Stadler, Z., Yaeger, R., Smith, J. J., Rousseau, B., Argiles, G., Patel, M., Desai, A., Saltz, L. B., Widmar, M.,…Diaz, L. A. (2022). PD-1 Blockade in Mismatch Repair–Deficient, Locally Advanced Rectal Cancer. *New England Journal of Medicine*, *386*(25), 2363-2376. https://doi.org/doi:10.1056/NEJMoa2201445

Cercek, A., Wilson, C., & Xu, R.-H. (2026). Announcing the Lancet Commission on colorectal cancer: addressing the rising global burden. *The Lancet*, *407*(10539), 1582-1584. https://doi.org/10.1016/S0140-6736(26)00418-6

Chauvin, J. M., & Zarour, H. M. (2020). TIGIT in cancer immunotherapy. *J Immunother Cancer*, *8*(2). https://doi.org/10.1136/jitc-2020-000957

Compton, C. C. (2003). Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Mod Pathol*, *16*(4), 376-388. https://doi.org/10.1097/01.Mp.0000062859.46942.93

Cords, L., Engler, S., Haberecker, M., Rüschoff, J. H., Moch, H., de Souza, N., & Bodenmiller, B. (2024). Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer. *Cancer Cell*, *42*(3), 396-412.e395. https://doi.org/https://doi.org/10.1016/j.ccell.2023.12.021

Cuyle, P. J., & Prenen, H. (2017). Current and future biomarkers in the treatment of colorectal cancer. *Acta Clin Belg*, *72*(2), 103-115. https://doi.org/10.1080/17843286.2016.1262996

Dang, H., Harryvan, T. J., Liao, C.-Y., Danen, E. H. J., Spalburg, V. N. L. N., Kielbasa, S. M., Mei, H., Goeman, J. J., de Jonge-Muller, E. S., Janson, S. G. T., van der Reijden, J. J., Crobach, S., Hardwick, J. C. H., Boonstra, J. J., de Miranda, N. F. C. C., & Hawinkels, L. J. A. C. (2023). Cancer-Associated Fibroblasts Are Key Determinants of Cancer Cell Invasion in the Earliest Stage of Colorectal Cancer. *Cellular and Molecular Gastroenterology and Hepatology*, *16*(1), 107-131. https://doi.org/https://doi.org/10.1016/j.jcmgh.2023.04.004

Dawson, H., Christe, L., Eichmann, M., Reinhard, S., Zlobec, I., Blank, A., & Lugli, A. (2020). Tumour budding/T cell infiltrates in colorectal cancer: proposal of a novel combined score. *Histopathology*, *76*(4), 572-580. https://doi.org/https://doi.org/10.1111/his.14006

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., & Wallace, M. B. (2019). Colorectal cancer. *The Lancet*, *394*(10207), 1467-1480. https://doi.org/https://doi.org/10.1016/S0140-6736(19)32319-0

Deng, J.-J., Zhang, W., Xu, X.-M., Zhang, F., Tao, W.-P., Ye, J.-J., & Ge, W. (2016). Twist mediates an aggressive phenotype in human colorectal cancer cells. *Int J Oncol*, *48*(3), 1117-1124. https://doi.org/10.3892/ijo.2016.3342

Dienstmann, R., Mason, M. J., Sinicrope, F. A., Phipps, A. I., Tejpar, S., Nesbakken, A., Danielsen, S. A., Sveen, A., Buchanan, D. D., Clendenning,

M., Rosty, C., Bot, B., Alberts, S. R., Milburn Jessup, J., Lothe, R. A., Delorenzi, M., Newcomb, P. A., Sargent, D., & Guinney, J. (2017). Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. *Annals of Oncology*, *28*(5), 1023-1031. https://doi.org/https://doi.org/10.1093/annonc/mdx052

Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P. L., Nagai, J. S., Boys, C., Ramirez Flores, R. O., Kim, H., Szalai, B., Costa, I. G., Valdeolivas, A., Dugourd, A., & Saez-Rodriguez, J. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nature Communications*, *13*(1), 3224. https://doi.org/10.1038/s41467-022-30755-0

Fearon, E. R. (2011). Molecular genetics of colorectal cancer. *Annu Rev Pathol*, *6*, 479-507. https://doi.org/10.1146/annurev-pathol-011110-130235

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, *61*(5), 759-767. https://doi.org/10.1016/0092-8674(90)90186-i

Feig, C., Jones, J. O., Kraman, M., Wells, R. J., Deonarine, A., Chan, D. S., Connell, C. M., Roberts, E. W., Zhao, Q., Caballero, O. L., Teichmann, S. A., Janowitz, T., Jodrell, D. I., Tuveson, D. A., & Fearon, D. T. (2013). Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc Natl Acad Sci U S A*, *110*(50), 20212-20217. https://doi.org/10.1073/pnas.1320318110

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., & Pagès, F. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, *313*(5795), 1960-1964. https://doi.org/10.1126/science.1129139

Gerling, M., Büller, N. V. J. A., Kirn, L. M., Joost, S., Frings, O., Englert, B., Bergström, Å., Kuiper, R. V., Blaas, L., Wielenga, M. C. B., Almer, S., Kühl, A. A., Fredlund, E., van den Brink, G. R., & Toftgård, R. (2016). Stromal Hedgehog signalling is downregulated in colon cancer and its restoration restrains tumour growth. *Nature Communications*, *7*(1), 12321. https://doi.org/10.1038/ncomms12321

Goossens, S., Vandamme, N., Van Vlierberghe, P., & Berx, G. (2017). EMT transcription factors in cancer development re-evaluated: Beyond EMT and MET. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1868*(2), 584-591. https://doi.org/https://doi.org/10.1016/j.bbcan.2017.06.006

Graham, R. P., Vierkant, R. A., Tillmans, L. S., Wang, A. H., Laird, P. W., Weisenberger, D. J., Lynch, C. F., French, A. J., Slager, S. L., Raissian, Y., Garcia, J. J., Kerr, S. E., Lee, H. E., Thibodeau, S. N., Cerhan, J. R., Limburg, P. J., & Smyrk, T. C. (2015). Tumor Budding in Colorectal Carcinoma: Confirmation of Prognostic Significance and Histologic Cutoff in a Population-based Cohort. *Am J Surg Pathol*, *39*(10), 1340-1346. https://doi.org/10.1097/pas.0000000000000504

Grigore, A. D., Jolly, M. K., Jia, D., Farach-Carson, M. C., & Levine, H. (2016). Tumor Budding: The Name is EMT. Partial EMT. *J Clin Med*, *5*(5). https://doi.org/10.3390/jcm5050051

Grothey, A., Sobrero, A. F., Shields, A. F., Yoshino, T., Paul, J., Taieb, J., Souglakos, J., Shi, Q., Kerr, R., Labianca, R., Meyerhardt, J. A., Vernerey, D., Yamanaka, T., Boukovinas, I., Meyers, J. P., Renfro, L. A., Niedzwiecki,

D., Watanabe, T., Torri, V.,…Iveson, T. (2018). Duration of Adjuvant Chemotherapy for Stage III Colon Cancer. *New England Journal of Medicine*, *378*(13), 1177-1188. https://doi.org/doi:10.1056/NEJMoa1713709

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.,…Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11), 1350-1356. https://doi.org/10.1038/nm.3967

Hatse, S., Lambrechts, Y., Antoranz Martinez, A., De Schepper, M., Geukens, T., Vos, H., Berben, L., Messiaen, J., Marcelis, L., Van Herck, Y., Neven, P., Smeets, A., Desmedt, C., De Smet, F., Bosisio, F. M., Wildiers, H., & Floris, G. (2024). Dissecting the immune infiltrate of primary luminal B-like breast carcinomas in relation to age. *J Pathol*, *264*(3), 344-356. https://doi.org/10.1002/path.6354

Hav, M., Libbrecht, L., Ferdinande, L., Geboes, K., Pattyn, P., & Cuvelier, C. A. (2015). Pathologic Assessment of Rectal Carcinoma after Neoadjuvant Radio(chemo)therapy: Prognostic Implications. *Biomed Res Int*, *2015*, 574540. https://doi.org/10.1155/2015/574540

Hernandez, S., Rojas, F., Laberiano, C., Lazcano, R., Wistuba, I., & Parra, E. R. (2021). Multiplex Immunofluorescence Tyramide Signal Amplification for Immune Cell Profiling of Paraffin-Embedded Tumor Tissues [Review]. *Frontiers in Molecular Biosciences*, *Volume 8 - 2021*. https://doi.org/10.3389/fmolb.2021.667067

Hsu, M. S., & Leung, S. Y. (2026). Implementing Immunoscore in colorectal cancer: lessons from cohort studies, limitations and barriers to adoption. *The Journal of Pathology: Clinical Research*, *12*(1), e70071. https://doi.org/https://doi.org/10.1002/2056-4538.70071

Hu, M., Chen, Z., Hu, D., Xi, S., Wang, D., Zhang, X., Fong, W. P., Wen, L., Cai, Y., Yuan, Y., Li, B., Wu, X., Lu, Z., Chen, G., Li, L., Ding, P., Pan, Z., Wan, D., Du, Z.,…Li, Y. (2022). Delineating the molecular landscape of different histopathological growth patterns in colorectal cancer liver metastases [Original Research]. *Frontiers in Immunology*, *Volume 13 - 2022*. https://doi.org/10.3389/fimmu.2022.1045329

Issa, N., Arfanis, E., Hager, T., Aigner, C., Dietz-Terjung, S., Theegarten, D., Kühl, H., & Welter, S. (2019). A prospective comparison of growth patterns with radiomorphology in 232 lung metastases-basis for patient tailored resection planning? *J Thorac Dis*, *11*(7), 2822-2831. https://doi.org/10.21037/jtd.2019.07.04

Iveson, T. J., Kerr, R. S., Saunders, M. P., Cassidy, J., Hollander, N. H., Tabernero, J., Haydon, A., Glimelius, B., Harkin, A., Allan, K., McQueen, J., Scudder, C., Boyd, K. A., Briggs, A., Waterston, A., Medley, L., Wilson, C., Ellis, R., Essapen, S.,…Paul, J. (2018). 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *Lancet Oncol*, *19*(4), 562-578. https://doi.org/10.1016/s1470-2045(18)30093-7

Jass, J. R., Ajioka, Y., Allen, J. P., Chan, Y. F., Cohen, R. J., Nixon, J. M., Radojkovic, M., Restall, A. P., Stables, S. R., & Zwi, L. J. (1996). Assessment of invasive growth pattern and lymphocytic infiltration in colorectal cancer. *Histopathology*, *28*(6), 543-548. https://doi.org/10.1046/j.1365-2559.1996.d01-467.x

Jass, J. R., Love, S. B., & Northover, J. M. (1987). A new prognostic classification of rectal cancer. *Lancet*, *1*(8545), 1303-1306. https://doi.org/10.1016/s0140-6736(87)90552-6

Joyce, J. A., & Fearon, D. T. (2015). T cell exclusion, immune privilege, and the tumor microenvironment. *Science*, *348*(6230), 74-80. https://doi.org/doi:10.1126/science.aaa6204

Karamitopoulou, E., Lugli, A., Panayiotides, I., Karakitsos, P., Peros, G., Rallis, G., Patsouris, E. S., Terracciano, L., & Zlobec, I. (2010). Systematic assessment of protein phenotypes characterizing high-grade tumour budding in mismatch repair-proficient colorectal cancer [https://doi.org/10.1111/j.1365-2559.2010.03615.x]. *Histopathology*, *57*(2), 233-243. https://doi.org/https://doi.org/10.1111/j.1365-2559.2010.03615.x

Karamitopoulou, E., Zlobec, I., Koelzer, V. H., Langer, R., Dawson, H., & Lugli, A. (2015). Tumour border configuration in colorectal cancer: proposal for an alternative scoring system based on the percentage of infiltrating margin. *Histopathology*, *67*(4), 464-473. https://doi.org/10.1111/his.12665

Karapetis, C. S., Khambata-Ford, S., Jonker, D. J., O'Callaghan, C. J., Tu, D., Tebbutt, N. C., Simes, R. J., Chalchal, H., Shapiro, J. D., Robitaille, S., Price, T. J., Shepherd, L., Au, H. J., Langer, C., Moore, M. J., & Zalcberg, J. R. (2008). K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med*, *359*(17), 1757-1765. https://doi.org/10.1056/NEJMoa0804385

Keum, M. A., Lim, S. B., Kim, S. A., Yoon, Y. S., Kim, C. W., Yu, C. S., & Kim, J. C. (2012). Clinicopathologic factors affecting recurrence after curative surgery for stage I colorectal cancer. *J Korean Soc Coloproctol*, *28*(1), 49-55. https://doi.org/10.3393/jksc.2012.28.1.49

Koelzer, V., & Lugli, A. (2014). The Tumor Border Configuration of Colorectal Cancer as a Histomorphological Prognostic Indicator [Review]. *Frontiers in Oncology*, *4*. https://doi.org/10.3389/fonc.2014.00029

Koelzer, V. H., Zlobec, I., & Lugli, A. (2016). Tumor budding in colorectal cancer—ready for diagnostic practice? *Human Pathology*, *47*(1), 4-19. https://doi.org/https://doi.org/10.1016/j.humpath.2015.08.007

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kopetz, S., Grothey, A., Yaeger, R., Cutsem, E. V., Desai, J., Yoshino, T., Wasan, H., Ciardiello, F., Loupakis, F., Hong, Y. S., Steeghs, N., Guren, T. K., Arkenau, H.-T., Garcia-Alfonso, P., Pfeiffer, P., Orlov, S., Lonardi, S., Elez, E., Kim, T.-W.,…Tabernero, J. (2019). Encorafenib, Binimetinib, and Cetuximab in BRAF V600E–Mutated Colorectal Cancer. *New England Journal of Medicine*, *381*(17), 1632-1643. https://doi.org/doi:10.1056/NEJMoa1908075

Korphaisarn, K., Morris, V., Davis, J. S., Overman, M. J., Fogelman, D. R., Kee, B. K., Dasari, A., Raghav, K. P. S., Shureiqi, I., Trupti, M., Wolff, R. A., Eng, C., Menter, D. G., Hamilton, S., & Kopetz, S. (2019). Signet ring cell colorectal cancer: genomic insights into a rare subpopulation of colorectal adenocarcinoma. *British Journal of Cancer*, *121*(6), 505-510. https://doi.org/10.1038/s41416-019-0548-9

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Latacz, E., Verheul, S. M. L., Sillis, Y., van Dam, P.-J., Doukas, M., Grunhagen, D. J., Nyström, H., Dirix, P., Dirix, L., Van Laere, S., Verhoef, C., & Vermeulen, P. (2024). Molecular characterization of the histopathological growth patterns of colorectal cancer liver metastases by RNA sequencing of targeted samples at the tumor-liver interface. *Clinical & Experimental Metastasis*, *42*(1), 1. https://doi.org/10.1007/s10585-024-10319-w

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., Skora, A. D., Luber, B. S., Azad, N. S., Laheru, D., Biedrzycki, B., Donehower, R. C., Zaheer, A., Fisher, G. A., Crocenzi, T. S., Lee, J. J., Duffy, S. M., Goldberg, R. M., Chapelle, A. d. l.,…Diaz, L. A. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Medicine*, *372*(26), 2509-2520. https://doi.org/doi:10.1056/NEJMoa1500596

Lee, J. H., & Massagué, J. (2022). TGF-β in developmental and fibrogenic EMTs. *Seminars in Cancer Biology*, *86*, 136-145. https://doi.org/https://doi.org/10.1016/j.semcancer.2022.09.004

Leggett, B., & Whitehall, V. (2010). Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology*, *138*(6), 2088-2100. https://doi.org/10.1053/j.gastro.2009.12.066

Liu, Y., Dai, Y., & Wang, L. (2026). Spatial omics at the forefront: emerging technologies, analytical innovations, and clinical applications. *Cancer Cell*, *44*(1), 24-49. https://doi.org/10.1016/j.ccell.2025.12.009

Liu, Z., Ba, Y., Shan, D., Zhou, X., Zuo, A., Zhang, Y., Xu, H., Liu, S., Liu, B., Zhao, Y., Weng, S., Wang, R., Deng, J., Luo, P., Cheng, Q., Hu, X., Yang, S., Wang, F., & Han, X. (2025). THBS2-producing matrix CAFs promote colorectal cancer progression and link to poor prognosis via the CD47-MAPK axis. *Cell Reports*, *44*(4), 115555. https://doi.org/https://doi.org/10.1016/j.celrep.2025.115555

Loughrey, M. B. Q., Philip / Shepherd, Neil A. (2023). *Dataset for histopathological reporting of colorectal cancer* (G049).

Luca, B. A., Steen, C. B., Matusiak, M., Azizi, A., Varma, S., Zhu, C., Przybyl, J., Espín-Pérez, A., Diehn, M., Alizadeh, A. A., van de Rijn, M., Gentles, A. J., & Newman, A. M. (2021). Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell*, *184*(21), 5482-5496.e5428. https://doi.org/10.1016/j.cell.2021.09.014

Lugli, A., Karamitopoulou, E., & Zlobec, I. (2012). Tumour budding: a promising parameter in colorectal cancer. *British Journal of Cancer*, *106*(11), 1713-1717. https://doi.org/10.1038/bjc.2012.127

Lugli, A., Kirsch, R., Ajioka, Y., Bosman, F., Cathomas, G., Dawson, H., El Zimaity, H., Fléjou, J. F., Hansen, T. P., Hartmann, A., Kakar, S., Langner, C., Nagtegaal, I., Puppa, G., Riddell, R., Ristimäki, A., Sheahan, K., Smyrk, T., Sugihara, K.,…Quirke, P. (2017). Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Mod Pathol*, *30*(9), 1299-1311. https://doi.org/10.1038/modpathol.2017.46

Martinez-Ordoñez, A., Duran, A., Ruiz-Martinez, M., Cid-Diaz, T., Zhang, X., Han, Q., Kinoshita, H., Muta, Y., Linares, J. F., Kasashima, H., Nakanishi, Y., Omar, M., Nishimura, S., Avila, L., Yashiro, M., Maeda, K., Pannellini, T., Pigazzi, A., Inghirami, G.,…Moscat, J. (2023). Hyaluronan driven by epithelial aPKC deficiency remodels the microenvironment and creates a vulnerability in mesenchymal colorectal cancer. *Cancer Cell*, *41*(2), 252-271.e259. https://doi.org/10.1016/j.ccell.2022.11.016

Matly, A. A. M. (2024). *Investigating the association between epithelial-mesenchymal transition pathwayrelated proteins and patients' outcomes in colorectal cancer* University of Glasgow]. https://theses.gla.ac.uk/id/eprint/84744

McCarthy, J. B., El-Ashry, D., & Turley, E. A. (2018). Hyaluronan, Cancer-Associated Fibroblasts and the Tumor Microenvironment in Malignant Progression. *Front Cell Dev Biol*, *6*, 48. https://doi.org/10.3389/fcell.2018.00048

Meyer, Y., Bohlok, A., Olthof, P., Donckier, V., Doukas, M., Lucidi, V., Vermeulen, P., Grünhagen, D., & Verhoef, C. (2023). Histopathological growth patterns of neuroendocrine tumor liver metastases. *Clin Exp Metastasis*, *40*(3), 227-234. https://doi.org/10.1007/s10585-023-10211-z

Min, K. W., Kim, D. H., Noh, Y. K., Son, B. K., Kwon, M. J., & Moon, J. Y. (2021). Cancer-associated fibroblasts are associated with poor prognosis in solid type of lung adenocarcinoma in a machine learning analysis. *Sci Rep*, *11*(1), 16779. https://doi.org/10.1038/s41598-021-96344-1

Mitrovic, B., Handley, K., Assarzadegan, N., Chang, H. L., Dawson, H. A. E., Grin, A., Hutchins, G. G. A., Magill, L., Quirke, P., Riddell, R. H., Gray, R. G., & Kirsch, R. (2021). Prognostic and Predictive Value of Tumor Budding in Colorectal Cancer. *Clinical Colorectal Cancer*, *20*(3), 256-264. https://doi.org/https://doi.org/10.1016/j.clcc.2021.05.003

Monteran, L., & Erez, N. (2019). The Dark Side of Fibroblasts: Cancer-Associated Fibroblasts as Mediators of Immunosuppression in the Tumor Microenvironment. *Front Immunol*, *10*, 1835. https://doi.org/10.3389/fimmu.2019.01835

Moretto, R., Rossini, D., Catteau, A., Antoniotti, C., Giordano, M., Boccaccino, A., Ugolini, C., Proietti, A., Conca, V., Kassambara, A., Pietrantonio, F., Salvatore, L., Lonardi, S., Tamberi, S., Tamburini, E., Poma, A. M., Fieschi, J., Fontanini, G., Masi, G.,…Cremolini, C. (2023). Dissecting tumor lymphocyte infiltration to predict benefit from immune-checkpoint inhibitors in metastatic colorectal cancer: lessons from the AtezoT RIBE study. *J Immunother Cancer*, *11*(4). https://doi.org/10.1136/jitc-2022-006633

Morikawa, T., Kuchiba, A., Qian, Z. R., Mino-Kenudson, M., Hornick, J. L., Yamauchi, M., Imamura, Y., Liao, X., Nishihara, R., Meyerhardt, J. A., Fuchs, C. S., & Ogino, S. (2012). Prognostic significance and molecular associations of tumor growth pattern in colorectal cancer. *Ann Surg Oncol*, *19*(6), 1944-1953. https://doi.org/10.1245/s10434-011-2174-5

Murray, P. J., Allen, J. E., Biswas, S. K., Fisher, E. A., Gilroy, D. W., Goerdt, S., Gordon, S., Hamilton, J. A., Ivashkiv, L. B., Lawrence, T., Locati, M., Mantovani, A., Martinez, F. O., Mege, J. L., Mosser, D. M., Natoli, G., Saeij, J. P., Schultze, J. L., Shirey, K. A.,…Wynn, T. A. (2014). Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity*, *41*(1), 14-20. https://doi.org/10.1016/j.immuni.2014.06.008

Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J.,…disease working, g. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330-337. https://doi.org/10.1038/nature11252

Nearchou, I. P., Lillard, K., Gavriel, C. G., Ueno, H., Harrison, D. J., & Caie, P. D. (2019). Automated Analysis of Lymphocytic Infiltration, Tumor Budding, and Their Spatial Relationship Improves Prognostic Accuracy in Colorectal Cancer. *Cancer Immunology Research*, *7*(4), 609-620. https://doi.org/10.1158/2326-6066.Cir-18-0377

Nia, H. T., Munn, L. L., & Jain, R. K. (2020). Physical traits of cancer. *Science*, *370*(6516). https://doi.org/10.1126/science.aaz0868

Nurmik, M., Ullmann, P., Rodriguez, F., Haan, S., & Letellier, E. (2020). In search of definitions: Cancer-associated fibroblasts and their markers. *International Journal of Cancer*, *146*(4), 895-905. https://doi.org/https://doi.org/10.1002/ijc.32193

O'Connell, J. B., Maggard, M. A., & Ko, C. Y. (2004). Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst*, *96*(19), 1420-1425. https://doi.org/10.1093/jnci/djh275

Offermans, K., Reitsam, N. G., Simons, C. C. J. M., Grosser, B., Zimmermann, J., Grabsch, H. I., Märkl, B., & van den Brandt, P. A. (2024). The relationship between Stroma AReactive Invasion Front Areas (SARIFA), Warburg-subtype and survival: results from a large prospective series of colorectal cancer patients. *Cancer & Metabolism*, *12*(1), 21. https://doi.org/10.1186/s40170-024-00349-z

Omatsu, M., Nakanishi, Y., Iwane, K., Aoyama, N., Duran, A., Muta, Y., Martinez-Ordoñez, A., Han, Q., Agatsuma, N., Mizukoshi, K., Kawai, M., Yamakawa, G., Namikawa, M., Hamada, K., Fukunaga, Y., Utsumi, T., Sono, M., Masuda, T., Hata, A.,…Seno, H. (2023). THBS1-producing tumor-infiltrating monocyte-like cells contribute to immunosuppression and metastasis in colorectal cancer. *Nat Commun*, *14*(1), 5534. https://doi.org/10.1038/s41467-023-41095-y

Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F. S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T. T., Berger, M. D., Nagtegaal, I. D., Vink-Börger, E., Hartmann, A., Geppert, C., Kolwelter, J., Merkel, S., Grützmann, R., Van den Eynde, M., Jouret-Mourin, A.,…Galon, J. (2018). International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet*, *391*(10135), 2128-2139. https://doi.org/10.1016/s0140-6736(18)30789-x

Park, J. H., Powell, A. G., Roxburgh, C. S., Horgan, P. G., McMillan, D. C., & Edwards, J. (2016). Mismatch repair status in patients with primary operable colorectal cancer: associations with the local and systemic tumour environment. *Br J Cancer*, *114*(5), 562-570. https://doi.org/10.1038/bjc.2016.17

Pearce, O. M. T., Delaine-Smith, R. M., Maniati, E., Nichols, S., Wang, J., Böhm, S., Rajeeve, V., Ullah, D., Chakravarty, P., Jones, R. R., Montfort, A., Dowe, T., Gribben, J., Jones, J. L., Kocher, H. M., Serody, J. S., Vincent, B. G., Connelly, J., Brenton, J. D.,…Balkwill, F. R. (2018). Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers. *Cancer Discov*, *8*(3), 304-319. https://doi.org/10.1158/2159-8290.Cd-17-0284

Provenzano, P. P., Cuevas, C., Chang, A. E., Goel, V. K., Von Hoff, D. D., & Hingorani, S. R. (2012). Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. *Cancer Cell*, *21*(3), 418-429. https://doi.org/10.1016/j.ccr.2012.01.007

Pula, B., Witkiewicz, W., Dziegiel, P., & Podhorska-Okolow, M. (2013). Significance of podoplanin expression in cancer-associated fibroblasts: A

comprehensive review. *Int J Oncol*, *42*(6), 1849-1857. https://doi.org/10.3892/ijo.2013.1887

Puré, E., & Blomberg, R. (2018). Pro-tumorigenic roles of fibroblast activation protein in cancer: back to the basics. *Oncogene*, *37*(32), 4343-4357. https://doi.org/10.1038/s41388-018-0275-3

Qi, J., Sun, H., Zhang, Y., Wang, Z., Xun, Z., Li, Z., Ding, X., Bao, R., Hong, L., Jia, W., Fang, F., Liu, H., Chen, L., Zhong, J., Zou, D., Liu, L., Han, L., Ginhoux, F., Liu, Y.,…Su, B. (2022). Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. *Nat Commun*, *13*(1), 1742. https://doi.org/10.1038/s41467-022-29366-6

Rajaganeshan, R., Prasad, R., Guillou, P. J., Chalmers, C. R., Scott, N., Sarkar, R., Poston, G., & Jayne, D. G. (2007). The influence of invasive growth pattern and microvessel density on prognosis in colorectal cancer and colorectal liver metastases. *Br J Cancer*, *96*(7), 1112-1117. https://doi.org/10.1038/sj.bjc.6603677

Riihimäki, M., Hemminki, A., Sundquist, J., & Hemminki, K. (2016). Patterns of metastasis in colon and rectal cancer. *Scientific Reports*, *6*(1), 29765. https://doi.org/10.1038/srep29765

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D. G., Egeblad, M., Evans, R. M., Fearon, D., Greten, F. R., Hingorani, S. R., Hunter, T., Hynes, R. O., Jain, R. K., Janowitz, T., Jorgensen, C., Kimmelman, A. C., Kolonin, M. G., Maki, R. G., Powers, R. S., Puré, E.,…Werb, Z. (2020). A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer*, *20*(3), 174-186. https://doi.org/10.1038/s41568-019-0238-1

Salem, M. E., Weinberg, B. A., Xiu, J., El-Deiry, W. S., Hwang, J. J., Gatalica, Z., Philip, P. A., Shields, A. F., Lenz, H. J., & Marshall, J. L. (2017). Comparative molecular analyses of left-sided colon, right-sided colon, and rectal cancers. *Oncotarget*, *8*(49), 86356-86368. https://doi.org/10.18632/oncotarget.21169

Sandberg, T. P., Stuart, M., Oosting, J., Tollenaar, R., Sier, C. F. M., & Mesker, W. E. (2019). Increased expression of cancer-associated fibroblast markers at the invasive front and its association with tumor-stroma ratio in colorectal cancer. *BMC Cancer*, *19*(1), 284. https://doi.org/10.1186/s12885-019-5462-2

Sautès-Fridman, C., Petitprez, F., Calderaro, J., & Fridman, W. H. (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat Rev Cancer*, *19*(6), 307-325. https://doi.org/10.1038/s41568-019-0144-6

Schürch, C. M., Bhate, S. S., Barlow, G. L., Phillips, D. J., Noti, L., Zlobec, I., Chu, P., Black, S., Demeter, J., McIlwain, D. R., Kinoshita, S., Samusik, N., Goltsev, Y., & Nolan, G. P. (2020). Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell*, *182*(5), 1341-1359.e1319. https://doi.org/10.1016/j.cell.2020.07.005

Shah, A. H., Gami, A. J., Desai, N. H., Gandhi, J. S., & Trivedi, P. P. (2022). Tumor budding as a prognostic indicator in colorectal carcinoma: a retrospective study of primary colorectal carcinoma cases in a tertiary care center. *Indian Journal of Surgical Oncology*, *13*(3), 459-467. https://doi.org/10.1007/s13193-022-01498-7

Su, Z.-F., Wu, C.-Z., Ding, H.-R., Zhan, Q., Ba, Y.-B., Cheng, L., Li, J.-Y., Zhou, X.-D., Li, L.-J., Zhang, L.-X., Li, J., Li, Y., & Ren, B. (2026). Spatially resolved multiomics reveals the self-enforcing property of the leading-edge multicellular ecosystem of head and neck cancer. *Proceedings of the*

National Academy of Sciences, *123*(2), e2519474123. https://doi.org/doi:10.1073/pnas.2519474123

Sung, H., Siegel, R. L., Laversanne, M., Jiang, C., Morgan, E., Zahwe, M., Cao, Y., Bray, F., & Jemal, A. (2025). Colorectal cancer incidence trends in younger versus older adults: an analysis of population-based cancer registry data. *Lancet Oncol*, *26*(1), 51-63. https://doi.org/10.1016/s1470-2045(24)00600-4

Tampakis, A., Tampaki, E. C., Nonni, A., Kostakis, I. D., Posabella, A., Kontzoglou, K., von Flüe, M., Felekouras, E., Kouraklis, G., & Nikiteas, N. (2021). High fascin-1 expression in colorectal cancer identifies patients at high risk for early disease recurrence and associated mortality. *BMC Cancer*, *21*(1), 153. https://doi.org/10.1186/s12885-021-07842-4

Tang, F., Zhu, Y., Shen, J., Yuan, B., He, X., Tian, Y., Weng, L., & Sun, L. (2025). CD44+ cells enhance pro-tumor stroma in the spatial landscape of colorectal cancer leading edge. *British Journal of Cancer*, *132*(8), 703-715. https://doi.org/10.1038/s41416-025-02968-9

Taskin, O. C., Reid, M. D., Bagci, P., Balci, S., Armutlu, A., Demirtas, D., Pehlivanoglu, B., Saka, B., Memis, B., Bozkurtlar, E., Leblebici, C. B., Birceanu, A., Xue, Y., Erkan, M., Kapran, Y., Baygul, A., Sokmensuer, C., Scarpa, A., Luchini, C.,…Adsay, V. (2022). Infiltration pattern predicts metastasis and progression better than the T-stage and grade in pancreatic neuroendocrine tumors: a proposal for a novel infiltration-based morphologic grading. *Modern Pathology*, *35*(6), 777-785. https://doi.org/10.1038/s41379-021-00995-4

Thanki, K., Nicholls, M. E., Gajjar, A., Senagore, A. J., Qiu, S., Szabo, C., Hellmich, M. R., & Chao, C. (2017). Consensus Molecular Subtypes of Colorectal Cancer and their Clinical Implications. *Int Biol Biomed J*, *3*(3), 105-111.

Tokodai, K., Narimatsu, H., Nishida, A., Takaya, K., Hara, Y., Kawagishi, N., Hashizume, E., & Ohuchi, N. (2016). Risk factors for recurrence in stage II/III colorectal cancer patients treated with curative surgery: The impact of postoperative tumor markers and an infiltrative growth pattern. *Journal of Surgical Oncology*, *114*(3), 368-374. https://doi.org/https://doi.org/10.1002/jso.24320

Trejo, C. L., Babić, M., Imler, E., Gonzalez, M., Bibikov, S. I., Shepard, P. J., VanSteenhouse, H. C., Yeakley, J. M., & Seligmann, B. E. (2019). Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue. *PLOS ONE*, *14*(2), e0212031. https://doi.org/10.1371/journal.pone.0212031

Trujillo, J. A., Sweis, R. F., Bao, R., & Luke, J. J. (2018). T Cell-Inflamed versus Non-T Cell-Inflamed Tumors: A Conceptual Framework for Cancer Immunotherapy Drug Development and Combination Therapy Selection. *Cancer Immunol Res*, *6*(9), 990-1000. https://doi.org/10.1158/2326-6066.Cir-18-0277

Unal Kocabey, D., & Cakir, I. E. (2024). The prognostic significance of growth pattern, tumor budding, poorly differentiated clusters, desmoplastic reaction pattern and tumor-stroma ratio in colorectal cancer and an evaluation of their relationship with KRAS, NRAS, BRAF mutations. *Annals of Diagnostic Pathology*, *73*, 152375. https://doi.org/https://doi.org/10.1016/j.anndiagpath.2024.152375

van Dam, P. J., van der Stok, E. P., Teuwen, L. A., Van den Eynden, G. G., Illemann, M., Frentzas, S., Majeed, A. W., Eefsen, R. L., Coebergh van den Braak, R. R. J., Lazaris, A., Fernandez, M. C., Galjart, B., Laerum, O. D.,

Rayes, R., Grünhagen, D. J., Van de Paer, M., Sucaet, Y., Mudhar, H. S., Schvimer, M.,…Vermeulen, P. B. (2017). International consensus guidelines for scoring the histopathological growth patterns of liver metastasis. *Br J Cancer*, *117*(10), 1427-1441. https://doi.org/10.1038/bjc.2017.334

Van, T. M., & Blank, C. U. (2019). A user's perspective on GeoMxTM digital spatial profiling. *Immuno-Oncology Technology*, *1*, 11-18. https://doi.org/https://doi.org/10.1016/j.iotech.2019.05.001

van Wyk, H. C., Park, J., Roxburgh, C., Horgan, P., Foulis, A., & McMillan, D. C. (2015). The role of tumour budding in predicting survival in patients with primary operable colorectal cancer: a systematic review. *Cancer Treat Rev*, *41*(2), 151-159. https://doi.org/10.1016/j.ctrv.2014.12.007

van Wyk, H. C., Park, J. H., Edwards, J., Horgan, P. G., McMillan, D. C., & Going, J. J. (2016). The relationship between tumour budding, the tumour microenvironment and survival in patients with primary operable colorectal cancer. *British Journal of Cancer*, *115*(2), 156-163. https://doi.org/10.1038/bjc.2016.173

Wang, L. M., Kevans, D., Mulcahy, H., O'Sullivan, J., Fennelly, D., Hyland, J., O'Donoghue, D., & Sheahan, K. (2009). Tumor budding is a strong and reproducible prognostic marker in T3N0 colorectal cancer. *Am J Surg Pathol*, *33*(1), 134-141. https://doi.org/10.1097/PAS.0b013e318184cd55

Wang, W., Jin, J., Zhou, Z., Wang, Y., Min, K., Zuo, X., Jiang, J., Zhou, Y., & Shi, J. (2023). Snail inhibits metastasis via regulation of E-cadherin and is associated with prognosis in colorectal cancer. *Oncol Lett*, *25*(6), 271. https://doi.org/10.3892/ol.2023.13857

Wei, J., Chen, Z., Hu, M., He, Z., Jiang, D., Long, J., & Du, H. (2021). Characterizing Intercellular Communication of Pan-Cancer Reveals SPP1+ Tumor-Associated Macrophage Expanded in Hypoxia and Promoting Cancer Malignancy Through Single-Cell RNA-Seq Data [Original Research]. *Frontiers in Cell and Developmental Biology*, *Volume 9 - 2021*. https://doi.org/10.3389/fcell.2021.749210

Wikberg, M. L., Edin, S., Lundberg, I. V., Van Guelpen, B., Dahlin, A. M., Rutegård, J., Stenling, R., Oberg, A., & Palmqvist, R. (2013). High intratumoral expression of fibroblast activation protein (FAP) in colon cancer is associated with poorer patient prognosis. *Tumour Biol*, *34*(2), 1013-1020. https://doi.org/10.1007/s13277-012-0638-2

Williams, C. J. M., Gray, R., Hills, R. K., Shires, M., Zhang, L., Zhao, Z., Gardner, T., Sapanara, N., Xu, X. M., Bai, I., Yan, D., Muranyi, A., Dance, S., Aghaei, F., Hemmings, G., Hale, M., Kurkure, U., Guetter, C., Richman, S. D.,…Quirke, P. (2024). Evaluation of CD3 and CD8 T-Cell Immunohistochemistry for Prognostication and Prediction of Benefit From Adjuvant Chemotherapy in Early-Stage Colorectal Cancer Within the QUASAR Trial. *J Clin Oncol*, *42*(29), 3430-3442. https://doi.org/10.1200/jco.23.02030

Xin, L., Gao, J., Zheng, Z., Chen, Y., Lv, S., Zhao, Z., Yu, C., Yang, X., & Zhang, R. (2021). Fibroblast Activation Protein-α as a Target in the Bench-to-Bedside Diagnosis and Treatment of Tumors: A Narrative Review. *Front Oncol*, *11*, 648187. https://doi.org/10.3389/fonc.2021.648187

Yamanashi, T., Nakanishi, Y., Fujii, G., Akishima-Fukasawa, Y., Moriya, Y., Kanai, Y., Watanabe, M., & Hirohashi, S. (2009). Podoplanin Expression Identified in Stromal Fibroblasts as a Favorable Prognostic Marker in Patients with Colorectal Carcinoma. *Oncology*, *77*(1), 53-62. https://doi.org/10.1159/000226112

Yeh, Y.-H., Hsiao, H.-F., Yeh, Y.-C., Chen, T.-W., & Li, T.-K. (2018). Inflammatory interferon activates HIF-1α-mediated epithelial-to-mesenchymal transition via PI3K/AKT/mTOR pathway. *Journal of Experimental & Clinical Cancer Research*, *37*(1), 70. https://doi.org/10.1186/s13046-018-0730-6

Zhao, Z., Li, T., Yuan, Y., & Zhu, Y. (2023). What is new in cancer-associated fibroblast biomarkers? *Cell Commun Signal*, *21*(1), 96. https://doi.org/10.1186/s12964-023-01125-0

Zlobec, I., Baker, K., Minoo, P., Hayashi, S., Terracciano, L., & Lugli, A. (2009). Tumor border configuration added to TNM staging better stratifies stage II colorectal cancer patients into prognostic subgroups. *Cancer*, *115*(17), 4021-4029. https://doi.org/10.1002/cncr.24450

Zwing, N., Failmezger, H., Ooi, C.-H., Hibar, D. P., Cañamero, M., Gomes, B., Gaire, F., & Korski, K. (2020). Analysis of Spatial Organization of Suppressive Myeloid Cells and Effector T Cells in Colorectal Cancer—A Potential Tool for Discovering Prognostic Biomarkers in Clinical Research [Original Research]. *Frontiers in Immunology*, *Volume 11 - 2020*. https://doi.org/10.3389/fimmu.2020.550250