

UNIVERSITY OF
BIRMINGHAM



Exploring Immunogenomic Influences on the Microenvironment of Colorectal Cancer

by
Neeraj Lal

**A thesis submitted to the University of Birmingham for the
degree of
DOCTOR OF PHILOSOPHY**

Institute of Immunology and Immunotherapy
College of Medical and Dental Sciences
University of Birmingham
September 2016

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis focussed on the immunobiology of colorectal cancer (CRC). It explored the role of the $\gamma\delta$ T cell ligand Endothelial Protein C Receptor (EPCR) in tumourigenesis, and subsequently characterised the relationship between intra-tumoural immunity and tumour genetics. *In silico* analyses and immunohistochemistry indicated EPCR was commonly overexpressed in epithelial cancers including CRC. EPCR was upregulated due to gene amplification and DNA hypomethylation alongside neighbouring genes on chromosome 20q, a region previously implicated in tumourigenesis. These results clarify why EPCR is upregulated in diverse epithelial malignancies, with implications for EPCR-focussed clinical studies and understanding of $\gamma\delta$ T cell immunity.

TCGA analyses revealed that a novel immune signature, termed The Co-ordinate Immune Response Cluster (CIRC), comprising 28 genes, was co-ordinately regulated across CRC patients. Four patient subgroups were delineated based on CIRC expression. Microsatellite instability and *POLE/POLD1* mutations were associated with high mutational burden and immune infiltration. Immune checkpoint molecules were highly co-ordinated in expression. *RAS* mutation was associated with lower CIRC expression. Further analyses revealed that *RAS*-associated immunosuppression was greatest in the most immunosuppressed transcriptional subtype, CMS2. These findings have implications for design of stratified immunotherapy approaches and highlight factors contributing to the particularly poor outcome of *RAS* mutant CRC.

Acknowledgements

I would first like to thank my primary supervisor Professor Benjamin Willcox who has guided me with great enthusiasm. I am extremely grateful to him for giving me the opportunity to undertake this PhD in his group, and for allowing me to lead and develop my project. I learned a huge amount from our discussions that will benefit me in the future. I would also like to thank Dr Chris Tselepis, my co-supervisor, for his valued guidance and support throughout the years. I'm grateful for his encouragement before and during my time in the department. In addition, I would like to thank Professor Gary Middleton, my other co-supervisor, for his impact on this work. I really enjoyed our discussions about this project and the field, and they have provided me with valuable insights into clinical research and tumour immunology.

In addition to my supervisors, I would like to thank and acknowledge a number of other individuals who have helped me in various aspects of this work. Dr Andrew Beggs provided regular expert help on the bioinformatics and sequencing aspects. Dr Carrie Willcox provided invaluable guidance on my *in vitro* studies in particular. Dr Mahboob Salim, Dr Daniel Stones, Dr Jonathan James and Dr Chris Dawson provided guidance on various techniques. Dr Philippe Taniere was very generous with his time for the optimisation and validation of my histopathology analyses. Christopher Bagnall's expertise in immunohistochemistry proved crucial to this work. I enjoyed my collaborations with Dr Sebastian Boegel and Professor Ugur Sahin on the neoepitope studies, Dr Justin Guinney on the CMS analyses, and Dr Aarti Shikotra and Professor Peter Bradding on the EPCR mast cell studies. I also learned much from working with Professor Hisham Mehanna and his team. Mr Thomas Pinkney was very supportive in my clinical work. Dr Fiyaz Mohammed's help and guidance were much appreciated. I would also like to thank Dr Oliver Pickles, Jo Stockton and Louise Tee for their contributions to the RNAseq work, Dr Ghaleb Goussous for his work on antibody validation, and Dr Martin Davey and the rest of the Willcox and Tselepis groups for their help over the years.

I would also like to acknowledge and thank Cancer Research UK for funding this project. Finally, I would like to thank my wife Ekta, my parents, my parents-in-law and my brother-in-law who have all been hugely supportive over the years.

Table of Contents

CHAPTER 1: INTRODUCTION	1
1.1 Introduction	2
1.1.1 Colorectal cancer development	2
1.1.2 Management of colorectal cancer	6
1.2 Cancer immunology	8
1.3 Cancer immunotherapy	9
1.3.1 Adoptive cell therapy	10
1.3.2 Engineered T cell receptors and chimeric antigen receptors	11
1.3.3 Vaccination	12
1.3.4 Checkpoint blockade	13
1.4 Colorectal cancer immunology	16
1.5 EPCR and the immune recognition of cancer	19
1.6 EPCR structure and signalling	21
1.7 EPCR's role in blood coagulation	23
1.8 Protease Activated Receptors and EPCR signalling	25
1.9 The study of EPCR in cancer	30
1.9.1 <i>In vitro</i> studies	30
1.9.2 <i>In vivo</i> and clinical studies	32
1.10 Project aims	41
 CHAPTER 2: METHODS	 43
2.1 Bioinformatic analyses	44
2.1.1 EPCR bioinformatics	44
2.1.2 Immunity bioinformatics	44
2.1.2.1 Gene identification	44
2.1.2.2 Data extraction	46
2.1.2.3 Data analysis	46
2.1.2.4 Mutation rate analysis	47
2.1.2.5 Neoantigen analysis	47
2.1.2.6 Cancer cell line encyclopaedia analysis	48

2.1.2.7 Consensus Molecular Subtypes (CMS) analysis.....	48
2.2 Specimen collection.....	49
2.2.1 EPCR specimens.....	49
2.2.2 RAS cohort specimens.....	49
2.2.2.1 Sample size calculation and sample collection.....	49
2.2.2.2 Sample processing.....	50
2.3 Immunohistochemistry.....	51
2.3.1 EPCR and PAR1 immunohistochemistry.....	51
2.3.1.1 Immunohistochemistry protocol.....	51
2.3.1.2 EPCR immunohistochemistry scoring.....	52
2.3.2 RAS cohort immunohistochemistry.....	54
2.3.2.1 Immunohistochemistry protocol.....	54
2.3.2.2 RAS cohort Immunohistochemistry scoring.....	55
2.3.2.3 HLA Class II antibody validation.....	57
2.4 Survival analyses.....	59
2.4.1 EPCR survival analysis.....	59
2.4.2 RAS cohort survival analysis.....	60
2.5 shRNA transfection.....	61
2.6 Flow cytometry.....	62
2.6.1 EPCR cell line expression.....	62
2.6.2 EPCR expression in mast cells.....	63
2.7 MTT assays.....	64
2.8 BrdU assays.....	64
2.9 Migration assays.....	65
2.10 Invasion assays.....	66
2.11 Activated Protein C ERK phosphorylation.....	66
2.12 Western blotting.....	67
2.13 Microarray.....	68
2.14 sEPCR ELISA.....	70
2.15 CEA ELISA.....	71
2.16 RNAseq.....	72
2.16.1 RNA extraction.....	72

2.16.2 Truseq panel.....	73
2.17 Microsatellite status.....	76

CHAPTER 3: THE ROLE OF EPCR IN COLORECTAL TUMOURIGENESIS.....78

3.1 Introduction.....	79
3.1.1 The physiological role of EPCR.....	79
3.1.2 EPCR as a ligand for $\gamma\delta$ T-cells.....	79
3.1.3 EPCR in cancer.....	80
3.1.4 Chapter aims.....	81
3.2 Results.....	83
3.2.1 EPCR mRNA expression in multiple cancer types.....	83
3.2.2 EPCR expression in colorectal cancer.....	85
3.2.3 The association between EPCR expression in colon cancer and gene amplification and hypomethylation.....	90
3.2.4 The association between EPCR expression and chromosome 20q amplification.....	94
3.2.5 EPCR expression in colorectal cancer cell lines.....	101
3.2.6 Exogenous Activated Protein C (APC) induction of ERK phosphorylation in colon cancer cells.....	103
3.2.7 Detection of EPCR-related proteins in CRC.....	105
3.2.8 Microarray analysis.....	108
3.2.9 EPCR shRNA knockdown assays.....	114
3.2.10 Serum sEPCR correlation with CEA and clinical stage.....	118
3.2.11 EPCR expression and chemotherapy or cetuximab responsiveness	120
3.3 Discussion.....	123

CHAPTER 4: AN IMMUNOGENOMIC STRATIFICATION OF COLORECTAL

CANCER.....	129
4.1 Introduction.....	130
4.2 Results.....	133
4.2.1 Co-ordinate expression of immune response-related genes in colorectal cancer	133

4.2.2	Molecular determinants of the co-ordinate immune response cluster.....	142
4.2.3	MEK activation signature.....	150
4.2.4	RAS mutant subtype analysis.....	152
4.2.5	Consensus Molecular Subtypes analysis.....	155
4.2.6	The CIRC compared to other immune signatures.....	157
4.2.7	The relationship between immunity and total mutation rates.....	159
4.2.8	Neoantigen prediction analysis.....	171
4.3	Discussion	182
CHAPTER 5: THE IMPACT OF RAS MUTATION ON THE IMMUNE PHENOTYPE OF COLORECTAL CANCER		191
5.1	Introduction	192
5.2	Results	194
5.2.1	Cohort characteristics.....	194
5.2.2	Immunohistochemistry.....	195
5.2.3	Targeted RNAseq panel.....	207
5.2.4	Survival analysis.....	209
5.2.5	Cancer Cell Line Encyclopaedia analysis	211
5.2.6	Consensus Molecular Subtypes analysis.....	213
5.3	Discussion	219
CHAPTER 6: OVERALL DISCUSSION		231
APPENDIX: ASSOCIATION OF IMMUNITY IN PAIRED CRC BIOPSY AND RESECTION SPECIMENS		243
COLLABORATIVE STATEMENT		243
LIST OF REFERENCES		246

List of Figures

CHAPTER 1

Figure 1.1 The genetic model for colorectal tumourigenesis.....	4
Figure 1.2 EGFR pathways.....	7
Figure 1.3 The structure of EPCR and CD1d.....	22
Figure 1.4 PC induced signalling pathways.....	29

CHAPTER 2

Figure 2.1 HLA Class II antibody validation.....	58
Figure 2.2 Summary of RNAseq library preparation workflow.....	75

CHAPTER 3

Figure 3.1 Validation of EPCR antibody.....	86
Figure 3.2 Scanned image of EPCR-stained CRC section.....	87
Figure 3.3 Digital software analysis of EPCR-stained CRC section.....	88
Figure 3.4 EPCR expression in mast cells.....	89
Figure 3.5 PROCR amplification, deletion and mutation rates across a range of cancer types and databases.....	91
Figure 3.6 The association between EPCR expression and <i>PROCR</i> gene amplification and hypomethylation in the TCGA CRC dataset.....	92
Figure 3.7 Relationship between PROCR methylation, GISTIC copy number scores and expression.....	93
Figure 3.8 Copy numbers of chromosome 20 genes in 195 TCGA colorectal cancer patients.....	97
Figure 3.9 PROCR correlations with chromosome 20 genes.....	98
Figure 3.10 Genomic copy number variations across a range of cancers.....	99
Figure 3.11 Expression of EPCR on colorectal cancer cell lines.....	102
Figure 3.12 Western blotting investigating the effect of APC on ERK phosphorylation.....	104
Figure 3.13 Detection of Protein C in HCT116 cell lysate.....	106
Figure 3.14 PAR1 immunohistochemistry in CRC and normal colon.....	107
Figure 3.15 Gene set enrichment analyses.....	113

Figure 3.16 shRNA EPCR knockdown in HCT116 cells.....	116
Figure 3.17 Functional assays after shRNA EPCR knockdown in HCT116 cells.....	117
Figure 3.18 ELISA for sEPCR and CEA on serum from CRC patients.....	119
Figure 3.19 Kaplan-Meier curves for progression free survival in EPCR high and low patients.....	122

CHAPTER 4

Figure 4.1 Two-dimensional hierarchical clustering reveals a closely co-ordinated immunological gene expression cluster.....	136
Figure 4.2 Significant correlation between LAG3 and PDCD1 (PD1) expression.....	139
Figure 4.3 Significant correlation between CTLA4 and PDCD1 (PD1) expression.....	140
Figure 4.4 Correlation between ACTB (Beta Actin) and PDCD1 (PD1) expression.....	141
Figure 4.5 Two-dimensional hierarchical clustering delineates distinct immunological CRC patient groups.....	143
Figure 4.6 Mutation rate by Microsatellite status.....	160
Figure 4.7 CIRC expression by MSS status.....	161
Figure 4.8 Relationship between mutation rate and CIRC expression, by microsatellite and POL subtypes.....	162
Figure 4.9 Variability of CIRC expression in the MSS POLE+POLD1 wild type patient group.....	164
Figure 4.10 Percentage of all tumours within each mutation rate group.....	166
Figure 4.11 CIRC expression in each mutation rate group.....	167
Figure 4.12 CIRC expression in relation to number of mutations per tumour.....	168
Figure 4.13 Percentage of patients with high mutation rates (>15/Mbp) by CIRC expression group.....	170
Figure 4.14 Number of total presented Class I neoepitopes per patient by MSS status.....	173
Figure 4.15 CIRC score in patients with top binding neoantigens versus patients with bottom binding neoantigens.....	175
Figure 4.16 Class I neoantigens analysis.....	176
Figure 4.17 Class II neoantigens analysis.....	177
Figure 4.18 Neoepitope prediction for recurrent KRAS mutations.....	181

CHAPTER 5

Figure 5.1 Cell-based IHC scoring algorithm.....	196
Figure 5.2 Pixel-based IHC scoring algorithm.....	197
Figure 5.3 Stromal CD8 staining.....	200
Figure 5.4 Stromal CD4 staining.....	201
Figure 5.5 Stromal Tbet staining.....	202
Figure 5.6 Epithelial CXCL10 staining.....	203
Figure 5.7 Epithelial PD-L1 staining.....	204
Figure 5.8 Epithelial Class II HLA staining.....	205
Figure 5.9 Epithelial STAT1 staining.....	206
Figure 5.10 Kaplan-Meier curves for overall survival.....	210
Figure 5.11 CIRC expression z-scores in each CMS grouping by RAS mutation status.....	214

List of Tables

CHAPTER 2

Table 2.1 Final gene list for hierarchical clustering analysis.....	45
Table 2.2 List of antibodies.....	54
Table 2.3 Definiens scoring method and thresholds for each marker.....	56
Table 2.4 shRNA clones IDs.....	61
Table 2.5 Taqman Probe IDs.....	68
Table 2.6 Genes analysed in targeted RNAseq panel.....	73
Table 2.7 Microsatellite status PCR reagents.....	76

CHAPTER 3

Table 3.1 ONCOMINE expression data.....	84
Table 3.2 Genes most highly co-expressed with EPCR in TCGA CRC dataset, and their chromosomal locations.....	96
Table 3.3 Correlation of the Carvalho et al. putative chromosome 20q oncogenes and PROCR.....	100
Table 3.4 Genes differentially expressed with APC treatment of HCT116 cells.....	109
Table 3.5 Top 30 gene sets in Gene Set Enrichment Analysis of APC-treated HCT116 cells versus control HCT116 cells.....	112

CHAPTER 4

Table 4.1 Initial gene group for analysis.....	134
Table 4.2 Genes within the co-ordinate immune response cluster (CIRC).....	137
Table 4.3 R ² values correlating expression of inhibitory molecules in colorectal cancer	138
Table 4.4 Characteristics of patient groups.....	144
Table 4.5 Multivariate analysis of gene expression changes in key mutation group.....	147
Table 4.6 Frequencies of mutations in total patient population and distribution of mutations across patient groups A to D.....	149
Table 4.7 Mean MEK signature, CIRC, and other selected immune genes in NRAS mutant, KRAS mutant, and RAS wild type patients.....	151

Table 4.8 Mean CIRC expression within each RAS mutation subtype, and number and percentage of patients per group.....	153
Table 4.9 Expression of key genes by RAS subtypes	154
Table 4.10 Characteristics of each CMS group.....	155
Table 4.11 Distribution of Consensus Molecular Subtype groups across CIRC patient groups.....	156
Table 4.12 Expression of various immune signatures/metagenes by RAS mutation status	158
Table 4.13 Expression of immune signatures by RAS subtype in MSS patients	158
Table 4.14 KRAS subtype neoantigens predicted to present on HLA A and B Figure.....	179
Table 4.15 KRAS subtype neoantigens predicted to present on HLA A and B – selected mutations.....	180
Table 4.16 Analysis of HLA-restriction and presentation frequency of predicted strong binding KRAS G12V-derived neoantigens	180

CHAPTER 5

Table 5.1 Median Histological scores or Percental scores in epithelial regions.....	199
Table 5.2 Median Histological scores or Percental scores in stromal regions.....	199
Table 5.3 Logistic regression of RAS mutant versus wild type cases using a custom targeted TRUSEQ RNAseq panel	208
Table 5.4 Cell line encyclopaedia analysis Figure 5.4 Stromal CD4 staining.....	212
Table 5.5 CIRC, MEK, T cell, cytotoxic cell and iDC signatures by CMS grouping.....	216

List of Abbreviations

µl:	Microlitre
µm:	Micrometre
5FU:	Fluorouracil
APC:	Activated Protein C
APColi:	Adenomatous polyposis coli
BRAF:	v-Raf murine sarcoma viral oncogene homolog B
BrdU:	Bromodeoxyuridine
CAR:	Chimeric Antigen Receptor
CCLE:	Cancer Cell Line Encyclopaedia
CD:	Cluster of Differentiation
CEA:	Carcinoembryonic Antigen
CIMP:	CpG Island Methylator Phenotype
CIN:	Chromosomal Instability
CIRC:	Co-ordinate Immune Response Cluster
CMS:	Consensus Molecular Subtype
CMV:	Cytomegalovirus
COIN:	Clinical trial - COntinuous chemotherapy plus cetuximab or INtermittent chemotherapy with standard continuous palliative combination chemotherapy
CRC:	Colorectal Cancer
CTLA4:	Cytotoxic T-Lymphocyte-Associated Protein 4
DC:	Dendritic Cell
DMEM:	Dulbecco's Modified Eagle Medium
DNA:	Deoxyribonucleic Acid
EGF:	Epidermal Growth Factor
EGFR:	Epidermal Growth Factor Receptor
ELISA:	Enzyme-Linked Immunosorbent Assay
EPCR:	Endothelial Protein C Receptor
eQTLs:	Expression Quantitative Trait Loci
ERK:	Extracellular Signal-regulated Kinase
GSEA:	Gene Set Enrichment Analysis
HLA:	Human Leukocyte Antigen
HRP:	Horseradish Peroxidase
IFN:	Interferon
IHC:	Immunohistochemistry
IL:	Interleukin
KFSYSCC:	Koo Foundation Sun-Yat-Sen Cancer Center dataset
KRAS:	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
mAb:	Monoclonal Antibody
MAP:	Mitogen-Activated Protein
MEK:	Mitogen-activated Protein Kinase Kinase
MHC:	Major Histocompatibility Complex

ml:	Millilitre
mm:	Millimetre
mRNA:	Messenger Ribonucleic Acid
MSI-H:	Microsatellite Instability (high)
MSI-L:	Microsatellite Instability (low)
MSS:	Microsatellite Stable
MT:	Mutant
MTT:	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium
NK:	Natural Killer
NKG2D:	Natural Killer Group 2D
NRAS:	Neuroblastoma RAS viral oncogene homolog
nsSNV:	Non-synonymous Single-Nucleotide Variation
PAR1:	Protease-Activated Receptor 1
PC:	Protein C
PCR:	Polymerase Chain Reaction
PD-1:	Programmed Cell Death Protein 1
PD-L1:	Programmed Death Ligand 1
PD-L2:	Programmed Death Ligand 2
PFS:	Progression-Free Survival
PI3K:	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIK3CA:	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform
PROCR:	Protein C Receptor (Gene encoding EPCR)
PTEN:	Phosphatase and Tensin Homolog
qRT-PCR:	Quantitative Real-Time Polymerase Chain Reaction
RNA:	Ribonucleic Acid
RT:	Room Temperature
sEPCR:	Soluble Endothelial Protein C Receptor
shRNA:	Short Hairpin RNA
TCGA:	The Cancer Genome Atlas
TCR:	T Cell Receptor
Th:	T helper
TNF:	Tumour Necrosis Factor
WT:	Wild Type

Chapter 1

Introduction

1.1 Introduction

This thesis explores the immunobiology of colorectal cancer (CRC). CRC is the fourth commonest cancer in the UK and the second largest cause of cancer-related death [1]. Though there has been an increase in survival over the last 30 years, 5 year survival for metastatic (stage IV) disease remains as low as 12% [2]. It therefore represents a significant challenge for both healthcare systems and for the research community.

1.1.1 Colorectal Cancer Development

CRC development follows a well characterised progression from normal colonic mucosa to adenoma to carcinoma, with known molecular aberrations responsible for each step (Figure 1.1). This paradigm was first proposed by Fearon and Vogelstein [3].

Mutation in the Adenomatous Polyposis Coli (APC_{Coli}) gene is a key early step in CRC development [4]. This gene is crucial in the degradation of β -catenin, which is a component of the Wnt signalling pathway. Wnts produced by stromal cells maintain the proliferating undifferentiated stem cell population at the base of colonic crypts (the functional subunits of the colon). As stem cells migrate up the crypt, they receive fewer Wnt signals and therefore differentiate, cease proliferating and eventually apoptose. Following APC_{Coli} mutation, β -catenin levels remain high even in the absence of Wnt signalling, due to the decrease in its degradation. Therefore cells with APC_{Coli} mutation remain undifferentiated and do not migrate normally up the crypt. This leads to the formation of an adenomatous polyp. Mutations of APC_{Coli}

may be inherited, causing familial adenomatous polyposis (FAP), a familial syndrome that leads to the development of a large number of colonic polyps. However, the majority of CRCs are associated with a spontaneous mutation of APColi or another defect in the Wnt signalling pathway such as β -catenin mutation [4].

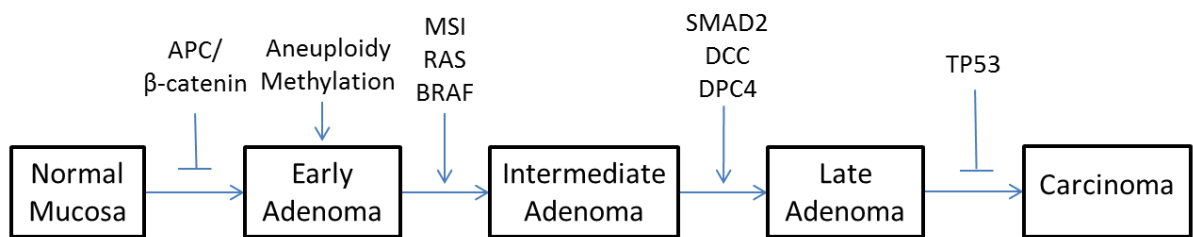


Figure 1.1 The genetic model for colorectal tumourigenesis (adapted from Fearon and Vogelstein [3]), showing the key mutations and abnormalities responsible for progression from normal mucosa to carcinoma.

Following APC_{Coli} mutation and the development of an adenoma, further mutations occur which drive the progression to invasive carcinoma [3]. These can include mutations in the *KRAS*, *NRAS*, *HRAS*, *PIK3CA*, *PTEN*, *BRAF*, and *TP53* genes. Biologically, CRCs are of one of two major subtypes – microsatellite unstable (MSI-H) and microsatellite stable (MSS) [5]. MSI-H tumours, as seen in the hereditary condition Lynch syndrome and in a proportion of sporadic tumours, are hypermutated. This hypermutation is caused by defects in the DNA mismatch repair genes (including *MLH1*, *MLH2* and *MSH2*). Conversely, MSS tumours have low mutation rates but have a high degree of chromosomal instability, leading to aneuploidy and large gene copy number variations. These two subtypes tend to have differing clinical courses. MSI-H tumours tend to metastasise less frequently than MSS, and produce a stronger intra-tumoural immune response [5]. In addition, the overall patient survival is superior with MSI-H tumours. However, patients with MSI-H tumours that recur have a poor prognosis [6, 7].

The mutation profile of a tumour has implications not only for its progression but also for its sensitivity to emerging targeted therapies. One particularly important pathway in this regard is the Epidermal Growth Factor Receptor (EGFR) pathway [8, 9] (Figure 1.2). Mutations in the tyrosine kinases RAS and RAF can lead to constitutive activation of the pathway and MEK/ERK phosphorylation, even in the absence of EGFR activation by Epidermal Growth Factor (EGF). This pathway is particularly important in cancer cell growth, and is therefore targeted in many emerging therapies. Mutations within these genes have therapeutic implications. For example, the EGFR monoclonal antibodies cetuximab and panitumumab are ineffective in

patients with RAS mutant tumours, as inhibition of EGFR alone fails to inhibit RAS signalling [10].

1.1.2 Management of colorectal cancer

Currently, the management of CRC is dependent on the stage of disease at presentation, and whether the disease is colonic or rectal [11]. In brief, early stage I and II disease that is localised to the colon is usually treated solely with surgical resection. Patients with stage III disease (with lymphatic metastases) may also require adjuvant (post-operative) chemotherapy for 6 months. Finally, the treatment of patients with stage IV disease (distant metastases) is dependent on whether the intent is curative or palliative. Curative intent may be feasible if the metastases are surgically resectable. Otherwise, palliative treatment is offered which may consist of primary chemotherapy combined with targeted therapy, such as with EGFR monoclonal antibody therapy. In contrast, rectal cancers may require preoperative chemoradiotherapy to downstage the tumours and enable surgical resection.

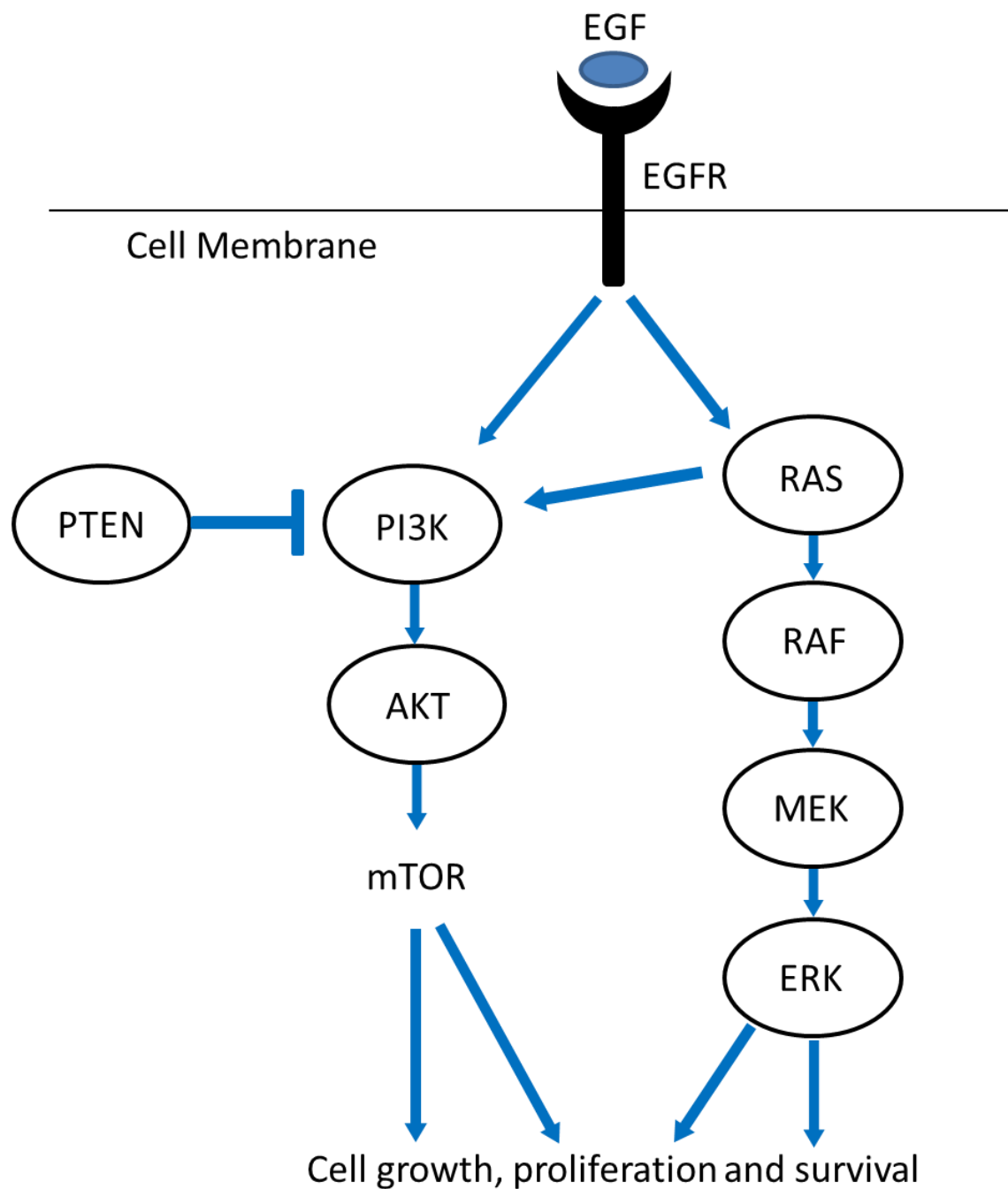


Figure 1.2 EGFR pathways. Schematic of key EGFR-associated signalling pathways. EGFR activation (by its cognate ligand epidermal growth factor (EGF)), leads to activation of key signalling cascades including the RAS-RAF-MEK-ERK and PI3K-AKT-mTOR pathways. This leads to increased cell growth, proliferation and survival. PTEN negatively regulates the PI3K-AKT-mTOR pathway. In malignancy, mutations can occur in several of these genes, including *EGFR*, *RAS*, *RAF*, *PI3K* and *PTEN*, which can lead to constitutively active signalling in the absence of EGFR activation.

1.2 Cancer Immunology

Cancer immunology is the study of the interactions of cancer cells and the immune system. It is a rapidly developing field, and recent therapeutic advances have given hope that cancer immunotherapy will become a key treatment modality in a range of cancer types [12-16].

The immune system has long been perceived to play an important role in preventing the development of tumours – the concept of immunosurveillance – where the immune system constantly surveys the body for developing tumours and eliminates them before they become clinically apparent [17, 18]. Immunosurveillance is thought to comprise three phases – the elimination phase, the equilibrium phase and the escape phase. In the elimination phase, the immune system identifies and destroys cancerous cells. In some cases, where elimination is not complete, this is followed by the equilibrium phase. During this phase, the immune system maintains residual cancerous cells in a state of dormancy, which can persist for many years, such as in the case of the latent tumour cells that lead to recurrent or metastatic disease [19]. During this dormancy, the selection pressure created by the immune response encourages the growth of tumour cells that are able to avoid immune killing – a process also known as immunoediting [18]. This process may lead to the development of tumour cells that are resistant to the immune attack. These cells may then proliferate and dominate the tumour mass – the escape phase. The concept of immunosurveillance remains controversial in the field [18]. However, it is clear that tumour cells escape immunity through several established mechanisms and potentially some unknown ones. Some of the key mechanisms include down-regulation of Class I MHC expression, repression of tumour antigen expression and

presentation, the release of suppressive cytokines (such as IL-10, TGF- β), and the repression of NKG2D ligands (to evade NK attack) [17]. Subsequently, the immune system may no longer effectively eliminate tumour cells. Immunotherapy strategies need to overcome this loss of tumour immunogenicity – a challenge made considerably more difficult by the fact that tumour cells are effectively ‘altered self’, rather than foreign such as infectious pathogens.

1.3 Cancer immunotherapy

Cancer immunotherapy involves the harnessing of immunity for anti-cancer effect. Various approaches are in development, some of which are now used in patient care. Briefly, these include approaches that induce or increase neoantigen (mutated antigens) or tumour-associated antigen directed T cell responses (such as vaccination, chimeric antigen receptors (CARs) and adoptive T cell therapies), strategies that induce stimulatory signalling and proliferation (such as IL-4, GM-CSF and TNF-based (cytokine) therapies), modalities that utilise native or engineered viruses that kill cancer cells (oncolytic virus therapy) and finally methodologies that inhibit suppressive pathways or restore responses of exhausted/inactivated T cells through the inhibition of immune checkpoints (checkpoint blockade) [17, 20, 21]. In addition, conventional cytotoxic regimes may improve the anti-tumour immune response by releasing tumour neoantigens and modifying the immunosuppressive environment of the tumour [22-24]. Here, I introduce some of the key emerging immunotherapy modalities.

1.3.1 Adoptive cell therapy

Adoptive cell therapy exploits the anti-tumour activity of lymphocytes to eradicate tumour cells [20]. It requires the isolation of lymphocytes from the patient's peripheral blood, tumour tissue or draining lymph nodes, followed by *ex vivo* expansion and stimulation of the lymphocytes with a mixture of cytokines, followed by reinfusion back into the patient. This approach expands tumour-specific T cells away from the immunosuppressive tumour microenvironment. The lymphocytes reinfused are generally mixtures of CD4 and CD8 positive T cells. It is often administered following lymphodepletion, which may improve the functionality of the infused T cells by eliminating immunosuppressive cells such as Tregs and myeloid-derived suppressor cells [25].

The key advantage with this strategy is that it enables the generation of a large number of high avidity T cells (avidity is the accumulated binding strength of the TCR with the peptide-MHC complex, determined by the accumulated strength of multiple individual interactions/affinities). Also there is no need to break tolerance against tumour antigens as is often the case with other approaches [20]. Disadvantages include the cost and time required to isolate and expand the T cell populations. Furthermore, the side effects, particularly with lymphodepletion, can be life-threatening.

In addition, this approach has been successful mostly in melanoma thus far. However, results in this tumour type have been promising, with high response rates and long-lasting tumour regression [26]. Additionally, a durable response was

obtained in a patient with cholangiocarcinoma on infusion of autologous mutation-specific CD4⁺ cells which adopted a poly-functional Th1 phenotype [27].

1.3.2 Engineered T cell receptors and Chimeric Antigen Receptors

There are two main strategies through which T cells are engineered to improve their anti-tumour activity. The first is TCR engineering. In this approach, the TCR α and β chains are designed or modified to increase the TCR's antigen specificity and avidity [20]. The synthetic TCR is then expressed by the T cell [21].

The second approach is the creation of chimeric antigen receptors (CARs). CARs use a single-chain variable fragment (scFv) derived from the variable heavy and light chains of an antibody to target an extracellular antigen [21]. The advantage of this approach is that it can target antigens independently of the peptide-HLA complex [28]. CARs have evolved significantly since the generation of the initial functional CAR T cells. T cell activation in the lymph node requires both 'signal 1', normally derived from the TCR-peptide-MHC interaction, and 'signal 2', by stimulation of CD28 with B7 costimulatory molecules [20]. First generation CARs only provide activation of signal 1 to T cells [21]. Therefore, first generation CARs could lead to T cell anergy, due to repeated antigen stimulation without signal 2. Second generation CARs address this by including an additional co-stimulatory domain that provides signal 2, such as the stimulatory domains of CD28 [21]. Third generation CARs include two co-stimulatory domains, such as CD28 and OX40. Finally, 'armoured' CAR T cells include a further gene that provides the resulting T cell with a survival or

cytotoxicity advantage, or modulates the microenvironment. This may include IL-12 or CD40L.

Responses have been seen with genetically engineered T cell therapies in various cancers. However, the clinical use of highly avid TCRs have been associated with off-target effects on healthy tissues that express the same target antigen [20].

1.3.3 Vaccination

Vaccination in cancer includes preventative vaccines against cancer-causing infectious diseases such as Hepatitis B or the Human Papilloma Virus, and therapeutic vaccines which are being developed to treat cancers. One of the main obstacles to the latter approach has been the identification of suitable antigens. The ideal target antigens should be expressed at high levels in the target tumour in a high proportion of patients, and at low levels in normal tissue to minimise off-target effects [21]. In addition, the antigen should be important for the tumour's growth or survival, to reduce the risk of immune escape due to downregulation of antigen expression [29].

Suitable antigens may include tumour antigens, which are expressed at much higher levels in tumour tissue than normal tissue, such as NY-ESO-1, MUC-1 and CA-125 [20, 21]. A further possibility is the targeting of neoantigens, which derive from the peptides encoded by tumour mutations. An issue with this approach is that it is unclear to what extent neoantigens overlap between patients. Therefore, efforts have been made to develop personalised vaccines, which target the specific neoantigens found in an individual patient's tumour [16]. This requires genetic

sequencing of the patient's tumour, *in silico* prediction of neoantigen-HLA binding, and then vaccine production – in this case a polytope mRNA vaccine (containing 5 epitopes). This approach is being investigated in an early phase clinical study. In addition, both tumour antigens and neoantigens may be utilised in cell therapies.

Dendritic cells (DCs) play a crucial role in vaccination approaches, which initial vaccine therapies did not fully exploit. DCs, as professional antigen presenting cells, process the delivered antigens and present them upon Class II HLA molecules to CD4 T cells. For this reason, attempts have been made to create DC-based vaccines, where DCs are isolated from patients and are loaded with antigens *ex vivo* [21]. One such DC vaccine, sipuleucel-T, has been associated with a 4-month median survival improvement in metastatic prostate cancer [30, 31].

1.3.4 Checkpoint blockade

Of the various immunotherapies in development, checkpoint blockade has been the most successful thus far [14, 15]. The rationale of this therapy is to unleash pre-existing anti-tumour responses. CTLA-4, an inhibitory receptor expressed on T cells, prevents T cell activation by DCs through its interaction with the B7 co-stimulatory molecule, blocking the crucial 'signal 2' discussed above [20]. Ipilimumab, an anti-CTLA-4 antibody, which has been approved for clinical usage in advanced melanoma, was a significant step forward for the field of immunotherapy. This approval followed phase III trials that demonstrated significant prolongation of overall survival [12]. In addition, 20% of patients who survived went on to have long term durable benefit, which seems to be a particular strength of checkpoint blockade in

comparison to targeted therapies. However, response rates were low (12%), and a significant proportion of patients developed severe immune-related adverse events [21]. Of note, responses to this drug took longer to manifest than existing therapies, reflecting its mode of action.

PD-1 is another inhibitory receptor expressed on T cells. The ligands for PD-1, PD-L1 and PD-L2, are expressed on tumour cells and antigen presenting cells [21]. The downstream signalling of PD-1, rather than preventing the activation of T cells like CTLA-4, inhibits T cell proliferation, cytokine release, and cytotoxicity [32, 33]. Therefore, CTLA-4 regulates *de novo* immune responses, whereas PD-1 affects ongoing immune responses. Two anti-PD-1 mAbs, pembrolizumab and nivolumab, have been approved for the treatment of advanced melanoma. Clinical trials showed that up to 40% of patients with advanced melanoma had objective responses with these drugs – a considerable improvement over the 12% response rate observed with ipilimumab monotherapy [34, 35]. This included patients who previously had no response to ipilimumab.

Nivolumab has also been approved for treatment of squamous-cell lung cancer refractory to platinum based therapy, based on a study showing a 3 month improvement on overall survival, and a 17% improvement in 2 year survival over those receiving docetaxel [15].

A key improvement of PD-1 over CTLA-4 blockade is the toxicity profile. The rates of severe toxicity were considerably reduced with PD-1 therapy [12, 14, 35], perhaps reflecting the benefit of specifically targeting cancer-induced immunosuppressive pathways rather than non-specifically activating the immune system [20].

Monoclonal antibodies have also been developed against PD-L1 and PD-L2, which have shown responses in a range of human cancers [20]. In addition, other immune checkpoints such as TIM-3 and LAG-3 have also been targeted [21].

A key priority in the field is the development of biomarkers for these therapies. Currently, PD-L1 is used as a biomarker for PD-1 therapy. However, patients with low PD-L1 expression can occasionally respond to PD-1 blockade, and conversely, patients with high PD-L1 expression may not respond [20], demonstrating that PD-L1 may be a suboptimal biomarker.

Due to these highly promising results, checkpoint blockade therapies are being investigated in a range of cancer types. However, results in CRC have been relatively disappointing [14, 36], despite the fact that CRC is a tumour type where intratumoural immunity is known to be predictive for patient outcome [37].

1.4 Colorectal cancer immunology

Though immunotherapies have had limited success in CRC thus far, this disease has been the prototype cancer for studying the significance and the impact of tumour immunity and the tumour microenvironment [38]. Over a decade ago, multiple groups demonstrated that tumour infiltrating lymphocytes (TILs) are associated with improved prognosis in CRC [39-41]. Then, a seminal paper by Galon and colleagues in 2006 revealed that the density of TILs and the expression of certain immune-related genes are of prognostic and predictive value in CRC [37]. The group identified a dominant cluster of genes for Th1 adaptive immunity, the expression of which were inversely associated with tumour recurrence. Furthermore, immunohistochemistry analysis of CRC tissue microarrays revealed that densities of CD3, CD8, GZMB and CD45RO in the tumour core (CT) and invasive margin (IM) could stratify patients into groups with different disease-free survival and overall survival rates. Combining the CT and IM scores increased the accuracy of the survival prediction. CD3 density in these regions remained associated with overall survival after multivariate analysis adjusting for tumour and lymph node stage. This led to the development of the 'Immunoscore', which is based on CD3 and CD8 densities within the CT and IM, and is currently undergoing worldwide validation as a prognostic marker in CRC [38]. These two markers were selected as they are the two easiest membrane stains, and to avoid the technical difficulties with CD45RO and GZMB staining. The aim of the validation is to ensure standardisation of the assay, which is crucial to ensure that this test can be applicable on a worldwide multicentre basis. The relative simplicity of the test, requiring only two IHC stains, is likely to aid its implementation and reproducibility. However, though it is prognostic

in CRC, and is potentially superior to the existing TNM classification in predicting outcome, it remains unclear how well it will predict responses to chemotherapy. It is likely that this will be crucial in determining its clinical utility in the stratification of patients into treatment groups, rather than as an additional prognostic marker with only limited value. This will be particularly relevant to Stage II patients, where the benefit of adjuvant chemotherapy is unclear [38]. Predicting which patients have a high recurrence risk could help target chemotherapy appropriately in this group. To determine its predictive value, clinical trials will likely be necessary.

In their key 2006 paper, Galon and colleagues initially highlighted a cluster of 7 genes that were correlated with disease-free survival [37] – *TBX21/Tbet*, *IRF1*, *IFN γ* , *CD3z*, *CD8*, granulysin and granzyme B. Subsequently, they further confirmed that the Th1 genes, *TBX21/Tbet*, *IFNG*, *IRF1* as well as *STAT1* were all individually associated with disease-free survival [42]. Other independent immune gene predictors in this analysis included *IL18RAP*, *ICOS*, PD-L1, PD-L2 and PD-1. Additionally, the group found that patients with high expression of genes associated with Th1 T helper cells (*TBX21/Tbet*, *IRF1*, *IL12RB2* and *STAT4*) exhibited prolonged disease-free survival, whereas those with high expression of Th17-associated genes (*RORC*, *IL17A*) had a poor prognosis [43]. This highlights the potential significance of Th1 CD4 cells, alongside the previously discussed cytotoxic CD8 cells, in mediating effective anti-tumour immunity.

Despite the known impact of immunity on patient outcome in CRC, the factors that determine a patient's immune phenotype are still unclear. These are likely to include tumour, host and environmental factors [44]. Within tumour and host factors, few systematic analyses have investigated the somatic and germline molecular drivers of

immune infiltration. Although microsatellite unstable (MSI-H) cancers are known to be associated with increased TIL density [5, 45-50], the nature of the immune infiltration and the molecular drivers of the immune phenotype in microsatellite stable (MSS) CRCs are poorly understood. In particular it is unclear whether defined molecular subsets (such as *RAS* mutant, *BRAF* mutant, *PIK3CA* mutant, quadruple wildtype (*BRAF*, *PIK3CA*, *NRAS*, *KRAS* all wildtype)) are associated with high or low immune infiltration. In addition, in both MSI-H and MSS cancers the extent to which therapeutically targetable inhibitory immune checkpoint receptors are represented are unclear and is of substantial interest, particularly considering recent checkpoint blockade failures in colorectal cancer [13, 14]. Given the prognostic and predictive relevance of CRC immunophenotype, a clearer understanding of the link between immunophenotype with tumour genotype is crucial.

1.5 EPCR and the immune recognition of cancer

As discussed above, the conventional adaptive immune response plays an important role in preventing the development and progression of cancers. However, the role of the unconventional immune response has yet to be fully explored. Recently, $\gamma\delta$ T cells, an unconventional T cell subset, have been suggested to correlate closely with prognosis across different cancers [51]. $\gamma\delta$ T cells are less well understood than conventional $\alpha\beta$ T cells and are thought to have an innate-like function similar to NK cells [52]. One of the primary aims of this project is to determine the significance of Endothelial Protein C Receptor (EPCR), a $\gamma\delta$ T cell ligand, in epithelial tumourigenesis. EPCR had no known immunological role until a recent study by the Willcox and Déchanet-Merville groups [53], who aimed to identify possible ligands for the $\gamma\delta$ T cell receptor. $\gamma\delta$ T cells have important roles in bacterial and viral infection. However, they are also thought to be important in cancer immunosurveillance. Mice deficient in $\gamma\delta$ T cells have increased rates of carcinogenesis. Expansion of this set of T cells in blood is associated with better outcome in cancer in humans [54]. As $\gamma\delta$ T cells are not MHC/HLA-restricted, the HLA haplotype of patients is not relevant. This makes them attractive for potential immunotherapies. Early stage clinical studies involving $\gamma\delta$ T cells in cancer have shown promising results [54-57]. However, the manner in which the V δ 2-negative $\gamma\delta$ T cell receptor (TCR) in particular recognises infected or transformed cells had not been determined.

The work stemmed from a patient who underwent lung transplantation and developed acute cytomegalovirus (CMV) infection whilst on immunosuppressive medication. In this patient, a $\gamma\delta$ clone with a specific TCR called LES had expanded and comprised 40% of blood $\gamma\delta$ T cells and 20% of all blood T cells. It had

previously been suggested that V δ 2-negative $\gamma\delta$ T cells could recognise CMV infected cells and tumour lines [58]. Therefore, the authors hypothesised that the TCR of the LES clone recognised a ligand present on both CMV infected cells and transformed cells.

The group created cDNA from the LES TCR's V γ 4 and V δ 5 chains, and induced expression of this TCR in JRT3 cells (which lack a TCR). These cells now reacted to several cancer cell lines (including HT29) and CMV infected cells. Antibodies were produced against the LES TCR's ligand by immunising mice with HT29 colorectal cancer cells. One antibody, 2E9, stained all LES targets, including the tumour lines, after which it also inhibited all LES recognition. To identify the ligand of 2E9 (2E9L), the group immunoprecipitated proteins from 2E9L positive cells (including HT29 and HeLa). They separated the proteins with SDS-PAGE and then trypsinised the 2E9 reactive band. Using mass spectrometry, they were able to identify the protein as EPCR.

The group confirmed that EPCR and 2E9L were the same. EPCR antibodies prevented 2E9 binding in previously positive cells. Using surface plasma resonance, the group were able to confirm direct and specific binding between EPCR and 2E9. Through mutational analysis, they identified specific binding sites. Significantly, EPCR did not have an antigen-presenting role like MHC but directly bound to the TCR. However, EPCR alone was necessary but insufficient for TCR activation.

In view of the role of $\gamma\delta$ T cells in cancer immunosurveillance, the discovery of EPCR as a ligand in a specific subset was significant. EPCR is not upregulated upon CMV infection, but EPCR's upregulation on various tumour lines [59] suggested the

possibility that it may have relevance for $\gamma\delta$ T cell recognition of cancer as part of a “multimolecular stress signature”. This highlighted the role of EPCR in tumorigenesis as an interesting avenue to explore.

1.6 EPCR structure and signalling

EPCR is a type I transmembrane protein with 20% homology to MHC class I and CD1 (Figure 1.3), though it lacks an $\alpha 3$ domain [60]. It was initially described in relation to Protein C (PC), an important modulator of coagulation [61]. EPCR increases the conversion of the zymogen PC to Activated Protein C (APC), and is also a receptor for both PC and APC. APC cleaves coagulation factors V and VIII and therefore down modulates coagulation [60].

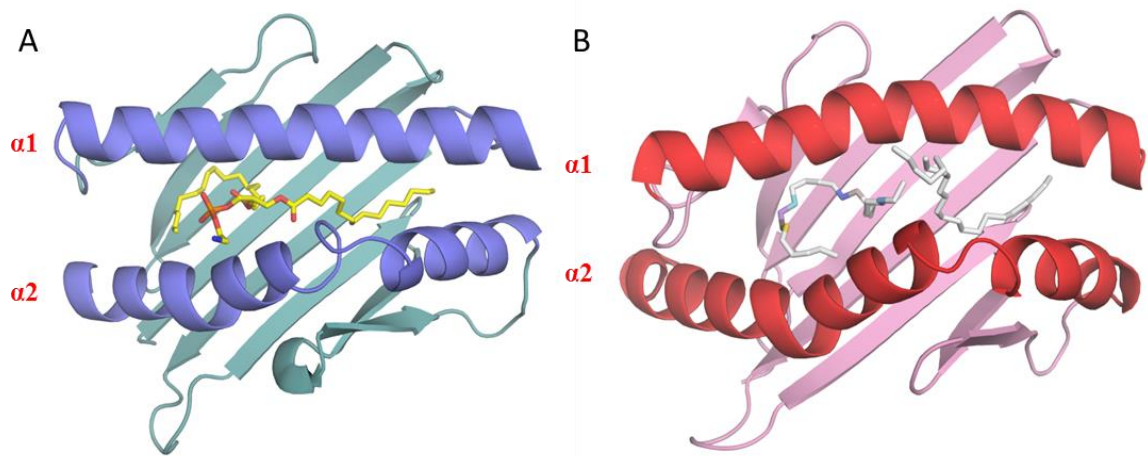


Figure 1.3 The structure of a) EPCR and b) CD1d [62, 63]. The $\alpha1$ and $\alpha2$ platforms are highly homologous. The $\alpha3$ domain of CD1d has been omitted for clarity. The lipids within the binding grooves are P-Phosphatidylmethanolamine or phosphatidylcholine (EPCR) and α -galactosylseramide (CD1d).

Figure courtesy of Dr Fiyaz Mohammed.

1.7 EPCR's role in blood coagulation

EPCR's most significant role is in preventing the excess clotting of blood. Seegers and collaborators initially identified autoproteolytic II-A, now termed Protein C, in the 1960s [64]. The discovery of PC partially explained the remarkable capacity of the haemostatic system to clot in areas of endothelial injury but maintain sufficient flow elsewhere. Esmon and colleagues at the University of Oklahoma (USA) have made significant contributions to the understanding of PC and its pathways, and have expanded the known roles of PC from the inhibition of clotting factors to the wide range of functions known today. It became apparent that whilst PC activation requires thrombin, the rate of activation with thrombin alone is insufficient for effective coagulation modulation [65]. Esmon postulated that there could be an important cofactor in the vascular endothelium. Prior to this, endothelium was thought to only play a passive role in clotting. This was confirmed by perfusing an isolated rabbit heart's coronary arteries with thrombin and PC. When both of these were perfused, the effluent from the heart possessed anticoagulant activity. Without perfusing through the vessels, the thrombin and PC failed to produce anticoagulant activity. When the perfusate contained only thrombin, the effluent failed to activate PC. This experiment was crucial in demonstrating that the endothelium was central to the prevention of potentially harmful clotting.

It took several more years for the group to identify EPCR [61], and demonstrate that it directly activates PC. In the interim period, a study suggested that PC protected against *E. coli* septicaemia in baboons [66]. Therefore, it seemed increasingly likely that PC affected inflammation and intracellular signalling via an unknown receptor. The Esmon group showed that APC bound to Human Umbilical Vein Endothelial

Cells (HUVECs) [61]. From here, the group were able to determine the likely final primary structure of the APC-binding protein, which they termed Endothelial Protein C Receptor. They correctly predicted that it was a type 1 transmembrane protein, with 4 glycosylation sites in the extracellular domain and a short cytoplasmic tail. This cytoplasmic tail is significant in the cellular distribution of EPCR [60]. Interestingly, the sequence of EPCR is highly homologous to the CD1 family (Figure 1.3). Esmon and colleagues did make an astute observation that CD1d promotes T cell responses, and therefore EPCR could interact in some manner with the immune system. Although EPCR lacks the $\alpha 3$ domain of CD1, the $\alpha 1$ and $\alpha 2$ domains are highly analogous. The binding of PC to HUVECs was compared to several other cell lines [61]. Esmon noted that the osteosarcoma cell line HOS and the epidermoid carcinoma line HEph-2 both bound to APC, albeit much more weakly than the HUVEC line. The HOS line and the monocyte cell line U937 both had low levels of detectable EPCR mRNA. These findings may have been the earliest indications of EPCR expression in cancers. However, clearly the emphasis of Esmon's work was to investigate the receptor structure and function in terms of coagulation and inflammation, perhaps explaining why EPCR's expression in those cancer cell lines was not investigated further. Esmon correctly suggested that EPCR would be involved in intracellular signalling, and that the EPCR/APC complex might activate other proteins. However, at that stage, it was not linked directly to protease activated receptor 1 (PAR1), a type of proteolytically-activated receptor that was identified later [67]. Thus, although it was clear that EPCR significantly augmented the activation of PC, the intracellular signalling initiated as a result of EPCR activation was poorly understood.

1.8 *Protease Activated Receptors and EPCR signalling*

Protease activated receptors (PARs) are group of receptors activated by proteolytic cleavage, and are crucial for EPCR-mediated signalling. The first to be described was PAR1, which is activated by thrombin and factor X [67]. The Ruf group utilised PAR1 deficient murine fibroblasts to determine whether APC had proteolytic activity upon PARs [68]. PAR1 deficient fibroblasts were responsive to APC only when EPCR was co-expressed with a PAR. PAR1 cleavage-blocking antibodies prevented APC-induced intracellular signalling but not that induced by the direct PAR1 agonists. Blocking PAR1 also inhibited any PAR2 signalling. This was evidence that the APC/EPCR complex activated PAR1 by proteolytic cleavage. This work provided the important insight that PC mediated intracellular effects, including anti-apoptotic and anti-inflammatory effects, are entirely dependent on PAR1 activation. Gene induction by APC and direct PAR agonists was highly concordant. Anti-apoptotic genes that were upregulated by APC included the Bcl-2 related protein A1 and inhibitor of apoptosis protein 1 (IAP1). Other upregulated genes include negative regulators of multiple pro-inflammatory pathways, including mitogen activated protein kinase (MAPK) and the nuclear factor kappa B (NF-kb) pathways. The gene profile matched closely with the expression profile of PC defined in an earlier study [69]. The Ruf group's work substantially progressed the understanding of PC function by linking EPCR and PC to PAR1, a receptor that can initiate a diverse set of intracellular pathways.

Grinnell and collaborators utilised RNA microarrays to identify the broad range of genes modulated by APC [69]. However, the group did not link APC's effects with EPCR or PAR1, as this was established subsequently [68]. Yet, the work by Grinnell

was the earliest evidence that the PC/APC/EPCR pathway had a significant impact on gene transcription, independently of APC's inhibition of thrombin production, and explained the *in vitro* and *in vivo* results found in the sepsis setting, where APC was found to have an anti-inflammatory effect [66, 70, 71]. The genes regulated by APC were related to inflammation, apoptosis, cell survival and cell adhesion (which is significant in early cancer metastasis [72]). The inflammatory gene NF- κ B2 was downregulated by APC. Genes related to cell adhesion included intracellular adhesion molecule-1 (ICAM-1), vascular cell adhesion molecule 1 (VCAM-1) and E-Selectin. Two pro-apoptotic genes were downregulated by APC – calreticulin and TRPM-2. Anti-apoptotic genes that were upregulated included A1 and IAP. All of these genes have potential relevance in the cancer setting, although APC and EPCR were not thought to be relevant in cancer at that time. Physiologically produced PC could increase cell and organ survival. *In vitro* assays using Staurosporine (an apoptosis inducer) confirmed that recombinant APC (rhAPC) reduced the proportion of cells undergoing apoptosis, as confirmed by caspase-3 staining. This effect was dose dependent, which in light of the now known signalling pathway, suggests progressively greater EPCR and PAR1 signalling with higher doses of APC. With this work, APC's potential as an anti-inflammatory and anti-apoptotic agent were substantiated with transcriptional data.

Other than A1 and IAP, other pro- and anti-apoptotic proteins have been found to be modulated by EPCR, including Bax and Bcl-2 [73]. Cheng and collaborators investigated the effects of APC in ischaemic brain injury. Hypoxia induced apoptosis in endothelial cells through an increase in p53 and subsequent alteration of the Bax/Bcl-2 ratio. APC produced a cytoprotective effect in hypoxic conditions by

reducing p53 (by 77% in 2 hours) and altering the Bax/Bcl-2 ratio to reduce apoptosis. APC reduced TUNEL and caspase-3 (a key executioner caspase) staining. This was independent of thrombin effects. PAR1 and EPCR-blocking antibodies completely inhibited this effect, but non-blocking EPCR antibodies did not prevent the effect. IAP or A1 expression was unaltered, which differs from Joyce's findings in HUVECs [69]. The finding that APC/EPCR/PAR1 can modulate p53 is particularly significant with regard to cancer. p53 is a well-described and crucial tumour suppressor protein. EPCR reducing p53 activity could potentially affect apoptosis in cancer. Therefore it is feasible that EPCR expression on cancer cells could confer an intrinsic advantage on them.

In addition to reducing apoptosis, Xue et al. provided *in vitro* evidence that EPCR activation can induce cellular proliferation [74]. After demonstrating EPCR expression in skin keratinocytes using immunohistochemistry, they established that APC stimulation of keratinocytes causes phosphorylation of ERK2. This phosphorylation was dependent on both EPCR and PAR1, as confirmed by the lack of phosphorylation when using blocking antibodies. ERK2 is an example of a mitogen-activated protein kinase (MAPK), which is important in growth factor signalling via the EGFR pathway. The discovery that EPCR activation leads to MAPK phosphorylation suggests that it could have a role in cell proliferation in endothelium and skin. Alongside the anti and pro-apoptotic genes that are regulated by EPCR [69], this is another mechanism by which EPCR could encourage uncontrolled growth in cancer.

Several EPCR-activated pathways overlap with pathways known to be important in cancer, including the ERK pathway. Figure 1.4 provides a summary of EPCR-associated signalling.

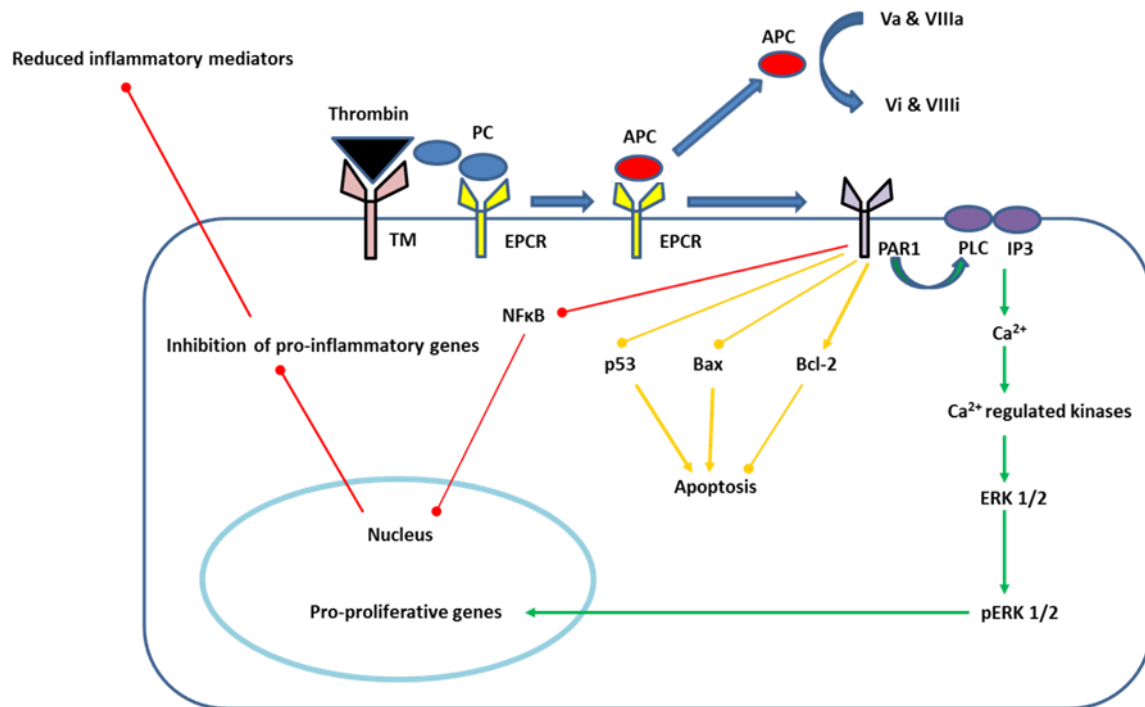


Figure 1.4 PC induced signalling pathways (adapted from Thiyagarajan et al. 2007 [75]).

The thrombin/thrombomodulin (TM) complex presents PC to EPCR. PC is then activated (APC). APC inactivates factors V and VIII. In addition the APC/EPCR complex cleaves PAR1, a G-protein coupled receptor. This leads to activation of phospholipase C. Inositol triphosphate (IP3) mediates influx of calcium, causing activation of kinases and phosphorylation of ERK (green arrows). This induces expression of pro-proliferative genes in the nucleus. Additionally, PAR1 activation downregulates p53 and alters the Bax/Bcl-2 ratio, thus inhibiting apoptosis (yellow arrows). PAR1 also mediates APC's anti-inflammatory effect through downregulation of NFkB and inhibition of pro-inflammatory genes (red arrows).

1.9 The study of EPCR in cancer

1.9.1 *In vitro* studies

The earliest evidence that EPCR may have a role in cancer came from Fukudome and collaborators in 2001 [76]. The group approached their work on the basis of PC's and EPCR's anticoagulant activity. Their initial aim was to explain the coagulopathic state in cancer. Around the same period, other groups were investigating EPCR and APC's anti-inflammatory effects [68-70]. Using techniques including immunohistochemistry and western blotting, the Fukudome group determined that EPCR was expressed on a range of glioblastoma and haematological cancer cell lines. Importantly, they demonstrated that PC activation was inhibited by an EPCR-blocking antibody, but not by an antibody that allowed EPCR/PC binding. Positive EPCR staining was demonstrated on paraffin-embedded sections of invasive ductal breast carcinoma (96% of samples). The group postulated that the high frequency of EPCR expression in these aggressive cancers may have benefitted tumour progression. Yet at the time very few viable explanations for this observation could be provided. Nevertheless, by closely scrutinising for factors that explained coagulopathy in cancer, the group identified aberrant expression of EPCR, a protein that we now know has pleiotropic effects [69].

The Scheper group suggested that EPCR may have a role in chemoresistance [59]. The group generated a panel of antibodies against proteins highly expressed in multi-drug resistant (MDR) cancer cell lines. One of these antibodies, LMR-42 mAb (monoclonal antibody), reacted with a protein that was induced by selecting doxyrubicin resistant cancer cell lines. The group noted that LMR-42 mAb reacted

with several MDR lines, but not their parental lines nor non-cancer cell lines. The highest expression was in cell lines most poorly responsive to chemotherapy. Of note, the commonly used HCT116 colon cancer line was strongly positive. IHC staining with the antibody was positive on endothelium and some cancer tissue. The group deduced through sequence analysis that LMR-42 was identical to EPCR. This work demonstrated that EPCR was associated with chemoresistance, without providing a mechanistic explanation for this effect. EPCR expression was a stable trait in these cells even after several weeks without doxorubicin treatment, and was therefore not transiently induced by the chemotherapy. The group suggested that EPCR might provide a coagulation-free environment, but this cannot fully explain its effects, especially in view of these cell line studies, where coagulation does not occur. Therefore, this work gives a strong indication that EPCR is important potentially in cancer and may have relevance to chemotherapy treatment, but does not provide a mechanism of action.

A study from the University of Carolina (USA) was the first to investigate the role of EPCR in cancer cell migration and invasion [72]. Around this time, evidence was accumulating that APC increased migration in endothelial cells [77] and keratinocytes [74]. This group used a transwell assay system to investigate migration in endothelial cells (HUVECs) and breast cancer cell lines (MDA-MB-231 and MDA-MB-435). APC was found to increase breast and endothelial cell migration and invasion by 125-375%. This effect was dose dependent. PC did not have the same effect. Anti-EPCR and anti-PAR1 antibodies both attenuated migration and invasion. The anti-EPCR antibody inhibited production of APC, but EPCR was also noted to be crucial to APC's effects on migration. As the effect was seen both in endothelial cells and

cancer cells, it suggested that a common APC/EPCR/PAR1 mediated pathway was present in both cells. APC was not itself a chemotactic agent. This suggests that the pathway increases invasion and chemotaxis towards other chemotactic factors, which may have relevance in early metastasis. The group could not fully explain the mechanism of the increased migration and invasion. They suggested that APC's protease activity was crucial, both in the activation of PAR1 (intracellular signalling) and perhaps in the activation of extracellular matrix proteases, as had been shown previously [74]. This paper brought developments of EPCR understanding in the endothelial setting to cancer.

1.9.2 *In vivo and clinical studies*

Several groups have performed *in vivo* studies into EPCR's role in various cancers, but clinical data are limited. A key study by a group based at the University of Navarra (Spain) extensively investigated EPCR's role in cancer metastasis using *in vivo* models [78]. Building on previous findings relating APC/EPCR to apoptosis, cell survival and migration, the group sought to determine if EPCR expression aids the metastatic process in lung cancer. They identified an adenocarcinoma cell line (A549) that expressed EPCR based on flow cytometry. APC was able to induce intracellular signalling via EPCR, leading to phosphorylation of ERK and Akt. They therefore established that key EPCR-relevant pathways in endothelial cells were also functional in this tumour cell line. On microarray analysis, the genes upregulated by APC treatment included several that have previously been identified [69], and as expected had anti-apoptotic and cyto-protective roles. EPCR expression was

knocked down in A549 cells using shRNA (small hairpin RNA). Rather than proceeding with *in vitro* assays, the group utilized these knockdown cell lines in an *in vivo* mouse model of lung cancer. Controls or shRNA treated A549 cells were injected into Athymic nude-*Foxn1^{nu}* mice and subsequent development of bone metastases was assessed. The key finding was that the shRNA group of mice developed significantly fewer hind limb bone metastases. Similar results were obtained by inoculating mice with control A549 cells and then regularly injecting them with EPCR-blocking antibodies. Antibody administration significantly decreased the metastatic tumour burden. Taking a non-EPCR expressing parental A549 line and inducing EPCR expression (using viral vectors) increased the development of metastasis in this mouse model. Administration of EPCR-blocking antibodies ameliorated this effect. This is strong evidence that EPCR has an important role in metastasis development.

Anton and collaborators also made attempts to characterise EPCR's expression in a cohort of lung cancer patients (n=295). These patients underwent standard treatment, i.e. surgery with or without adjuvant chemotherapy. EPCR expression was determined post-operatively by immunohistochemistry. Pathologists blinded to the details of each case scored each patient's EPCR expression. The group performed a multivariate Cox regression analysis to determine the relationship between EPCR expression and progression, recurrence, chemotherapy response and survival. In multivariate analysis, high EPCR expression was associated with a shorter time to progression (TP, P=0.021), suggesting that EPCR could be an independent predictor of disease progression. There was no significant relationship with overall survival. However, when only the stage I patients (those with cancer

confined to the lung) were analysed, high EPCR expression significantly correlated with disease progression ($p < 0.001$) and decreased survival ($p < 0.05$). When stage I patients were removed from the overall analysis, no significant correlations were observed in the stage II-IV patients alone. This analysis suggests that EPCR may be important in the development of early metastasis rather than in the continued growth of established metastases.

Another important finding in this study relates to patient stratification and prediction of chemoresponsiveness, which are both increasingly important in cancer. Patients with stage I cancer were grouped according to whether they had adjuvant chemotherapy post-surgery or not. In patients with low EPCR expression, chemotherapy was not associated with any change in disease progression or survival. In contrast, for patients with high EPCR expression, chemotherapy was associated with decreased progression and increased survival. This suggests that EPCR expression could be used to predict chemotherapy response, as well as serving as a prognostic marker for early stage patients. This would require validation in prospective trials. Overall however, this work has progressed the understanding of EPCR's role in cancer through a multi-faceted approach, and makes the protein a strong candidate for investigation in a range of cancers. It also demonstrates its possible clinical uses as a predictive and prognostic biomarker, or in the modulation of EPCR to control disease.

An earlier study by Bezuhly et al. [79] led to observations that were contrary to the work published by Anton et al, albeit in a different tumour type. The group aimed to relate endothelial EPCR activity with melanoma liver and lung metastases. Bezuhly and colleagues examined whether APC, with its established anti-inflammatory

activity, could reduce adhesion molecule expression and therefore metastasis development. APC decreased both adhesion and migration significantly in *in vitro* models. However, this effect was lost at the higher doses of APC, for reasons that were unclear.

The group used a mouse model in which endothelial EPCR was overexpressed (Tie2-EPCR). These mice and WT mice were inoculated with the melanoma cells. Two weeks later, surviving mice were sacrificed and spleen and liver were taken for analysis. In the two week period, 4/10 of the WT mice died, but all of the Tie2-EPCR mice survived. The liver and lungs had significant reductions in tumour nodules in the Tie2-EPCR mice respectively. qRT-PCR revealed that P-selectin and TNF were significantly reduced in the EPCR overexpressing mice, suggesting that a reduction in adhesion molecule expression accounted for the decreased metastases. Injecting APC in WT mice led to a less marked but still significant reduction in tumour nodule development. Therefore in this study, EPCR expression limited metastasis. The group did not perform any experiments where EPCR expression was reduced, to assess whether the opposite effect occurred.

This study contradicts Anton's findings, where reducing EPCR in lung cancer cells using shRNA reduced the development of bone metastases, as did an EPCR blocking antibody. There are several possible reasons for this discrepancy. The metastatic sites investigated differ, and it is possible that crucial pathways vary in different anatomical sites. The B16-F10 melanoma line is a mouse cancer line, whereas the A549 lung cancer line is a human line. The melanoma line did not express EPCR, whereas the lung cancer line did. This means that the anti-human antibodies used in the Anton study only targeted the engrafted cancer, whereas in

the Bezuhy study the anti-mouse antibody only targeted the endothelium. Therefore the different findings may arise from inherent differences in the biology of different tumour types, and also in the specific targeting of EPCR in cancer or in endothelium. The effects of EPCR may be context dependent.

Schaffner, Esmon and colleagues investigated the relevance of EPCR expression in stem cell-like cancer cell (CSC) populations [80]. CSCs are postulated to be important in cancer chemoresistance, recurrence and metastasis. They may have the ability to differentiate into all the cell types within a cancer. Previous work had shown that EPCR is expressed in aggressive breast cancer CSCs [81]. Schaffner and colleagues confirmed EPCR's importance in breast cancer CSC niches and investigated its consequences in breast cancer *in vitro* and *in vivo* models. Like Beaulieu and colleagues [72], this group utilized the breast cancer cell line MDA-MB-231. However, they selected a highly aggressive triple-negative (oestrogen receptor, progesterone receptor, Her2/neu negative) derivative line by using cells that were passed through mouse mammary fat pad. They separated the derived cells into EPCR⁺ and EPCR⁻ cells using fluorescence-activated cell sorting (FACS). Similarly to previous groups [69, 78], they extracted RNA and performed microarray analyses to determine differential gene expression. In addition to genes previously described such as MAPK1/ERK2 [69], markers associated with CSCs were overexpressed in EPCR positive cells. These included ALDH1B1 and ALDH1A3, and integrins $\alpha 4$ and $\alpha 6$.

The group used a polyoma middle T (PyMT) mouse model that spontaneously develops breast cancer in a fashion similar to the clinical scenario. They were able

to breed these mice with EPCR^{low/low} mice to produce an EPCR deficient model for breast cancer. These EPCR^{low/low} mice had a significantly smaller tumour burden and increased survival over the experimental period compared to controls. However, the group took further steps to ensure that only the EPCR gene was differentially expressed, by breeding PyMT EPCR^{flox/flox} mice. EPCR deletion led to an increase in survival and decrease in tumour load.

Two findings from these experiments were particularly important. Firstly, EPCR expression in the cancer cells rather than the host determined outcome. Secondly, the tumours resulting from high EPCR-expressing cell xenografts mostly contained cells that were EPCR negative. This demonstrates that the EPCR cells either reduced EPCR expression or differentiated into other cell types, as would be expected from CSCs. The EPCR-expressing cells may therefore have an increased capacity to initiate tumour growth. This initiating capacity was reduced by using blocking antibodies that were injected alongside cancer cells in one experiment and intra-peritoneally in another. In the intra-peritoneal group, mice were administered 5 injections of blocking antibody (or control) over 14 days. In both experiments, the blocking antibodies were associated with decreased tumour growth and burden.

This study demonstrated that EPCR may have broad relevance across epithelial cancer groups. The work links EPCR to the stem-cell like phenotype and attempted to identify the source of ligands in the tumour environment. The group discussed the possibility of therapeutic EPCR blockade. The safety of latter approach has not been explored, though the Schaffner group note that no adverse effects have been observed in animal models. This point should be treated cautiously, as there are EPCR-associated examples such as APC in sepsis where animal models did not

reveal the serious complications of treatment [66, 70, 71, 82], and the universal expression of EPCR on endothelium could be a major obstacle.

Keshava and colleagues investigated EPCR expression in mesothelioma and their results were seemingly contradictory to what has been observed in other cancers [83]. Their work was centred on Tissue Factor (TF), an important factor in the extrinsic coagulation pathway that, like EPCR, has been linked to poor outcomes in cancer. In investigating TF, they also studied EPCR, due to their shared pathways. *In vitro*, the group found that a more aggressive mesothelioma cell line (REN) did not express EPCR, whereas EPCR was present in less aggressive lines (MK9 and MS-1). This is contrary to the results in other cancers [78, 80]. In their xenograft models, REN cells that had been treated with a pZeoSV plasmid containing human EPCR cDNA were injected into mouse pleural cavities. The EPCR cDNA decreased tumour growth and burden significantly. The Keshava group found that apoptosis markers (TUNEL) were increased in the tumours with increased EPCR expression, which again differs from the known actions of EPCR and PAR1 [69]. Conversely, grafting EPCR shRNA-treated MK9 and MS-1 cells significantly increased tumour growth and burden, but only when TF was overexpressed concurrently. There was no effect observed on parental MK9 and MS-1 cells that were treated with the shRNA.

The authors acknowledged that their findings differ from others groups, and suggested that this may be due to the particular cell types that they have used or the particular environment of these mesotheliomas. The biology of these cancers may differ. Like EPCR, TF (in complex with FVIIa and Xa) activates PAR1 [67]. The group noted that PAR1 activation can be pro or-anti apoptotic depending on the amplitude and duration of the signal [73, 84, 85]. In the presence of TF mediated

activation of PAR1, EPCR may have a pro-apoptotic/anti-proliferative effect. This study demonstrates the complexity of these signalling pathways, where EPCR may have distinct effects in different cancer types.

A French group, Ducros and collaborators, investigated EPCR's potential role as a biomarker and in cancer associated coagulopathy [86]. The group obtained blood samples from 79 patients with ovarian cancer. They found that soluble EPCR (sEPCR) correlated closely with CA-125, the established ovarian tumour marker. The group also found that sEPCR and membranous EPCR were expressed in the ascitic fluid of these patients. The study is most notable for its extensive genotyping. They sequenced the sEPCR from ovarian cancer patients and membranous EPCR from several cancer lines, including ovarian (OVCAR-3), lung (A549), colorectal (HT-29) and breast (MDA-MB-231) lines. The group detected 13 single nucleotide polymorphisms (SNPs), which were consistent with established SNPs in endothelial EPCR. Cells carrying haplotype A3 shed a higher quantity of sEPCR into their growth medium. This is consistent with work suggesting that the A3 haplotype in cancer patients is associated with higher levels of sEPCR, which is associated with increased venous thrombosis [87]. The mechanism of this effect is thought to be due to sEPCR binding with PC in blood, leading to less APC production by endothelial EPCR. This study demonstrates that EPCR may be a potential biomarker for at least one type of cancer, and may be predictive for the development of thromboembolism. Potentially, sEPCR measurement could allow more appropriate and personalised thromboprophylaxis.

It is evident that recently there has been increasing interest in EPCR's role in cancer. The recent finding that it can act as a $\gamma\delta$ TCR ligand suggests possible interaction

with anti-tumour immune responses. The tumour biology of EPCR therefore becomes very relevant. However, several key questions remain unanswered. The majority of these studies have investigated EPCR-intrinsic functions, and have not investigated the wider biological context of EPCR dysregulation. Though many studies have utilised mouse models, there are very few that have investigated EPCR in the human setting [78, 86]. Furthermore, there have been diverse effects of EPCR in different tumours, and the reason for this has not been explored. The breadth of its expression in different cancers also needs to be determined. In view of its recently discovered function as a $\gamma\delta$ T cell ligand, it is now increasingly crucial to understand EPCR's role in tumourigenesis.

1.10 Project aims

Emerging immunotherapies such as checkpoint blockade are now impacting on patient care. However, the results have been variable in different cancer types [13, 14]. Responses to checkpoint blockade have been linked to genetic factors [88], and though genetics and immunity are both relatively well understood, the link between the two is less well characterised. CRC is a good model to develop this understanding, as it has well defined molecular subtypes with variable mutation rates and immune infiltration (which has been linked to patient outcome [37]). To understand why immunotherapies have had limited success thus far in CRC, I shall investigate the genetic and pathological factors that impact on the immune microenvironment (Chapter 4). This could potentially enable better targeting of existing immunotherapies, and could facilitate the development of novel approaches. To characterise immunity in different patient groups, I shall establish a gene expression signature for the immune microenvironment. In addition, I will explore the immune microenvironment of the key *RAS* mutant subgroup of patients (Chapter 5).

In addition, I shall explore the role of the $\gamma\delta$ T cell ligand EPCR in colorectal tumourigenesis (Chapter 3). The known functions of EPCR have expanded since its discovery. Initially significant mainly for its role as an anti-coagulant factor, it is now implicated in inflammation and potentially immunity as a $\gamma\delta$ T cell ligand. There is also growing evidence that EPCR has a role in tumourigenesis. Since this project was conceived, EPCR has been studied in a range of tumour types. However, it has not as yet been investigated in CRC, the third commonest cancer in this country and the second most common cause of cancer-related death. In view of the well-

established developmental pathway and pathological characterisation of CRC, this tumour type is a promising model in which to explore the role of EPCR in tumourigenesis. Therefore, I shall investigate the extent and mechanism of EPCR expression and its impact in CRC, both *in vitro* and in a clinical cohort, to determine its significance in this setting.

Chapter 2

Methods

2.1 Bioinformatic analyses

2.1.1 EPCR bioinformatics

EPCR mutation, methylation, copy number and expression data were retrieved from the The Cancer Genome Atlas (TCGA) project via the cBioportal tool [89-92]. Pathological and clinical data were also downloaded. Cancer cell line encyclopaedia [93] data were retrieved using the CCLE portal and cBioportal. Oncomine (Compendia Bioscience) was used for analysis and visualization of EPCR expression in multiple tumour types [94].

Data were tabulated and analysed with The Integrative Genomics Viewer [95] and Excel (Microsoft Corp.). Pearson correlations were performed when the data distributions were parametric, and Spearman when they were non-parametric. Significance tests were performed in Minitab (Minitab Inc.) using T-tests for parametric data and Mann-Whitney U for non-parametric data. Normality of distributions were confirmed using the Anderson-Darling Test.

2.1.2 Immunity bioinformatics

2.1.2.1 Gene identification

Genes associated with innate and adaptive tumour immunity in cancer [37, 42] and genes relevant to established colon cancer pathways were shortlisted as initial genes. This list was expanded using molecular network pathway analysis (using cBioportal) [89, 90, 92]. This revealed further molecules and genes with known genetic, pathway and functional associations with the initial genes. Genes revealed

through this approach were screened for inclusion into the final gene list based on known immunological relevance (Table 2.1).

Gene IDs	
ACTB	HLA-DQA2
CCL11	HLA-DRA
CCL2	HLA-DRB5
CCL5	ICAM1
CD247	ICOS
CD274	IFNG
CD276	IL12RB2
CD3D	IL17A
CD3E	IL18RAP
CD3G	IL7R
CD4	IRF1
CD80	KLRK1
CD86	LAG3
CD8B	MADCAM1
CTLA4	MICB
CX3CL1	PDCD1
CXCL10	PDCD1LG2
CXCL9	PROCR
GNLY	RAET1E
GZMB	RAET1G
HAVCR2	STAT1
HLA-A	STAT3
HLA-B	TBX21
HLA-C	TNFRSF14
HLA-DMA	TNFSF4
HLA-DMB	ULBP1
HLA-DOA	ULBP2
HLA-DOB	ULBP3
HLA-DPA1	VCAM1
HLA-DPB1	VTCN1
HLA-DQA1	

Table 2.1 Final gene list for hierarchical clustering analysis

2.1.2.2 Data extraction

Normalised Agilent microarray and RNAseq z-score data for initial genes and mutation data for *TP53*, *KRAS*, *BRAF*, *NRAS*, *HRAS*, *PIK3CA* and *PTEN* were extracted from The Cancer Genome Atlas (TCGA) colorectal dataset [91] using the cBioportal tool [89, 90, 92]. Where expression data were unavailable, genes were excluded from the analysis. Clinical data for these patients were retrieved from the TCGA portal (<https://tcga-data.nci.nih.gov/tcga/>) and were tabulated with genetic data.

2.1.2.3 Data analysis

Data were tabulated in Excel (Microsoft Corp.) and unsupervised two-dimensional hierarchical clustering was performed using MeV (Dana-Farber Cancer Institute, Boston, MA, USA) and the Pearson correlation. Patient and gene clusters were identified by varying gene-tree distance-thresholds and visual analysis. Normality of distributions was confirmed with the Anderson-Darling test. Pearson Coefficient of determination (R^2) values were calculated in Excel and Minitab (Minitab Inc.) to investigate correlations in gene expression. Gene expression was correlated against beta actin as a control gene. T tests were performed for univariate analyses of normally distributed data, and Mann Whitney U tests for non-normally distributed data. Multivariate linear regression analyses of gene expression against mutations in key tumour associated genes (*P53*, *KRAS*, *BRAF*, *NRAS*, *HRAS*, *PI3KCA*, *PTEN*) were completed in Stata 12.2 (Statacorp.). For two by two comparisons, a Chi squared test was used; if any groups contained less than 5 values, the Fishers Exact

test was used in preference. MEK activation signature data were retrieved for the TCGA dataset from the authors [16] and statistical analyses were performed in Excel.

2.1.2.4 Mutation rate analysis

Mutation rate and number data for 12 TCGA cancer types, including CRC, produced by Kandoth et al. [96] were downloaded from the Synapse platform (<https://www.synapse.org/>, Sage Bionetworks). Expression data for CIRC genes were extracted using cBioportal, and were combined with mutation data in Excel. CIRC expression was calculated for each case, and correlations with mutation rate were made both by mutational group comparisons and Pearson and Spearman tests. Graphs were produced in Excel and Minitab.

2.1.2.5 Neoantigen analysis

Neoantigen analysis was undertaken in collaboration with Sebastian Boegel and Ugur Sahin (TRON, University of Mainz, Germany). Full CRC TCGA RNAseq data and sequencing data were used. Initially, the number of non-synonymous single nucleotide variations (nsSNVs) was determined per patient. Using RNAseq, the number of reads per mutation was calculated to determine whether the neoantigen was expressed. HLA-type was predicted from RNAseq data using seq2HLA and HLA-binding predictions were made using the IEDB MHC binding prediction algorithm v2.9 (consensus method), as described by Boegel et al. [97, 98]. A cutoff for strong binding (and thus “likely immunogenic”) was 1.0. This approach was used to determine how the strength of neoantigen binding to HLA and the total number of neoantigens related to CIRC expression in TCGA patients.

From the total neoantigen data, the number of neoantigens per case and top rank neoantigen data were calculated and combined with CIRC expression in Excel, and correlation tests were performed. Individual analysis of *RAS*-associated neoantigens were performed by extracting *RAS* neoantigen data from total neoantigen data for all cases and determining the number and frequency of these neoantigens as well as binding strength in each case.

2.1.2.6 Cancer cell line encyclopaedia analysis

CIRC microarray z-score data for 51 colorectal cell lines were retrieved from the CCLE database using cBioportal. These were combined with mutation data for *KRAS*, *NRAS*, *HRAS*, *BRAF*, *PIK3CA* and *PTEN*. Mean expression of the CIRC and key immune genes in each mutational subtype were calculated, and significance testing was performed using T-Tests for one by one comparisons.

2.1.2.7 Consensus Molecular Subtypes (CMS) analysis

Consensus molecular subtype (CMS) data were obtained from the Colorectal Cancer Subtyping Consortium through the Synapse platform (<https://www.synapse.org/>, Sage Bionetworks).

These data were utilized to perform intra-CMS group analysis of *RAS* mutation and CIRC. In addition, RNAseq data were downloaded for the extended TCGA RNAseq dataset and the Koo Foundation Sun-Yat-Sen Cancer Center (KFSYSCC) dataset. Single sample enrichment of pathways was assessed using the Gene Set Variation Analysis (GSVA) algorithm [99]. Data from TCGA were downloaded as level 3 processed data; genes with no variation were removed, and remaining genes expression values were log2 transformed. CMS labels for these datasets were

downloaded from Synapse at <https://www.synapse.org/crcsc>. Immune cell type gene sets used in this analysis are described in Bindea et al. [100]. Differential enrichment using *KRAS* mutation status within and across CMS groups was tested using the Wilcoxon rank sum test, and corrected for multiple testing using Benjamini-Hochberg procedure.

2.2 Specimen collection

2.2.1 EPCR specimens

Patient tumour and serum specimens were collected from the University of Birmingham Human Biomaterials Resource Centre (HBRC) (ethical approval HBRC 11-058). Survival analysis utilised tumour samples from the MRC COIN study (ISRCTN27286448), under ethical approval 13/WM/0339 [10]. All patients had provided consent for use of their tissue. Specimens were cut to a thickness of 4 microns.

2.2.2 RAS cohort specimens

2.2.2.1 Sample size calculation and sample collection

To determine the required sample size, a power calculation was performed using Altman's nomogram. This confirmed that a total of 100 samples (50 RAS mutant, 50 RAS wildtype) were required for a power of 80%, a significance level (α) of 0.05 and a minimum detectable standardised difference of 0.55. This power provided an 80% probability of detecting a difference between the *RAS* mutant and wildtype groups if one existed. The significance level was the threshold below which the null

hypothesis (that there was no difference between the *RAS* mutant and wildtype groups) would be rejected, in this case corresponding to a p value <0.05.

The specimens were selected to represent the range of *RAS* mutations observed in the original TCGA microarray data set. The final cohort comprised 28 *RAS* G12D/G13D mutants (24.3%), 38 *RAS* non-G12D/G13D mutants (33.0%), and 49 *RAS* wild types (42.6%). Therefore, there were 66 *RAS* mutants and 49 *RAS* wildtypes (total = 115). Samples were obtained from the completed CRUK Stratified Medicine Programme One pilot study and CRC patients from the Queen Elizabeth Hospital, Birmingham. These patients had their *RAS* mutation status determined as part of their routine clinical workup. Samples were collected under ethical approval HBRC 14-205 (Sponsor: University of Birmingham).

2.2.2.2 Sample processing

Suitable formalin-fixed, paraffin-embedded (FFPE) blocks were identified from pathology reports and were retrieved and processed at the HBRC biobank. Twenty sections of 4 micron thickness were cut from each block and were mounted onto slides for staining. Additionally, six 10 micron scrolls were cut from each block and were put into 2 eppendorfs (containing 3 scrolls each) for RNA and DNA extraction.

2.3 Immunohistochemistry

2.3.1 EPCR and PAR1 immunohistochemistry

2.3.1.1 Immunohistochemistry protocol

Paraffin-embedded colorectal cancer sections and matched adjacent normal sections were deparaffinised and heat-induced epitope retrieval (HIER) was performed by adding the slides (in a slide rack) to a tray containing Novocastra Epitope Retrieval Solution pH8 (Leica, RE7116), which was incubated in a water bath heated to 97°C for 40 minutes. Slides were then washed twice (5 minutes per wash) in diluted DAKO wash buffer (S3006, containing 0.05 mol/L Tris/HCl, 0.15 mol/L NaCl, 0.05% Tween 20 at pH7.6). A hydrophobic mark was drawn around the tissue section using a DAKO Delimiting pen (DAKO, S2002). Two to three drops of dual endogenous enzyme inhibitor were applied (component of Dako Envision Detection System peroxidase/DAB, Rabbit/Mouse kit, K406511-1). Slides were then washed twice as above. Non-specific binding was blocked by adding 100µl of 1% normal goat serum (NGS) (50µl concentrated NGS diluted in 5000µl of antibody diluent), with 30 minute incubation at room temperature (RT). Primary antibody was then added immediately after tipping off the NGS without washing the slides - 100µl 1:75 mouse anti-human EPCR antibody (R&D Systems, clone 304519, targeting human EPCR Ser18-Ser210, mouse anti-human), diluted with antibody diluent (DAKO, S0809). The slides were incubated with the primary antibody overnight at 4°C. Slides were then washed three times with wash buffer and incubated for 30 minutes with 4 drops of polyclonal anti-mouse secondary antibody (Dako, component of Envision Detection System) at RT. Slides were then washed three times with wash buffer as above. Peroxidase

activity was revealed by adding 100µl diaminobenzidine (DAB – component of Dako Envision Detection System) for 5 minutes at RT. The slides were counterstained with haematoxylin (100%, for 40 seconds) and were then washed under warm running tap water for 2 minutes. Slides were dehydrated in 70% ethanol for 2 minutes, then 100% industrial methylated spirit (IMS) for 2 minutes, and were then added to Histoclear for 2 hours at RT (Fisher Chemical). Coverslips were mounted using 2-3 drops of DPX (National Diagnostics), and were left to dry overnight. Guidance on antibody optimisation was provided by a histopathologist (Dr Philippe Taniere, Queen Elizabeth Hospital, Birmingham). Controls that omitted the primary antibody were also performed to rule out non-specific binding of the secondary antibody. For PAR1 IHC the same protocol was used with a primary antibody that targeted PAR1 (NBP1-71770, clone N2-11 Novus Biologicals), at a dilution of 1:100.

2.3.1.2 EPCR immunohistochemistry scoring

EPCR-stained slides were imaged on a Vectra 2.0 (Perkin Elmer) system using custom imaging algorithms developed on Inform software (Perkin Elmer). Analysis of slides was performed on Inform software, using trained tissue and cell segmentation algorithms that were validated by a histopathologist (Dr Philippe Taniere). Staining intensities were determined on a per cell basis, and H-scores were created by the Inform software. H scores were calculated by the following formula: $[1 \times (\% \text{ cells with weak staining}) + 2 \times (\% \text{ cells with moderate staining}) + 3 \times (\% \text{ cells with heavy staining})]$. This produced a score ranging from 0 to 300. Tumour region and Stroma region H-scores across all slides were collated and compared using the unpaired T Test, after the normality of distributions were confirmed using the Anderson Darling test.

COIN study slides were stained for EPCR at the HBRC biobank using a Bondmax Autostainer. The primary antibody (R&D systems, clone 304519) incubation time was 10 minutes at a dilution of 1:200. Antigen retrieval was performed at pH9. Stained slides were scanned using a Leica SCN400 slide scanner. Scanned slides were analysed on Definiens Tissue Studio software. Tumour regions of each slide were manually identified, and trained segmentation algorithms were used to separate epithelium and stromal regions. Quantification of EPCR staining was performed on a regional basis, with the software quantifying the percentage of pixels that had strong staining, moderate staining, weak staining or no staining in each area. These data were used by the software to create a percental score for each region in each slide.

2.3.2 RAS cohort immunohistochemistry

2.3.2.1 Immunohistochemistry protocol

Immunohistochemical staining was performed using a Leica Bondmax autostainer. Clinically validated (IVD-CE) antibodies were chosen where available (Table 2.2). Otherwise, knockdown/knockout validated antibodies were selected. If these were not available, in-house validation was performed, as described below.

Marker	Vendor	Antibody clone	Dilution	Primary antibody incubation time (minutes)	Conditions	Validation
CD4	Leica	4B12	Prediluted, ready to use	30	pH9	IVD-CE (clinically validated)
CD8	Leica	4B11	Prediluted, ready to use	20	pH9	IVD-CE (clinically validated)
Tbet	Cell Signalling Technology (CST)	D6N8B	1:100	20	pH9	Validated by CST*
PD-L1	Cell Signalling Technology (CST)	E1L3N	1:200	20	pH9	Validated by CST*
STAT1	Cell Signalling Technology (CST)	D1K9Y	1:500	20	pH9	Validated by CST*
CXCL10	Abcam	ab9807 (vendor ID)	1:100	40	pH9	In-house (data not shown)
Class II HLA	Abcam	CR3/43	1:100	20	pH6	In-house

Table 2.2 List of antibodies.

*CST antibody validation includes western blot analysis, staining of paraffin-embedded cell pellets with known expression, the use of blocking peptides, and the staining of gene knockdown and knockout cell lines.

Initial staining conditions and concentrations were based on vendor recommendations, and were iteratively optimised in conjunction with a histopathologist (Dr Philippe Tanriere).

2.3.2.2 RAS cohort Immunohistochemistry scoring

After staining of the CRC sections for the markers above, slides were scanned at 40x magnification using a Leica SCN400 slide scanner. Scanned slides were digitally analysed using Definiens Tissue Studio software. Analysis algorithms were created and optimised for each individual marker. Regions of interest (ROIs) were created in the tumour regions of each slide, in order to enable an inter-tumoural analysis. All tumours were digitally segmented into tumour epithelium and stroma regions using trained segmentation algorithms. Depending on the marker type, staining was quantified on a per cell basis or on an area basis. Class II HLA alone was quantified using the area method as the staining was strong and therefore the Definiens software could not reliably identify haematoxylin-stained nuclei (which is necessary for cell based scoring). Percentages of cells or pixels with high, medium, low or no immunoreactivity were quantified in each region. This produced either histological scores for cell-based scoring, which is a function of the number and intensity of immunoreactive cells as described above, or percental scores for area-based scoring, which is a function of the number and intensity of positive pixels on the scanned specimen. Thresholds for negative/low, low/medium and medium/high were set to maximise the dynamic range of results between samples and to reduce false positive results (Table 2.3). Haematoxylin thresholds (the staining intensities at which haematoxylin was recognised) were set individually and differed for each

antibody due to differences in DAB staining. Haematoxylin thresholds were set to ensure accurate identification of individual cells.

Marker	Scoring method	Haematoxylin threshold	Negative/low threshold	Low/moderate threshold	Moderate/high threshold
CD4	Cell based	0.05	0.28	0.30	0.40
CD8	Cell based	0.05	0.16	0.25	0.35
Tbet	Cell based	0.02	0.15	0.20	0.25
PDL1	Cell based	0.03	0.20	0.40	0.60
STAT1	Cell based	0.01	0.14	0.20	0.35
CXCL10	Cell based	0.01	0.14	0.20	0.30
Class II HLA	Area based	0.02	0.25	0.60	1.00

Table 2.3 Definiens scoring method and thresholds for each marker

After analysis, segmentation was manually validated for each slide. Analysis was performed using Excel (Microsoft Corp) and Minitab (Minitab Inc). The normality of the distribution of Histological scores in each group (*RAS* mutant or *RAS* wildtype) was determined by performing the Anderson-Darling test. All data were non-parametrically distributed. Therefore, for one by one comparisons, Mann Whitney U tests were performed for significance testing. A p-value <0.05 was considered significant.

2.3.2.3 HLA Class II antibody validation

To validate the HLA class II mAb (Anti-HLA DR + DP + DQ antibody [CR3/43] (ab17101)) (Abcam) a stably CIITA (Class II major histocompatibility complex transactivator) lentiviral transduced HEK293T cell line was used. The parental HEK293T and K562 cell lines served as negative controls. Expression of MHC class II was initially confirmed by flow cytometry. Cells were then pelleted by centrifugation and fixed in 10% formaldehyde. The pellets were embedded into paraffin blocks and cut into 4µm sections and mounted onto glass slides. Immunohistochemistry was performed on those sections using the Leica Bond-RX automated platform. Images of the sections were scanned using a Leica SCN400 Slide scanner (Figure 2.1).

Flow cytometry was performed. Cells in the cell suspension were counted and the volume containing 1×10^5 transferred to a 96-well plate. Cells were then pelleted by centrifugation and re-suspended in 1:50 dilution of BioLegend allophycocyanin (APC)-conjugated anti-human HLA-DR (clone L243) (BioLegend). A further two wells were used for isotype control (BioLegend APC Mouse IgG1 isotype) and negative control (MACS buffer). The cells were incubated with the antibody for 30 minutes at 4°C. They were then washed with MACS buffer and fixed in 2% paraformaldehyde (PFA). Flow cytometry was performed on a BD LSR II Flow Cytometer running the BD FACSDiva software (BD Biosciences). Flow cytometry demonstrated high levels of MHC class II in transfected cells in comparison to complete lack of expression in the negative controls.

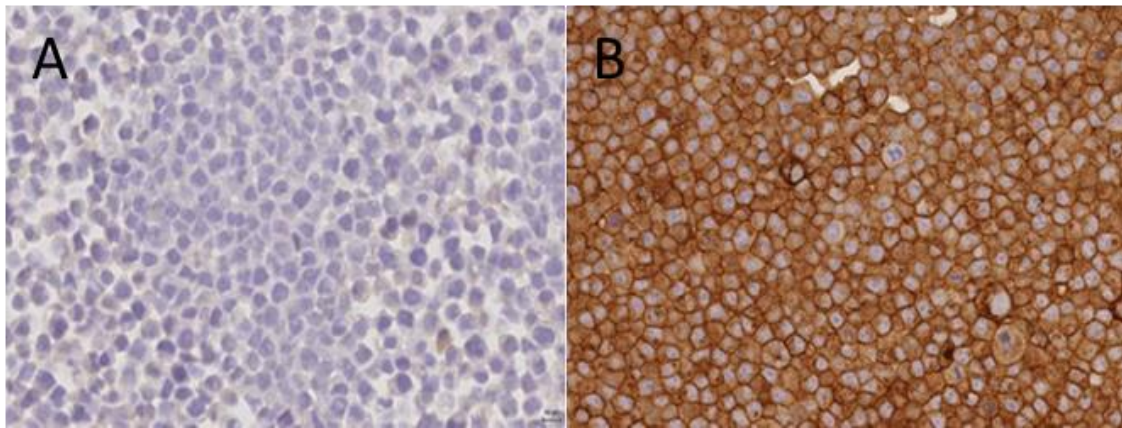


Figure 2.1 HLA Class II antibody validation. The left panel shows staining of the parental HEK293T cell line, and the right panel shows the Class II HLA-expressing transfected HEK293T cell line.

(Figure courtesy of Dr Ghaleb Goussous)

2.4 Survival analyses

2.4.1 EPCR survival analysis

Clinical data for the patients in the COIN study were obtained from the MRC under REC approval 13/WM/0339. Data obtained included treatment arm (chemotherapy or chemotherapy plus cetuximab), demographic details including age and sex, clinical data including WHO performance status, number of metastases, timing of metastases, site of primary (within colorectum), serum carcinoembryonic antigen (CEA), microsatellite status and mutation status for key genes including *KRAS*, *BRAF*, *NRAS* and *PI3K*. Survival data included 12-week response, best response, death, censored time to death, progression and censored time to progression (PFS time).

The median EPCR percental score was used to categorise patients into EPCR high and low groupings. Additionally, a separate analysis was performed for the top 20% of EPCR expressors against the bottom 20%. Statistical modelling was performed using Stata 12.1 (Stata Corp). Both the prognostic and predictive analyses utilized the Cox proportional hazards model, and adjusted for factors that significantly associated with PFS to $p < 0.05$. The proportional hazards assumption was tested by calculating Schoenfeld residuals (estat phtest).

Stata commands:

```
stset pfstime, failure(pfsstatus)
```

```
estat phtest
```

```
stcox i.factorx - (for each factor)
```

```
xi:stepwise, pr(0.05):stcox factor x factor y i.EPCRfactor
```

sts graph, by(factorx)

2.4.2 *RAS cohort survival analysis*

Clinical data for the patient cohort were extracted from patient notes. Clinical information included patient age, gender, ethnicity, tumour stage, *RAS* mutation status and overall survival. Microsatellite status data (as determined by fragment analysis) were added, as were the results of IHC scoring. Using Stata, survival analysis was performed (Cox regression) as above to determine the association between *RAS* mutation/patient stage and survival.

2.5 shRNA transfection

5 clones of EPCR shRNA (Sigma MISSION lentiviral pLKO.1-puro transduction particles) and 1 control shRNA (Sigma MISSION pLKO.1-puro Control high titre transduction particles - SCH001H) were obtained from Sigma (Table 2.4).

	TRC number	Clone ID	Sequence
1	TRCN0000061378	NM_006404.3-253s1c1	CCGGTGGCCTCCAAAGACTTCATATCTCGAGATATG AAGTCTTTGGAGGCCATTTTTG
2	TRCN0000061382	NM_006404.3-706s1c1	CCGGTCGGTATGAACTGCGGGAATTCTCGAGAATT CCCGCAGTTCATACCGATTTTTG
3	TRCN0000300553	NM_006404.3-679s21c1	CCGGGCAGCAGCTCAATGCCTACAACCTCGAGTTGTA GGCATTGAGCTGCTGCTTTTTG
4	TRCN0000369969	NM_006404.3-1123s21c1	CCGGTTTGCTGAATTAGTCTGATAACTCGAGTTATC AGACTAATTCAGCAAATTTTTG
5	TRCN0000061379	NM_006404.3-743s1c1	CCGGGTGCAGTATGTGCAGAAACATCTCGAGATGT TTCTGCACATACTGCACTTTTTG

Table 2.4 shRNA clone IDs.

1.6×10^4 HCT116 cells or HT29 cells were plated in each well of a 96 well plate and grown to 70-80% confluence in Dulbecco's Modified Eagle Medium (DMEM) with 10% Foetal Calf Serum (FCS). The media was removed, and 110µl media and hexadimerine bromide (polybrene, final concentration 8µg/ml) was added to each well. shRNA lentiviral particles were added to appropriate wells at multiplicity of infections (MOIs) of 5, 1.5 and 0.5. Cells were incubated with shRNA for 20 hours at 37° with 5% carbon dioxide. The media containing lentivirus was then removed, and fresh media was added (120µl per well), and cells were incubated for a further 24 hours. Triplicate wells for each lentiviral construct and control were used. The media was then removed from wells, and fresh media containing puromycin (1µg/ml) was

added, which was used to select shRNA transfected cells, as the shRNA construct contained a puromycin resistance gene. A puromycin titration curve was previously produced to determine the minimum concentration that caused complete cell death after 3-5 days (1µg/ml). Puromycin-containing media was replaced every 3-4 days until resistant colonies could be identified. Colonies were expanded into a 12 well plate. All 5 constructs and the controls for MOI 5 and MOI 0.5 were expanded. They were subsequently expanded into 6 well plates, and then into 10cm culture dishes.

EPCR expression was quantified with flow cytometry. Maintenance of EPCR knockdown throughout the experimental period was confirmed with flow cytometry after completion of assays.

EPCR-overexpressing HT29 cells were a generous gift from Julie Dechanet-Merville (University of Bordeaux, France).

2.6 Flow cytometry

2.6.1 EPCR cell line expression

After incubation of HCT116 and HT29 cells with shRNA, media were removed from the wells and cells were washed with PBS buffer. Cells were trypsinised with 500µl trypsin for 2 minutes at 37°C. Trypsin was neutralised with media, and 500µl of cell suspension were aliquoted into flow cytometry tubes. These were centrifuged and the supernatant was discarded. EPCR (anti-human EPCR APC, eBioscience, clone RCR-227) and isotype control (Rat IgG1 K Isotype Control APC, eBioscience, clone eBRG1) antibodies were diluted with MACs buffer (5µl EPCR antibody in 50µl MACs

buffer per test, 0.83µl isotype control antibody in 66.6µl MACs buffer per test). MACs buffer comprised 1 x PBS (pH 7.2) + 2nM EDTA + 0.5% BSA. Cells were suspended in 50µl of diluted antibody, and the tubes were incubated in a dark ice box for 30 minutes. MACs buffer was then added to the tubes using transfer pipettes, cells were centrifuged, and the supernatant was discarded. The cells were resuspended with MACs buffer, and were again centrifuged. Supernatant was discarded, and cells were resuspended in MACs buffer. The cells were fixed with 250µl of 1% PFA, and were stored at 4°C until analysis. Cells were analysed on an Accuri C6 or LSR II (BD Biosciences) flow cytometer using standard protocols, and data were analysed using FlowJo and Excel.

2.6.2 EPCR expression in mast cells

The human Mast Cell (MC) line HMC-1.1 (V560G) [101] was cultured in IMDM with 10% FBS. Human Lung Mast Cells (HLMCs) were isolated from healthy lung obtained at surgery using anti-CD117-coated Dynabeads at ~99% purity [102]. HLMCs were cultured in DMEM (Dulbecco's Modified Eagle Medium) supplemented with 10% FCS, and cytokines (100 ng/ml SCF, 50 ng/ml IL-6, and 10 ng/ml IL-10) as described previously [103]. HMC-1 cells and HLMCs were stained with 2E9 antibody (1µg/mL) [53] which recognises EPCR protein or stained with IgM isotype control. Cells were indirectly labelled with FITC (Dako) and analysed by one-colour flow cytometry on a FACSCanto (BD Biosciences).

EPCR expression was then confirmed on lung FFPE tissue using IHC. Briefly, 10 nM citrate buffer was used for antigen retrieval on 4 µm sequential paraffin sections, followed by immunostaining with mouse anti-human EPCR monoclonal antibody

(R&D systems, clone 304519, 10 µg/mL) and mouse anti-human tryptase (DAKO, clone AA1, 0.1 µg/mL). Sections were counterstained with Mayer's haematoxylin.

2.7 MTT assays

HCT116 and HT29 shRNA EPCR knockdown cells were cultured in DMEM with 10% FCS. Equal numbers of cells (4×10^3) were plated in each well of a 96 well plate. Cells were incubated at 37°C with 5% carbon dioxide. After 48 hours of incubation, chemotherapy drugs were added to the plates (either 5-Fluorouracil (5FU) or epirubicin) in concentrations determined by serial dilution (5FU - 0 to 256 µM, epirubicin – 0 to 8 µM). After a further 48 hours of incubation, 10µl MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, 5mg/ml in PBS) was added to each well and after 2 hours cells were lysed with 50µl DMSO and absorbance was calculated using a Victor plate reader at 490nm for 0.1 seconds. Colour intensity was normalised against controls untreated with chemotherapy. All experiments were performed in triplicate.

2.8 BrdU assays

Assays were performed using the BrdU (bromodeoxyuridine) cell proliferation ELISA kit (Roche). Cells were plated and treated with epirubicin or 5FU as with the MTT assays above. After 48 hours of incubation, cells were labelled with 10µl per well of BrdU labelling solution and incubated for 4 hours at 37°. The media was then removed and the plate dried with absorbent paper. Cells were then denatured by

incubation with 200µl FixDenat solution for 30 minutes at RT. Cells were incubated with 50µl Anti-BrdU-POD antibody solution for 90 minutes at RT (100µl per well) and then washed with wash buffer. The wash buffer was then removed and absorbent paper was used to remove excess wash fluid. This was repeated a further two times. 100µl of substrate solution was added to each well, and plates were incubated for 15 minutes at RT and were then analysed on a Victor plate reader at 405nm for 0.1 seconds. Colour intensity was normalised against untreated controls.

2.9 Migration assays

70-80% confluent EPCR-perturbed and control HCT116 and HT29 cells were serum-starved by incubation in serum-free medium (DMEM) for 24 hours. 600µl of DMEM with 10% FCS was added to the wells of a 24 well plate. 2×10^5 cells in 100µl serum-free DMEM were added to a Corning Transwell insert (Sigma, pore size 0.4 µm). The insert was then added to a well of the 24 well plate. Cells were incubated from 3 - 48 hours. The cells on the lower side of the Transwell insert were fixed with 30% methanol and stained with 2% crystal violet for 2 hours at 4°C. The Transwell inserts were removed and the upper side of the membrane was cleaned with a cotton bud to remove the non-migrated cells. The inserts were photographed with a light microscope and camera in 5 set regions (upper left, upper right, lower left, lower right and centre) using Axiovision software (Carl Zeiss). Cells were counted and analysed using Excel (Microsoft Corp). The Mann-Whitney U test was used to compare migration in control versus knockdown lines.

2.10 Invasion assays

70-80% confluent EPCR-perturbed and control HCT116 and HT29 cells were serum-starved as with the migration assays. Cells were suspended at 1.0×10^6 cells/ml in serum-free DMEM. 250 μ l of this cell suspension (containing 2.5×10^5 cells) was added to a QCM cell invasion assay insert (Millipore). The insert was then added to a well of a 24 well plate that contained DMEM with 10% FCS. Cells were incubated for 24-48 hours at 37°C and 5% CO₂. The cells and media from the top side of the insert were removed by pipetting out the remaining cell suspension. The cells on the lower side of the insert were then detached by incubation for 30 minutes at 37°C with cell detachment solution (component of Millipore QCM cell invasion assay kit). The insert was removed and detached cells were lysed and fluorescently labelled with 200 μ l lysis buffer/dye solution (CyQuant GR Dye 1:75 with 4x lysis buffer, 15 minute incubation at RT). 200 μ l of the mixture was transferred to a new plate. The wells were read with a fluorescent plate reader (Victor) using the 480/520 nm filter set. Results were analysed in Excel (Microsoft Corp).

2.11 Activated Protein C ERK phosphorylation

Activated Protein C (APC, Sigma) was added at a concentration of 180nM to confluent HCT116 cells after 24 hours of serum starvation. APC was added for 2 minutes, 5 minutes or 10 minutes, with or without prior 30 minute incubation with EPCR function-blocking antibodies (RCR-252, Novus Biologicals, 20 μ g/ml) after which cells were lysed with sodium dodecyl sulfate (SDS) containing phosphatase inhibitor (Phosphatase Inhibitor Cocktail II, Sigma).

2.12 Western blotting

Western blotting for ERK and pERK was performed using HCT116 cell lysates with Biorad TGX precast gels (Miniprotean TGX precast gel 12% 10 wells). The transfers from gels onto membranes were performed using the Biorad Trans-blot Turbo system (using Biorad Transblot Turbo transfer pack PDVF midi). After blockade with 3% BSA in TBST, separate membranes were incubated with primary antibody (ERK/pERK, New England Biolabs PhosphoPlus (Thr202/Tyr204) Antibody Duet (4370S + 4695S)) overnight at 4°C (pERK antibody 1:2000 dilution, ERK antibody 1:1000 dilution, both into 20ml diluent (TBST with 5% BSA and 0.01% sodium azide)). Membranes were then incubated with goat anti-rabbit horseradish peroxidase (HRP) conjugated secondary antibodies (1:20000 dilution, 1µl antibody diluted in 20ml BSA). Development was completed with an EZ-ECL kit (Biological Industries) and images of membranes were taken using the ChemiDoc MP system (Biorad). Images were analysed using Image Lab 4.1 (Biorad).

Western blotting was also performed to determine whether PC and APC were present in culture media (DMEM) or HCT116 cell lysate or supernatant. Exogenous APC (Sigma), was used as a positive control in this analysis. The western blotting protocol was as described above, except that a protein C primary antibody (clone NBP1-58065, Novus Biologicals, 1:1000 dilution) was used. This antibody detected both PC and APC, as it targeted the light chain of PC from amino acids 1 to 236, which are found in both PC and the cleaved form APC.

2.13 Microarray

HCT116 cells were treated with APC after 48 hours of serum starvation. Optimal APC concentrations and treatment durations were determined by assessing APC-induced expression changes of 4 genes (*NFkB2*, *PCNA*, *BCL2A1*, *EFNA1*, which were chosen based on previous findings in endothelium [69]) using quantitative real time polymerase chain reaction (qRT-PCR) (probes from Invitrogen (Table 2.5)).

Genes	Probe IDs	NCBI reference Sequence
NFkB2	Hs01028901_g1	NM_001077494.2
PCNA	Hs00427214_g1	NM_002592.2
BCL2A1	Hs00187845_m1	NM_004049
EFNA1	Hs00358886_m1	NM_004428.2

Table 2.5 Taqman Probe IDs

APC concentrations of 22.5nM, 45nM, 90nM and 180nM and treatment durations of 24 and 48 hours were tested, based on a previous microarray study [69]. RNA was extracted from APC-treated cells using the RNeasy Plus Kit (Qiagen), as per the manufacturer's instructions, alongside negative controls. Reverse transcription was performed using the superscript VILO cDNA synthesis kit (Life Technologies), as per the manufacturer's instructions. qPCR was performed using a Sensimix II probe kit (Bioline) on a ABI 750 qPCR machine (Applied Biosystems). The following program cycle was used: 50°C for 2 minutes; 95°C for 10 minutes; (95°C for 15 seconds; 60°C for 1minute) x 40. Data were analysed in Excel. The mean cycle threshold

values (Ct) were calculated using triplicate reactions, and were normalised against 18S controls to calculate gene expression. A concentration of 180nM for 24 hours was associated with maximal alteration in gene expression.

RNA was extracted from cells treated with 180nM APC for 24 hours using the RNeasy Plus Kit (Qiagen) alongside negative controls. Extracted RNA purity and concentration was confirmed using a spectrophotometer (NanoDrop 3300, Thermo Scientific). RNA was sent for two-colour Agilent Microarray analysis at the Functional Genomics Facility at the University of Birmingham, using Agilent SurePrint G3 Human Gene Expression v3 8x60K Microarrays. Three experimental replicates were performed. A minimum of two technical replicates were performed for each specimen. APC treated cells were assigned to Cy5 and controls to Cy3. Differential gene expression data was produced in R using the Limma package (Bioconductor) after 'Loess' normalisation [104-106]. Genes with a Bayes factor greater than 5 or an adjusted p value of <0.05 (Benjamini and Hochberg's method) were considered for further analysis. Further analysis was completed using DAVID [107, 108] and GSEA [109, 110].

Limma commands for microarray analysis:

```
library (limma)
```

```
targets <- readTargets("Targets.txt")
```

```
RG <- read.maimages(targets, path=".", source="agilent.median")
```

```
RG <- backgroundCorrect(RG, method="normexp", offset=16)
```

```
MA <- normalizeWithinArrays(RG, method="loess")
```

```

noControls <- MA$genes$ControlType == 0

MA2 <- MA[noControls,]

MA.avg <- avereps(MA2, ID=MA$genes$ProbeName)

design <- modelMatrix(targets, ref="Control")

fit <- lmFit(MA2, design)

fit2 <- eBayes(fit)

output <- topTable(fit2, adjust="BH", coef="Test", number=100000)

write.table(output, file="Results.txt", sep="\t", quote=FALSE)

```

2.14 sEPCR ELISA

EPCR ELISA was performed using a Human EPCR ELISA kit (Elabscience, E-EL-H0065). Serum from 21 early stage (Dukes A or B) and 19 late stage (Dukes C or D) CRC patients were obtained from the HBRC biobank. Reference standards were created with serial dilution. 100µl of serum or standard was added to a well of the ELISA plate which was then incubated at 37°C for 90 minutes. The solution was removed, and 100µl of EPCR biotinylated detection antibody was added and incubated for 60 minutes at 37°C. Each well was aspirated with wash buffer (350µl per well). The wash buffer was removed at each step by patting the plate dry on absorbent paper. The wells were washed three times in total. 100µl of HRP conjugate solution was added to each well and the plate was covered and incubated for 30 minutes at 37°C. The wells were washed five times as above. 90µl of

substrate solution was added to each well. The plate was sealed and incubated in darkness at 37°C for 15 minutes. The reaction was terminated by adding 50µl of Stop solution (containing sulphuric acid) to each well. The optical density was measured on a Victor plate reader at 450nm. Results were analysed on Excel.

2.15 CEA ELISA

CEA ELISA was performed using a CEA human ELISA kit (Abcam, ab99992). Reference standards were created with serial dilution. 100µl of serum or standard was added to a well of the ELISA plate which was then incubated at RT for 150 minutes. Each well was aspirated and washed four times with wash buffer (300µl per well). The wash buffer was removed at each step and the plate dried using absorbent paper. 100µl of biotinylated CEA detection antibody was added and incubated for 60 minutes at RT. The wells were washed four times as above. 100µl of HRP-streptavidin solution was added to each well and the plate was covered and incubated for 45 minutes at RT. The wells were washed four times as above. 100µl of substrate solution was added to each well. The plate was sealed and incubated in darkness at 37°C for 30 minutes. 50µl of Stop solution was then added to each well to terminate the reaction. The optical density was measured on a Victor plate reader at 450nm. Results were analysed on Excel.

2.16 RNAseq

2.16.1 RNA extraction

RNA was extracted from 3 FFPE scrolls per tumour using the Purelink FFPE total RNA isolation kit (Invitrogen) as per the manufacturer's instructions. After extraction, the samples were treated with 1µl/sample DNase (DNase I Amplification Grade, Thermo Fisher (1 unit/µl)) to eliminate any extracted DNA. The RNA concentration was determined using the Qubit RNA Broad Range assay kit (Life Technologies), as per the manufacturer's instructions.

2.16.2 Truseq panel

A custom Illumina Truseq library/panel was designed in collaboration with Dr Andrew Beggs. This panel targeted the genes in Table 2.6. The genes selected included the 28 genes in the CIRC metagene, a host of additional immune-related genes, and genes associated with the Dry et al MEK signature. The CIRC and Dry et al MEK signature are described in chapter 4.

ALOX5AP	ELF1	ICOS	MAP2K3
BIN2	ETV4	IDO1	NCKAP1L
BNIP3	ETV5	IFNG	ORM1
CCL18	EVI2A	IGHA1	PDCD1
CCL5	EVI2B	IGLL1	PDCD1LG2
CD247	FOXP3	IL10	PHLDA1
CD274	FXYP5	IL10RA	PIAS1
CD4	GNLY	IL12A	PROS1
CD48	GZMH	IL12B	PTPRC
CD53	GZMK	IL17A	S100A6
CD74	HAVCR2	IL18	SASH3
CD80	HLA-A	IL18RAP	SERPINB1
CD8A	HLA-B	IL6	SLCO4A1
CIITA	HLA-C	IRF1	SOCS1
CSF1R	HLA-DMA	ITGB2	SPRY2
CTLA4	HLA-DMB	JAK2	SRGN
CXCL10	HLA-DOA	KANK1	STAT1
CXCL11	HLA-DPA1	LAG3	TBX21
CXCL13	HLA-DPB1	LAPTM5	TRIB2
CXCL9	HLA-DQA2	LAT2	VEGFA
CYBB	HLA-DRA	LCP2	WAS
DOCK2	HLA-DRB1	LGALS3	
DUSP4	HLA-DRB5	LST1	
DUSP6	ICAM1	LZTS1	

Table 2.6 Genes analysed in targeted RNAseq panel.

Figure 2.2 contains a summary of the library preparation workflow. In brief, 400ng of the RNA extracted from each tumour was used for the protocol. The RNA was reverse transcribed into first strand cDNA using Protoscript II reverse transcriptase and random primers. The Oligo pool was then hybridised to the cDNA. This hybridized an oligo pool containing upstream and downstream oligos specific to the targeted regions of interest to the cDNA samples and bound them to paramagnetic streptavidin beads. The bound oligos were then washed, extended and ligated. This removed unbound oligos from cDNA. A DNA polymerase extended from the upstream oligo through the targeted region, followed by ligation to the 5' end of the downstream oligo using a DNA ligase. This resulted in the formation of products containing the targeted regions of interest flanked by sequences required for amplification. Then PCR amplification was performed, amplifying the extension-ligation products using primers that added index sequences for sample multiplexing. The resultant output was cleaned using AMPure XP beads, to purify the PCR products from the other reaction components. The library was then pooled and quantified. This combined, quantified, denatured, and diluted equal volumes of library in hybridisation buffer in preparation for sequencing. The sequencing was performed on a Miseq sequencing system (Illumina). The full protocol is available online:

<http://support.illumina.com/downloads/truseq-targeted-rna-expression-protocol-guide-1000000005011.html>.

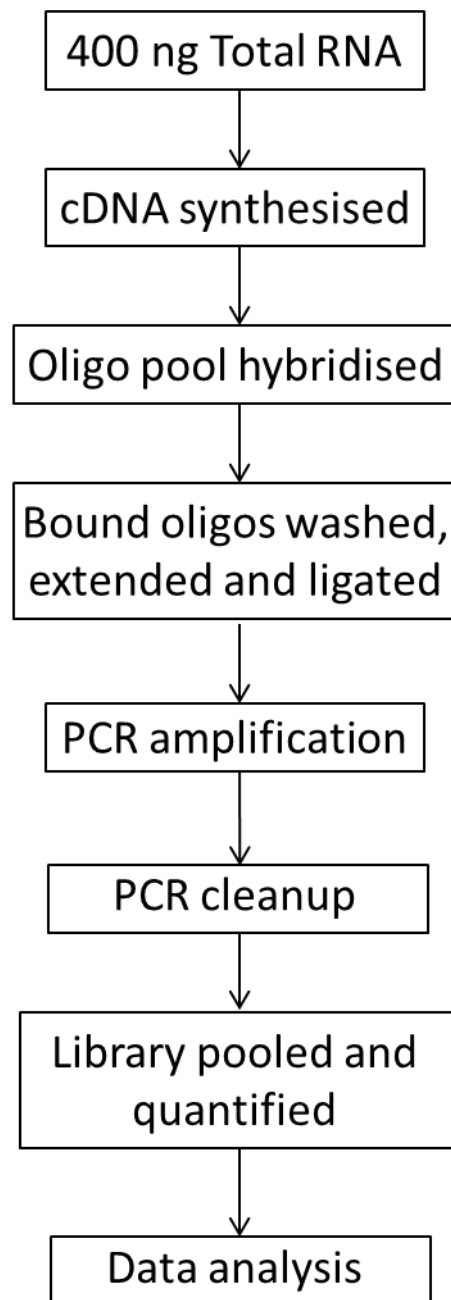


Figure 2.2 Summary of RNAseq library preparation workflow.

2.17 Microsatellite status

DNA was extracted from one FFPE scroll per tumour using the Maxwell RSC DNA FFPE kit (Promega) as per the manufacturer's instructions. DNA concentration was determined using the Qubit dsDNA BR assay kit (Life Technologies).

Microsatellite status was determined with a protocol adapted from Buhard et al. [111]. This technique uses a refined version of the 'Bethesda panel' of five markers for the analysis of MSI. The Bethesda panel includes two mononucleotide (BAT-25 and BAT-26) and three dinucleotide (D5S346, D2S123, and D17S250) repeats. Tumours with instability at two or more of these markers are defined as being MSI-H, whereas those with instability at one repeat or showing no instability are defined as MSI-low and MSS tumours, respectively.

Extracted DNA from tumour samples underwent polymerase chain reaction (PCR) with the reagent volumes in Table 2.7.

Reagent	Volume per reaction (µl)
Water	19
5x	5
Primer Mix	0.6
MyTaq	0.2
Sample (DNA)	1.0

Table 2.7 Microsatellite status PCR reagents

PCR reaction conditions were 95°C for 2 minutes, then 33 cycles of 95°C for 15 seconds, 55°C for 15 seconds and 72°C for 50 seconds. Successful amplification was confirmed by running the product on an agarose gel.

Fragment analysis was conducted on an Applied Biosystems 3730 DNA analyser set to run in 'fragment analysis' mode. 1µl sample (from a 1/100 dilution of PCR product), 9µl of formamide and 1µl LIZ-600 size standard (Applied Biosystems) were loaded onto each well of the plate. Sizes of microsatellite markers were determined using Peak Scanner software and microsatellite status in each sample was assessed.

Chapter 3

The role of EPCR in colorectal tumourigenesis

3.1 Introduction

3.1.1 The physiological role of EPCR

Endothelial protein C receptor (EPCR) is a type I transmembrane protein that is homologous to Major Histocompatibility Complex (MHC) molecules. It was initially described as a receptor for Protein C (PC), which is crucial in the down modulation of blood coagulation [61]. EPCR increases the conversion of the zymogen PC to Activated Protein C (APC), and is a receptor for both PC and APC. In endothelium, APC cleaves factors V and VIII, which is central to its anti-coagulative effect [60]. In addition to coagulation, EPCR expression on endothelium is known to play important roles in inflammation, apoptosis and cell proliferation [69, 73, 74, 112]. APC binding to EPCR initiates intracellular signalling and alters gene transcription [69]. This signalling is dependent on Protease-Activated Receptor 1 (PAR1). Downstream effects of PAR1 activation include ERK and Akt phosphorylation and subsequent alteration in the expression of inflammatory genes [69, 74]. Pro and anti-apoptotic proteins such as Bax and Bcl2, as well as p53, are modulated by PAR1 cleavage [73]. Subsequently, it is thought that APC/EPCR, through cleavage of PAR1, has anti-inflammatory and anti-apoptotic effects [113].

3.1.2 EPCR as a ligand for $\gamma\delta$ T-cells

EPCR has been shown to be upregulated in various cancer cell lines [59]. Moreover, work in our laboratory has highlighted EPCR as a direct ligand for V δ 2-negative $\gamma\delta$ T-cells [53], the predominant tissue subset of these unconventional T cells. $\gamma\delta$ T cells are able to recognise 'altered self', by detecting upregulation of host components *via* germline-encoded receptors (eg NKG2D) or potentially *via* the $\gamma\delta$ TCR, in either

microbial or non-microbial stress. EPCR is a direct ligand for the V γ 4 δ 5 LES TCR, originally isolated from a patient with acute CMV. EPCR, expressed on endothelial cells infected by CMV *in vivo*, was not upregulated by CMV infection, but TCR recognition of EPCR in the context of a CMV-induced 'multimolecular stress signature' was sufficient to induce LES recognition of target cells. LES was able to recognise various cell lines expressing EPCR in an EPCR-dependent manner. These studies highlighted the possibility that EPCR might act as a signal of epithelial tumourigenesis to $\gamma\delta$ T cells. To understand this more fully, a greater knowledge of EPCR's role in epithelial tumourigenesis was required. I therefore sought to understand the extent and significance of EPCR's expression in epithelial cancers, the cellular mechanisms of its expression, and its functional significance in transformed tumour cells. EPCR-associated signalling pathways have potential relevance in cancer, overlapping with key proliferative and apoptotic pathways in the tumour setting, raising the possibility that dysregulated EPCR expression on transformed epithelial tissue may directly effect similar mechanisms to promote tumour cell survival and growth.

3.1.3 EPCR in cancer

Recently, there has been increasing interest in EPCR's potential role in the epithelial tumour setting. However, there have been conflicting results in different tumour types. A number of studies suggest that EPCR expression increases tumour cell survival. In mouse xenograft models of lung cancer, treatment with EPCR-blocking antibodies reduces metastatic tumour burden [78]. Knockdown of EPCR in a gastric cancer cell line suppresses cell proliferation and migration [114]. In breast cancer, EPCR is a marker of cancer stem cell-like populations, which are thought to have a

high tumour-initiating capacity, and *in vivo* EPCR blockade attenuates tumour growth [80]. Conversely, mouse melanoma studies suggest that APC treatment and endothelial EPCR overexpression are associated with decreased metastasis development in this tumour type [79]. Additionally, in *in vivo* models of mesothelioma, EPCR knockdown significantly increases tumour growth and burden [83].

These opposing results demonstrate that there is ambiguity regarding EPCR's significance in the cancer setting, with varying effects in different tumour types. However, there are clearly suggestions that EPCR is clinically relevant. Significantly, EPCR has been shown to be a marker of chemoresistance in tumour cell lines [59], including colorectal cancer cell lines such as HCT116. Furthermore, EPCR expression is predictive for chemotherapy response in early stage non-small cell lung cancer [78]. Finally, in ovarian cancer, serum EPCR expression is correlated with the tumour marker CA-125, suggesting possible clinical relevance as a biomarker [86]. The role of EPCR in CRC, a tumour type with high mortality and prevalence [1], has not been explored, and in view of the well-established developmental pathway and pathological characterisation of CRC, this tumour type is a promising model in which to clarify the role of EPCR in tumourigenesis.

3.1.4 Chapter aims

In this chapter, I have characterised EPCR expression in CRC, and have provided potential explanations into the mechanisms of its expression. The role of APC induced signalling on ERK and gene transcription has been explored. Furthermore, I have shown that EPCR has a modest impact on tumour cell chemosensitivity and

influences migration. Finally, utilizing cancer tissue from the MRC COIN clinical trial [10], which investigated the efficacy of EGFR monoclonal antibody therapy in metastatic colorectal cancer, I provide insights into EPCR's likely clinical relevance. The data provide a compelling mechanism for EPCR expression in multiple epithelial tumour settings, and suggest that its upregulation could represent a wider dysregulation of chromosome 20q. These findings have implications for understanding the clinical significance of EPCR in different tumour settings. Furthermore, they highlight that single $\gamma\delta$ TCR ligands may have the potential to denote 'altered self' in different ways in distinct stress stimuli, including infection and tumourigenesis.

3.2 Results

3.2.1 EPCR mRNA expression in multiple cancer types

The Oncomine database, which enables comparative analysis of gene expression in tumours versus normal tissue, was interrogated to determine EPCR expression in multiple tumour types [94]. EPCR was upregulated in tumour versus normal in 126 separate datasets and downregulated in 50 datasets, which included 21 different cancer types. Of all cancer types, colorectal cancer had the highest level of EPCR overexpression, and the most datasets in total in which EPCR was overexpressed (Table 3.1). In CRC, across 9 separate non-overlapping datasets, EPCR was significantly overexpressed ($p < 0.0001$). This suggested amongst all tumours, colorectal cancer had one of the highest and most consistent levels of EPCR upregulation.

Interestingly in certain cancer types, particularly ovarian cancer, EPCR was amongst the most downregulated genes. This highlights that the expression of EPCR is highly variable across different cancers.

Analysis type by cancer	Cancer vs. normal	
Bladder cancer	4	
Brain and CNS cancer	13	
Breast Cancer	3	9
Cervical Cancer	3	
Colorectal cancer	22	
Oesophageal cancer	8	2
Gastric cancer	8	
Head and Neck cancer	14	
Kidney cancer	9	1
Leukaemia	8	1
Liver cancer	4	2
Lung cancer		5
Lymphoma	11	1
Melanoma		2
Myeloma	1	
Other cancer	12	4
Ovarian cancer		10
Pancreatic cancer	5	
Prostate cancer		9
Sarcoma	2	4

Percentile of underexpressed genes containing EPCR			Percentile of overexpressed genes containing EPCR		
1%	5%	10%	10%	5%	1%

Table 3.1 ONCOMINE expression data – number of datasets in which EPCR is significantly differentially expressed in cancer versus normal to $p < 0.0001$. Cell colour is determined by the highest overexpression (left column) or underexpression (right column) gene rank percentile for EPCR.

3.2.2 EPCR expression in colorectal cancer

To confirm EPCR's upregulation in colorectal cancer, sections from 30 colorectal cancers and adjacent normal regions were stained for EPCR using immunohistochemistry. The primary EPCR antibody was validated using a HCT116 EPCR knockdown cell line, and endothelium acted as a positive control (Figure 3.1). In addition, a control was performed with the primary antibody omitted (data not shown).

Stained sections were imaged on the Vectra platform and were analysed using Inform software (Figure 3.2 and 3.3). Segmentation and scoring algorithms were validated by a Histopathologist. This confirmed that EPCR reactivity was higher in tumour regions compared to normal regions ($p < 0.0001$). Reactivity was measured using the H-score, which is a function of reactivity intensity and percentage coverage. The mean H score in tumour regions was 244.2 (standard error 5.03), versus 87.6 (standard error 10.43) in normal regions. EPCR reactivity was observed in the cytoplasm and membrane compartments. Intense reactivity of what appeared to be mast cells (as confirmed by a Histopathologist) was observed throughout tissue sections (Figure 3.2D). Collaborative studies by Professor Peter Bradding confirmed EPCR expression on mast cells (Figure 3.4). This is the first report of immunoreactivity to EPCR on mast cells to my knowledge.

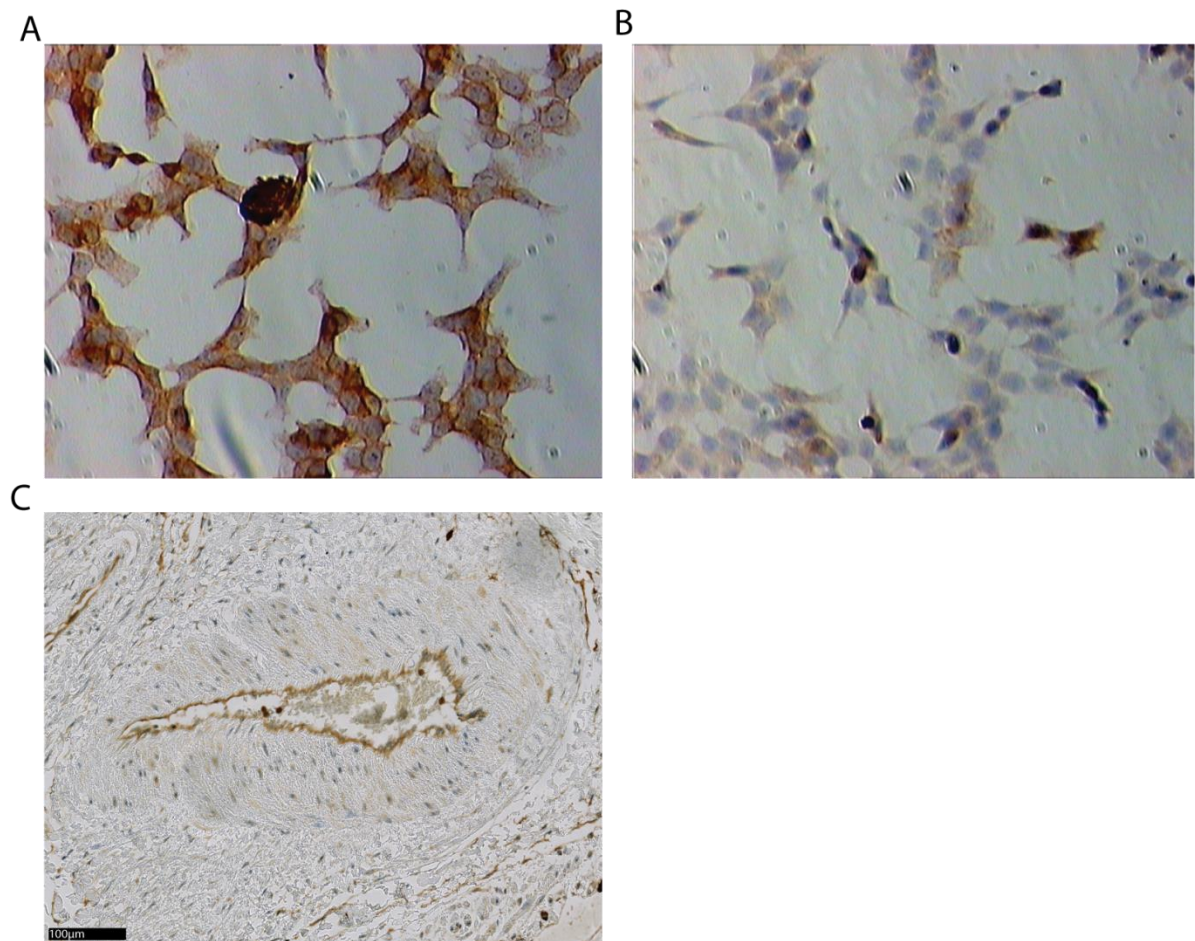


Figure 3.1 Validation of EPCR antibody.

- a) IHC staining of wild type HCT116 cells with anti-EPCR antibody.
- b) IHC staining of EPCR shRNA knockdown HCT116 cells with anti-EPCR antibody.
- c) Positive EPCR immunoreactivity in endothelium.

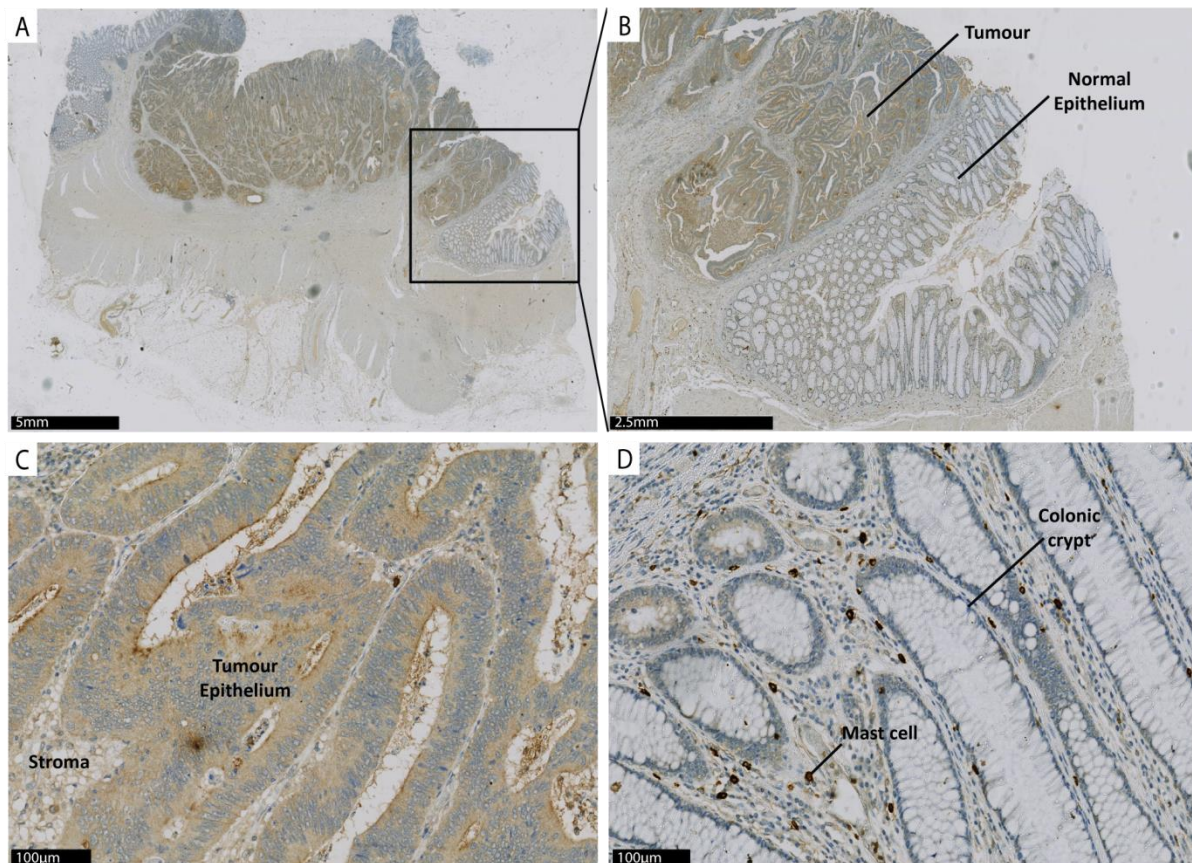


Figure 3.2 Scanned image of EPCR-stained CRC section.

- a) Whole slide scanned image of CRC region surrounded by normal colonic mucosa.
- b) 5x magnified image of tumour with adjacent normal colonic mucosa
- c) 20x magnified image of tumour region, demonstrating strong EPCR staining
- d) 20x magnified image of normal region, with negative staining of colonic crypts

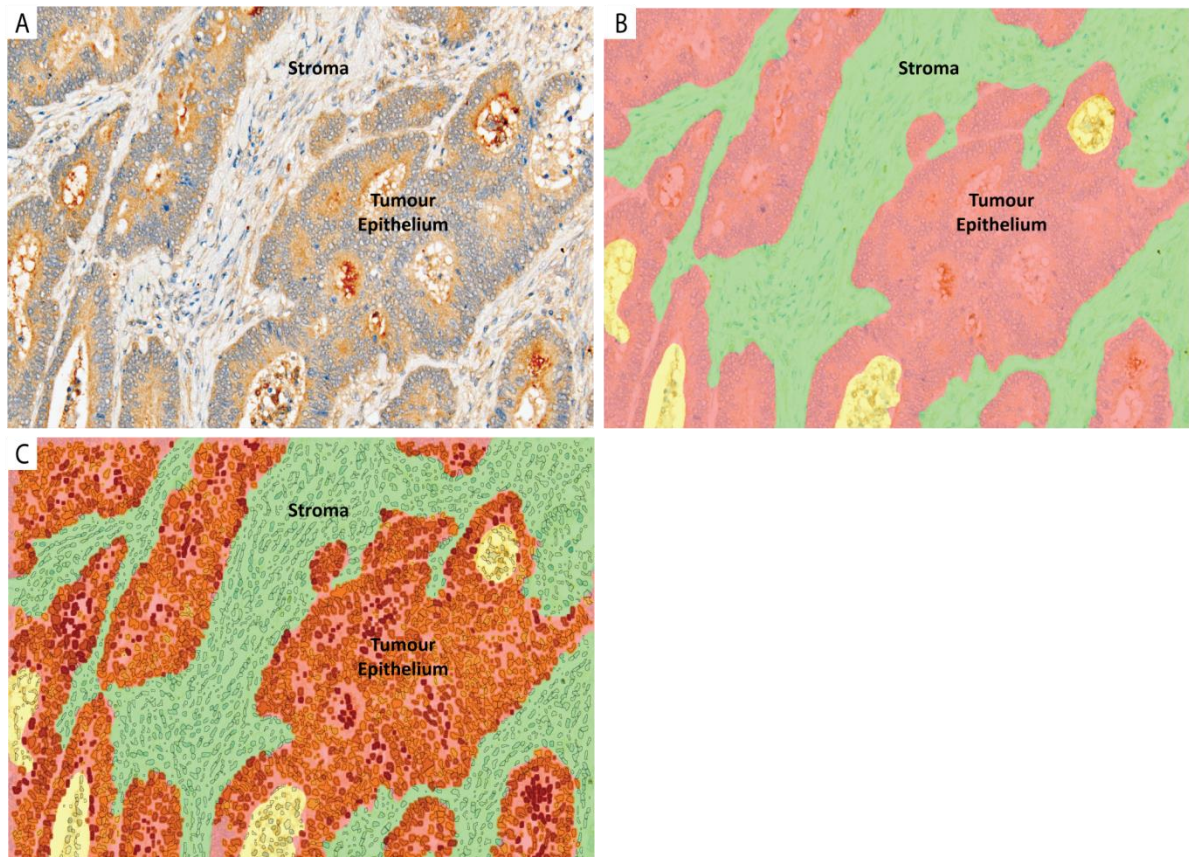


Figure 3.3 Digital software analysis of EPCR-stained CRC section.

- a) Area for digital analysis with Inform software (20x magnification)
- b) Image a) with Inform tissue-segmentation through trained automated algorithm. Red=epithelium, green=stroma, yellow=background
- c) Image b) with staining intensities within tumour cells. Red=strong, orange=moderate, yellow=weak

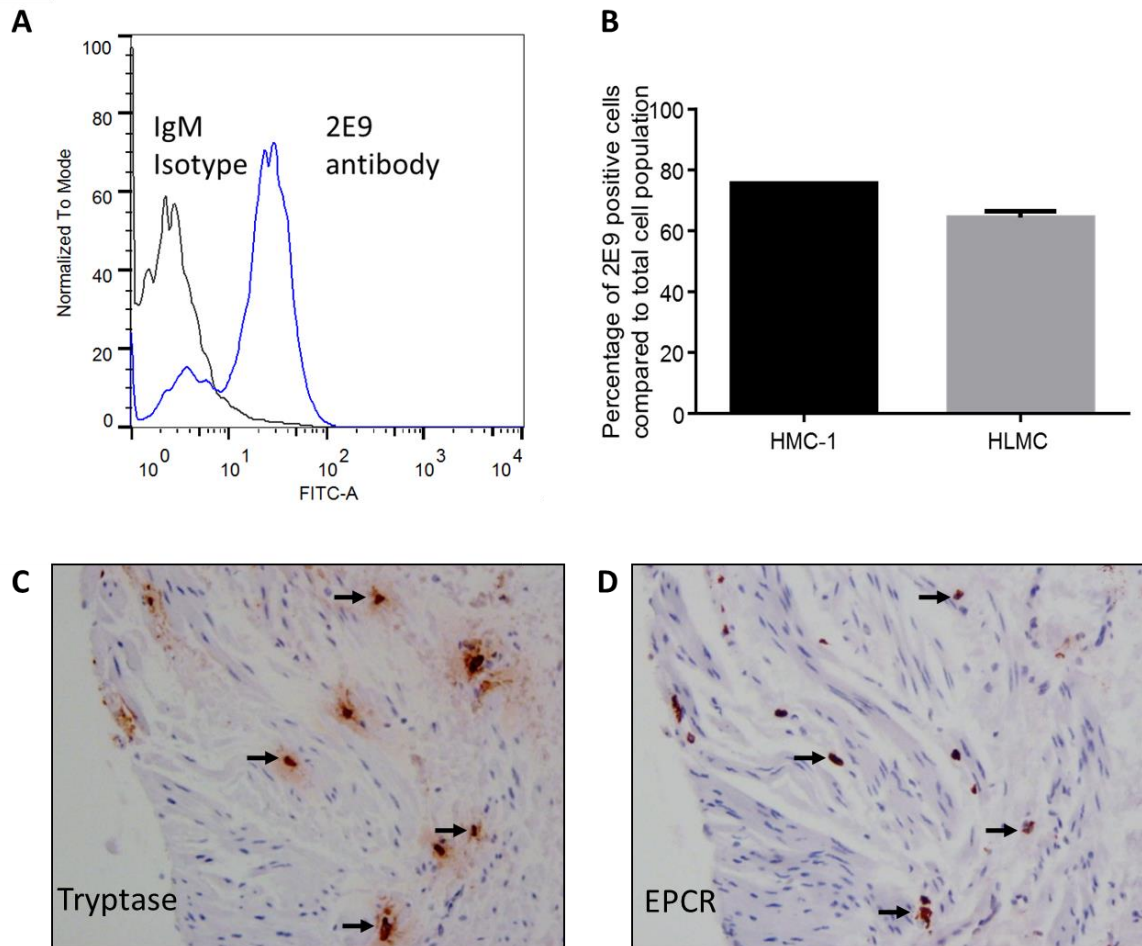


Figure 3.4 EPCR expression on mast cells.

- a) Flow cytometry demonstrating EPCR protein expression on mast cells using 2E9 antibody
- b) Flow cytometry data are presented as percentage of 2E9+ cells of the total cell population using the Overton method [115] for Human Mast Cell line 1 (HMC-1) and Human Basophil and Lung Mast Cell line (HLMC)
- c) and d) Sequential sections demonstrating co-localisation of EPCR to tryptase positive mast cells within airway tissue (*black arrows*, x20 magnification).

3.2.3 The association between EPCR expression in colon cancer and gene amplification and hypomethylation

To determine the possible mechanisms of EPCR overexpression, I utilised public bioinformatic genomics and transcriptomics datasets [91]. The EPCR gene (*PROCR*), though rarely mutated in any cancer type, was frequently amplified (Figure 3.5). In the CRC TCGA dataset, gene amplification was significantly associated with increased mRNA expression ($p < 0.05$). Overall, there was a strong correlation between *PROCR* copy number and mRNA expression (Spearman $\rho = 0.325$, $p < 0.00001$) (Figure 3.6A). In the same dataset, the chromosomal unstable (CIN) group of cancers had a significantly higher EPCR expression than the non-CIN group ($p < 0.01$). Despite this, some EPCR overexpressors were of the non-CIN microsatellite unstable (MSI) hypermutated subtype. These tumours were mostly diploid, thus gene amplification could not account for overexpression. I therefore suspected that epigenetic factors were also important in determining EPCR expression. Indeed, there was a strong inverse correlation between *PROCR* promoter methylation and gene expression across the entire patient dataset (Spearman $\rho -0.59$, $p < 0.001$) (Figure 3.6 B, C). When copy number data were combined with methylation data, it became clear that hypomethylation played a key role in determining expression across all patient groups (Figure 3.7). Patients that overexpressed EPCR had significantly lower methylation than those that did not ($p < 0.001$). Amplification and hypomethylation when combined associated with the highest expression. Having established that EPCR is overexpressed in a large number of colorectal tumours, and the likely mechanism, the wider genomic context of EPCR expression was then investigated.

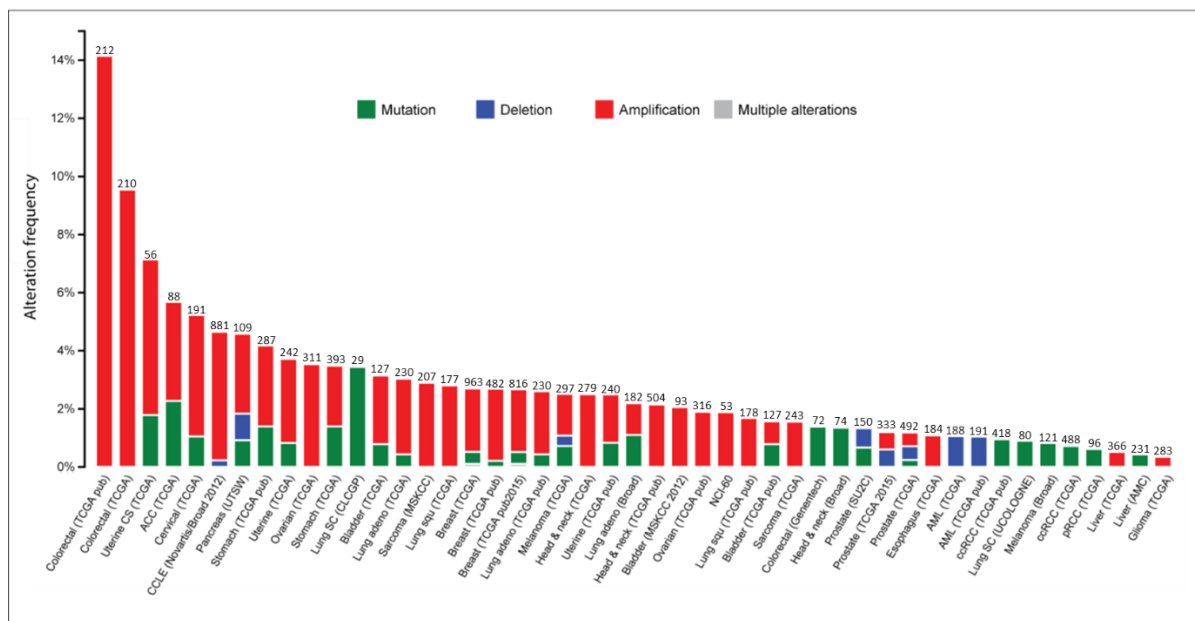


Figure 3.5 PROCR amplification, deletion and mutation rates across a range of cancer types and databases [89, 92]. The number above each bar represents the size of each dataset.

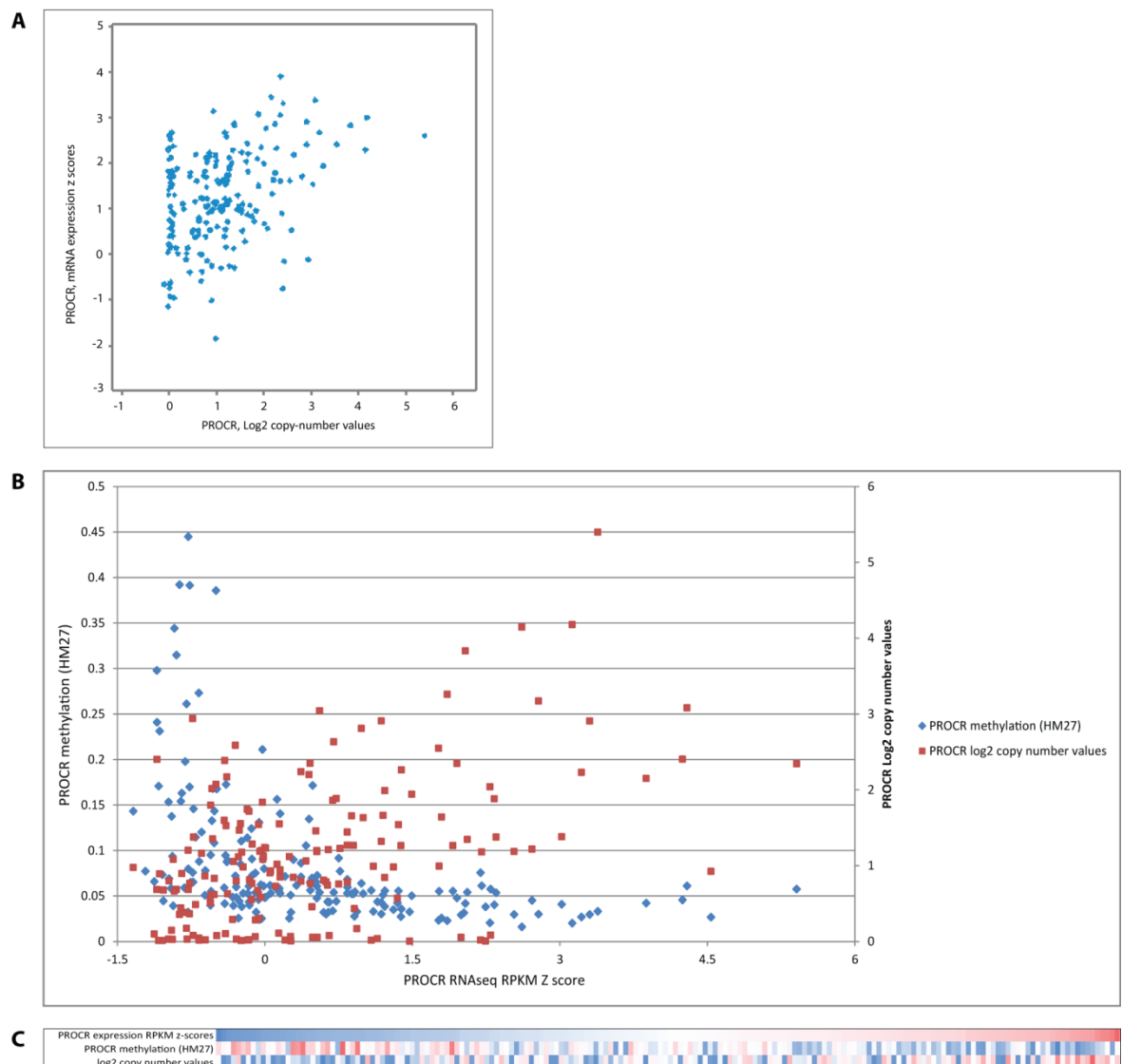


Figure 3.6 The association between EPCR expression and *PROCR* gene amplification and hypomethylation in the TCGA CRC dataset (n=195)

- PROCR* copy number versus mRNA expression
- Variation in *PROCR* copy number and promoter methylation with mRNA expression
- Heatmap showing correlation between *PROCR* mRNA expression, promoter methylation and copy number values (blue = low, red= high)

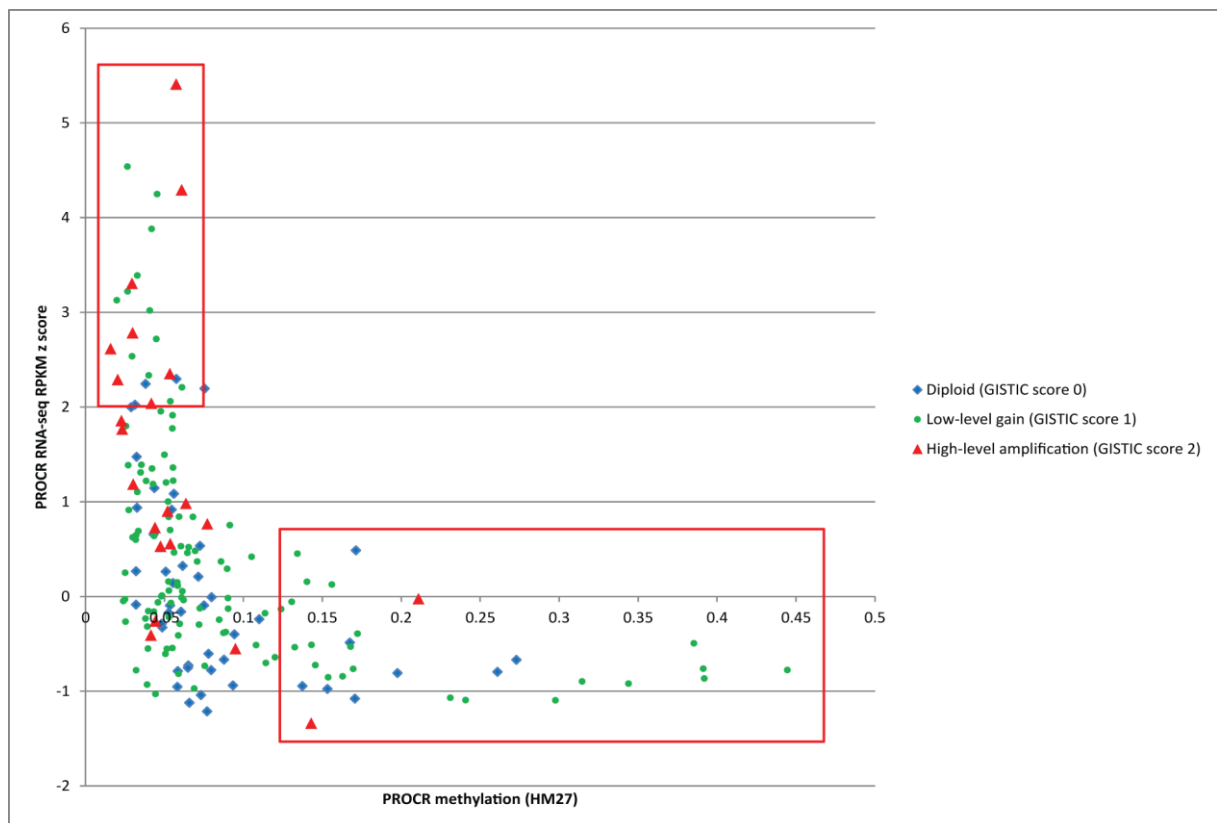


Figure 3.7 Relationship between PROCR methylation, GISTIC copy number scores and expression. GISTIC (Genomic Identification of Significant Targets in Cancer) scores of 0 (blue) represent tumours with diploid PROCR copy number, scores of 1 (green) represent tumours with low level gains in PROCR copy number, and scores of 2 (red) represent tumours with high level amplification of PROCR copy number. The upper box highlights patient group with low methylation and highest expression. The lower box highlights most highly methylated patients, who have low expression regardless of copy number status.

3.2.4 The association between *PROCR* expression and chromosome 20q amplification

To determine whether *PROCR* was co-expressed with other genes, I explored TCGA expression and genetic data. The majority of genes whose expression correlated most closely with *PROCR* were located in the same chromosomal region - chromosome 20q (Table 3.2).

On an individual patient basis, *PROCR* gene amplification was frequently associated with amplification of all genes in chromosome 20q (Figure 3.8). Across the entire 195 patient cohort, the copy numbers of the majority (99.42%) of genes located on chromosome 20q were positively correlated with *PROCR* copy number (Figure 3.9A), whereas a significantly smaller proportion of 20p genes were correlated (36.90%, $p < 0.05$ (chi-squared)). In addition, the expression of a high proportion (53.80%) of 20q genes were positively correlated with *PROCR* expression, though the correlations were not as strong as those seen with copy number (Figure 3.9C). A significantly smaller proportion of 20p genes were correlated in terms of expression (13.72%, $p < 0.05$ (chi-squared)). Finally, 55.33% of chromosome 20q genes correlated with *PROCR* in terms of promotor methylation. However, a similar proportion of 20p genes were also correlated in terms of methylation (51.30%, $p = 0.47$, chi-squared) (Figure 3.9B). This suggests that whilst *PROCR* copy number and expression co-regulation are a 20q regional phenomenon, regulation of methylation is less region-specific.

Across the entire colon cancer dataset, chromosome 20q was amongst the most frequently amplified chromosomal regions (Figure 3.10). Additionally, 20q was amplified across a range of cancer types, including several in which EPCR has been shown to be expressed, including melanoma [79], gastric cancer [114] and lung squamous cell cancer [78] (Figure 3.10). These data suggest that EPCR may be a marker of tumours with chromosome 20q amplification, which are known to have poorer outcome [116, 117].

Gene Symbol	Spearman Score vs PROCR (mRNA expression)	Chromosomal location
HM13	0.45	20q11.21
PDRG1	0.44	20q11.21
C20ORF24	0.43	20q11.23
PSMA7	0.43	20q13.33
TPD52L2	0.43	20q13.33
TRPC4AP	0.43	20q11.22
MRGBP	0.43	20q13.33
ADRM1	0.42	20q13.33
BRI3	0.42	7q21
EIF6	0.41	20q11.22
ATP6V1F	0.41	7q32.1
EFNA2	0.41	19p13.3
SSUH2	0.41	3p25.3
NDRG3	0.4	20q11.23
RHOD	0.4	11q13.2
EDEM2	0.39	20q11.22
ACTR5	0.39	20q11.23
DYNLRB1	0.39	20q11.22
CEBPA	0.39	19q13.11
FAM96B	0.39	16q22.1
UQCC1	0.38	20q11.22
PIGU	0.38	20q11.22
ROMO1	0.38	20q11.22
AHCY	0.38	20q11.22
NEU1	0.38	6p21.33
TOMM34	0.38	20q13.12
TCFL5	0.38	20q13.33
RPS21	0.38	20q13.33
TLDC2	0.37	20q11.23

Table 3.2 Genes most highly co-expressed with EPCR in TCGA CRC dataset, and their chromosomal locations

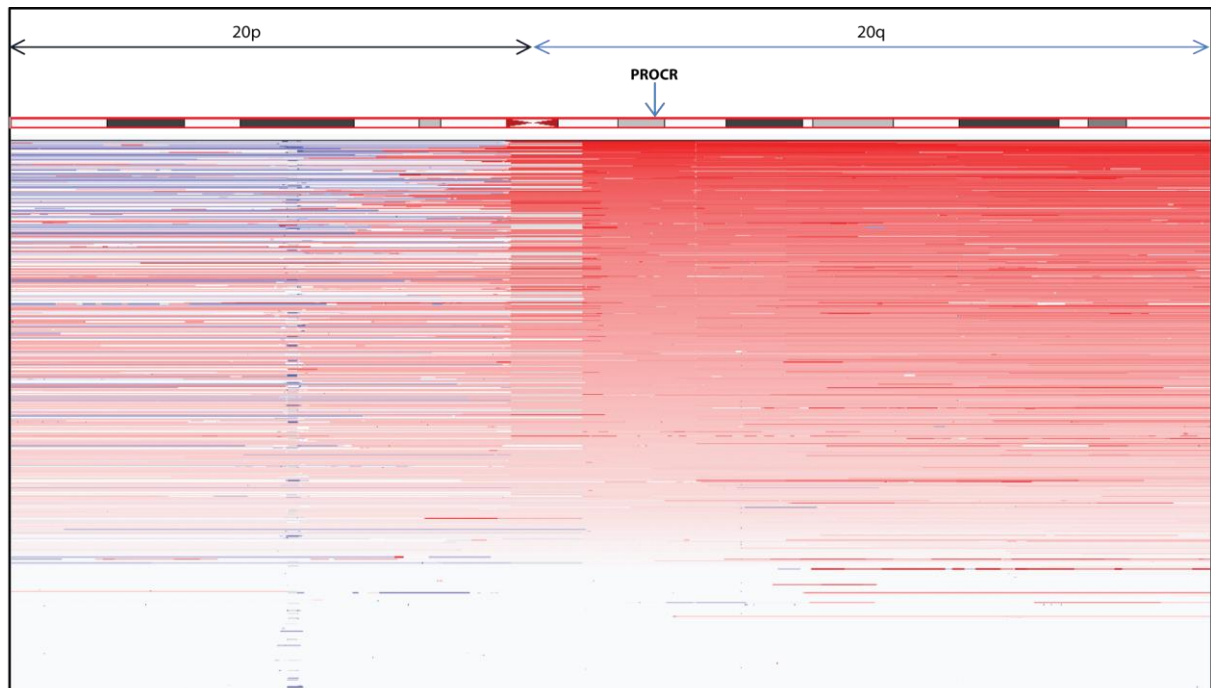


Figure 3.8 Copy numbers of chromosome 20 genes in 195 TCGA colorectal cancer patients. Genes are organised by chromosomal location (20p is left, 20q is right). Each individual horizontal coloured line represents a patient. Blue represents loss of gene, red represents gain of gene and white represents no change in copy number. The location of PROCR is indicated by the arrow.

Image created using Integrated Genomics Viewer.

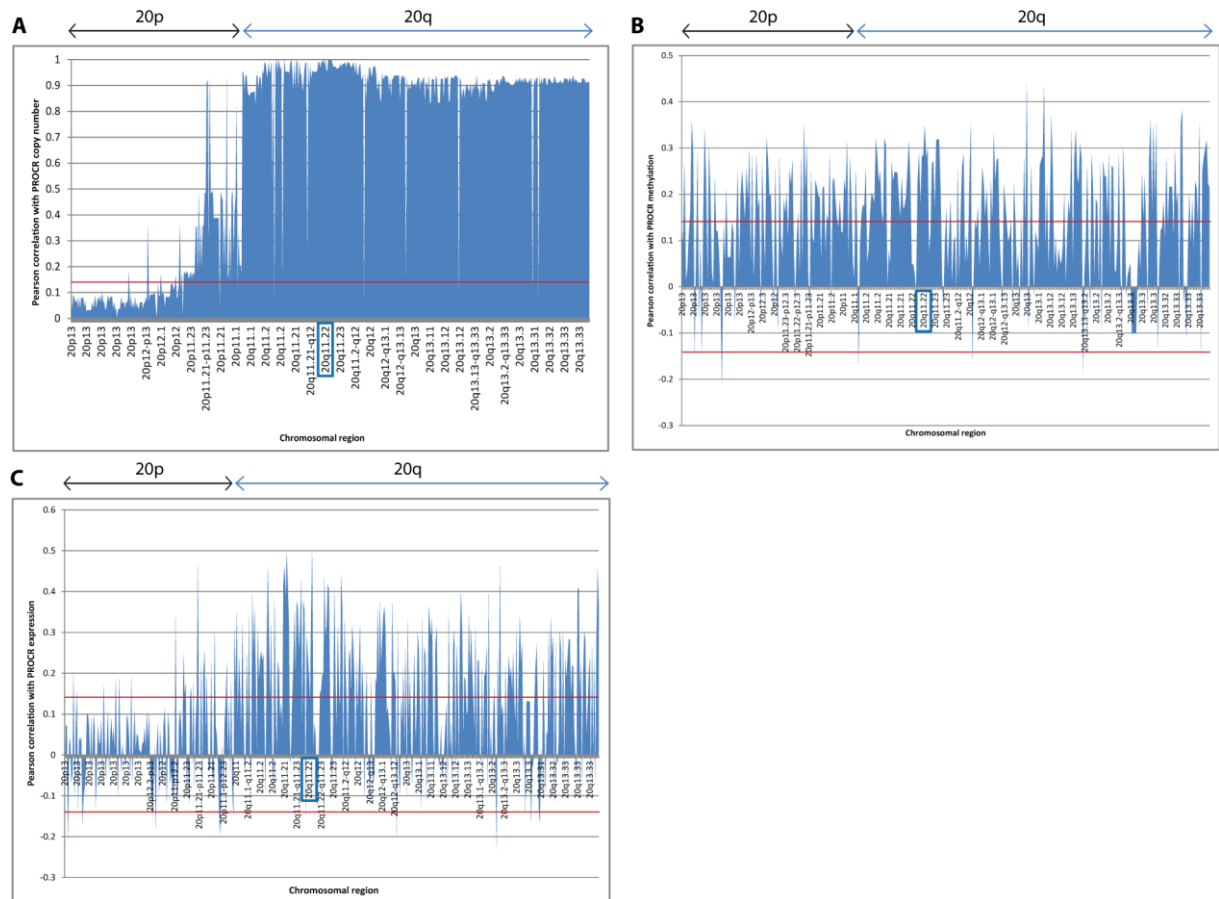


Figure 3.9 PROCR correlations with chromosome 20 genes.

Pearson correlations for PROCR gene (A) copy number, (B) methylation and (C) expression with genes on chromosome 20, arranged by chromosome region. PROCR is located on 20q11.2 (boxed). Red lines represent the thresholds of statistical significance (Pearson value >0.1405 or <-0.1405 , significant to $p<0.05$).

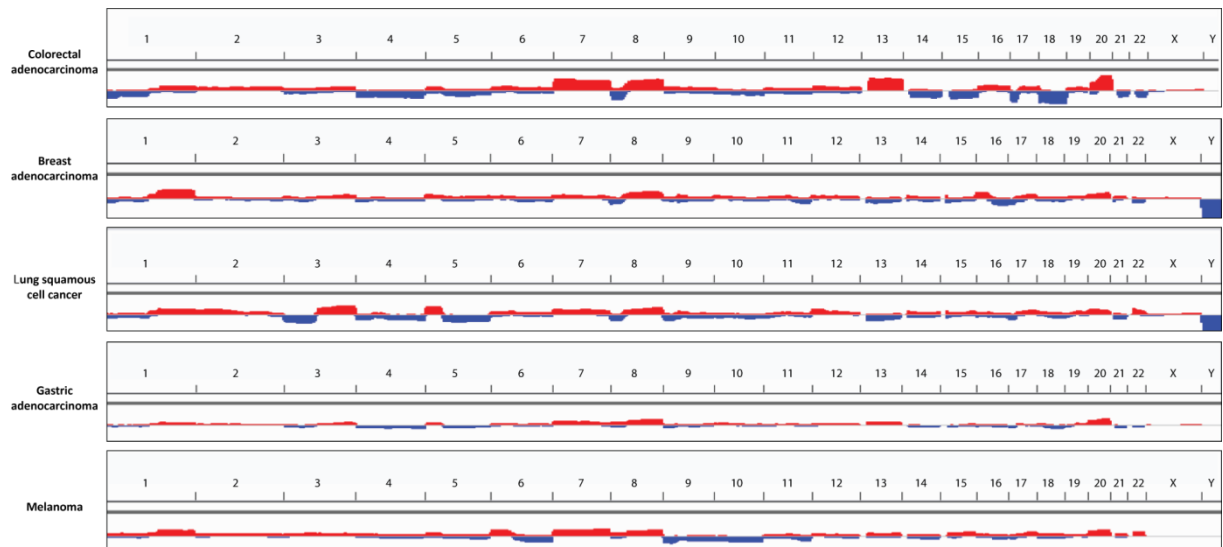


Figure 3.10 Genomic copy number variations across a range of cancers.

Red represents gene and chromosome amplification. Blue represents deletion of genes and chromosomes. Chromosomes are ordered from 1 to 23,X,Y. Chromosome 20q is one of the most frequently amplified regions in colon cancer and across a range of cancer types (colon adenocarcinoma, breast adenocarcinoma, lung squamous cell cancer, gastric adenocarcinoma and melanoma TCGA data).

Image created using Integrated Genomics Viewer.

Carvalho et al. previously identified 3 regions, called ‘smallest regions of overlap’ (SROs), of chromosome 20q that are commonly amplified in CRC [117]. *PROCR* is located within the first of these (SRO 1) which spans 32-36 Mb. The group identified 7 putative oncogenes located on 20q based on upregulation on carcinomas versus adenomas, association with 20q amplification, and correlation between copy number and gene expression. Of these 7 genes, 6 significantly correlated with EPCR expression (Table 3.3). Furthermore, EPCR expression was significantly correlated with 13/13 potential ‘cancer initiating genes’ located on 20q identified by Tabach et al. [116]. Finally, *PROCR* was significantly correlated with a gene that was previously ranked first in a microarray-based CIN signature (*TPX2*) [118], reinforcing the view that its expression is associated with chromosomal instability.

	Chromosomal location	Copy number	Methylation	Expression
C20orf24	20q11.22	0.989164	0.227405	0.498511
AURKA	20q13.2	0.907967	0.009653	0.239549
RBM38 (RNPC1)	20q13.32	0.928319	0.21619	0.053815
NELFCD (TH1L)	20q13	0.901242	0.155924	0.334379
ADRM1	20q13.33	0.920162	0.295276	0.454516
MRGBP (C20orf20)	20q13.33	0.909783	N/A	0.409151
TCFL5	20q13.3-qter	0.909783	0.214221	0.364161

Table 3.3 Correlation of the Carvalho et al. putative chromosome 20q oncogenes and *PROCR*. Table values for copy number, methylation and expression are Pearson scores (R) versus *PROCR* in the TCGA CRC dataset. $P < 0.05$ where $R > 0.1405$, highlighted red.

Having established that EPCR upregulation correlates with regional dysregulation, I then sought to understand the significance of EPCR expression in tumourigenesis.

3.2.5 EPCR expression in colorectal cancer cell lines

Using the Cancer Cell Line Encyclopaedia [93], *in silico* analyses of EPCR expression in multiple colorectal cancer cell lines were performed to inform *in vitro* studies. 5 cell lines were selected which were predicted to have high EPCR expression (HCT116, HT29, RKO, COLO320 and SW480). EPCR expression was confirmed with flow cytometry (Figure 3.11). All CRC carcinoma cell lines that were tested expressed EPCR. Both HCT116 and HT29 cell lines had high expression, and these lines were chosen for further *in vitro* studies as they represent the two main biological subtypes of CRC (MSI-H and CIN respectively). EPCR expression was also investigated in adenoma cell lines. The AA/C1 adenoma line exhibited low expression (Figure 3.11 F,G), but its malignant derivative AA/C1 10C had higher expression.

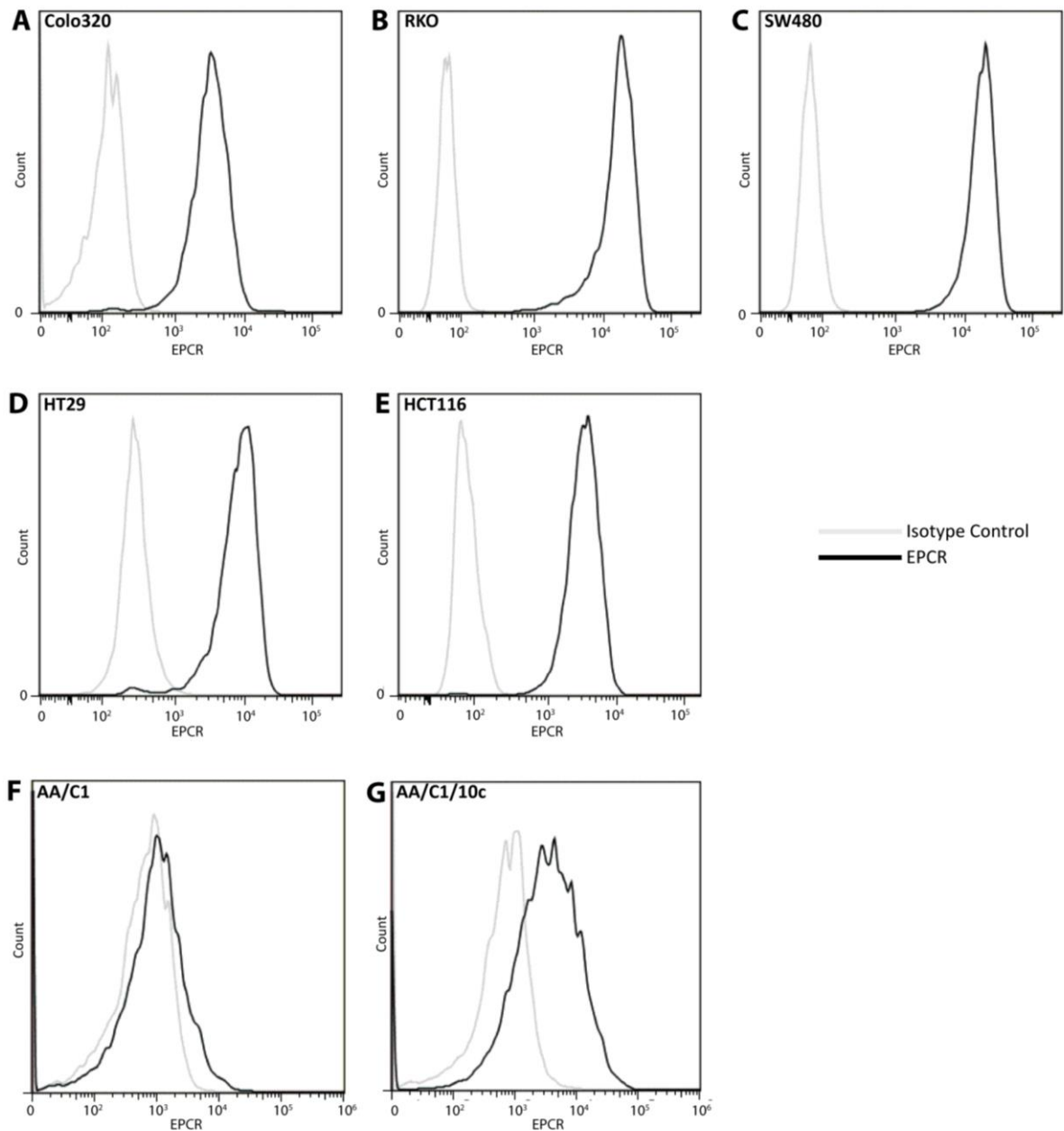


Figure 3.11 Expression of EPCR on colorectal cancer cell lines. Expression of EPCR on a) Colo320, b) RKO, c) SW480, d) HT29 and e) HCT116 cells by flow cytometry. Expression of EPCR on f) AA/C1 non-tumourigenic adenoma cells or g) the tumourigenic AA/C1/10C derivative. Isotype control staining is shown in grey and EPCR staining in black.

3.2.6 Exogenous Activated Protein C (APC) induction of ERK phosphorylation in colon cancer cells

ERK is a key MAP kinase that mediates several of the downstream effects of EGFR and RAS signalling, as well as other pathways [10, 119]. APC treatment of vascular endothelial cells has been shown to induce ERK phosphorylation [120]. Therefore, I wanted to investigate whether EPCR-mediated signalling could induce ERK phosphorylation in the CRC setting, as this could have implications for the effects of EPCR upregulation on tumour cells. HCT116 cells were therefore treated with exogenous Activated Protein C (APC) to determine whether APC could induce ERK phosphorylation in this colorectal cancer cell line (Figure 3.12). Treatment of serum-starved HCT116 cells with APC for 2-10 minutes was associated with an increase in ERK phosphorylation. This increase was inhibited by EPCR-blocking antibodies, suggesting that APC-induced ERK phosphorylation is EPCR dependent.

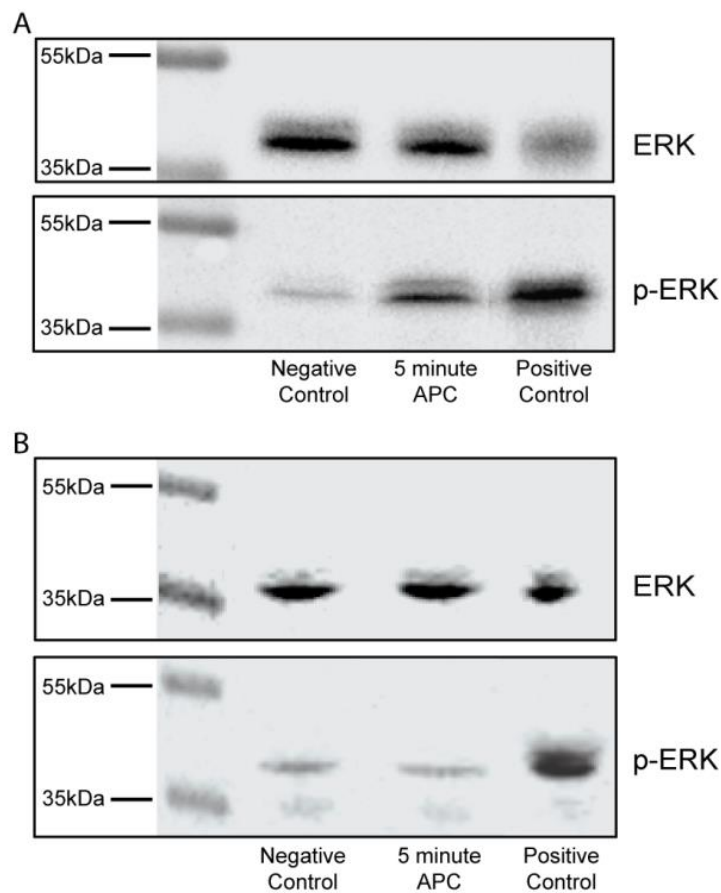


Figure 3.12 Western blotting investigating the effect of APC on ERK phosphorylation.

- Western blotting for pERK and Total-ERK of lysates from serum-starved HCT116 cells that had been treated with APC for 5 minutes. Negative controls were serum starved only. Positive controls were treated with 50% FCS for 10 minutes.
- Western blotting after pre-treatment with EPCR-blocking antibody. The protocol was otherwise identical to a), and was performed in parallel.

ERK phosphorylation assays and western blotting were performed in triplicate.

3.2.7 Detection of EPCR-related proteins in CRC

APC is produced through the activation of Protein C (PC), via EPCR [75]. To determine whether HCT116 cells produce endogenous PC, serum-starved HCT116 cells were lysed and western blotting for PC was performed (Figure 3.13). A faint band for PC was observed in the cell lysate, but no signal was detected in the supernatant or media. This suggests that these cells can produce PC endogenously, and therefore autocrine EPCR-mediated PC activation may be possible.

To determine if Protease-Activated Receptor 1 (PAR1) (which is crucial for EPCR-mediated signalling [75]) is present in primary CRC tissue, exploratory immunohistochemistry was performed on 10 CRC cases. This demonstrated PAR1 immunoreactivity in cancer cells (Figure 3.14) in all cases. The presence of PAR1 suggests that EPCR may be capable of initiating intracellular signalling in CRC.

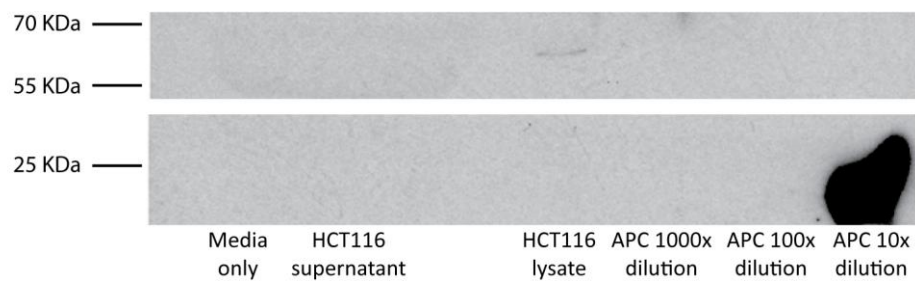


Figure 3.13 Detection of Protein C in HCT116 cell lysate. The anti-Protein C antibody detects both Protein C (62 KDa) and Activated Protein C (APC, 21 KDa). Exogenous APC was used as a positive control. A faint band was detected in the 60-65KDa range in the HCT116 lysate, suggestive of the presence of PC. Lysis of HCT116 cells and western blotting for APC were performed in triplicate. The faint band visible in the HCT116 lysate was present in all three experiments.

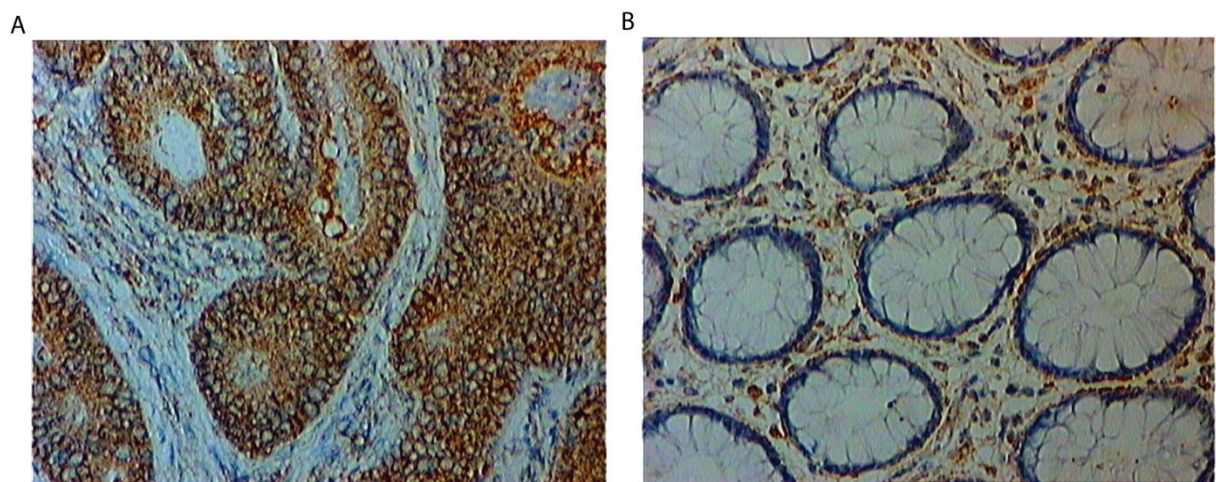


Figure 3.14 PAR1 immunohistochemistry in (A) CRC and (B) normal colon.

In addition, controls were performed with the primary antibody omitted (data not shown).

3.2.8 Microarray analysis

After establishing that APC can induce ERK phosphorylation in HCT116 cells, I wanted to determine the effect of APC treatment on gene transcription. Optimal APC concentrations and treatment durations were determined by assessing APC-induced expression changes of 4 genes (*NFkB2*, *PCNA*, *BCL2A1*, *EFNA1*, which were chosen based on previous findings in endothelium [69]) using qRT-PCR. A concentration of 180nM for 24 hours was associated with maximal alteration in gene expression. Therefore, serum-starved HCT116 cells were treated with 180nM APC for 24 hours. RNA was extracted from both APC-treated cells and control cells for two-colour Agilent microarray analysis. Differential gene expression was determined using the R Limma package. There was strong overlap between the three experimental replicates, with over 400 genes with differential gene expression across all three replicates. Table 3.4 shows genes with significant differential expression (Bayes factor > 5). Within the list of differentially expressed genes, a high proportion were ribosomal. Gene groupings were investigated by gene set enrichment analysis (GSEA) (Table 3.5, Figure 3.15). Several gene groupings were enriched, a large proportion of which were associated with gene transcription. Additionally, several gene sets associated with epidermal growth factor (EGF) signalling were enriched, consistent with the ERK phosphorylation data. Finally, a gene set containing genes upregulated by thrombin signalling in HUVEC cells was significantly enriched, consistent with thrombin and APC's common signalling pathways.

Gene ID	Gene name	LogFC (Fold Change)	P Value	Adjusted P Value	Bayes factor
RPL13A	ribosomal protein L13a	0.729831	1.32E-09	5.04E-05	11.4789
RPS2	ribosomal protein S2	0.75432	2.37E-09	5.04E-05	11.07495
RPS21	ribosomal protein S21	0.64925	7.57E-09	8.69E-05	10.22977
RPL38	ribosomal protein L38	0.640211	8.88E-09	8.69E-05	10.10872
RPS20	ribosomal protein S20	0.56367	1.23E-08	9.05E-05	9.857144
RPL12	ribosomal protein L12	0.589346	1.61E-08	0.000105	9.647333
RPL26	ribosomal protein L26	0.527112	2.52E-08	0.000119	9.293517
RPS27	ribosomal protein S27 (metallopanstimulin 1)	0.569845	2.63E-08	0.000119	9.257807
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	0.687543	3.33E-08	0.000125	9.067365
RPL23	ribosomal protein L23	0.691122	3.35E-08	0.000125	9.06284
RPL8	ribosomal protein L8	0.581833	3.41E-08	0.000125	9.047782
RPL35A	ribosomal protein L35a	0.614064	3.76E-08	0.00013	8.96871
RPL31	ribosomal protein L31	0.484977	7E-08	0.000188	8.453052
TACSTD2	tumor-associated calcium signal transducer 2	0.505105	8.88E-08	0.000226	8.251977
RPL37A	ribosomal protein L37a	0.457551	9.24E-08	0.000226	8.218734
UBC	ubiquitin C	0.505567	1.26E-07	0.000286	7.951056
RPL32	ribosomal protein L32	0.44282	1.27E-07	0.000286	7.948702
KITLG	KIT ligand	0.430458	1.36E-07	0.000289	7.885965
RPS18	ribosomal protein S18	0.430622	1.38E-07	0.000289	7.876527
DUT	dUTP pyrophosphatase	-0.4403	1.79E-07	0.000357	7.648652
GIPR	gastric inhibitory polypeptide receptor	0.665148	2.49E-07	0.000421	7.360759
RPS8	ribosomal protein S8	0.49689	2.51E-07	0.000421	7.35272
RPL7A	ribosomal protein L7a	0.431458	2.9E-07	0.00046	7.224366
RPS13	ribosomal protein S13	0.366457	4.07E-07	0.00052	6.922628
CLK1	CDC-like kinase 1	-0.38066	4.25E-07	0.000532	6.882765
RPL17	ribosomal protein L17	0.377933	4.57E-07	0.000555	6.818574
ING3	inhibitor of growth family, member 3	0.367639	5.26E-07	0.000596	6.692568
RPL3	ribosomal protein L3	0.446173	6.22E-07	0.000629	6.540546
RPS7	ribosomal protein S7	0.413569	8.01E-07	0.000677	6.309131
SLC1A3	solute carrier family 1 (glial high affinity glutamate transporter), member 3	0.36994	9.43E-07	0.000693	6.159936
LCN2	lipocalin 2 (oncogene 24p3)	0.483695	9.43E-07	0.000693	6.159605
MT2A	metallothionein 2A	0.448027	9.86E-07	0.000695	6.118647
RPL10A	ribosomal protein L10a	0.373973	9.91E-07	0.000695	6.113851
ANXA2	annexin A2	0.352575	1.02E-06	0.000695	6.089494
C10orf10	chromosome 10 open reading frame 10	0.396982	1.07E-06	0.000716	6.040645

HNRPH2	heterogeneous nuclear ribonucleoprotein H2 (H')	0.348645	1.17E-06	0.000757	5.958679
ARHGEF15	Rho guanine nucleotide exchange factor (GEF) 15	0.451955	1.24E-06	0.000776	5.910181
RPL36A	ribosomal protein L36a	0.459017	1.27E-06	0.000784	5.886451
RPS16	ribosomal protein S16	0.587827	1.33E-06	0.000806	5.841946
RPS3A	ribosomal protein S3A	0.38013	1.42E-06	0.000822	5.782664
RPL34	ribosomal protein L34	0.401456	1.5E-06	0.000845	5.732861
ATF6	activating transcription factor 6	0.330841	1.52E-06	0.000846	5.71839
FAU	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (fox derived); ribosomal protein S30	0.402734	1.54E-06	0.000846	5.70528
BRP44L	brain protein 44-like	0.389064	1.68E-06	0.000881	5.624935
RPL13	ribosomal protein L13	0.526131	1.85E-06	0.000946	5.534252
MIF	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	0.379488	1.85E-06	0.000946	5.533545
EIF4A3	eukaryotic translation initiation factor 4A, isoform 3	-0.33887	1.92E-06	0.000947	5.501359
DHTKD1	dehydrogenase E1 and transketolase domain containing 1	0.364546	1.94E-06	0.000951	5.489604
FAM49B	family with sequence similarity 49, member B	0.396678	1.97E-06	0.000955	5.477903
RPS26	ribosomal protein S26	0.454294	2.2E-06	0.001032	5.372796
RPL30	ribosomal protein L30	0.362797	2.21E-06	0.001032	5.367476
HRASLS5	HRAS-like suppressor family, member 5	0.456289	2.25E-06	0.001032	5.351502
TAPBP	TAP binding protein (tapasin)	0.312637	2.26E-06	0.001032	5.349157
CNOT7	CCR4-NOT transcription complex, subunit 7	0.352376	2.27E-06	0.001032	5.344529
RASD1	RAS, dexamethasone-induced 1	0.525178	2.31E-06	0.00104	5.326405
FKSG30	actin-like protein	0.315158	2.43E-06	0.001081	5.279509
SSBP4	single stranded DNA binding protein 4	-0.36137	2.67E-06	0.001143	5.192159
SPATA22	spermatogenesis associated 22	0.446524	2.86E-06	0.001206	5.125614
RPLP0	ribosomal protein, large, P0	0.376217	2.87E-06	0.001206	5.121352
DTL	denticless homolog (Drosophila)	-0.32126	2.91E-06	0.001212	5.10953
RPL10	ribosomal protein L10	0.320072	2.97E-06	0.001219	5.091869
RPS19	ribosomal protein S19	0.349232	2.97E-06	0.001219	5.090609

GJB3	gap junction protein, beta 3, 31kDa (connexin 31)	0.307522	3E-06	0.001225	5.08002
RPS12	ribosomal protein S12	0.319493	3.07E-06	0.001242	5.059895
HMBS	hydroxymethylbilane synthase	-0.31278	3.12E-06	0.001242	5.044439
DUSP1	dual specificity phosphatase 1	0.395831	3.12E-06	0.001242	5.043594
ATAD4	ATPase family, AAA domain containing 4	0.375962	3.15E-06	0.001242	5.034235
TES	testis derived transcript (3 LIM domains)	0.428908	3.27E-06	0.001244	5.000888

Table 3.4 Genes differentially expressed after APC treatment of HCT116 cells (Bayes Factor >5). Log FC shows fold change (base 2) in APC-treated cells versus controls. A logFC of 0 therefore represents equal expression in the APC-treated and control cells, >0 represents higher expression in the APC-treated cells, and <0 represents lower expression in the APC-treated cells.

NAME OF GENE SET	SIZE (Genes)	Enrichment score	FDR q- value	FWER p- value	RANK AT MAX
REACTOME_PEPTIDE_CHAIN_ELONGATION	83	0.830811	0	0	958
REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION	97	0.791672	0	0	958
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	106	0.767984	0	0	958
KEGG_RIBOSOME	85	0.838541	0	0	958
REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_REGULATION	91	0.791225	0	0	958
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX	102	0.748643	0	0	958
REACTOME_TRANSLATION	131	0.695274	0	0	958
REACTOME_INFLUENZA_LIFE_CYCLE	130	0.687707	0	0	958
BILANGES_SERUM_AND_RAPAMYCIN_SENSITIVE_GENES	61	0.72625	0	0	958
REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUENTLY_THE_43S_COMPLEX	35	0.815359	0	0	958
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S	43	0.737235	0	0	958
CHNG_MULTIPLE_MYELOMA_HYPERPLOID_UP	44	0.714276	0	0	958
REACTOME_METABOLISM_OF_MRNA	202	0.517838	0	0	958
FLOTHO_PEDIATRIC_ALL_THERAPY_RESPONSE_UP	50	0.642381	0	0	1173
NAGASHIMA_EGF_SIGNALING_UP	52	0.629965	0	0	2671
PECE_MAMMARY_STEM_CELL_DN	128	0.529283	0	0	1489
REACTOME_METABOLISM_OF_RNA	245	0.453754	0	0	958
NAGASHIMA_NRG1_SIGNALING_UP	160	0.485804	7.41E-05	0.001	2099
CHASSOT_SKIN_WOUND	10	0.878359	2.08E-04	0.003	1642
JECHLINGER_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_DN	63	0.526858	3.91E-04	0.006	886
HOLLEMAN_ASPARAGINASE_RESISTANCE_B_ALL_UP	22	0.674004	3.73E-04	0.006	551
REACTOME_OLFACTORY_SIGNALING_PATHWAY	102	0.474621	5.87E-04	0.01	3844
REACTOME_METABOLISM_OF_PROTEINS	382	0.395476	8.40E-04	0.015	2516
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	34	0.597898	8.05E-04	0.015	4451
TIEN_INTESTINE_PROBIOTICS_6HR_UP	50	0.533304	0.00148	0.028	1092
HSIAO_HOUSEKEEPING_GENES	363	0.389684	0.00247	0.049	1536
LEE_LIVER_CANCER_HEPATOBLAST	15	0.701492	0.00248	0.051	3202
AMIT_EGF_RESPONSE_40_HELA	39	0.545263	0.00395	0.083	2970
YAMASHITA_LIVER_CANCER_WITH_EPCAM_UP	46	0.535978	0.00390	0.085	1692
BILANGES_SERUM_RESPONSE_TRANSLATION	33	0.579625	0.00436	0.099	1489

Table 3.5 Top 30 gene sets in Gene Set Enrichment Analysis of APC-treated HCT116 cells versus control HCT116 cells.

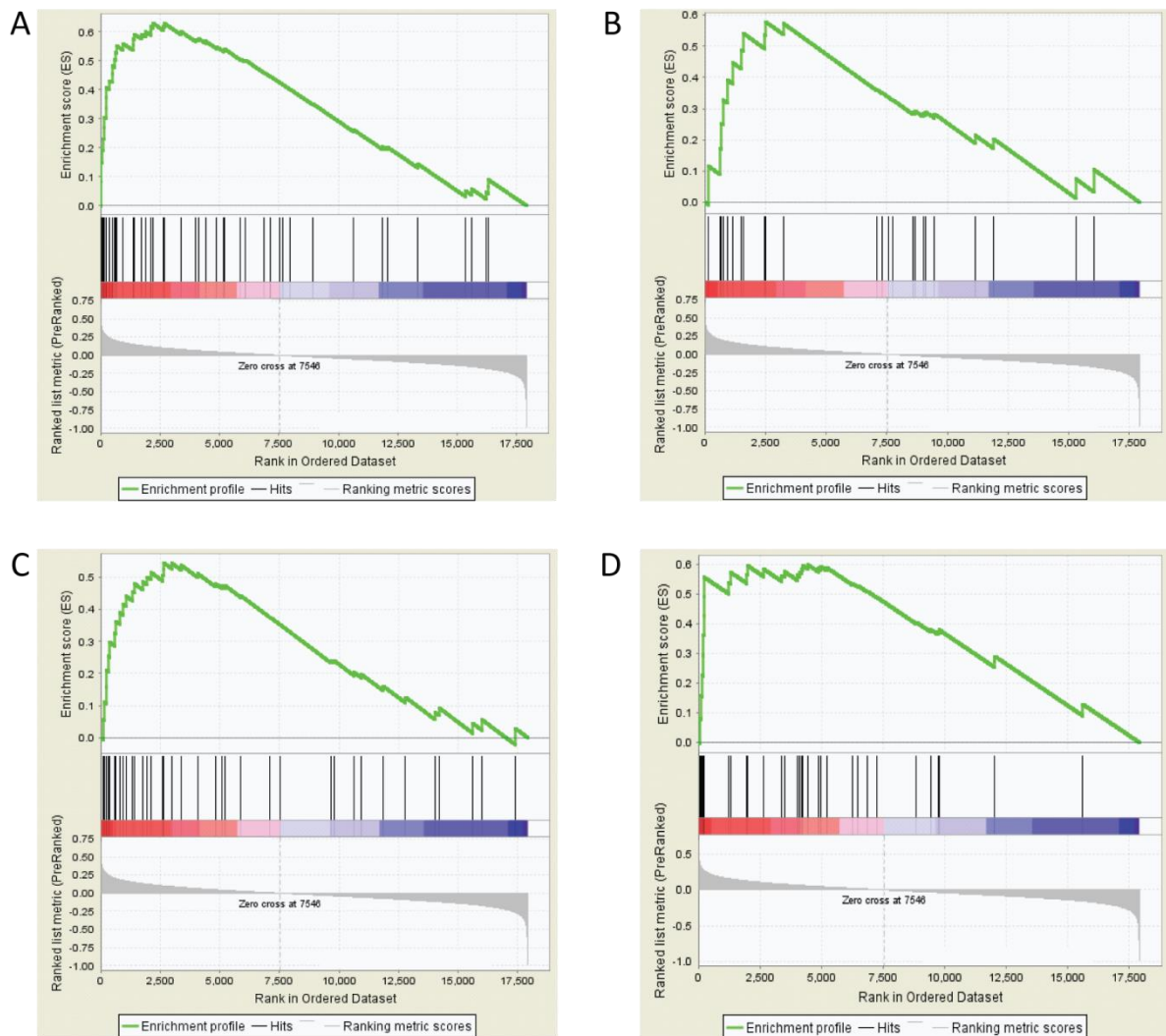


Figure 3.15 Gene set enrichment analyses. Enrichment plots for gene sets:

- a) Nagashima EGF signalling up
- b) Zwang EGF persistently up
- c) Amit EGF response 40 Hela
- d) Uzonyi response to leukotriene and thrombin

These enrichment plots rank all genes (bottom section) in order of most overexpression after APC treatment to most underexpression. 0 represents equal expression after treatment. The black vertical bars (middle section) represent genes from the relevant gene set. The green line (top section) represents the running enrichment score for the gene set as the analysis walks down the ranked list, increasing the running-sum statistic when a gene is in a gene set and decreasing it when it is not. As the peaks of these green lines occur towards the left (overexpression) side of the gene list, these gene sets are highly enriched.

3.2.9 EPCR shRNA knockdown assays

Having established that EPCR signalling can induce ERK phosphorylation and alter gene transcription, I wanted to determine whether the functional effect of EPCR provided any benefit to tumour cells. Previous work had demonstrated that EPCR is a marker of chemoresistant cell lines, including HCT116 [59]. To further examine the effect of EPCR expression on chemosensitivity, EPCR expression was knocked down in HCT116 cells using shRNA (Figure 3.16). Two shRNA constructs were chosen that induced high (clone 969 – 90-95%) and medium (clone 379 – 60-70%) levels of knockdown. EPCR knockdown significantly reduced the effectiveness of 5FU and epirubicin in MTT and BrdU assays (Figure 3.17 A,B).

Transwell assays were performed to determine how EPCR expression affected cellular migration. Over 48 hours, HCT116 cells with high EPCR knockdown had a significantly higher rate of migration than control cells ($p < 0.001$) or medium-level EPCR knockdown cells ($p < 0.001$) (Figure 3.17C). However, no significant differences between the groups were observed in QCM invasion assays (Figure 3.17D).

To determine whether the effect of EPCR perturbation was similar in a MSS/CIN cell line, the experiments were repeated in HT29 cells. In HT29 cells, EPCR knockdown did not have any consistent effect on MTT and BrdU chemosensitivity assays (Figure 3.17 A,B). In comparison to the HCT116 cells, the effect of the chemotherapy drugs was significantly smaller in both the control and knockdown cells. However, as with HCT116 cells, high EPCR knockdown ($>95\%$) correlated with an increase in cellular migration in the Transwell assay ($p < 0.01$) (Figure 3.17C). No difference was observed with knockdown in QCM invasion assays (Figure 3.17D). Both total cell

migration and invasion were less in the HT29 cells compared to the HCT116 cells. HT29 EPCR overexpression (to over 600% of wild type HT29) did not consistently affect chemosensitivity, migration or invasion. These data suggest that the effect of EPCR on chemosensitivity may differ in different cell lines, though the effects on migration were similar.

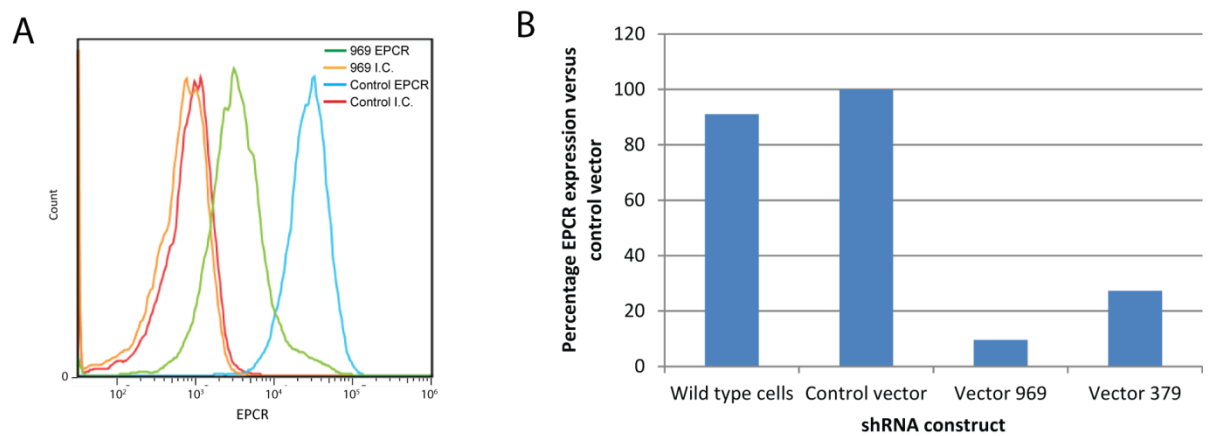


Figure 3.16 shRNA EPCR knockdown in HCT116 cells.

- Flow cytometry data confirming knockdown of EPCR expression with shRNA (vector 969) compared to control cells.
- EPCR expression in wild type cells and shRNA transfected cells (vectors 969 and 379), as a percentage of control vector transfected cells.

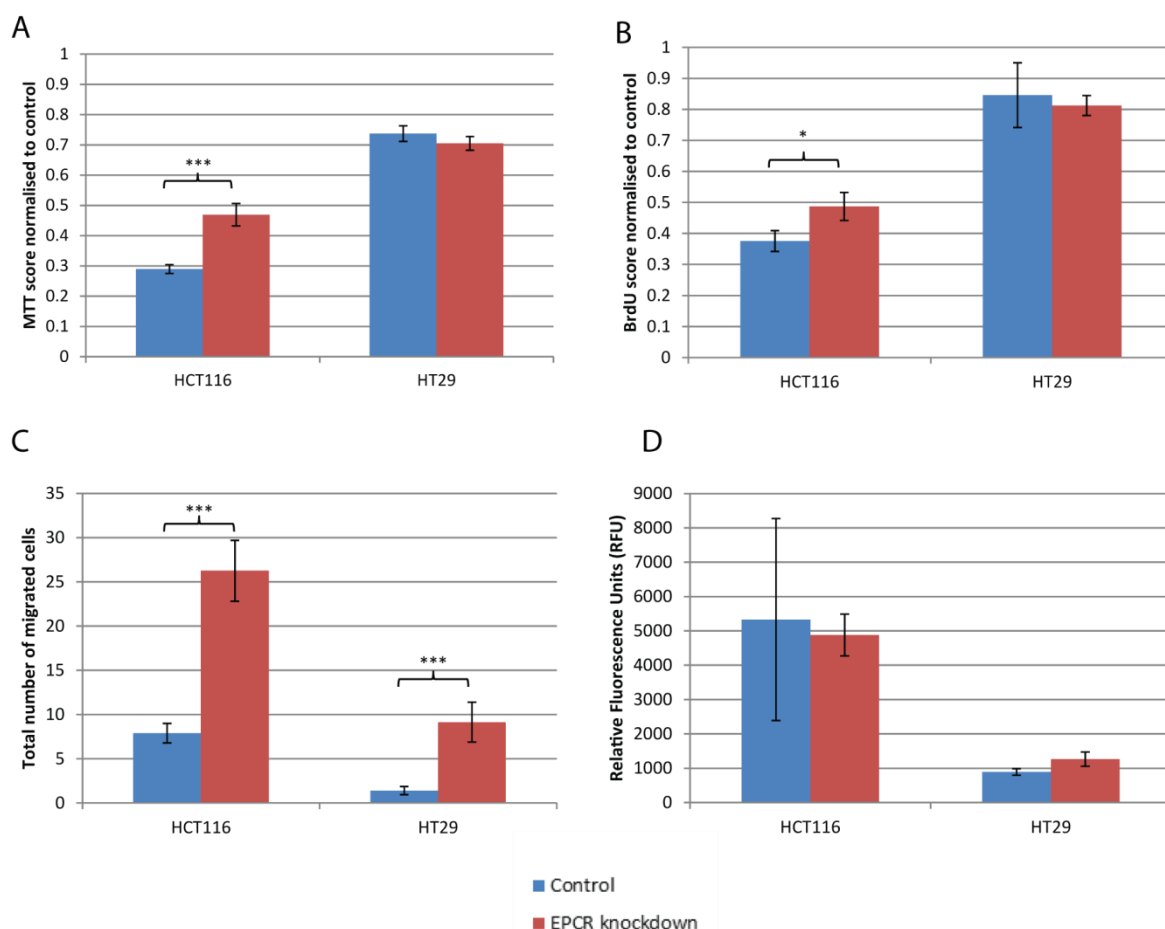


Figure 3.17 Functional assays after shRNA EPCR knockdown in HCT116 cells.

- MTT scores for control and shRNA knockdown (clone 969) HCT116 and HT29 cells after treatment with 5FU (32 μ M), shown as a percentage of control cells (cells untreated with chemotherapy).
- BrdU scores for control and shRNA knockdown (clone 969) HCT116 and HT29 cells after treatment with 5FU (32 μ M), shown as a percentage of control cells (cells untreated with chemotherapy).
- 48 hour Transwell migration assay – Number of cells that migrated through Transwell membrane for each clone in five assessed regions.
- 48 hour QCM invasion assay.

Asterixes represent statistical significance of EPCR-low cells versus control cells (*= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$). Bars represent standard error.

3.2.10 Serum sEPCR correlation with CEA and clinical stage

Serum soluble EPCR (sEPCR) has previously been shown to correlate with the ovarian tumour marker CA-125 in ovarian cancer [86]. To investigate whether sEPCR may increase with clinical stage or correlate with the colorectal tumour marker CEA in CRC patients, exploratory experiments were performed on serum retrieved from 21 patients with early stage colorectal cancer (Duke Stage A-B) and 19 patients with late stage disease (Duke Stage C-D). ELISAs for sEPCR and CEA were performed. Neither sEPCR nor CEA were significantly higher in patients with lymph node or distant metastases versus patients with localised disease ($p=0.371$ and 0.593 respectively) (Figure 3.18). There was a trend towards a correlation between sEPCR and CEA, but this did not reach significance (Spearman $Rho=0.271$, $p=0.091$). These studies were exploratory and underpowered, and further experiments with a sufficient numbers of samples will be required for confirmation of any association with stage or correlation between EPCR and CEA.

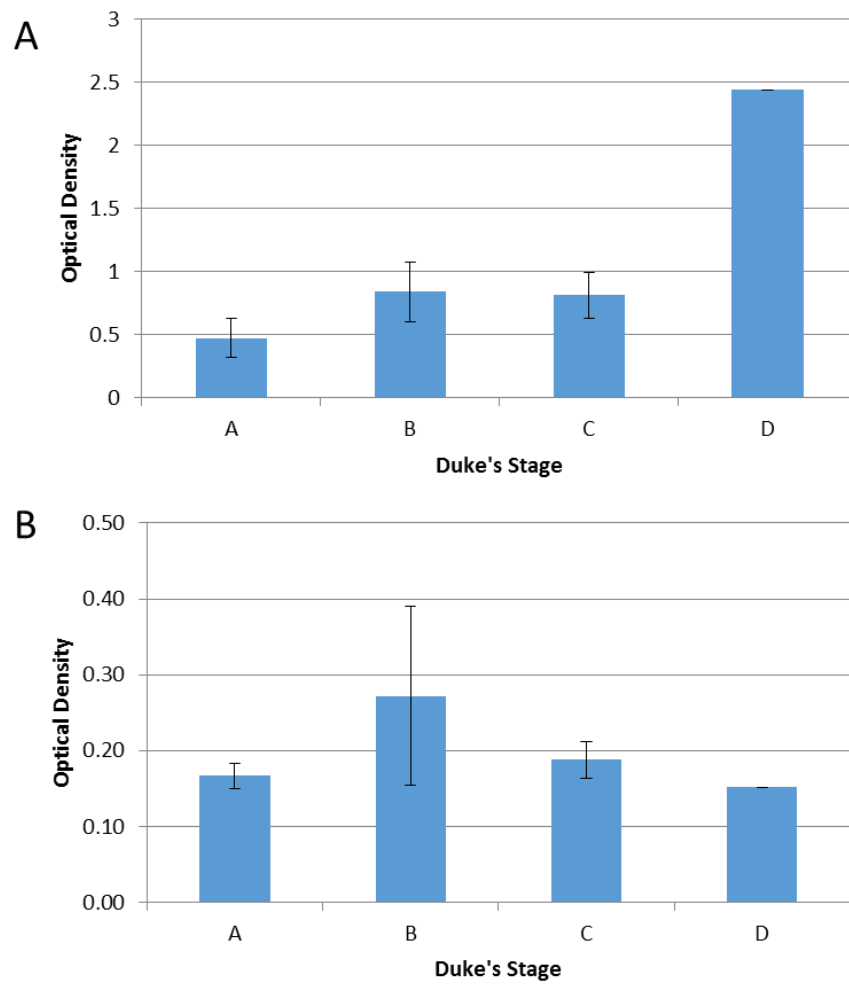


Figure 3.18 ELISA for (A) sEPCR and (B) CEA on serum from CRC patients.

3.2.11 EPCR expression and chemotherapy or cetuximab responsiveness

Previous studies have indicated that EPCR is a marker for chemoresistant cell lines [59], and conversely its expression may predict chemotherapy responses in early stage lung cancer [78]. Furthermore, I have shown that EPCR perturbation can marginally affect CRC cell line chemosensitivity. Therefore I wanted to determine whether EPCR could affect chemosensitivity in CRC patients. In addition, having established that EPCR can mediate APC-dependent ERK phosphorylation on CRC cells, I was also interested in the potential impact of EPCR upregulation on clinical responses to EGFR monoclonal antibodies (mAbs) in CRC patients. Patients with *RAS* mutation do not respond to EGFR mAbs due to 'bypass signalling' resulting from constitutively active MEK/ERK signalling [10]. As EPCR-mediated signalling induces ERK phosphorylation, I hypothesised that EPCR could also act as a bypass signalling pathway in an analogous manner.

To determine whether EPCR expression is associated with chemotherapy or EGFR monoclonal antibody (cetuximab) responsiveness in CRC, 153 CRC tumour samples from the MRC COIN study [10] were stained for EPCR using immunohistochemistry. Of these, 71 were from the chemotherapy arm, and 82 were from the chemotherapy plus cetuximab arm. These slides were analysed using Definiens Tissue Studio software. Percental scores for EPCR were calculated for each slide. Survival analyses were performed using the Cox proportional hazards model, adjusting for factors that significantly associated with progression-free survival (PFS) to $p < 0.05$ (*KRAS* mutation status and MSI status). No significant difference in PFS was observed between EPCR high and low cases (divided by median EPCR expression) across all patients (Figure 3.19) or when both treatment arms were analysed

individually. No difference in PFS was observed when the top and bottom 20% of EPCR expressors were compared. Finally, EPCR did not predict for survival in either treatment arm after exclusion of *RAS* mutant cases. These data suggest that EPCR expression does not predict for chemotherapy or cetuximab responsiveness in CRC.

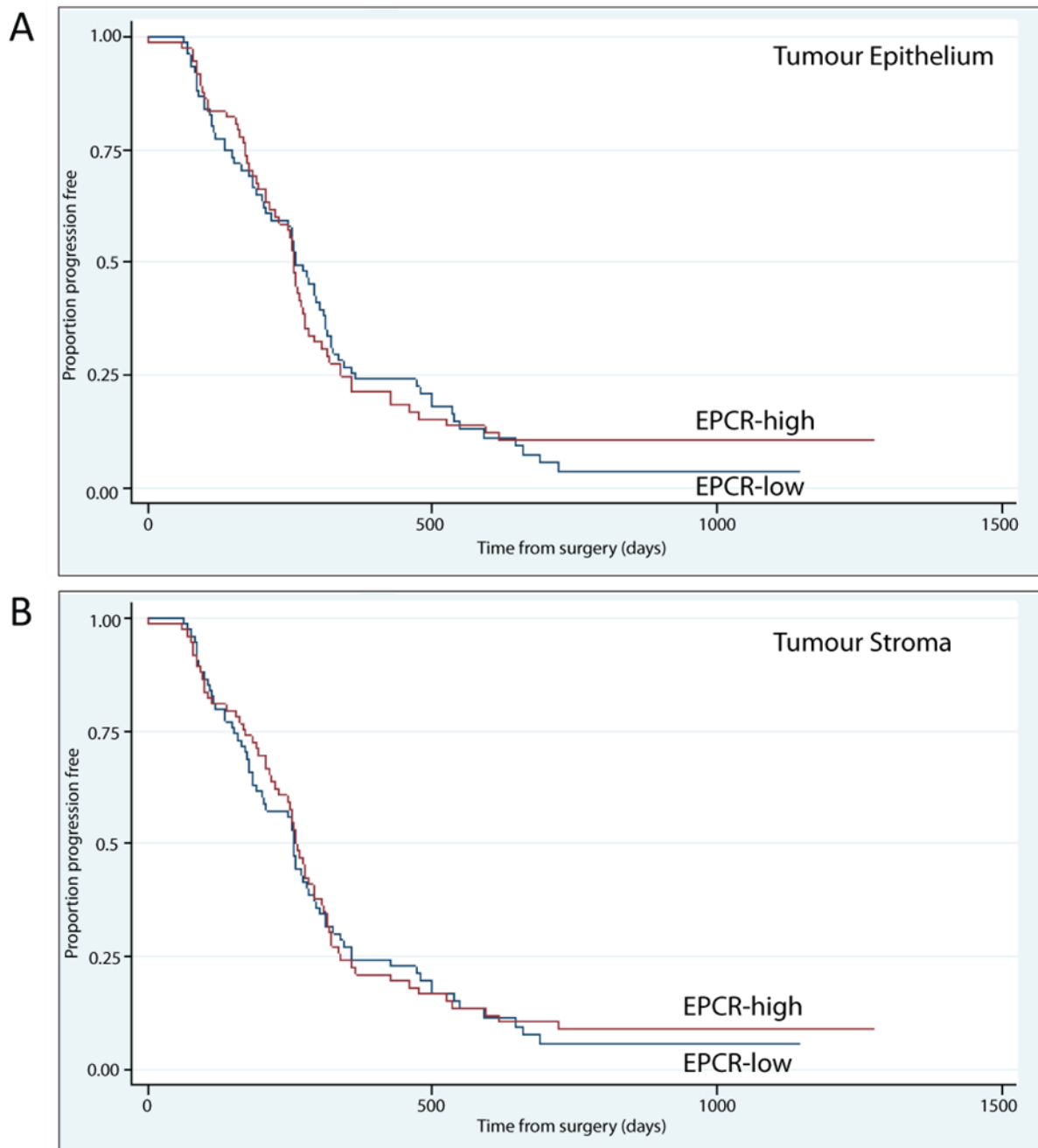


Figure 3.19 Kaplan-Meier curves for progression free survival in EPCR high and low patients. Patients have been separated by median EPCR expression in:

- a) Tumour epithelium
- b) Tumour stroma

The blue lines represent low EPCR expression, and the red lines represent high EPCR expression. This cohort contains 153 patients from the MRC COIN trial.

3.3 Discussion

EPCR, a receptor with anti-apoptotic and proliferative effects [69, 73, 74], is known to be expressed on various cancer cell lines [59, 72]. I have demonstrated, using robust digital immunohistochemistry algorithms and bioinformatic analyses, that EPCR is aberrantly expressed in colorectal cancer tissue, with expression being higher in cancer regions compared to normal regions in all cases tested.

Though EPCR expression has been investigated in various epithelial cancers, the mechanism of its upregulation has previously been unclear, and EPCR upregulation has not previously been linked to chromosome 20q amplification and promotor hypomethylation. 20q amplification occurs as a result of chromosomal instability (due to the loss of the CIN suppressor genes of Chr18q [121]) and occurs frequently in malignant transformation. The phenomenon occurs in multiple cancer types [116, 117, 122-124], and there is *in vitro* evidence that it occurs early in the natural history of transformation, promoting cancer initiation independently of other chromosomal abnormalities [116]. 20q amplification has been correlated with poor prognosis, aggressive tumour phenotype, progression, and metastasis formation, and Tabach et al. suggest that 20q amplification plays a causative rather than a bystander role in the process of tumourigenesis [116]. In colorectal cancer, 20q amplification has been associated with the progression of colonic adenomas into malignant carcinomas [117, 122, 125], and occurs in over 65% of CRC cases. Carvalho et al. confirmed that 20q amplification occurs in under 20% of non-progressed adenomas, but in over 60% of progressed adenomas [117]. This is consistent with my *in vitro*

findings, which demonstrated poor expression of EPCR in the adenoma cell line AA/C1. However, the AA/C1 10C derivative, a malignant variant of the AA/C1 cell line, had higher EPCR expression [126]. EPCR expression correlates strongly with 6/7 of the Carvalho et al. putative oncogenes and 13/13 of the Tabach et al. 'cancer initiating genes', all located on chromosome 20q [116, 117]. It is therefore possible that the characteristics of EPCR-expressing tumours may be a result of both EPCR and of co-expression of local genes on 20q. Previous studies have confirmed EPCR expression in breast cancer, ovarian cancer and lung cancer, but these studies have not related EPCR expression to chromosomal amplification or hypomethylation [78, 80, 86]. Significantly, all of these tumour types are associated with 20q amplification. I have shown that EPCR expression should be taken in the context of this phenomenon, as 20q amplification is itself associated with an aggressive and invasive phenotype [72, 78, 116, 117].

Though amplification of genes on 20q was uniform, promotor methylation was less co-ordinated, suggesting that this important epigenetic control mechanism may be one factor which accounts for the relatively greater variability in gene expression compared to copy number across the chromosomal region. Other epigenetic control mechanisms may explain the minority of cases where copy number and methylation status do not predict EPCR expression.

I then investigated the implications of aberrant EPCR expression on the functional phenotype of CRC cells, following studies highlighting EPCR's effects on cancer cell

migration, invasion and proliferation in different tumour settings [72, 78, 79, 83]. While the results establish that EPCR, via APC binding, can stimulate ERK signalling and elicit changes in gene transcription (as is the case in endothelial cells), the *in vitro* functional assays failed to identify compelling evidence that overexpression of EPCR *per se* might provide functional advantages for CRC cells. Interestingly, in both HCT116 and HT29 cells, EPCR expression appeared to decrease migration in *in vitro* assays. Furthermore, EPCR had variable effects on cancer cell phenotype across cell lines: specifically, it increased chemosensitivity of HCT116, consistent with a report that higher EPCR expression was associated with superior chemotherapy response in early stage lung cancer [78]. In contrast, no chemosensitivity effects were observed in HT29 cells, and no effects on invasion were observed in either cell line. However, an important limitation of these experiments and of approaches highlighting both beneficial and detrimental effects of EPCR perturbation in animal models [78, 79, 83, 114] is that they primarily address EPCR-intrinsic effects on cancer development, and typically focus on a small number of cell lines. My findings underline that the effects of EPCR upregulation on cancer cells may be heavily dependent on biological context, including the exact cancer cell line and potentially upregulation of functionally important neighbouring genes co-amplified on chromosome 20q. This demonstrates the need for models that represent the variety of biological variants observed in the clinical scenario.

Increased migration rates were observed in HCT116 and HT29 shRNA EPCR knockdown cells. As a potential inhibitor of migration, EPCR may impact on metastasis development. The migration data are contrary to the data observed in

breast cancer cells by Beaulieu and collaborators [72], who demonstrated that APC increased chemotaxis in a dose-dependent manner. However, rather than altering Activated Protein C (APC) concentrations as in Beaulieu's study, I perturbed EPCR directly. In addition, in lung cancer, EPCR knockdown decreased the development of bone metastases in a mouse xenograft model [78]. However, in mesothelioma, EPCR expression has been observed to suppress tumourigenicity and growth [83], which is consistent with my findings. As with chemosensitivity, it is possible that EPCR's effect on migration is variable in different tumour types, depending on the specific context of its expression.

Treatment of HCT116 with EPCR's natural ligand APC induced ERK phosphorylation, which was inhibited by EPCR blockade. Consistent with this, microarray analysis of APC-treated cells showed an increase in ribosomal RNA expression, which is a known effect of ERK phosphorylation [127]. An increase in cellular proliferation due to APC may explain why EPCR knockdown decreases chemosensitivity in these cells, as tumour cells that divide more rapidly are more sensitive to chemotherapy [4]. Multiple gene sets associated with EGF signalling were enriched, consistent with the suggestion that APC induces ERK phosphorylation through EPCR-mediated signalling. Additionally, a gene set associated with thrombin signalling was enriched. Thrombin in complex with thrombomodulin, in the endothelial setting, presents PC to EPCR, activating PC [75]. This enrichment therefore suggests that the effects of APC signalling on gene transcription in the endothelial and tumour cell settings overlap considerably.

Analysis of samples from the MRC COIN study did not reveal any association between EPCR expression and progression free survival in advanced CRC within the chemotherapy or chemotherapy plus cetuximab arms. There is a need for the identification of predictive biomarkers for EGFR monoclonal antibody therapy in metastatic CRC, as thus far, no predictive biomarkers have been identified in *RAS* wildtype patients. As EPCR mediates ERK phosphorylation, it could potentially have acted as an additional bypass pathway during EGFR inhibition with cetuximab, similar to *RAS* mutation. However, survival analysis of EPCR stained specimens did not show any association with PFS. Less than half of *RAS* wild type patients respond to EGFR monoclonal antibody therapy, and the response rate is only approximately 15% higher than with chemotherapy treatment alone [128]. Therefore there remains a considerable need to identify predictive biomarkers that will enable improved targeting of this therapy and to prevent toxicity in patients who will not derive benefit. Additionally, as EPCR is a marker for chemoresistant cell lines [59], and has been linked to chemotherapy responses in early stage lung cancer [78], I wanted to determine whether there was any association with PFS in CRC. No significant relationship with PFS was found, suggesting that the impact of the EPCR-high phenotype is insufficient to impact on chemosensitivity.

Previous work in our laboratory has demonstrated that EPCR is a ligand for a specific V δ 2-negative clone of the $\gamma\delta$ T-cell receptor [53]. EPCR is constitutively expressed on endothelium. In this context $\gamma\delta$ T-cells detect CMV via a multimolecular stress

signature. In epithelium however, where EPCR is not normally expressed, EPCR could act as *bona fide* marker of cellular stress as it is likely to be consistently upregulated in multiple transformed tissues. Therefore EPCR could represent a molecular exemplar of how $\gamma\delta$ T-cells can recognise stress in different settings. Further work is required to determine whether EPCR expression in epithelial tumours is relevant to $\gamma\delta$ T-cell responses.

EPCR may not be an attractive therapeutic target in CRC due to physiological expression throughout the vasculature and its variable effect on cancer cells. The *in vitro* and microarray data, taken together with the bioinformatics analysis and survival data, suggest that EPCR does not have a substantial impact on tumour phenotype. The analyses therefore raise the possibility that the EPCR-high phenotype may be due to regionally co-expressed genes, rather than EPCR itself. Work by Tsafir et al. has shown that 20q amplification is absent in normal colonic mucosa, occurs at low levels in adenoma, becomes more frequent as disease advances, and is found in the majority of metastatic samples [125]. As the group note, such wide scale chromosomal changes are likely to lead to the expression of genes which in themselves do not confer a selective advantage. EPCR may be such a bystander gene. However, EPCR may have value as a surrogate marker of 20q amplification. Furthermore, there have been suggestions that EPCR could predict for thrombotic risk [86]. Both of these aspects require further exploration in the CRC setting.

Chapter 4

An immunogenomic stratification of colorectal cancer

4.1 Introduction

In addition to investigating the role of EPCR, a ligand for unconventional T cells, I wanted to explore how conventional adaptive immunity could be used to stratify CRC patients. Patient stratification involves the grouping of patients for specific interventions or treatments. Bowel cancer treatment is currently stratified mainly by clinical stage and tumour site [11] as discussed in the thesis introduction. However, molecular stratification is becoming increasingly significant in clinical management. A key example of this, as highlighted in the previous chapter, is the mutation status of *RAS*, which is crucial in the prediction of EGFR mAb treatment responses [10]. Furthermore, stratification based on the mutation status of multiple genes is being investigated in multiple tumour types. In a key CRC stratified medicine clinical trial (FOCUS4 [129]), patients are stratified and recruited into specific treatment arms on the basis of their *BRAF*, *KRAS*, *NRAS*, *PIK3CA* and *PTEN* mutation status. In terms of immunological stratification, a tumour's microsatellite status affects patient survival, potentially due to differences in immune infiltration [5].

The extent of immune infiltration is now emerging as another key stratifier. The density of tumour infiltrating lymphocytes (TILs) and the expression of certain immune-related genes are of prognostic and predictive value in colorectal cancer (CRC) [37, 38, 42, 100, 130]. A worldwide task force is attempting to validate the 'immunoscore', an immunohistochemistry based scoring system centred on CD3 and CD8 density conceived by Galon and collaborators [38, 42]. TNM staging, when combined with immunoscore (TNM-I), is more prognostically accurate than traditional TNM staging alone [37, 130]. Whilst the simplicity of this scoring system will likely be

beneficial for clinical implementation, it is limited in the biological insights it can provide into the immune microenvironment in CRC. For instance, it does not incorporate key molecules, such as CD4, cytokines, class II molecules and inhibitory molecules, which are likely to play important roles in the tumour microenvironment. Furthermore, the factors that determine a patient's immune microenvironment are still unclear. These are likely to include tumour, host and environmental factors [44]. Within tumour and host factors, few systematic analyses have investigated the somatic and germline molecular drivers of immune infiltration. Although microsatellite unstable (MSI-H) cancers, which represent approximately 15% of colorectal cancers and have a superior prognosis [46], are known to be associated with increased TIL density [5, 45-50], the nature of the immune infiltration and the molecular drivers of the immune phenotype in microsatellite stable (MSS) CRC are poorly understood. In particular it is unclear whether defined molecular subsets (*RAS* mutant, *BRAF* mutant, *PIK3CA* mutant, quadruple wildtype (*BRAF*, *PIK3CA*, *NRAS*, *KRAS* all wildtype)) are associated with high or low immune infiltration. In addition, in both MSI-H and MSS cancers the extent to which therapeutically tractable inhibitory immune checkpoint receptors are represented are unclear and is of substantial interest, particularly considering recent checkpoint blockade failures in colorectal cancer [13, 14]. Given the prognostic and predictive relevance of CRC immunophenotype, a clearer understanding of the link between immunophenotype with tumour genotype is crucial.

To explore the determinants of immune phenotype I carried out a bioinformatic analysis of CRC data in The Cancer Genome Project (TCGA). These analyses defined a co-regulated cluster of immune related genes with a distinct Th1 bias,

expression of which defines four patient subgroups. Subsequently, the determinants of the immune response in colorectal cancer were explored. The findings have implications for our understanding of the immunobiology of colorectal carcinogenesis, and provide a potential framework for the development of stratified immunotherapy approaches for CRC. Much of this work has been published [131].

4.2 Results

4.2.1 Co-ordinate expression of immune response-related genes in colorectal cancer

I interrogated expression of immune response-related genes in the colorectal cancer (CRC) dataset in The Cancer Genome Atlas (TCGA). This dataset includes a total of 195 patients with complete transcriptomic, mutation and clinical data. Preliminary analyses focussed on an larger initial gene group (Table 4.1) that was based partly on previous studies on bio-molecular networks incorporating immune genes linked with disease-free survival in CRC [37, 42, 100]. These included those associated with Th1 subset function (*STAT1*, *IRF*, *IFNG*, *TBX21*, *IL18RAP*, *ICOS*, *GNLY*), certain chemokines (*CX3CL1*, *CXCL9*, *CXCL10*), adhesion molecules (*ICAM* and *MADCAM*) and an array of class II genes. I also included a number of immune checkpoint genes (PD-1, PD-L1, PD-L2, LAG3, TIM3, CTLA-4), two of which (PD-L1 and PD-L2) were previously associated with outcome [42]. Finally, the initial gene list was supplemented with class I genes, additional class II genes and genes involved in T cell activation. I also included NKG2D ligands (including ULBPs) and the $\gamma\delta$ T-cell ligand EPCR (*PROCR*).

Gene IDs	
ACTB	HLA-DQA2
CCL11	HLA-DRA
CCL2	HLA-DRB5
CCL5	ICAM1
CD247	ICOS
CD274	IFNG
CD276	IL12RB2
CD3D	IL17A
CD3E	IL18RAP
CD3G	IL7R
CD4	IRF1
CD80	KLRK1
CD86	LAG3
CD8B	MADCAM1
CTLA4	MICB
CX3CL1	PDCD1
CXCL10	PDCD1LG2
CXCL9	PROCR
GNLY	RAET1E
GZMB	RAET1G
HAVCR2	STAT1
HLA-A	STAT3
HLA-B	TBX21
HLA-C	TNFRSF14
HLA-DMA	TNFSF4
HLA-DMB	ULBP1
HLA-DOA	ULBP2
HLA-DOB	ULBP3
HLA-DPA1	VCAM1
HLA-DPB1	VTCN1
HLA-DQA1	

Table 4.1 Initial gene group for analysis.

Unsupervised two-dimensional hierarchical clustering was performed to assess the extent to which gene expression was co-ordinate or independent across the patient cohort. Visual analysis of the clustering highlighted a 28-gene subset (Table 4.2) which formed a clear gene grouping, expression of which was co-ordinately regulated (Figure 4.1). Gene-tree analysis of the dendrogram confirmed the validity of this grouping, identifying a subset of 24 highly co-ordinated genes (distance threshold 0.46) predominantly associated with Th1 immunity, including numerous class II MHC loci, and inhibitory molecules targeted in checkpoint blockade strategies. This 24-gene block was identical to my 28-gene cluster other than the absence of three additional class II MHC loci (*HLA-DRB5*, *HLA-DQA2* and *HLA-DQA1*) and one additional inhibitory molecule (*HAVCR2* (TIM3)), all of which correlated closely with the 24-gene block and were positioned directly adjacent on the gene cluster dendrogram. I therefore proceeded with the 28-gene cluster as it formed a clearly co-ordinated block on visual and correlation analysis. This grouping was termed the co-ordinate immune response cluster (CIRC).

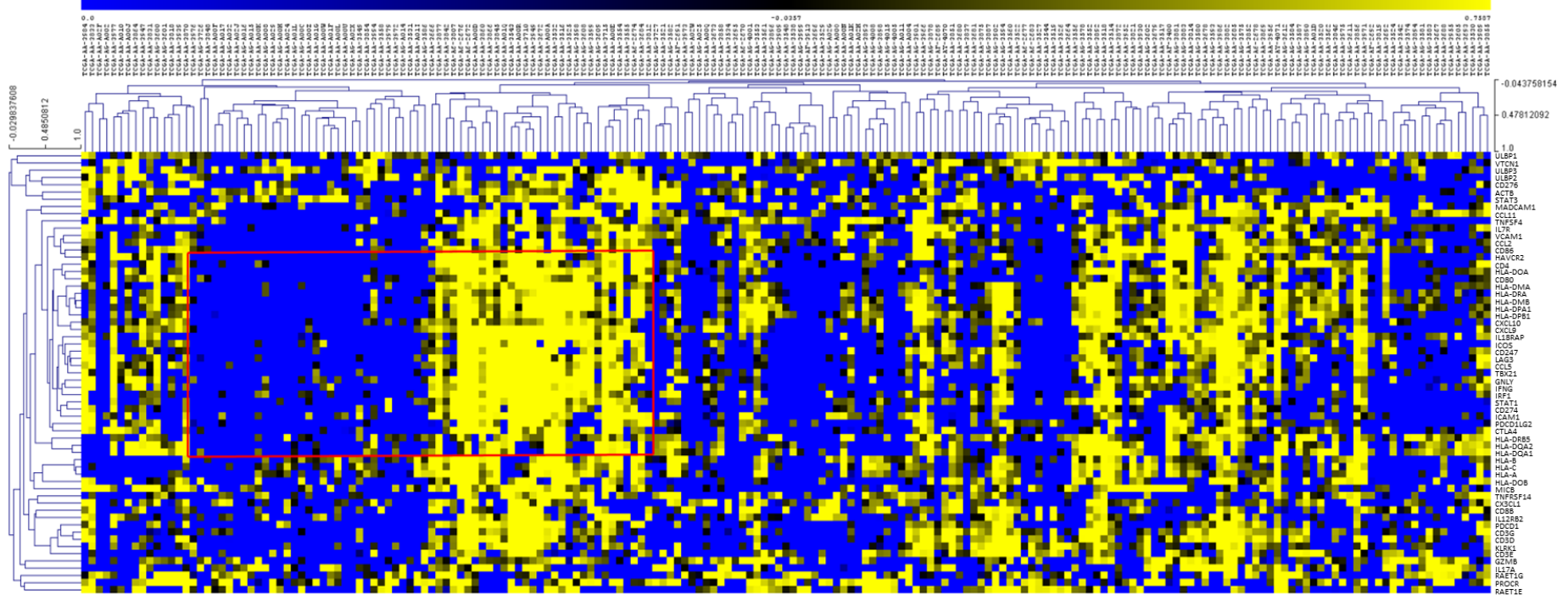


Figure 4.1 Two-dimensional hierarchical clustering reveals a closely co-ordinated immunological gene expression cluster – The Co-ordinate Immune Response Cluster (CIRC). Clustering was performed by gene expression (rows) and patients (columns) using the Pearson algorithm. Yellow represents high gene expression, black represents intermediate gene expression and blue represents low gene expression. The red box shows a group of closely associated genes (the CIRC) that had a co-ordinated expression pattern across the patient population.

Gene ID
HLA-DQA1
HLA-DQA2
HLA-DRB5
CTLA4
PDCD1LG2
ICAM1
CD274
STAT1
IRF1
IFNG
GNLY
TBX21
CCL5
LAG3
CD247
ICOS
IL18RAP
CXCL9
CXCL10
HLA-DPB1
HLA-DPA1
HLA-DMB
HLA-DRA
HLA-DMA
CD80
HLA-DOA
CD4
HAVCR2

Table 4.2 Genes within the co-ordinate immune response cluster (CIRC). These genes are presented in the order of the CIRC signature.

Of the 28 genes in the CIRC, 20 have previously been associated with outcome based on experimental data [37, 42, 100]. I have shown here that these prognostic immune response genes are highly co-ordinately expressed in CRC and correlate with other biologically associated genes including *HLA-DQA1*, *HLA-DRB5*, *HLA-DPB1*, *LAG3*, *TIM3*, *CTLA4* and *CCL5*. The degree of correlation between immune checkpoint receptor/ligand gene expression was notable (Table 4.3). Particularly striking was the correlation between PD1 and LAG3 ($r^2=0.62$, Figure 4.2). CTLA4 and PD1, both key targets of checkpoint blockade therapy, were also correlated (Figure 4.3). Beta actin acted as a control in this analysis (Figure 4.4).

		CTLA4	PD-1	PD-L1	PD-L2	TIM3	LAG3	Beta Actin
		CTLA4	PDCD1	CD274	PDCD1LG2	HAVCR2	LAG3	ACTB
CTLA4	CTLA4	X	0.36	0.42	0.35	0.35	0.42	0.04
PD-1	PDCD1	0.36	X	0.45	0.31	0.45	0.62	0.04
PD-L1	CD274	0.42	0.45	X	0.48	0.5	0.51	0.07
PD-L2	PDCD1LG2	0.35	0.31	0.48	X	0.49	0.31	0.08
TIM3	HAVCR2	0.35	0.45	0.5	0.49	X	0.52	0.06
LAG3	LAG3	0.42	0.62	0.51	0.31	0.52	X	0.03
Beta Actin	ACTB	0.04	0.04	0.07	0.08	0.06	0.03	X

Table 4.3 R² values correlating expression of inhibitory molecules in colorectal cancer. Analysis is controlled against beta actin.

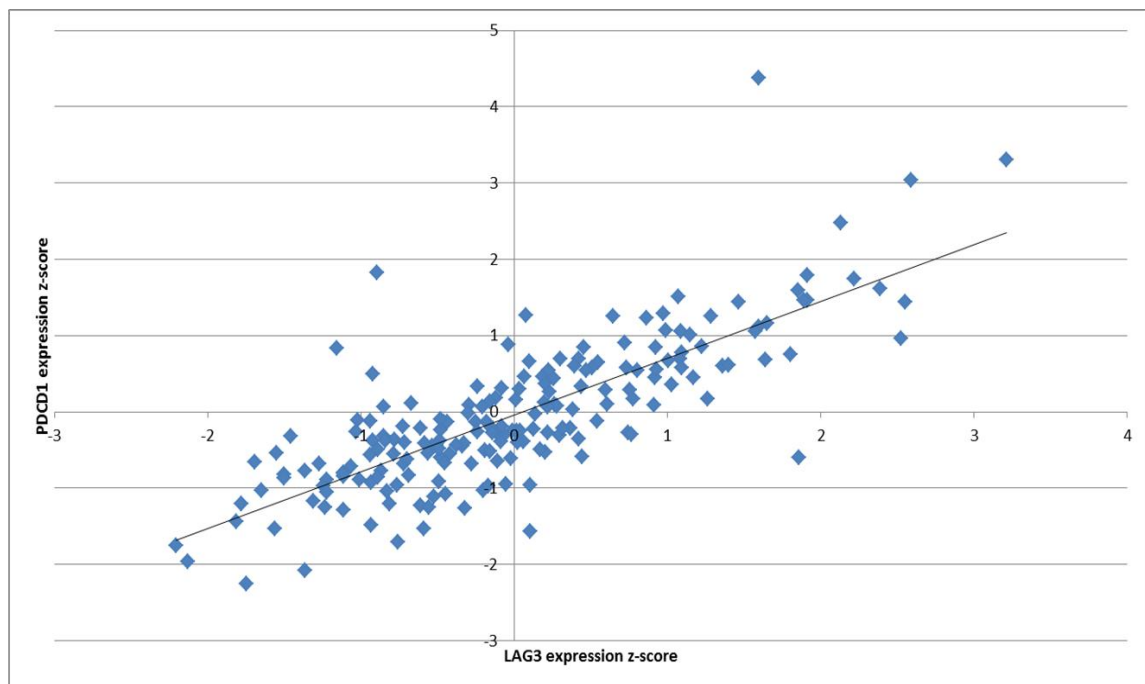


Figure 4.2 Significant correlation between LAG3 and PDCD1 (PD1) expression. Pearson correlation analysis indicates the inhibitory molecules LAG3 and PDCD1 (PD1) are highly coordinated in mRNA expression ($R^2=0.62$).

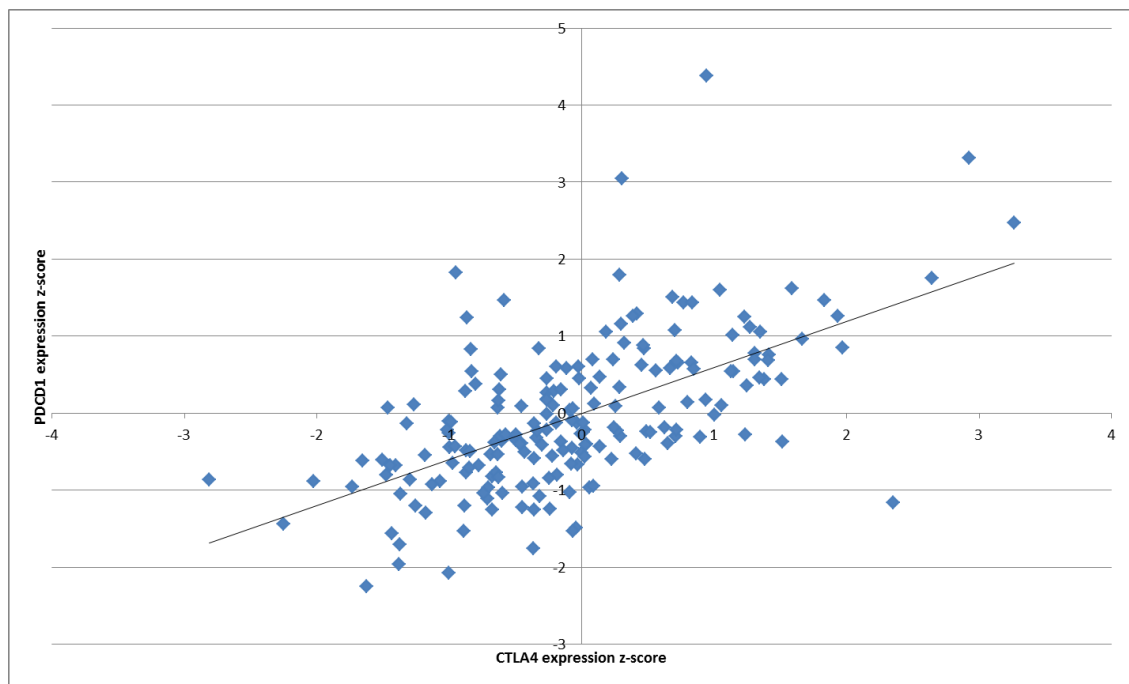


Figure 4.3 Significant correlation between CTLA4 and PDCD1 (PD1) expression. Pearson correlation analysis indicates the inhibitory molecules CTLA4 and PDCD1 (PD1) are highly coordinated in mRNA expression ($R^2=0.36$).

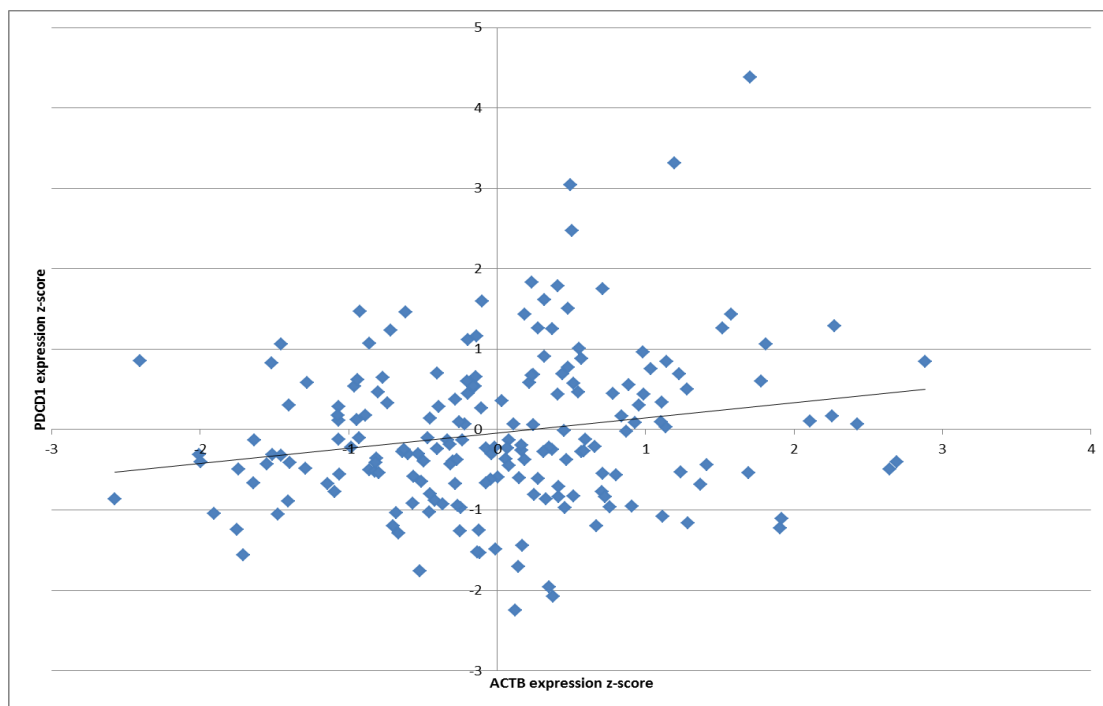


Figure 4.4 Correlation between ACTB (Beta Actin) and PDCD1 (PD1) expression ($R^2=0.04$). This acted as a control in the analysis.

Given the Th1 bias of the cluster I performed separate unsupervised hierarchical clustering analysis of Th2 cytokines IL-2, IL-3, IL-4, IL-5 and IL-6 and transcription factor GATA3, revealing that these genes were excluded from the CIRC. I also noted that the Th17 cytokine IL-17A was also excluded from the cluster (Figure 4.1). The exclusion of these genes confirmed that the CIRC was a Th1-centric cluster.

4.2.2 Molecular determinants of the co-ordinate immune response cluster

Two-dimensional hierarchical clustering (Figure 4.5) was performed for the entire gene list (Table 4.1) incorporating expression data together with molecular and clinical characteristics (*KRAS*, *BRAF*, *NRAS*, *TP53*, *PIK3CA* & *PTEN* mutations; microsatellite status, methylation subtype, tumour stage, tumour site & recurrence data). The microsatellite status of the total patient population was as follows: 11.8% were MSI-H, 15.9% were MSI-L (microsatellite-low) and 71.8% were MSS. 6.7% of patients were *BRAF* mutant (MT)/MSI-H, and 3.1% were *BRAF* MT/MSS. The mutational status of the cohort was as follows - *KRAS* MT (39.5%), *NRAS* MT (8.2%), *BRAF* MT (9.7%), *PIK3CA* MT (18.4%) and *TP53* MT (51.8%). 38.5% of patients were quadruple wild type (*BRAF* WT, *KRAS* WT, *NRAS* WT and *PIK3CA* WT). I also investigated mutations in the novel genes *POLE* (7.2% of patients) and *POLD1* (2.6% of patients), which have recently been highlighted as key drivers of colorectal carcinogenesis for a minority of CRC patients [132]. As in MSI-H tumours, the mutational burden in *POLE* and *POLD1* mutant tumours is high. In the case of *POLE* and *POLD1* mutants, this is due to a defect in the correction of mispaired bases inserted during DNA replication [132].

Hierarchical clustering delineated four distinct patient groups (Figure 4.5, Table 4.4)

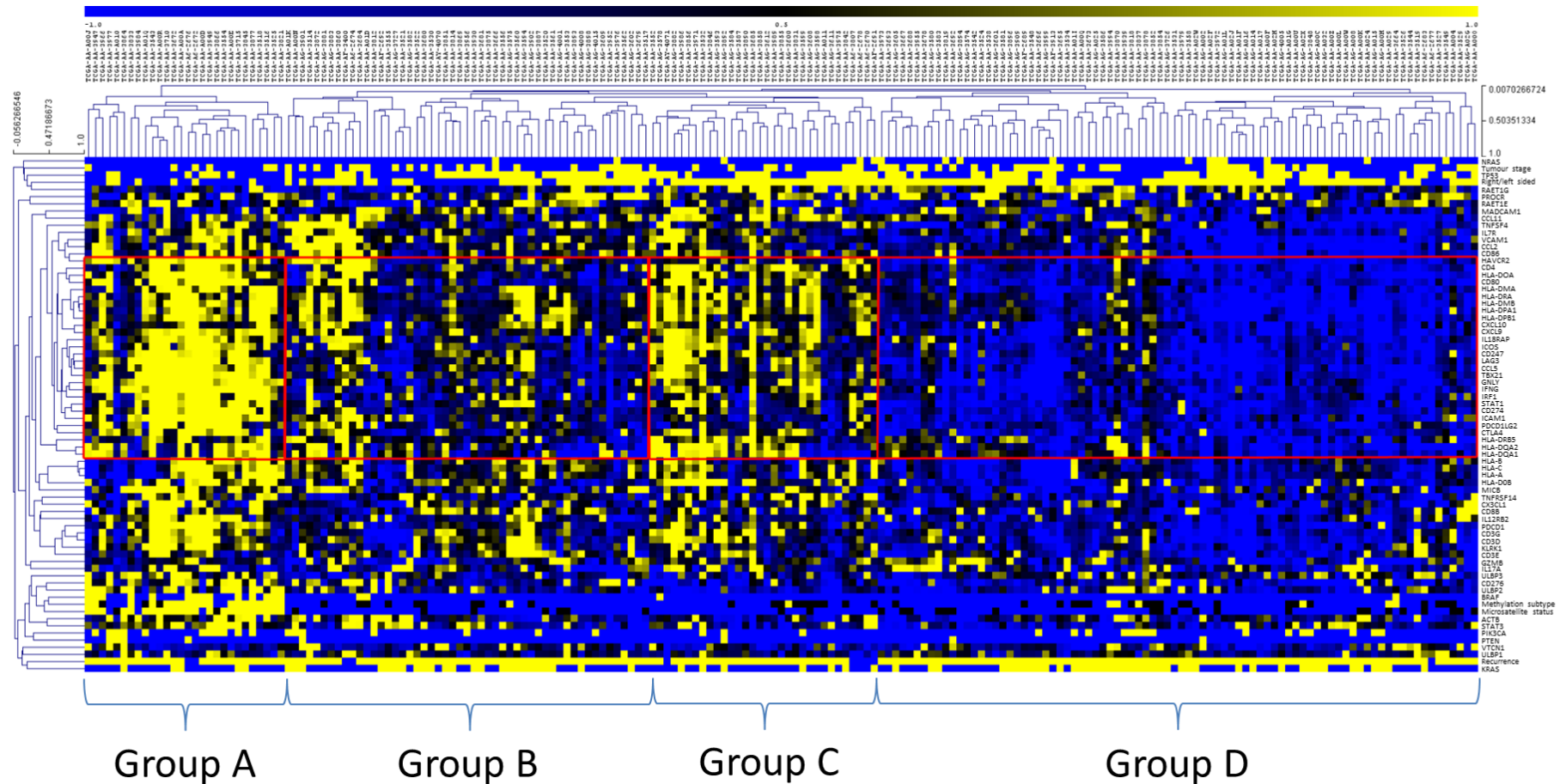


Figure 4.5 Two-dimensional hierarchical clustering delineates distinct immunological CRC patient groups. Gene expression (yellow, high expression; black, intermediate; blue, low expression) was clustered together with mutation data of key genes (*TP53*, *KRAS*, *BRAF*, *NRAS*, *PI3KCA* and *PTEN* (yellow, mutant; blue, wildtype)) and clinical data (microsatellite status (yellow, MSI-H; black, MSI-L; blue, MSS), recurrence data (yellow, recurred/progressed; blue, disease-free), tumour site (yellow, left sided; blue, right sided), tumour stage (yellow, stage III/IV; blue, stage I/II), methylation subtype (yellow, CIMP-H; black, CIMP-L; blue, CIMP-negative). Clustering was performed by genes/mutations/clinical data (rows) and patients (columns) using the Pearson algorithm. Red boxes indicate groups of patients with strong clustering of the co-ordinate immune response cluster. Patients were delineated into four distinct groups (A to D) on the basis of the dendrogram and the cluster expression.


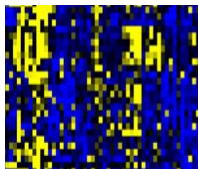
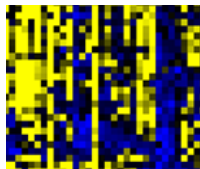
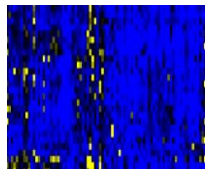
	Group A	Group B	Group C	Group D
Microsatellite status	MSI-H (82%)	MSS (86%)	MSS (94%)	MSS (75%) MSI-L (25%)
Methylation	CIMP-High (68%)	CIMP-Neg (77%)	CIMP-Neg (66%)	CIMP-Neg (69%)
Side of tumour	Right (82%)	Left (63%)	Left (94%)	Left (82%)
Stage I+II	73%	55%	72%	49%
TP53 MT	35%	65%	62%	48%
BRAF MT	50%	4%	3%	1%
KRAS MT	18%	47%	22%	49%
NRAS MT	0%	4%	9%	13%
PIK3CA MT	39%	14%	9%	18%
Quadruple WT	18%	39%	69%	33%
Percentage of patients	14%	26%	16%	43%
Mean cluster expression z-score	+0.98	+0.13	+0.50	-0.62
Cluster expression pattern				

Table 4.4 Characteristics of patient groups. For microsatellite status, methylation and tumour side, the most frequent result is stated. Tumour side refers to the right or left side of the colon. The cluster expression pattern displays the expression pattern of the CIRC cluster in each patient group.

Group A patients (Table 4.4) demonstrated strong CIRC expression (mean expression 0.98). Notably, all of the MSI-H patients were included in this group, who constituted 82.1% of all group A patients, whereas only 10.7% were MSS and 7.1% were MSI-L. Across the entire cohort, expression of the CIRC signature was significantly higher in the MSI-H cancers versus MSS ($p < 0.001$) and MSI-L cancers ($p < 0.001$). Multivariate analysis revealed that expression of *HLAA*, *HLAB* and *HLAC* were all significantly less in MSI-H than MSS cancers ($p = 0.027$, $p = 0.017$ and $p = 0.018$ respectively for HLA-A, B and C), consistent with previous observations [133]. Group A was characterised by the CpG island methylator phenotype (CIMP-H) (67.9%), right-sided tumour site (82.1%), *TP53* wild type (65%) and *KRAS* wild type (82.1%). 50% of patients in this group were *BRAF* mutant and of these *BRAF* mutants, 92.9% were MSI-H. Notably, *POLE* and *POLD1* mutant tumours were associated with higher expression of the CIRC ($p < 0.05$). 42.9% of patients in group A were either *POLE* or *POLD1* MT, and 70.6% of all patients with *POLE* or *POLD1* mutations were found in group A; 100% of *POLD1* mutants were assigned to group A. Of the non-MSI-H patients in group A, 3/5 were *POLE* mutant, two of which were MSS and one MSI-L.

I then analysed patient groups B, C and D (Table 4.4), which displayed lower expression of the CIRC signature than group A, to determine whether any molecular characteristics were associated with low CIRC expression. Group D comprised cancers with the lowest expression of the cluster (mean expression -0.62) and represented 43% of the entire patient population; group B had a lower mean cluster

expression (0.13) than group C (0.50). Significantly, *RAS* mutation (*KRAS/NRAS*), occurring in 47.7% of patients, was associated with lower CIRC expression ($p<0.001$). This was confirmed in the provisional TCGA dataset, in which mutation and expression data for a further 30 patients (225 total) was available ($p<0.001$). The frequency of *RAS* mutation was significantly higher in group B and D than other patient groups ($p=0.01$), with group D having the highest frequency (D (61.9%)>B (51.0%)>C (28.1%)>A (21.4%)) ($p=0.01$). Consistent with these observations, within MSS cancers (enriched in groups B-D), there was a strong trend for *RAS* mutation to be associated with low CIRC expression ($p=0.076$). Furthermore, in the MSS group, *NRAS* mutant cancers had significantly lower CIRC expression than *RAS* wild type cancers ($p<0.05$). In group A, 83.3% of *RAS* mutations were MSI-H and 16.7% were MSS. In *RAS* mutant MSI-H patients, the expression of *HLA-DRA* was lower (0.518) than in *RAS* wild types (1.027) but this difference did not reach significance ($p=0.209$). Finally, multivariate analysis indicated that *KRAS* and *NRAS* mutation, as well as *PIK3CA* and *PTEN* mutation, were associated with decreased CD4 expression ($p<0.05$) (Table 4.5).

Mutation	Gene Expression	Direction of change	P-value
BRAF	CD247	Down	<0.001
	CD80	Up	0.017
	GNLY	Up	0.031
	HAVCR2	Up	0.039
	HLAA	Up	0.004
	HLAB	Down	0.003
	HLADQA	Down	0.017
	LAG3	Up	<0.001
	CD4	Down	0.008
KRAS	CCL5	Up	0.002
	CD247	Down	0.007
	CD4	Down	0.036
	CXCL10	Down	0.037
PIK3CA	CD4	Down	0.001
	HAVCR2	Up	0.005
	STAT3	Up	0.002
TP53	CD274	Up	0.028
	CD276	Down	0.001
	CD4	Up	0.005
	CX3C11	Down	0.012
	CXCL10	Up	0.008
	HLAB	Up	0.024
	HLADQ2	Down	0.002
	IFNG	Down	0.004
	IL7R	Down	0.018
PTEN	CD4	Down	0.003
	CTLA4	Up	<0.001
	HLADMB	Up	<0.001
	HLADRA	Down	0.027
	IL7R	Down	0.004
	STAT1	Down	0.001
	STAT3	Up	0.026
	VTCN1	Up	0.017

Table 4.5 Multivariate analysis of gene expression changes in key mutation groups. p-values are derived from multivariate linear regression analysis. This analysis includes all genes in the initial gene list.

The association of low expression of the cluster and *RAS* mutation was particularly strong in the case of *NRAS* mutation where 76.9% of these cancers were in group D and over 90% of these tumours clustered in the two lowest expression groups (B and D).

Thus *RAS* mutation, unless linked with microsatellite instability, was commonly associated with low levels of Th1 infiltration and activation, class II expression and inhibitory checkpoint expression in this dataset. This depletion appeared to be stronger in *NRAS* cancers in particular. Table 4.6 shows the distribution of key mutations.

Expression of the CIRC correlated inversely with tumour stage. Stage IV cancers had significantly lower CIRC expression than stage I ($p < 0.05$) or stage II ($p < 0.001$) cancers. In analysis of MSS patients alone, stage IV cancers had lower expression than stage II ($p < 0.05$).

CIRC expression was higher in disease-free patients than those with progressive or recurrent disease ($p < 0.05$). This is despite the fact that only 22.2% of disease-free patients were MSI-H. Formal survival analysis on a per-stage basis was not possible due to the lack of sufficient survival data.

	% of total patient population	% of total in patient group A	% of total in patient group B	% of total in patient group C	% of total in patient group D
TP53 MT	51.8%	6.9%	32.7%	19.8%	40.6%
BRAF MT	9.2%	72.2%	11.1%	5.5%	11.1%
KRAS MT	41.0%	7.5%	30%	10%	52.5%
NRAS MT	6.7%	0%	15.4%	7.7%	76.9%
KRAS/NRAS MT	47.7%	6.5%	28.0%	9.7%	55.9%
PIK3CA MT	4.1%	50%	37.5%	0%	12.5%
Quadruple Wildtype	39.0%	6.6%	26.3%	28.9%	38.2%

Table 4.6 Frequencies of mutations in total patient population and distribution of mutations across patient groups A to D.

To prevent overlap of patients, in this analysis patients were classified initially on the basis of *KRAS* status, followed by *NRAS*, *BRAF*, and finally *PIK3CA*. According to these criteria, *KRAS* mutations may also have *NRAS*, *BRAF* and *PIK3CA* mutations (*KRAS* MT +/- *NRAS* MT +/- *BRAF* MT +/- *PIK3CA* MT). *NRAS* mutations are *KRAS* wildtype but may have mutations in *BRAF* and *PIK3CA* (*KRAS* WT + *NRAS* MT +/- *BRAF* MT +/- *PIK3CA* MT). *BRAF* mutants are *KRAS* and *NRAS* wildtype but may have mutations in *PIK3CA* (*KRAS* WT + *NRAS* WT + *BRAF* MT +/- *PIK3CA* MT). *PIK3CA* mutants are wildtype for *KRAS*, *NRAS* and *BRAF*. Quadruple wildtype patients are *KRAS*, *NRAS*, *BRAF* and *PIK3CA* wildtype. *TP53* mutation status is independent of other mutations.

4.2.3 MEK activation signature

To determine how *RAS* mutation impacts on immunity, I wanted to investigate the contribution of the RAS-RAF-MEK-ERK pathway on immunosuppression. To do this, the MEK activation signature characterised by Dry and colleagues was utilised [16]. This signature was initially developed to predict responses to the MEK inhibitor selumetinib *in vitro*, and has been validated in multiple tumour types including colon cancer. It comprises 18 genes that are upregulated by MEK activation, and this upregulation is a strong indicator of MEK and ERK phosphorylation, which occurs downstream to both *RAS* and *RAF*. The signature's authors also characterised a 5-gene CRC optimised MEK signature (MEK CRC), comprising *DUSP6*, *PHLDA1*, *SPRY2*, *DUSP4* and *ETV4*, which was used in this analysis.

In the TCGA dataset, *NRAS* and *KRAS* mutant tumours had significantly higher expression of the MEK CRC signature than *RAS* wild type tumours ($p < 0.05$ for both). The expression of the signature in *NRAS* mutants was borderline significantly higher compared with *KRAS* mutants ($p = 0.055$). The *NRAS* mutant group therefore served as a group where any MEK driven effects were likely to be most obvious. Indeed, the expression of the CIRC was significantly lower in the *NRAS mutants* compared with the *RAS* wild types ($p < 0.05$).

	KRAS MT	NRAS MT	RAS WT
MEK CRC signature	0.089	0.251	-0.096
CIRC	-0.187	-0.604	-0.011
Class II HLA	-0.104	-0.625	0.013
IL-12	-0.128	-0.513	0.322
CXCL9	-0.203	-0.646	-0.029
CXCL10	-0.075	-0.724	0.025

Table 4.7 Mean MEK signature, CIRC, and other selected immune genes in *NRAS* mutant, *KRAS* mutant, and *RAS* wild type patients. Values represent mean microarray z-scores.

Class II HLA expression is an important component of the CIRC and the differences in class II expression mirrored that of the MEK signature where expression was significantly lower in *NRAS* cancers compared with wild type (-0.625 vs 0.013, $p < 0.05$) and borderline significantly lower in *NRAS* compared with *KRAS* mutant tumours (-0.625 vs -0.104, $p = 0.067$). Another key molecule in Th1 immunity is IL-12, which is the key cytokine in the decision of CD4 cells to polarise towards Th1 and is produced by DCs, monocytes and macrophages. IL-12 expression was significantly lower in both *NRAS* and *KRAS* mutant cancers compared with wild type ($p < 0.05$ for both). There was a high degree of negative correlation between IL-12 expression and expression of the MEK signature ($R = -0.359$, $p < 0.00001$) and a positive correlation between IL-12 expression and expression of the CIRC ($R = 0.217$, $p < 0.01$), in all patients and in MSS patients only. When *KRAS* mutant MSS cancers were considered, there was a highly significant difference in IL-12 expression between those with MEK signature expression in the top and bottom quartiles

($p < 0.0001$). These data show strong correlations and suggest the possibility that *RAS* mutation could inhibit Th1 immunity through inhibition of IL-12.

4.2.4 *RAS* mutant subtype analysis

It is established that the degree of RAF/MEK/ERK pathway activation following *KRAS* mutation is dependent on the specific amino acid substitution that occurs [134]. For example, G12D/G13D *KRAS* mutations have low affinity for RAF and a fast hydrolysis rate, which should lead to relatively low levels of RAF activation. Consistent with this, G12D/G13D mutations have been shown to signal through the PI3K/Akt pathway, with no activation of RAF [135]. In contrast, G12V *KRAS* mutations, which are associated with more aggressive tumours, strongly signal through the canonical RAF/MEK/ERK pathway [135]. As a strong inverse relationship between the Dry MEK activation signature and CIRC expression was observed, I hypothesised that differing subgroups of *KRAS* mutants would also vary in immune profile. Table 4.8 shows the CIRC expression in each *KRAS* and *NRAS* subgroup, with the number of patients per group.

Substitution (KRAS unless otherwise stated)	CIRC	Number of patients per group	Percentage of patients per group
NRAS Q61L	0.556	1	1.45
G12R	0.146	1	1.45
G12D	0.007	20	28.99
G13D	-0.084	6	8.70
A146T	-0.087	6	8.70
R68S	-0.170	1	1.45
A146V	-0.286	1	1.45
G12V	-0.368	14	20.29
G12F	-0.428	1	1.45
G12S	-0.511	3	4.35
NRAS Q61K	-0.521	6	8.70
Q22K	-0.541	1	1.45
Q61L	-0.667	2	2.90
G12C	-0.724	4	5.80
NRAS Q61H	-1.100	1	1.45
NRAS G13R	-1.763	1	1.45

Table 4.8 Mean CIRC expression within each *RAS* mutation subtype, and number and percentage of patients per group. Mutations are ordered from highest to lowest CIRC expression.

It was clear that *KRAS* G12D/G13D mutants made up a large proportion of the *KRAS* mutants, but these patients had relatively high immunity. *KRAS* mutant patients were therefore divided into G12D/G13D and non-G12D/G13D mutants. Table 4.9 shows the expression of the CIRC and other key parameters in the *RAS* subgroups of MSS *POLE* and *POLD1* WT patients.

	G12D/G13D KRAS MT (n=26)	Other KRAS MT (n=37)	NRAS MT (n=9)	RAS WT (n=68)
KRAS expression	-0.177	0.706	-0.083	-0.327
NRAS expression	0.014	-0.543	0.369	0.222
CIRC	-0.014	-0.374	-0.604	-0.018
CXCL9	-0.009	-0.485	-0.646	-0.040
CXCL10	0.246	-0.421	-0.724	0.030
IL12A	-0.076	-0.167	-0.513	0.288
IL18	-0.104	-0.026	0.108	-0.071
IL6	-0.286	-0.533	-0.046	-0.139

Table 4.9 Expression of key genes by *RAS* subtypes in MSS *POLE+POLD1* WT patients. *KRAS* mutants are divided into G12D/G13D mutants and other *KRAS* mutants. Values represent mean microarray z-scores.

These data suggest that G12D/G13D *KRAS* mutants are very similar to *RAS* wild type in terms of immune profile, whereas the non-G12D/G13D *KRAS* mutants have significantly lower CIRC expression than *RAS* wild type cancers (-0.374 vs -0.018, $p < 0.05$). This may be due to the lower RAF/MEK/ERK activation seen in G12D/G13D *KRAS* mutants, and also lower *KRAS* expression than in non-G12D/G13D mutants. Therefore, it is likely that subcategorization of *KRAS* mutants will be essential when exploring the association between *RAS* mutation and immunity.

4.2.5 Consensus Molecular Subtypes (CMS) analysis

Several groups have identified molecular subtypes of colorectal cancer based on expression data. In an effort to create consensus and to facilitate clinical translation, and in parallel with my efforts to create an immunological classification, the colorectal cancer subtyping consortium (CRCSC) created an integrated classification - consensus molecular subtypes (CMS) [136]. This consortium of groups utilised data from six independent CRC transcriptional subtyping systems to create a consensus classification which focussed on the core transcriptional groups in CRC (Table 4.10).

Group	Typical Clinical Features	Common Pathological features
CMS1	Females, older age, right colon	MSI-H, hypermutation, <i>BRAF</i> MT
CMS2	Left Colon	Epithelial, MSS, high CIN, <i>TP53</i> MT, WNT/MYC pathway activation
CMS3		Epithelial, heterogeneous CIN/MSI, <i>KRAS</i> MT, IGFBP2 overexpression
CMS4	Younger age, Stage III/IV	Mesenchymal, CIN/MSI, TGF β /VEGF activation, NOTCH3 overexpression

Table 4.10 Characteristics of each CMS group. Adapted from Guinney et al [136].

To investigate how closely my CIRC patient groupings (based predominantly on immunological gene expression) correlated with the CMS classifications, I retrieved CMS data for the TCGA dataset. Table 4.11 shows the distribution of CMS classifications across the four CIRC patient groups.

	Percentage of total patient population	CIRC Group A	CIRC Group B	CIRC Group C	CIRC Group D	CMS group Th1 signature rank
CMS1	13.8%	81.5%	7.4%	3.7%	7.4%	1
CMS2	40.0%	0.0%	15.4%	23.1%	61.5%	4
CMS3	10.3%	0.0%	30.0%	20.0%	50.0%	3
CMS4	18.5%	5.6%	63.9%	13.9%	16.7%	2
Unclassified	12.3%	12.5%	25.0%	12.5%	50.0%	
CIRC group Th1 signature Rank		1	3	2	4	

Table 4.11 Distribution of Consensus Molecular Subtype (CMS) groups across CIRC patient groups.

The cells show the percentage of each CMS group that fall within each CIRC patient group, and the respective rank for Th1 infiltration from high to low (1=highest Th1 immunity, 4=lowest Th1 immunity), by CIRC expression for CIRC groups and Th1 infiltration gene set enrichment for CMS groups [136]. 5.1% of TCGA patients did not have CMS classification data available.

As expected, CMS1, which is a group comprised of mostly MSI-H and BRAF MT tumours with strong immunity, falls predominantly within Group A. Interestingly, 80% of CMS3 tumours, which are mostly epithelial with *KRAS* mutations, fall within patient groups B and D, similar to my *KRAS* distribution data (Table 4.6). This reinforces the hypothesis that *KRAS* mutant tumours are immunologically impoverished. A large proportion of CMS4 tumours, which are predominantly mesenchymal with a poor prognosis, were found in patient group B (63.9%), and a large proportion of CMS2 tumours, which are epithelial with high CIN and WNT/MYC pathway activation, fall within patient group D (61.5%). This demonstrates that patient groups B and D are distinct despite both groups having low CIRC expression levels.

4.2.6 The CIRC compared to other immune signatures

To determine whether the observed immunosuppression in *RAS* mutant cases was specific to my particular immune metagene/signature (CIRC), or whether it was observable with other immune signatures, I retrieved expression data from the TCGA CRC dataset for a variety of published immune metagenes and determined how well their expression correlated with the CIRC [49, 137-139]. These metagenes overlapped little with each other, with only 14.8% of genes replicated in two or more separate metagenes. Strong correlations were found between the CIRC and the Nagalla T/NK (T cell and NK cell) and M/D (macrophage and dendritic cell) signatures, as well as the class II-centric 14-gene Fehlker signature ($p < 0.0001$ for all). The expression of several of these immune signatures were significantly lower in *NRAS* mutant CRC compared to *RAS* wild type (Table 4.12), providing further evidence that *RAS* mutant tumours appear to be immunologically depleted relative to *RAS* wild type in this dataset. In addition, the expression of the signatures were highly concordant in each patient group.

	KRAS MT	NRAS MT	RAS WT
CIRC signature	-0.135	-0.646	0.157
Lym signature (3 gene version)	-0.149	-0.715	0.173
Fehlker signature (14 gene version)	-0.098	-0.538	0.132
Nagalla T/NK signature	-0.128	-0.528	0.096
Nagalla M/D signature	-0.132	-0.635	0.136

Table 4.12 Expression of various immune signatures/metagenes by *RAS* mutation status. Values represent mean microarray z-scores.

In MSS patients, when *KRAS* mutants were further subcategorised into G12D/G13D and non-G12D/G13D mutants, it was clear that non-G12D/G13D *KRAS* mutants had lower expression of all immune signatures than G12D/G13D *KRAS* mutants (Table 4.13).

	KRAS G12D/G13D MT	Other KRAS MT	NRAS MT	RAS WT
CIRC	-0.014	-0.291	-0.604	-0.011
Lym signature (3 gene version)	0.014	-0.309	-0.598	-0.015
Fehlker signature (14 gene version)	0.065	-0.226	-0.529	0.045
Nagalla T/NK signature	-0.008	-0.271	-0.455	-0.008
Nagalla M/D signature	0.063	-0.274	-0.635	0.003

Table 4.13 Expression of immune signatures by *RAS* subtype in MSS patients. *KRAS* has been divided into G12D/G13D *KRAS* mutants and other *KRAS* mutants. Values represent mean microarray z-scores.

4.2.7 The relationship between immunity and total mutation rates

MSI-H tumours, which have high mutational burden, had strong CIRC expression, whereas MSS tumours had lower CIRC expression generally. However, the degree of variability in immunity within the MSS group, as well as the reasons for this variability, is poorly understood. Oncogenic signalling by pathways such as RAS, as explored previously, may partly contribute. Neoantigens, which derive from non-synonymous mutations, are another possible contributor to this variability. Kandoth and colleagues determined mutation rates in 12 major cancer types using TCGA data, and made this data freely available [96]. To determine how mutation rate relates to immunity, I retrieved mutation rate (mutations per megabase pair (Mbp)) and total mutation (mutations per tumour) data for the colorectal TCGA patients and assessed their relationship with the CIRC. Overall, CIRC expression did not correlate with mutational rate ($R^2=0.0026$, $p=0.48$). However, those tumour types with high mutational load (MSI-H and *POLE/POLD1* mutant), had higher CIRC expression than MSS and *POLE+POLD1* WT tumours ($p<0.05$) (Figures 4.6, 4.7 and 4.8).

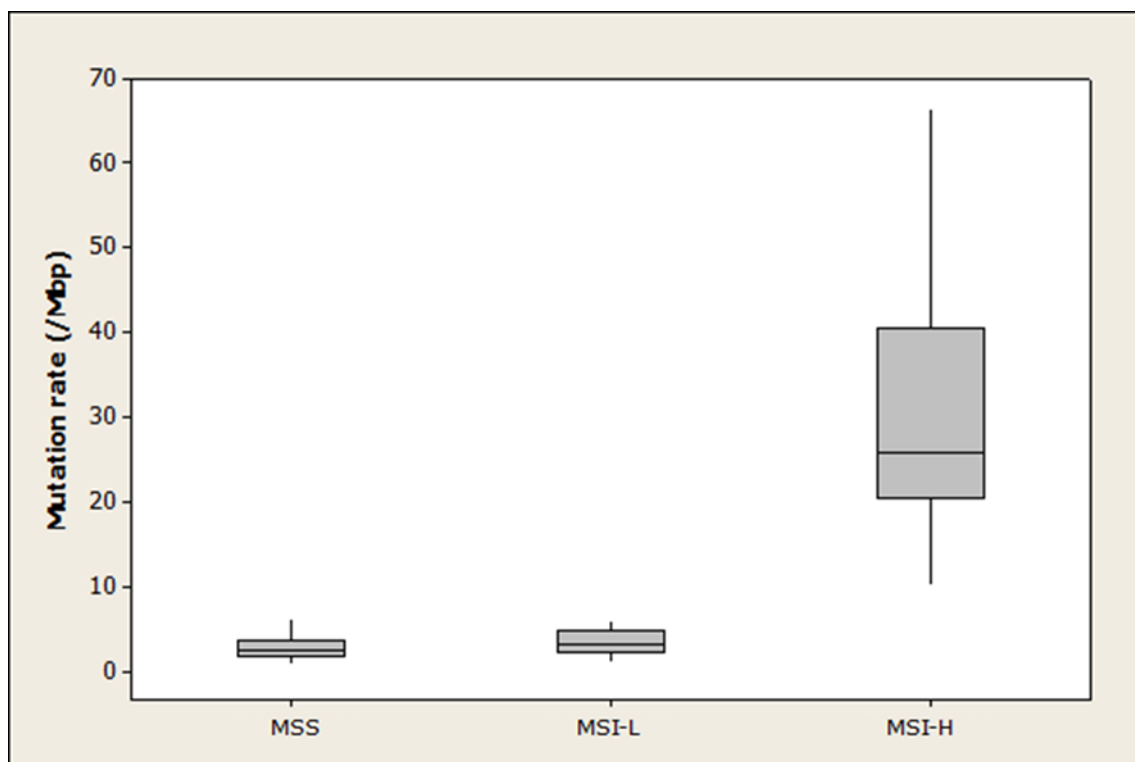


Figure 4.6 Mutation rate by Microsatellite status.

The boxes represent the 25th and 75th percentile values. The lines within the boxes represent median values. The lines emerging from the boxes represent the maximum and minimum values.

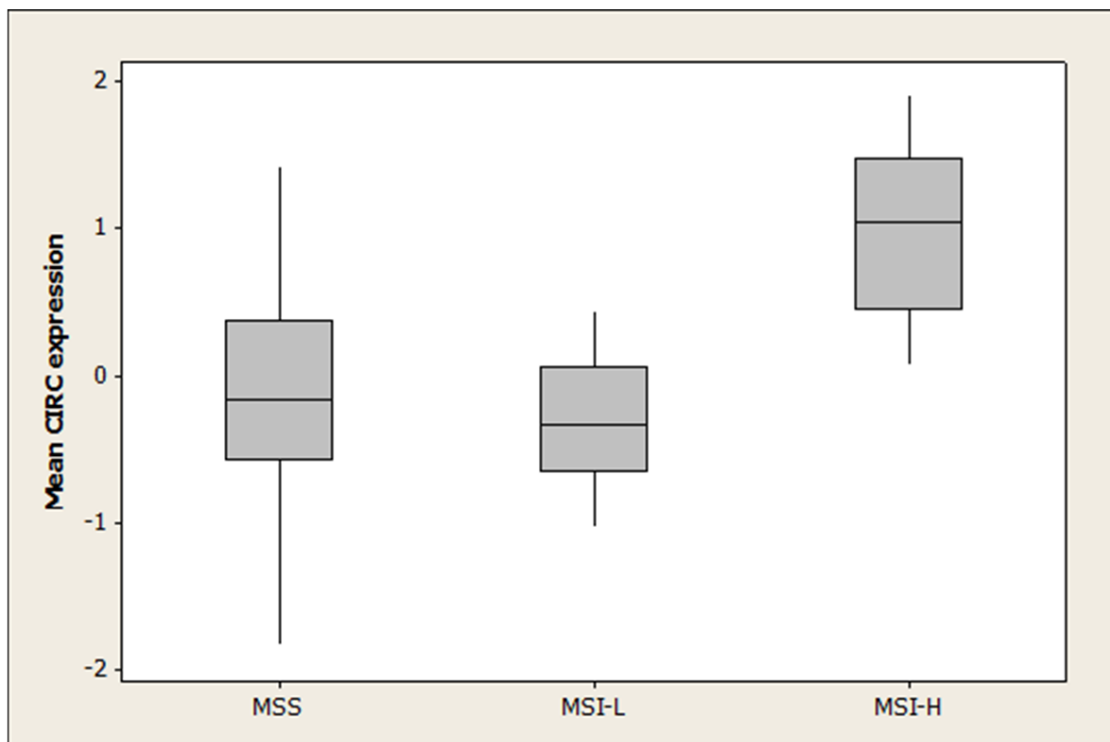


Figure 4.7 CIRC expression by microsatellite status.

The boxes represent the 25th and 75th percentile values. The lines within the boxes represent median values. The lines emerging from the boxes represent the maximum and minimum values.

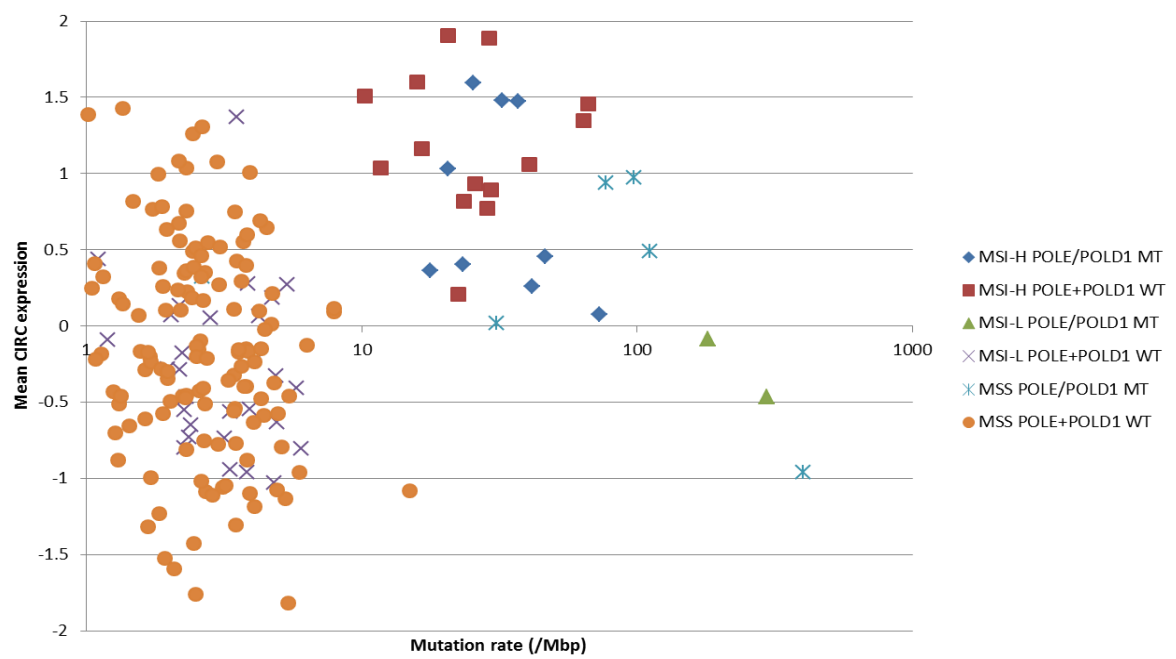


Figure 4.8 Relationship between mutation rate and CIRC expression, by microsatellite and POL subtypes.

All MSS *POLE* and *POLD1* WT patients had relatively low mutation rates (<15/Mbp) (Figure 4.9). In this group, there was no significant correlation between mutation rate and CIRC expression, and yet the CIRC expression was highly variable, with the patients with the highest CIRC expression in this group having similar expression to patients in the MSI-H group. Conversely, many patients in this group had very low CIRC expression. In figure 4.9, MSI-H patients above the horizontal blue line have a CIRC expression within the top 75% of all MSI-H patients. The box highlights MSS *POLE+POLD1* wild type patients (orange dots) with a CIRC signature equal to or above this cut off, representing 20.7% of this patient group. These patients could be classed as 'MSS-CIRC-high'. Like MSI-H patients, these immune-rich MSS patients may be amenable to checkpoint blockade approaches.

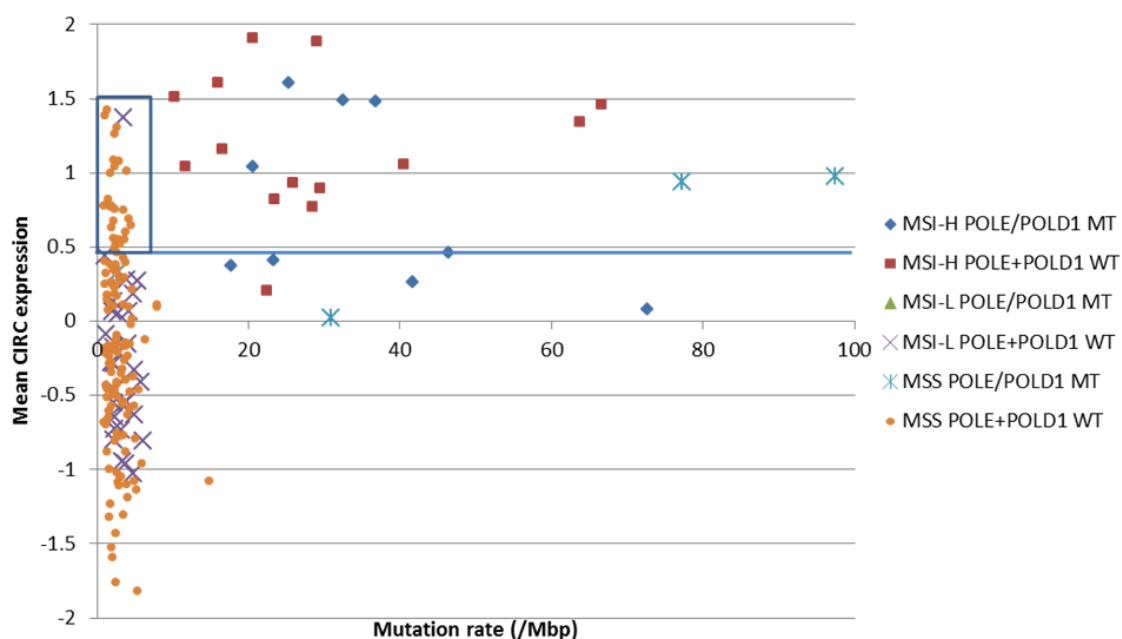


Figure 4.9 Variability of CIRC expression in the MSS *POLE+POLD1* wild type patient group. MSI-H patients above the horizontal blue line have a CIRC expression within the top 75% of all MSI-H patients. The box highlights MSS *POLE+POLD1* wild type patients (orange dots) with a CIRC signature equal to or above this cut off, representing 20.7% of MSS *POLE+POLD1* WT patients that could be classified as ‘MSS-CIRC-high’.

To further investigate how mutation rate correlates with the CIRC, the analysis was extended to a further 8 cancer types that had TCGA mutation rate data available (head and neck cancer, lung squamous cell cancer, lung adenocarcinoma, bladder cancer, glioblastoma, uterine cancer, ovarian cancer and breast cancer) [96].

When analysing mutation rate groups, it became clear that the majority of patients with these tumour types had mutation rates below 15/Mbp (Figure 4.10). However, when correlating mutation rate with CIRC expression, CIRC expression only began to rise markedly when the mutation rate proceeded above 15 mutations/Mbp (Figure 4.11).

Similar results were observed when total numbers of mutations were analysed rather than mutation rate (Figure 4.12). In this case, the CIRC only increased significantly when the number of mutations increased above approximately 500 mutations.

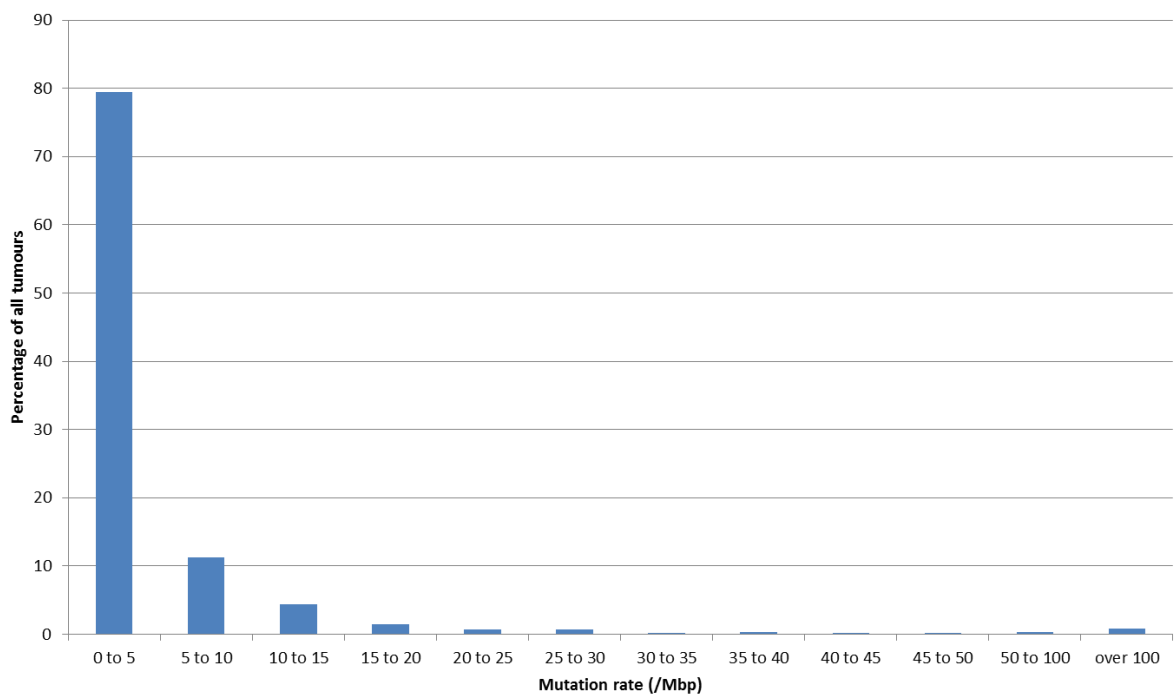


Figure 4.10 Percentage of all tumours within each mutation rate group.

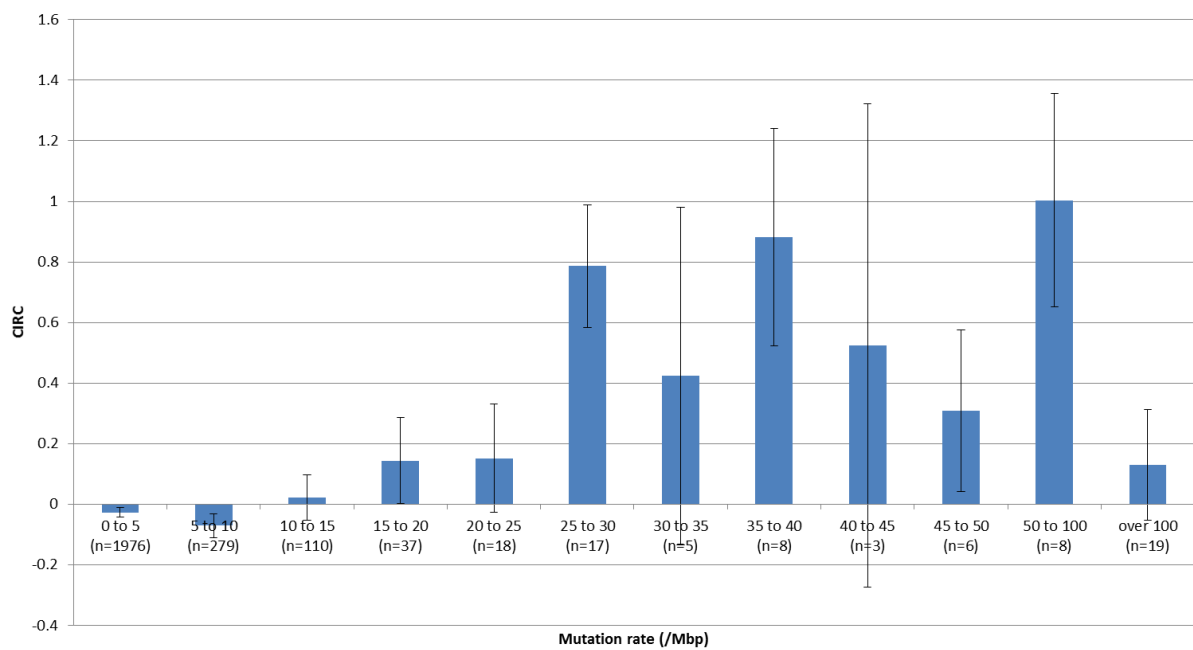


Figure 4.11 CIRC expression in each mutation rate group.

Error bars represent standard error.

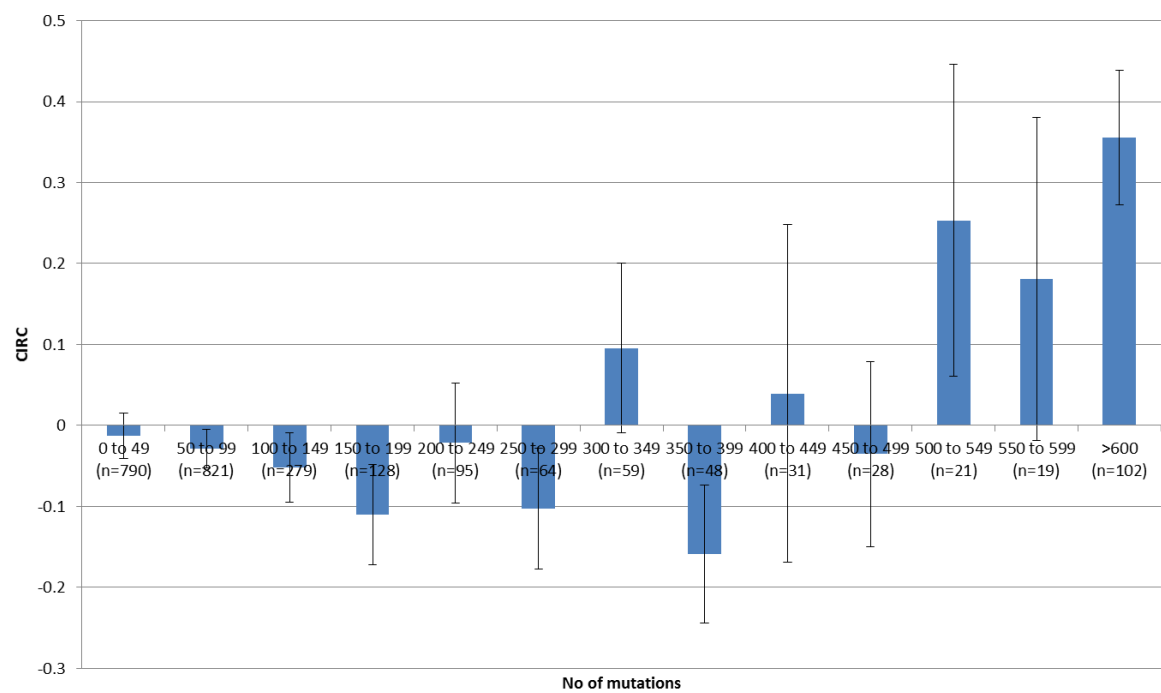


Figure 4.12 CIRC expression in relation to number of mutations per tumour.

Error bars represent standard error.

Since the vast majority of patients had a mutation rate below 15/Mbp (Figure 4.10), and a fraction of these patients had high immunity, only a small percentage of patients with high CIRC expression had high mutation rates (above 15/Mbp) (Figure 4.13). This suggests that in the majority of these patients' cancers, mutational rate was not the dominant determinant of intra-tumoural immunity. Mutation type may be important – patients with low mutation rate with high immunity may have particularly immunogenic mutations.

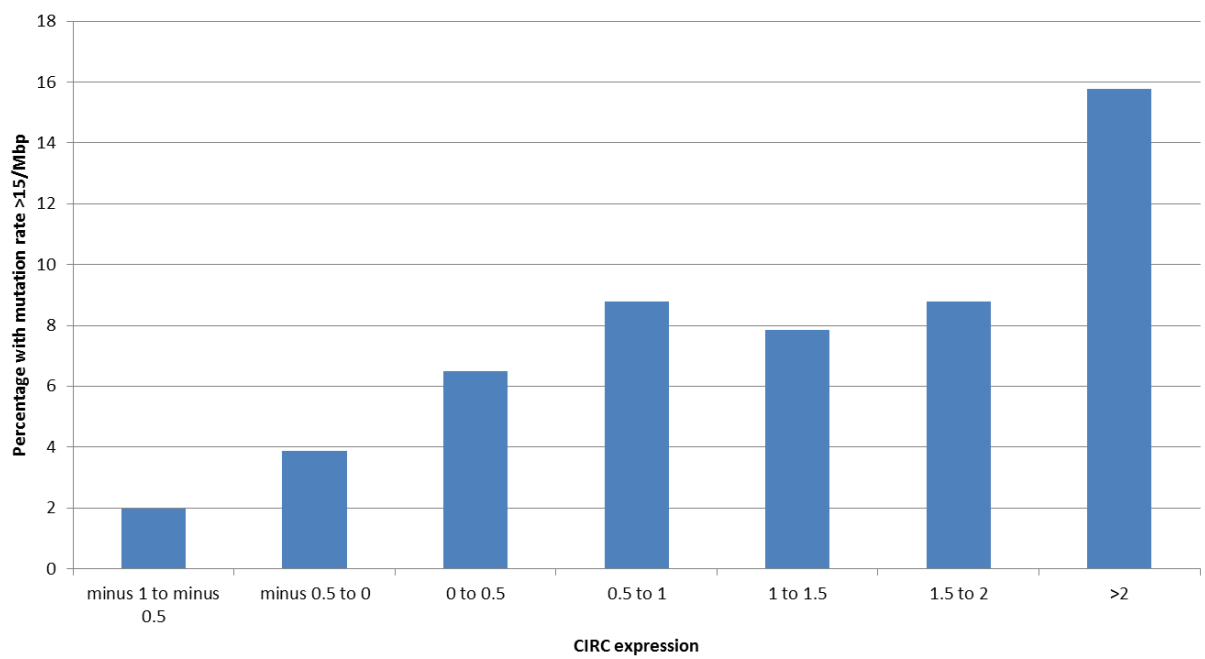


Figure 4.13 Percentage of patients with high mutation rates (>15/Mbp) by CIRC expression group.

4.2.8 Neoantigen prediction analysis

As the majority of patients in the TCGA CRC dataset had relatively low mutation rates but variable CIRC expression, I wanted to investigate the factors that determined the immune profile in these patients. One factor likely to have contributed, as shown above, is mutational profile and the effect of driver mutations. Another important possibility is that both the quality (in terms of antigenicity) and quantity of neoantigens were important. Certain mutations are likely to be more immunogenic than others, partly due to the strength of binding to the patient's particular class I and II HLA molecules. To explore this idea further, I collaborated with the Sahin Group at TRON (Translational Oncology, University of Mainz, Germany). This group had previously utilised Next Generation Sequencing (NGS) and RNAseq technology to predict the neoantigens present in individual cancers, and were using this approach in a personalised cancer vaccine trial [16]. In addition, they had performed a parallel analysis on the TCGA CRC dataset, and therefore had neoantigen data available. To perform the analysis the number of non-synonymous single nucleotide variations (nsSNVs), which are responsible for the alteration of the protein amino acid sequence, was determined for each patient. Then, using RNAseq data, the number of reads per mutation was calculated to determine whether the neoantigen was expressed. Highly expressed neoantigens are more likely to produce immune responses [16]. Finally, HLA-type was predicted from RNAseq data using seq2HLA, and HLA-binding predictions were made using the IEDB MHC binding prediction algorithm v2.9, "consensus" method [97, 98]. A cut-off for strong binding (and thus "likely immunogenic") was below 1.0. This approach was used to

determine how the strength of neoantigen binding to HLA and the total number of neoantigens related to CIRC expression in TCGA patients.

In addition, I wanted to explore how the neoantigen profile varied between patients and patient groups. A recent study demonstrated that there is little overlap in neoantigen profile between CRC patients and that individual neoantigens are rarely shared [140]. This group suggested that 70 peptides would be required in a generic vaccine to cover 50% of MSS CRC patients. However, what had not been explored is how the profile of neoantigens varies in key patient subgroups, such as *RAS* mutant patients. Increasing this understanding could be key to stratified vaccine approaches targeting such groups, which could have a key role in the window before personalised treatments are produced for patients.

Initially, the number of genomic, expressed and predicted to be presented nsSNVs were determined for each patient and were grouped by microsatellite status (Figure 4.14). As expected, MSI-H patients were predicted to present significantly more nsSNVs than MSI-L or MSS patients. However, the difference was larger than expected, with MSI-H tumours predicted to present a median of 359 nsSNVs, and MSS tumours predicted to present a median of only 26 nsSNVs ($p < 0.001$).

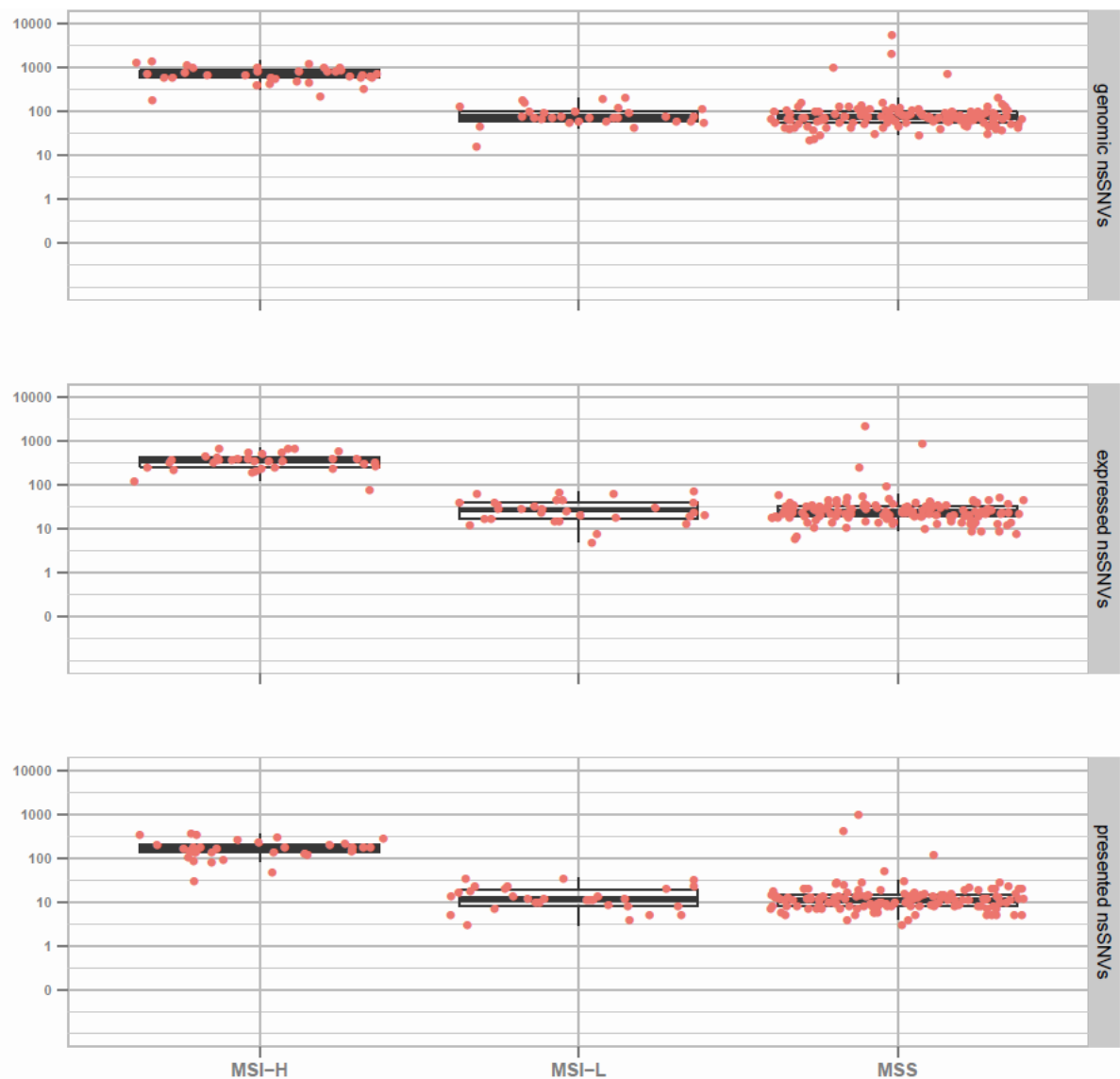


Figure 4.14 Number of total presented Class I neoepitopes per patient by MSS status. The top panel shows the total number of non-synonymous single nucleotide variations (nsSNVs) in each patient (red dots). The middle panel shows the number of these nsSNVs that were expressed (at mRNA level) in each patient. The bottom panel shows the number of these expressed nsSNVs that were predicted to be presented on that particular patient's predicted class I HLA type. Boxes represent median, 25th and 75th percentile of each group.

Figure courtesy of Sebastian Boegel, TRON.

In the overall population, patients with stronger binding Class I neoepitopes (in terms of the single strongest binder) had significantly higher CIRC scores than those with weaker binders ($p < 0.001$), which was consistent with the quality of neoantigens impacting significantly on the overall immune phenotype (Figure 4.15). This was also observed in relation to total number of neoantigens. However, in the MSS group alone, this correlation was not observed for neoantigen binding strength or quantity, suggesting that in the majority of patients with low mutation rates, strongest binding neoantigen or neoantigen number was not the main determinant of immune response. The same result was observed for Class II-binding neoantigens (Figure 4.15). Furthermore, the number of mutations in each sample did not correlate with CIRC expression within the MSS or MSI-H groups (though MSI-H patients had both higher mutation numbers and CIRC than MSS patients) (Figures 4.16 and 4.17). This is consistent with my findings above regarding mutation rates and immune responses across 8 tumour types. It is possible that the immune response only rises substantially with mutation rate above a certain level. This highlights the need for further work to explore and fully understand the determinants of immune response in this key patient group.

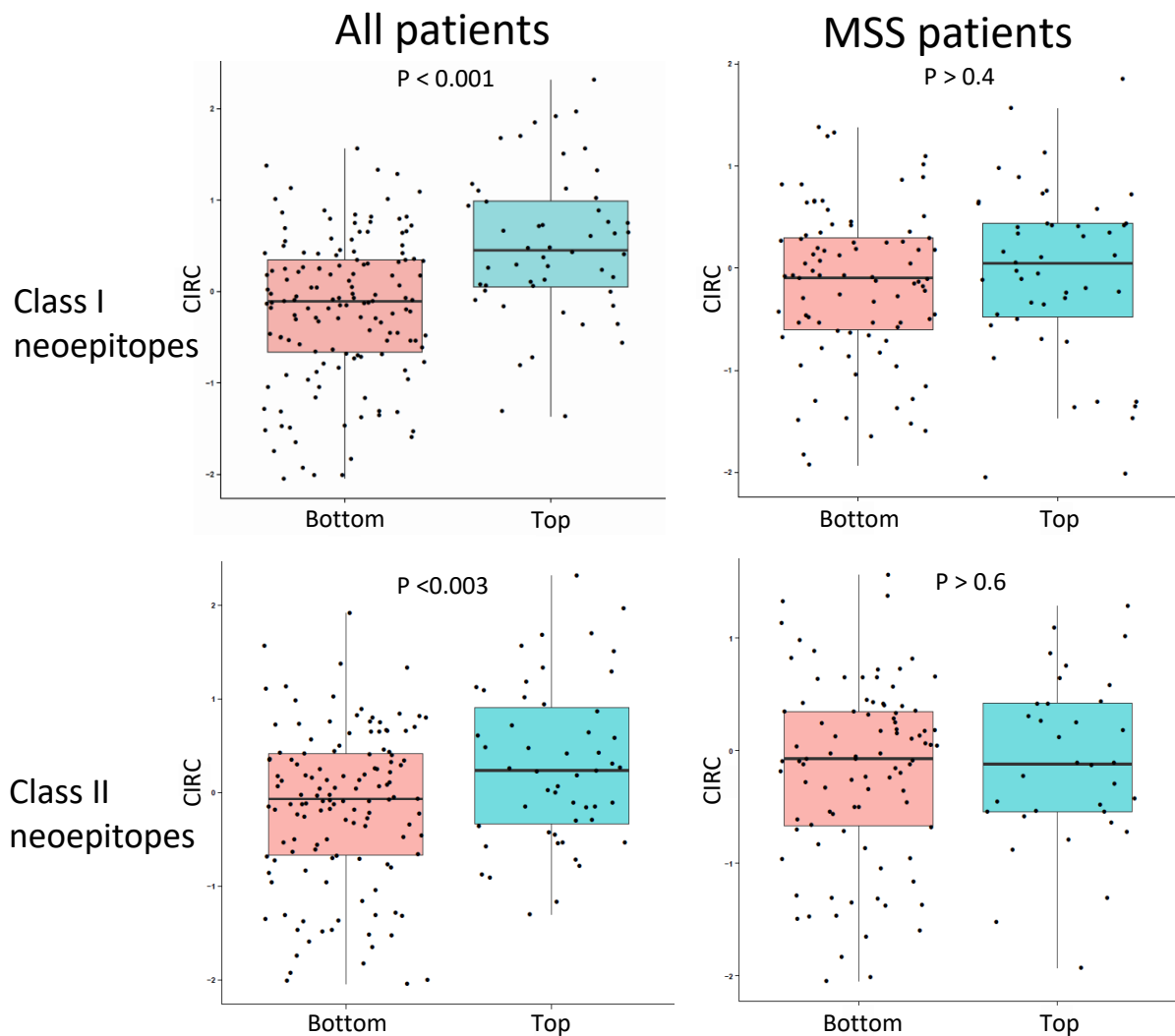


Figure 4.15 CIRC score in patients with top binding neoantigens versus patients with bottom binding neoantigens, for both Class I and II neo-epitopes in all patients and MSS patients only. The single strongest binding neoantigen in each case was considered.

Figure courtesy of Sebastian Boegel, TRON.

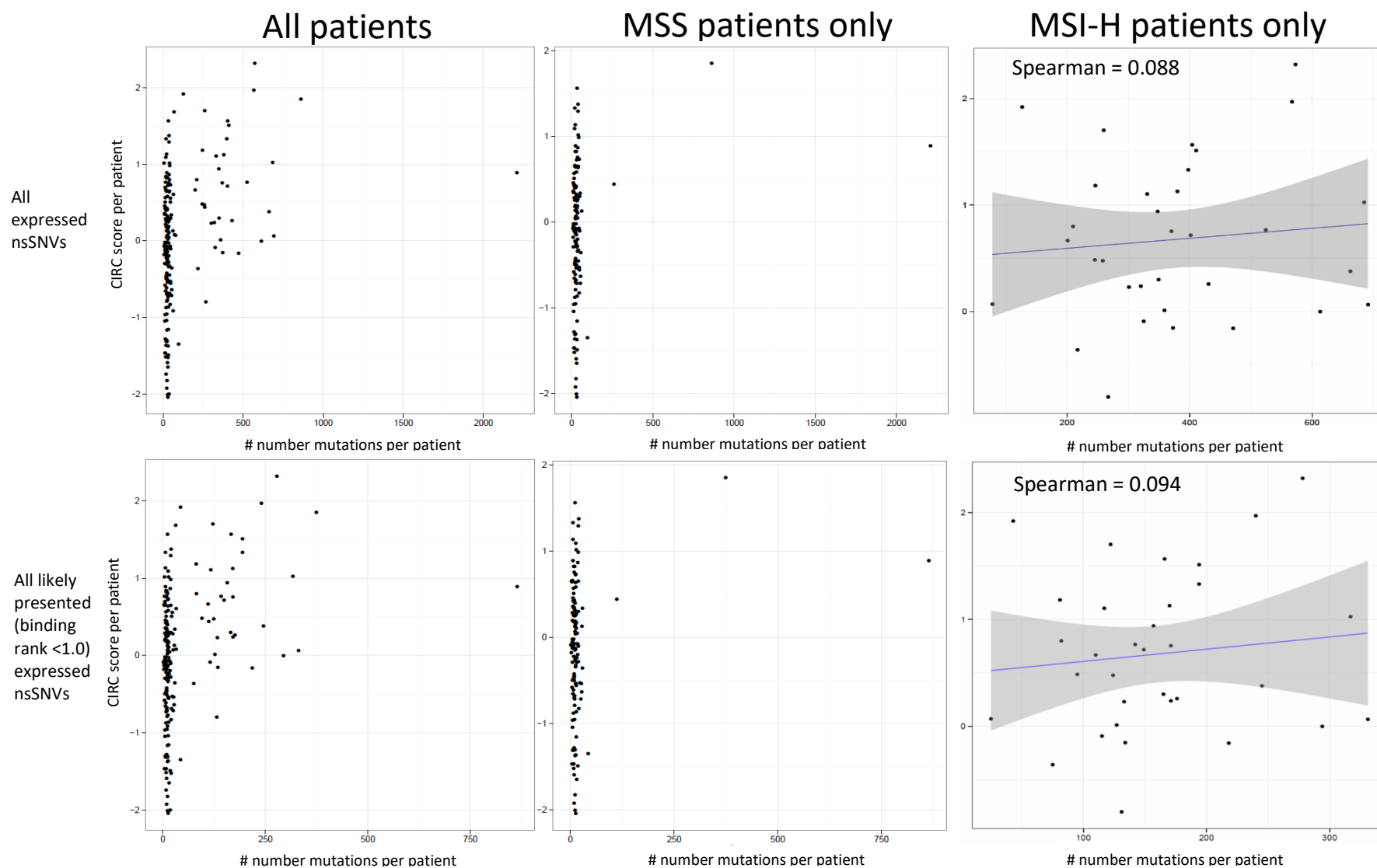


Figure 4.16 Class I neoantigens analysis. Correlation between CIRC and total number of expressed nonsynonymous single nucleotide variations (nsSNVs) and those likely to bind to class I HLA in all patients, MSS patients only, and MSI-H patients only. Spearman values are non-significant.

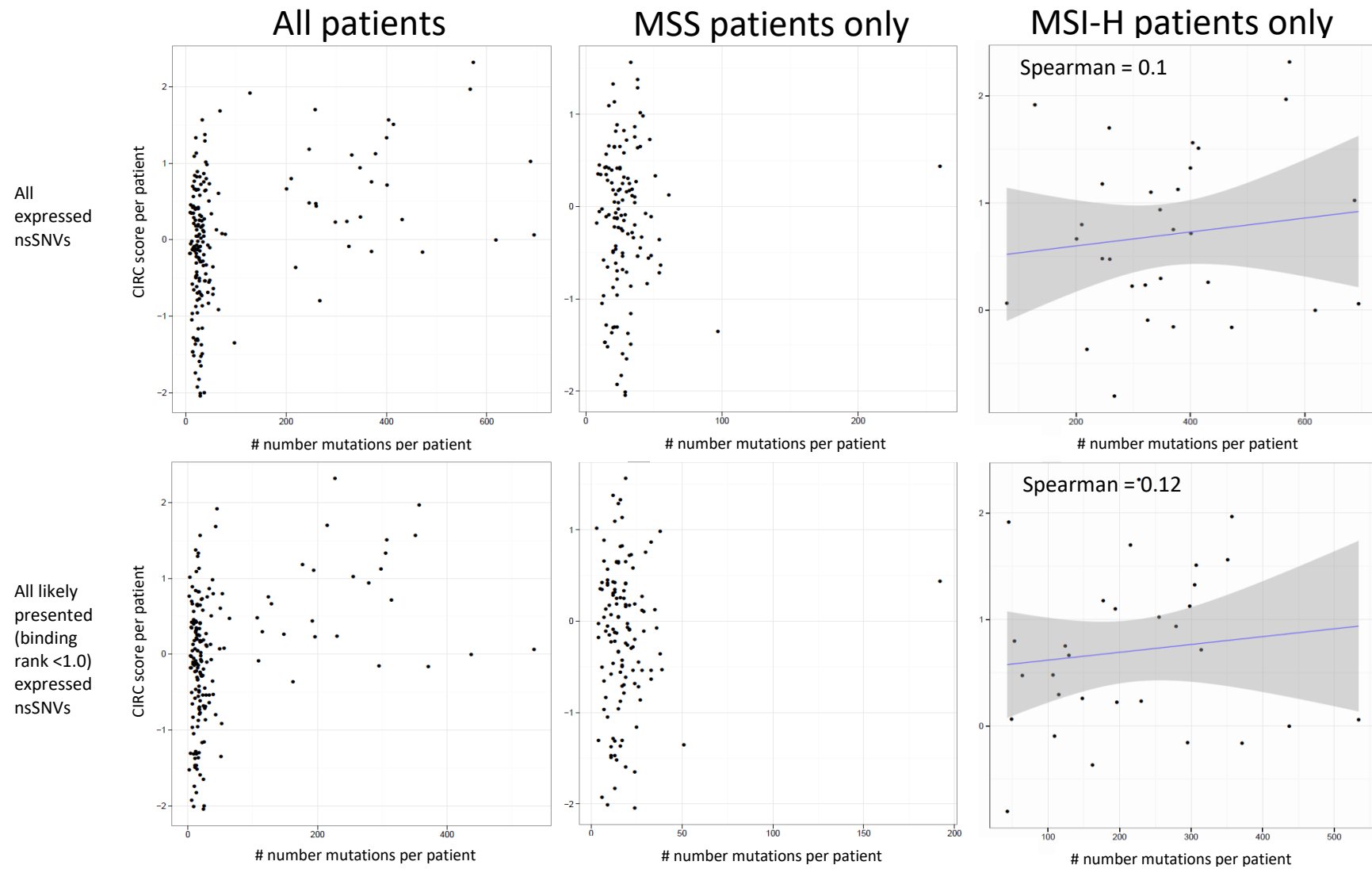


Figure 4.17 Class II neoantigens analysis. Correlation between CIRC and total number of expressed nonsynonymous single nucleotide variations (nsSNVs) and those likely to bind to class II HLA. Spearman values are non-significant.

In addition to investigating quality and quantity of neoantigens, and the association with the CIRC, I also wanted to investigate the antigenicity of *RAS*-associated neoantigens. This analysis demonstrated that approximately 94% of TCGA CRC *RAS* mutant patients were predicted to express a *RAS*-associated neoantigen on Class I HLA (Table 4.14). Approximately 23% of all *KRAS* mutant patients were predicted to present a 'likely immunogenic' strong-binding *KRAS*-associated neoantigen. When considering that MSS patients were only predicted to present a median of 26 Class I neoepitopes each, this represents strong overlap within this patient group. Furthermore, within certain *RAS* mutant subgroups (Table 4.15), such as the common G12V subgroup, 50% of patients were predicted to present an immunogenic *KRAS*-associated neoantigen (Figure 4.18). These analyses took into account the predicted HLA-types of the patients. Certain HLA and neoantigen combinations were recurrent (Table 4.16). The identification of recurrent *RAS*-associated neoantigens has implications for *RAS* as a target for vaccination and cell therapy approaches.

KRAS substitution	Total in group	Number predicted to present KRAS associated neoantigen (any rank)	Number predicted to present 'likely immunogenic' KRAS associated neoantigen (rank <1)	Percentage of group predicted to present a KRAS associated neoantigen (any rank)	Percentage of group predicted to present 'likely immunogenic' KRAS associated neoantigen (rank <1)
A146T	6	6	0	100.00	0
A146V	1	1	1	100.00	100
A155D	1	0	0	0.00	0
G12A	5	5	0	100.00	0
G12C	3	3	1	100.00	33.33
G12D	19	19	0	100.00	0
G12R	2	2	1	100.00	50
G12S	5	5	3	100.00	60
G12V	16	14	8	87.50	50
G13C	1	1	0	100.00	0
G13D	17	17	0	100.00	0
K117N	1	1	0	100.00	0
L19F	1	1	1	100.00	100
P34L	1	1	1	100.00	100
Q22K	1	1	1	100.00	100
Q61H	2	2	1	100.00	50
Q61K	2	0	0	0.00	0
Q61P	1	1	1	100.00	100
Q61R	1	1	1	100.00	100
All KRAS mutants	86	81	20	94.19	23.26

Table 4.14 *KRAS* subtype neoantigens predicted to present on HLA A and B.

KRAS substitution	Total in group	Percentage of group with 'likely immunogenic' KRAS associated neoantigen (rank <1)
A146T	6	0
G12A	5	0
G12C	3	33.33
G12D	19	0
G12S	5	60
G12V	16	50
G13D	17	0

Table 4.15 *KRAS* subtype neoantigens predicted to present on HLA A and B – selected mutations.

Predicted class I HLA subtype binding to KRAS neoepitope	Presented KRAS associated neo-epitope (substitution in red)	Percentage of total KRAS G12V mutant patient group (n=16)
B*49:01	TEYKLVVVGAV	6.25% (n=1)
B*40:02	TEYKLVVVGAV	6.25% (n=1)
B*18:01	TEYKLVVVGAV	18.75% (n=3)
A*03:01	VVGAVGVGK	18.75% (n=3)

Table 4.16 Analysis of HLA-restriction and presentation frequency of predicted strong binding *KRAS* G12V-derived neoantigens. Notably, this analysis predicts a promiscuous HLA-B-restricted neoepitope (TEYKLVVVGAV), as well as HLA-A derived G12V presentation by HLA-A*03 (VVGAVGVGK). In addition, it highlights the potential for presentation of some form of *KRAS* G12V-based neoantigen in up to 50% of the patient group.

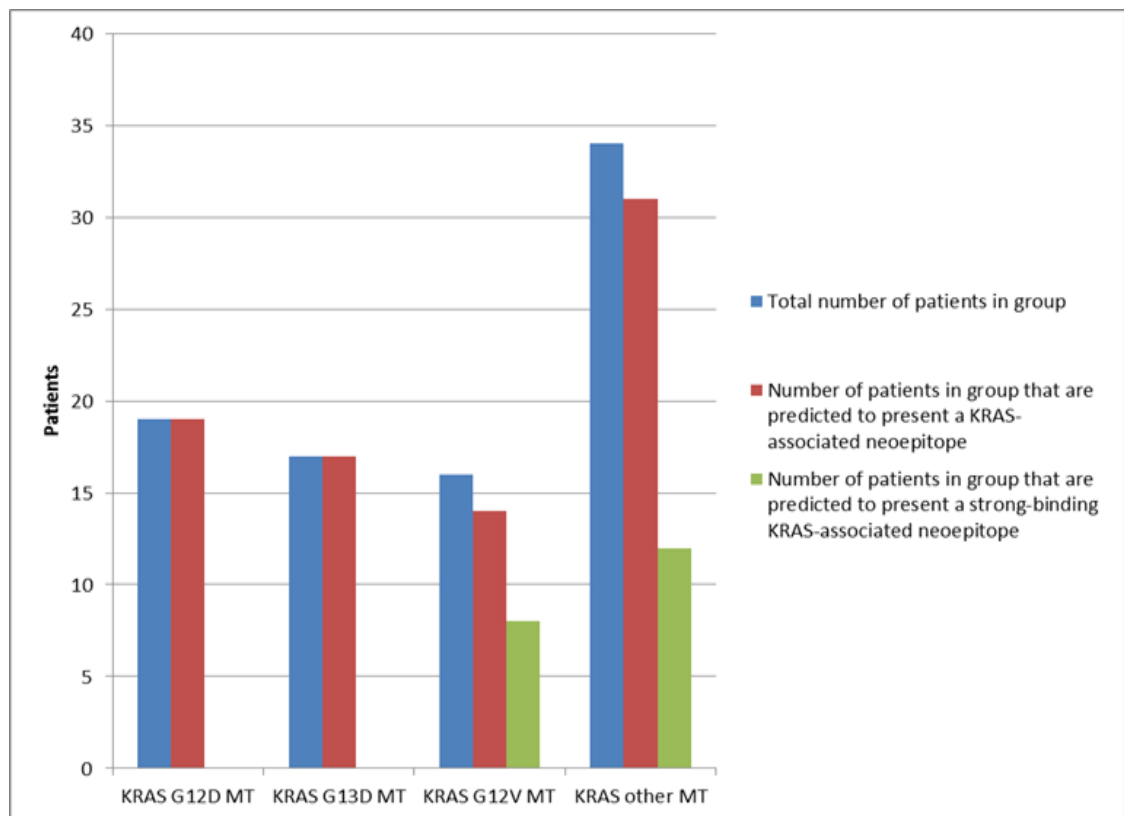


Figure 4.18 Neopeptide prediction for recurrent *KRAS* mutations.

All *KRAS* mutation subtypes (frequency indicated by the blue bars) were predicted to give rise to class I MHC epitopes (red bars). However the proportion of predicted strong-binding (score ≤ 1) *KRAS*-derived neoantigens (green bars) was highest for *KRAS* G12V and a miscellaneous grouping excluding the G12V, G12D, and G13D subtypes.

4.3 Discussion

In this chapter, I have identified a group of tightly co-regulated immune-related genes that was termed the Coordinate Immune Response Cluster (CIRC), and used this to assess differences in the intra-tumoural immune response in a molecularly characterised cohort of CRC patients. The CIRC signature I defined stems from the work of Galon and colleagues, who first established the prognostic impact of T cell infiltration in CRC, initially highlighting 7 co-modulated genes principally associated with Th1-associated immunity that correlated with outcome [37]. Subsequently, they demonstrated that the Th1 genes, *TBX21*, *IFNG*, *IRF1* and *STAT1* were all individually associated with outcome [42]. Other independent immune gene predictors in that analysis were *IL18RAP*, *ICOS*, PD-L1, PD-L2 and PD-1. I extended this gene list by adding the class I and class II genes, further immune checkpoint genes and a number of other genes related to innate immune recognition and T cell activation, generating an initial gene list of >50 genes. Hierarchical clustering then highlighted a core cluster of 28 tightly co-regulated genes comprising the CIRC signature. Importantly, 20 of the 28 genes had previously been highlighted by Galon and colleagues [37, 42, 141] as either prognostically or predictively relevant, including 5 of the original prognostic 7 gene cluster (with only *CD8A* (for which microarray data was unavailable) and *GZMB* not represented). In contrast, genes associated with Th2 and Th17 profiles were excluded from the cluster. Thus, applying a different approach to an independent dataset my analysis verified the high level of correlation of a number of genes based on Th1-associated immunity, and also highlighted involvement of additional immune genes.

An important feature of the CIRC signature is that it includes essentially all class II MHC loci, as well as CD4, whereas in contrast, expression of class I MHC molecules, CD8B and also GZMB are all excluded from the signature. In addition to the critical role of T helper cells for CD8 cell priming [142, 143] and expansion [144, 145], CD4 cells have also been suggested to be major mediators of immunological tumour cell death [146-148]. Adoptive CD4 T cell transfer has been found to up-regulate class II expression on tumour cells mediating protection from tumour progression [147]. Class II up-regulation was mediated via IFN γ and protection was attenuated using anti-IFN γ antibodies. Recent adoptive transfer approaches involving autologous CD4⁺ T cells have met with clinical success. A durable complete response was obtained in a patient with melanoma after infusion of NY-ESO-1 specific CD4⁺ cells recognizing an HLA-DP4 restricted epitope [149]. Additionally, a durable response was obtained in a patient with cholangiocarcinoma on infusion of autologous mutation-specific CD4⁺ cells which adopted a poly-functional Th1 phenotype [27]. These studies highlight the potential for CD4 cells to mediate clinically potent anti-tumour responses via Th1 mechanisms. The CD4-centric nature of the CIRC highlights that CD4 T cells may be important in CRC anti-tumour immunity.

The CIRC included the major immune checkpoint molecules. Not only PD-L1, PD-L2 but also LAG3, TIM3 and CTLA4 were all represented in the CIRC and there was a high degree of correlation between inhibitory checkpoint gene expression. This is consistent with the expected feedback sequelae of a pronounced Th1 infiltrate. IFN γ

is the canonical cytokine associated with Th1 T helper cells and expression of PD-L1 is significantly augmented by IFN γ [150]. IRF1 is of primary importance in the constitutive expression of PD-L1 and in IFN γ -driven up-regulation [151]. In animal models, significant percentages of infiltrating CD4+ and CD8+ cells co-express high levels of both PD-1 and LAG-3 [152]. Dual anti-LAG3/anti-PD-1 immunotherapy strikingly enhanced survival in the MC38 colon cancer model compared with animals treated with single antibodies alone [152]. My data provide justification for trialling combinations of checkpoint blockade agents in CRC.

One of the key aims of this study was to investigate the somatic factors associated with the immune response in CRC. Group A, which exhibited strong expression of the CIRC signature, was dominated by MSI-H tumours, all of which fell in this grouping. This complements very well the work of Brossière-Michot and colleagues, who revealed that MSI-H tumours have a high density of Tbet-positive Th1 T cells relative to MSS tumours, and higher expression of the chemokines CCL5, CXCL9 and CXCL10, all of which are found in the CIRC [50]. The Th1 response may be driven through activation of the CXCL9/CXCL10 signalling axis [50]. Previous studies have also shown that MSI-H phenotype is highly associated with class II expression [153]. Strongly DR positive tumours had a significantly higher TIL density than those with absent or weak staining and survival was significantly better in patients with high DR expression. Consistent with this, Bindea and colleagues demonstrated that the expression of several class II genes, including *HLA-DRA*, are individually associated with improved disease-free survival [141]. High DR expression in MSI-H CRC contrasts with that of class I molecules, which are

completely lost in 60% of sporadic MSI-H cases and only 16.7% of right sided MSS cancers [133]. Thus, the immune landscape of MSI-H CRC, which is characterised by a high mutational burden including frameshift mutations [48], is dominated by T helper cell infiltration and activation, class II expression and co-ordinated up-regulation of a range of immune checkpoint genes. My data suggested that these patients may be more responsive to checkpoint blockade – a hypothesis which has since been confirmed [154]. Similarly, *POLE/POLD1* mutant tumours, which also have a high mutational burden whether microsatellite stable or unstable [132], were also associated with high CIRC expression. This association has also now been confirmed in a retrospective clinical study [155]. These data in combination support the hypothesis that high mutational burden may translate into a strong immune response, possibly due to neoantigen presentation, leading to better outcomes. The finding that the gene encoding *TGF β R* undergoes frameshift mutation in 90% of MSI-H CRC, giving rise to a highly immunogenic promiscuous class II peptide, is entirely in keeping with this hypothesis [156]. Though *FOXP3* was not in the hierarchical clustering due to an absence of microarray data, I separately analysed RNA sequencing data which showed that *FOXP3* expression is highest in patient group A and lowest in group D, suggesting that the immune enrichment in MSI-H patients may include suppressive Tregs. However, the ratio of effector T cells to regulatory T cells, which was not determined in this analysis, may be important [157] and may differ between MSI-H and MSS patients.

A further aim of this work was the investigation of mutation rates and CIRC. I found that overall, the CIRC was higher in patients with high mutation rates, both in CRC

and across eight tumour types. Increased mutation rate is likely to increase the probability of immunogenic mutations, such as missense mutations. However, this relationship was not linear, and the CIRC only increased significantly above a certain level (15 mutations/Mbp, or 500 total mutations). MSI-H and MSS differed in both the number of neoantigens, as well as the strongest binding neoantigens. The frameshift mutations seen in MSI-H lead to multiple changes in amino acid sequence, whereas single point mutations may lead to single amino acid substitutions. It is likely that multiple amino acid changes are more immunogenic than single amino acid alterations, due to the higher number of neoantigens (each with potentially multiple amino acid changes) that will be derived from these mutated proteins. Factors related to this are likely to explain the difference in immunity between these two groups, as well as the difference in clinical outcomes. However, within both the MSS and MSI-H groups, there were no correlations between total mutations, total predicted neoantigens, or strongest-binding neoantigen and the CIRC. Therefore, it is possible that other undetermined factors may explain the differences between MSS patients. Specific immunogenic mutations may be important. Another possibility is that the cumulative strength of neoantigen binding is significant – this analysis was superficial as it only considered the single strongest binder. In addition, strong HLA-binding may not necessarily produce strong immune responses [20]. Other important factors may include tumour factors (including cell signalling, tumour antigen expression, microenvironmental factors), host factors (including germ line genetics, single nucleotide polymorphisms (SNPs), expression quantitative trait loci (eQTLs [158]), general health), and environmental factors (such as bowel flora [159, 160]). Determining how each of these factors and possibly as-yet-unknown factors

contribute to the overall immune phenotype is likely to be crucial to the appropriate targeting of existing therapies such as checkpoint blockade, as well as for the development of novel approaches that adjuvantise the microenvironment. As discussed, checkpoint blockade therapy is most effective in patients with high mutation load (MSI-H) [154]. This is likely to be due to the strong infiltration of exhausted T cells within these tumours, which are retained in the tumour by neoantigens presented on HLA molecules. Therefore, measures that can alter the immune environment to induce immune infiltration may improve efficacy of checkpoint blockade.

Interestingly, approximately 20% of MSS *POLE+POLD1* wild type patients had CIRC scores that fell within the top 75% of CIRC scores for MSI-H patients. These patients could be classified as 'MSS CIRC-high'. These patients, who have strong immune infiltration through an unknown mechanism, could, like MSI-H patients [154], be responsive to checkpoint blockade therapies. Potentially responsive patients could be identified by PD-L1 expression (which is a biomarker used to predict anti-PD1 mAb responses [14]), or high RNA or protein expression of various other molecules found within the CIRC, such as CD4 or Class II HLA, as my data suggest that these are upregulated in a co-ordinated fashion. This should be a priority for further clinical study, as MSS patients make up over 80% of CRC cases.

RAS mutation was significantly associated with lower CIRC expression, with over 60% of cancers in the group exhibiting very low expression (group D) being *RAS*

mutant. *KRAS* and *NRAS* mutant CRC had significantly lower levels of CD4+ T cells on multivariate analysis. Although 21.4% of cancers in group A were *RAS* mutant, over 80% of these were MSI-H. This is the first analysis to my knowledge to clearly define the immunological landscape of *RAS* mutant CRC. Ogino and colleagues examined the interaction between T cell infiltration and *KRAS*, *BRAF* and *PIK3CA* mutation status and found no significant associations with *KRAS* mutation [161], but this study was limited to the density of CD3+, CD8+, CD45RO+ and FoxP3+ cells. Morris and colleagues demonstrated that transfection of mutant *KRAS* into fibroblasts abrogated the IFN γ -mediated up-regulation of class II expression: effects on class I expression were minimal [162]. *RAS* transfection into fibroblasts inhibited proliferation and IFN γ production of alloreactive T cells, an effect mediated by loss of class II expression on target cells [163]. These data suggest that abrogation of IFN γ related signalling may be a possible mechanism for my observed paucity of class II expression and hence Th1-related gene expression in *RAS* mutant CRC. My data complement previous studies demonstrating widespread abnormalities of class I presentation in *RAS* mutant CRC [164]. Of the *RAS* subtypes, the less common *NRAS* mutant tumours had the lowest CIRC expression, potentially due to increased MEK activation. *NRAS* and *KRAS* mutant tumours had significantly higher expression of the Dry et al. MEK signature than *RAS* wild type cancers, as expected, and the expression of the MEK signature was borderline significantly higher in *NRAS* mutants compared with *KRAS* mutants, and thus served as a mutation set where any MEK driven differences would be likely most evident. I further investigated MEK signature expression and immune expression by the exact amino acid substitution that occurred. Certain *KRAS* mutations, such as G12D and G13D, have low affinity

for RAF and a fast hydrolysis rate, which leads to low levels of RAF activation [134]. G12D/G13D *KRAS* mutants had CIRC expression very similar to *RAS* wild type. The finding that *RAS*-associated immunosuppression is strongest in the non-G12D/G13D mutants, which have the highest MEK signatures, is consistent with the hypothesis that the immunosuppression is MEK-driven.

In MSI-H tumours, it is likely that any inhibitory effect of *RAS* mutation is overcome by the strong immunity resulting from neoantigens linked to mutations induced by microsatellite instability. Thus MSI-H *RAS* mutant tumours retain strong Th1 immunity, despite having numerically lower *HLA-DRA* expression than MSI-H *RAS* wildtype tumours. In sum, these data suggest that the micro-environment of MSS *RAS* mutant CRC is relatively immunologically unfavourable to conventional $\alpha\beta$ T cell responses.

The identification of recurrent and commonly presented strong-binding neoantigens derived from mutant *RAS* raises the possibility of utilising these antigens in vaccination approaches. Currently, *RAS*-associated neoantigens are frequently amongst the strongest identified, and are being targeted in personalised vaccine approaches [16]. However, the frequency of *RAS*-associated neoantigens in certain mutation groups, such as the G12V group, raises the possibility of stratified vaccination for patients with particular mutations. This is discussed further in the overall discussion.

The CIRC molecular associations relate to the CRCSC consensus molecular subtypes and their immune and molecular characteristics [136]. The majority of CMS1 patients (MSI high and immune activated) fall within CIRC group A. 82.5% of *KRAS* mutated patients were in CIRC groups D and B (the lowest two CIRC expression groups) and almost the same percentage of CMS3 patients, characterised by *KRAS* mutations and no significant immune infiltration and activation, were present in CIRC groups D and B. The relationship between CMS, *RAS* mutation and immunity is explored further in the next chapter.

The development of the CIRC provides a useful tool with which to explore the immunobiology of different cancers. As an example, I investigated CIRC expression in the TCGA head and neck squamous cell cancer dataset, in a study exploring the relationship between immunity and hypoxia. This revealed an inverse correlation between immunity and hypoxia signatures, and the combination of these two factors had a significant impact on patient survival. Those patients with high hypoxia and low immunity had significantly poorer survival than those with low hypoxia and high immunity. Patients with both low hypoxia and low immunity had intermediate survival. This work is now being prospectively validated, and has potential implications for the targeting of hypoxia modulators and immunotherapies to specific patient groups. Hypoxia and immunity could conceivably be combined into a new scoring system (for example, a 'hypoximunoscore'), which could potentially be prognostic and predictive in head and neck cancer.

Chapter 5

The impact of *RAS* mutation on the immune phenotype of colorectal cancer

5.1 Introduction

The *RAS* genes were the first identified oncogenes [165, 166], and encode a family of tyrosine kinases that includes KRAS, NRAS and HRAS. RAS is downstream of EGFR in the EGFR-RAS-RAF-MEK-ERK pathway, and its activation can mediate both RAF phosphorylation and PIK3CA phosphorylation [134]. RAS is therefore capable of activating two key oncogenic signalling pathways. After *RAS* mutation, which occurs in approximately 40% of CRC [91, 119], the RAS protein becomes constitutively active.

Patients with *RAS* mutant CRC have poorer outcomes than those with wild type *RAS*, and receive no benefit from EGFR inhibitors [167]. The outcome of *RAS* mutant CRC matches that of advanced non-small cell lung cancer [10, 168]. Although *KRAS* mutant tumours are critically dependent on intact MEK pathway signalling, clinical efforts to target this pathway either with or without parallel pathway inactivation of PI3K/AKT/mTOR signalling, through which mutant RAS can also signal [134], have been disappointing [169, 170]. Therefore, *RAS* mutant patients are a key group of unmet need in CRC. In addition to CRC, *RAS* mutation occurs in a range of common cancer types, with some, including pancreatic cancer, being predominantly *RAS* mutant [134].

Following my *in silico* transcriptomic analyses in the TCGA CRC dataset that revealed a link between *RAS* mutation and low expression of immunity related genes, I wanted to investigate the relationship between *RAS* and immunity further in

a separate, independent local sample. The TCGA analyses were based on transcriptomics, and, though RNA expression and protein expression are correlated, this association is not always linear or reliable. The correlation depends on multiple factors which vary according to the specific gene, including mRNA transcription and degradation, and protein translation and degradation [171, 172]. Therefore, for the analysis of the independent local sample set, in addition to transcriptomics, I wanted to investigate protein expression. Protein expression is not only more functionally relevant than RNA, but its quantification through immunohistochemistry is more clinically aligned, and allows the assessment of the tumour microenvironment.

In addition, I wanted to investigate how *RAS* mutation affects immunity within the key molecular subgroups of CRC. The recently published CMS analysis, introduced in the previous chapter, demonstrated that the classification of CRC could be significantly refined beyond one based on common single mutations [136]. The authors showed that *RAS* mutations are heterogeneous, being observed across all CMS groupings to varying extents and with these groupings displaying very different underlying biology. Secondly, although under-representation of a Th1 immune signature is observed in CMS3, which is the *RAS*-enriched sub-type, low Th1 signature expression is also seen in CMS2, characterised by high WNT and MYC pathway signalling and a lower *RAS* mutation rate. Indeed CMS2 is globally immunologically impoverished, with under-expression of multiple signatures of innate and adaptive immunity and to a much greater extent than CMS3. In this chapter, I show that the immunological impact of *RAS* mutation is dependent upon the transcriptional context of the tumour.

5.2 Results

5.2.1 Cohort characteristics

To determine the required sample size, a power calculation was performed using Altman's nomogram. This confirmed that a total of 100 samples (50 *RAS* mutant, 50 *RAS* wildtype) were required for a power of 80%, a significance level (α) of 0.05 and a minimum detectable standardised difference between the two groups of 0.55. This power provided an 80% probability of detecting a difference between the *RAS* mutant and wildtype groups if one existed. The significance level represented the threshold below which the null hypothesis (that there was no difference between the *RAS* mutant and wildtype groups) would be rejected, in this case corresponding to a p value <0.05.

The specimens were selected to represent the range of *RAS* mutations observed in the original TCGA microarray data set. The final cohort comprised 28 *RAS* G12D/G13D mutants (24.3%), 38 *RAS* non-G12D/G13D mutants (33.0%), and 49 *RAS* wild types (42.6%). Therefore, there were 66 *RAS* mutants and 49 *RAS* wildtypes (total = 115), which closely matched the requirements determined by the power calculation.

As microsatellite instability is associated with high immune infiltration [5], the microsatellite status for each tumour was confirmed by extracting total DNA from formalin-fixed paraffin-embedded (FFPE) tumour scrolls and performing fragment analysis (details in methods). This confirmed that 7 tumours (6.09%) were MSI-H. Of these 7 tumours, 3 were *RAS* mutant (2 G12D/G13D MT, 1 non-G12D/G13D mutant). Fragment analysis failed in 12.2% of cases due to low DNA yields.

5.2.2 Immunohistochemistry

To determine the relationship between *RAS* mutation status and protein expression of key markers of immune infiltration, IHC was performed targeting CD4, CD8, Tbet, Class II HLA, STAT1, PD-L1 and CXCL10 on the cohort of 115 primary CRC specimens. Clinically validated (IVD-CE) antibodies were chosen where available. Otherwise, knockdown/knockout validated antibodies were chosen. If these were not available, in-house antibody validation was performed (details in methods). Stained slides were digitally scanned, and were analysed using Definiens Tissue Studio software. Tumour regions were segmented into tumour epithelium and stroma regions during analysis, and percentages of cells or pixels with high, medium, low or no immunoreactivity were quantified in both regions. This produced either histological scores for cell-based scoring (Figure 5.1), which is a function of the number and intensity of immunoreactive cells, or percental scores for pixel-based scoring (Figure 5.2), which is a function of the number and intensity of positive pixels in the scanned specimen. The results for each slide were grouped by *RAS* mutant and *RAS* wild type, as well as by *RAS* mutant subtype. H-scores were compared for *RAS* mutant and *RAS* wild type across the entire sample set. To align the results with previous studies [153, 173], for PD-L1, STAT1 and Class II HLA, each patient was additionally categorised into expression groups according to the percentage of cells/pixels that were positive for the marker in each region. Finally, I calculated the number of Tbet positive cells in each group normalised by area (mm²), to enable comparisons with existing literature [174]. A separate correlation of immune markers in paired biopsy and resection samples was also performed (Appendix).

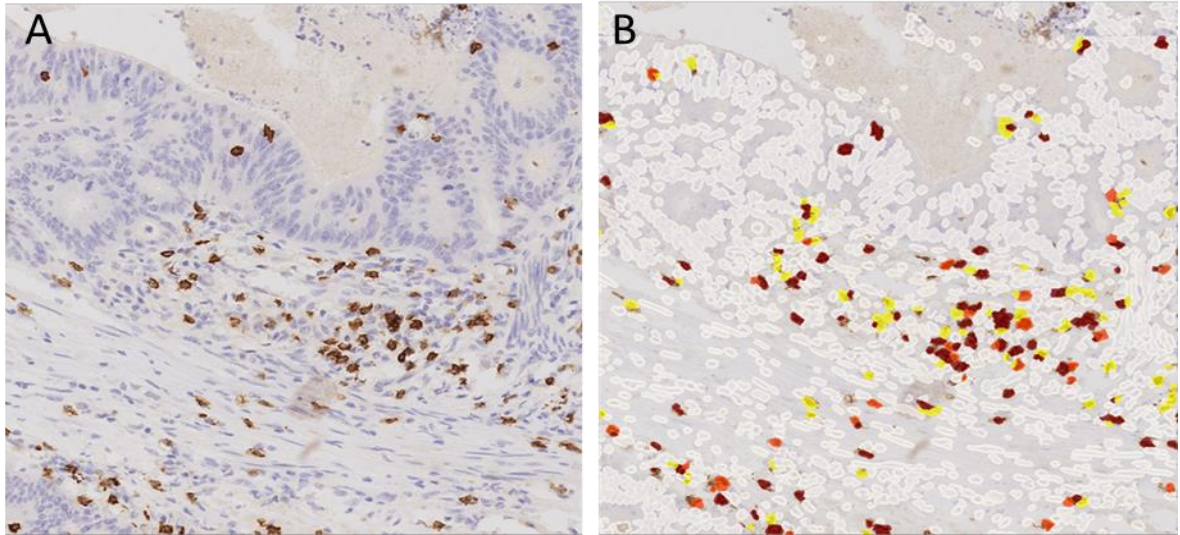


Figure 5.1 Cell-based IHC scoring algorithm.

Identification of CD8 immunoreactive cells using trained Definiens Tissue Studio cell-based scoring algorithm.

- a) Tumour region of a CD8-stained CRC slide (20x magnification)
- b) Image A, after analysis with cell identification and scoring algorithm. In this image, white cells are negative, yellow cells are lightly stained, orange cells are moderately stained, and dark red cells are heavily stained. This image is not segmented by regions.

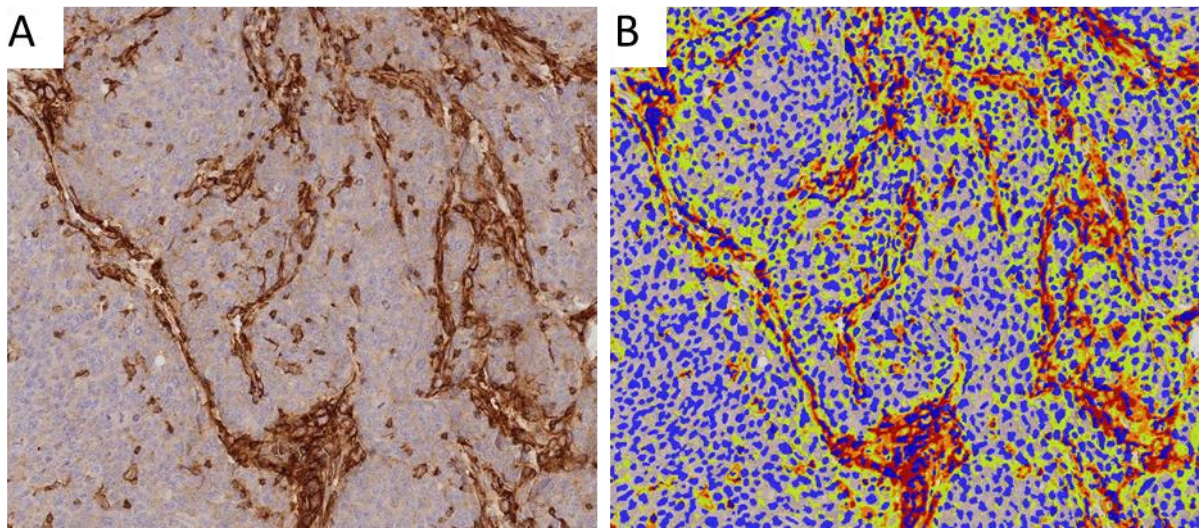


Figure 5.2 Pixel-based IHC scoring algorithm.

Identification of Class II HLA positive cells and pixels using trained Definiens Tissue Studio pixel-based algorithm.

- a) Tumour region of a Class II HLA-stained CRC slide (20x magnification)
- b) Image A, after analysis with cell identification and pixel scoring algorithm. In this image, blue represents cell nuclei. Yellow areas represent pixels with light staining, orange areas represent pixels with moderate staining and dark red areas represent pixels with heavy staining. This image is not segmented by regions.

In the epithelial compartment, STAT1 protein expression was significantly lower in *RAS* mutant cancer cells compared with their wild type counterparts thus recapitulating at the protein level the findings at the transcriptional level in the TCGA dataset (Table 5.1). This was the case whether samples were analysed by H-scores ($p=0.016$ in the epithelial compartment) or by analysing samples according to percentage of cells positive for STAT1 (Chi squared $p=0.033$ in the epithelial compartment). However, there were no significant differences in Tbet positive cells, CD4 positive cells, CD8 positive cells, PD-L1 reactivity or CXCL10 reactivity in either the epithelial or stromal compartments (Table 5.1 and 5.2). In the stromal compartment, class II HLA reactivity was borderline higher ($p=0.051$) in the *RAS* mutant group, contrary to my previous *in silico* findings. Just over 50% of both *RAS* mutant and *RAS* wild type tumours were negative for class II HLA and only 6.35% *RAS* mutant tumours had >50% class II HLA positive cells. Strikingly, more than 95% of both *RAS* mutant and *RAS* wild type cancers were cancer cell PD-L1 negative although stromal PD-L1 positivity was seen in greater than 50% of both sub-types. The median number of Tbet positive cells in the epithelial compartment of *RAS* mutant cancers was 146/mm² compared with 163/mm² in *RAS* wild type cancers ($p=0.46$) and in the stromal compartment 96/mm² and 104/mm² respectively ($p=0.98$). Figures 5.3 to 5.9 demonstrate staining in the tumours with the lowest, median and highest reactivity for each marker in the *RAS* mutant and wildtype groups (in stromal regions for CD8, CD4 and Tbet, and in epithelial regions for CXCL10, PD-L1, Class II HLA and STAT1).

No significant differences were observed in protein expression of any of the markers between MSI-H and MSS cases, likely reflecting the small number of MSI-H samples. In addition no significant differences were seen between G12D/G13D *RAS* mutant cases and non-G12D/G13D *RAS* mutant cases in terms of IHC scoring, though this study was not sufficiently powered for *RAS* subgroup analysis.

	RAS mutant	RAS wild type	P-value
CD8	3	6	0.633
CD4	1	0	0.462
Tbet	2	3	0.308
STAT1	180	238	0.016
PD-L1	0	0	0.960
CXCL10	270	224	0.175
Class II HLA	125.16	136.79	0.260

Table 5.1 Median Histological scores or Percental scores in epithelial regions. Class II HLA reactivity is represented by percental scores. All other markers are represented by histological scores.

	RAS mutant	RAS wild type	p-value
CD8	10	11	0.986
CD4	5	6	0.984
Tbet	5	5	0.897
STAT1	88	122	0.086
PD-L1	1	1	0.741
CXCL10	70	55	0.290
Class II HLA	143.87	135.82	0.051

Table 5.2 Median Histological scores or Percental scores in stromal regions. Class II HLA reactivity is represented by percental scores. All other markers are represented by histological scores.

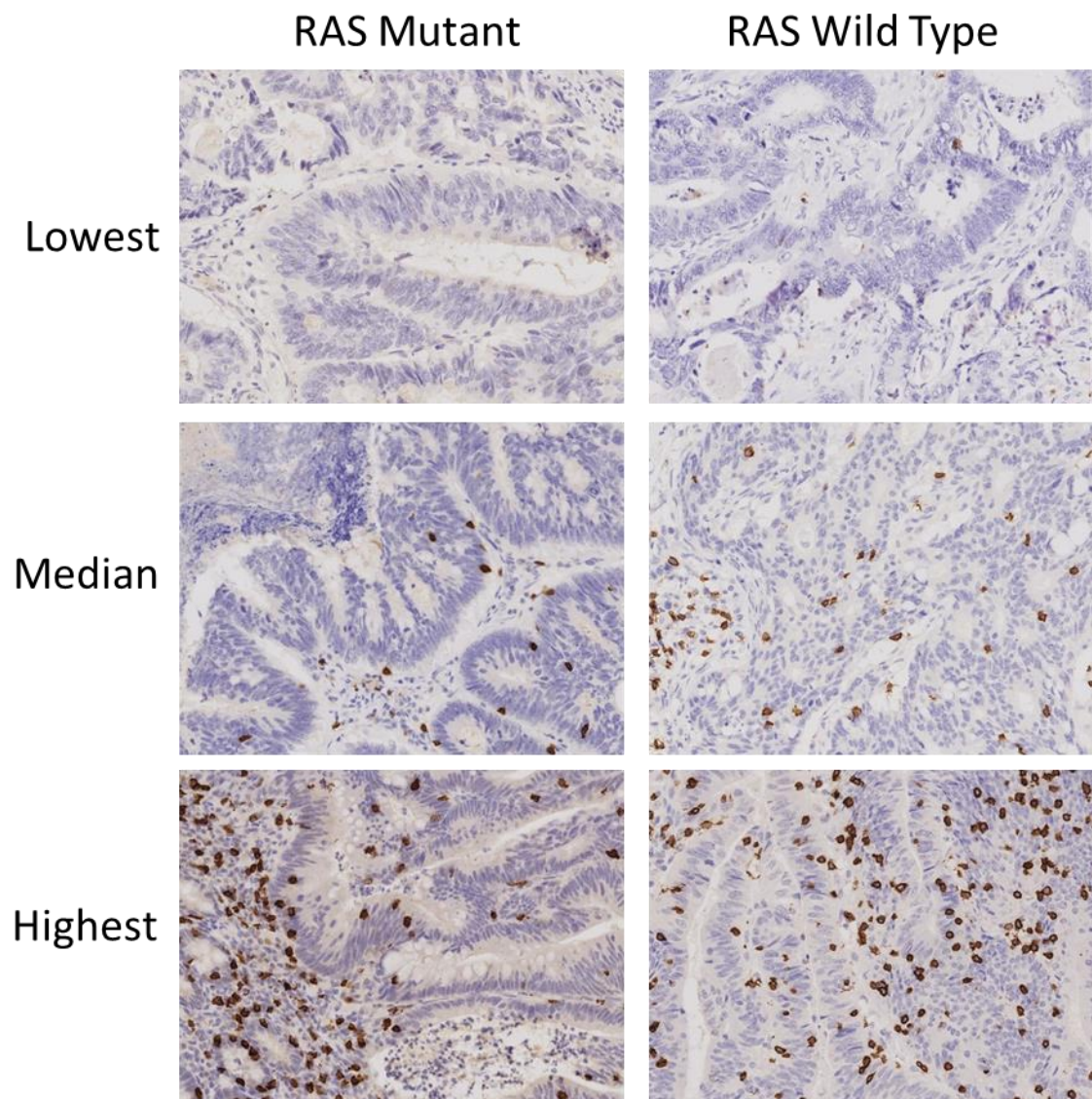


Figure 5.3 Stromal CD8 staining. Representative images (20x magnification) of the tumours with the lowest, median and highest stromal CD8 reactivity in the *RAS* mutant and *RAS* wildtype groups.

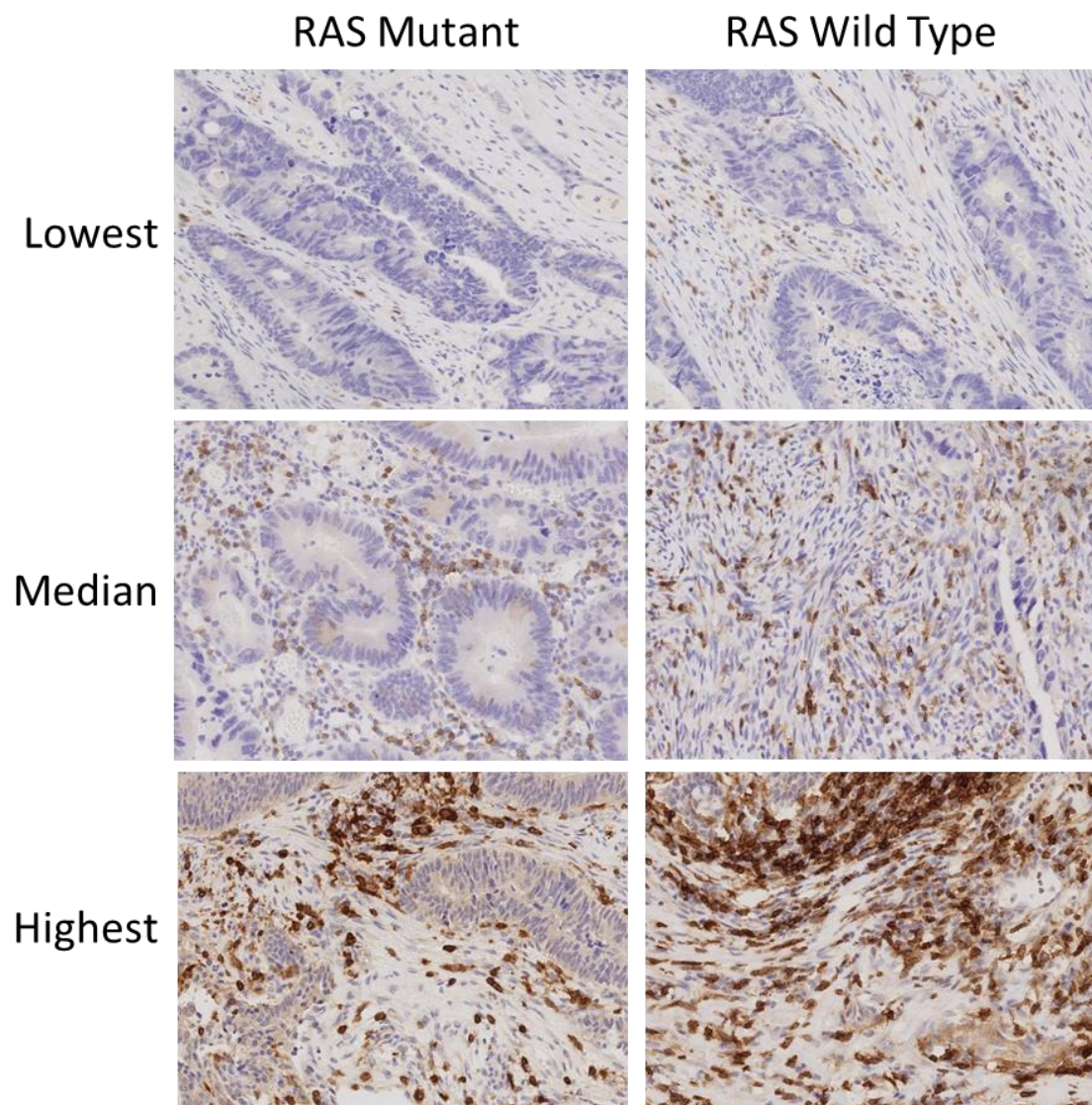


Figure 5.4 Stromal CD4 staining. Representative images (20x magnification) of the tumours with the lowest, median and highest stromal CD4 reactivity in the *RAS* mutant and *RAS* wildtype groups.

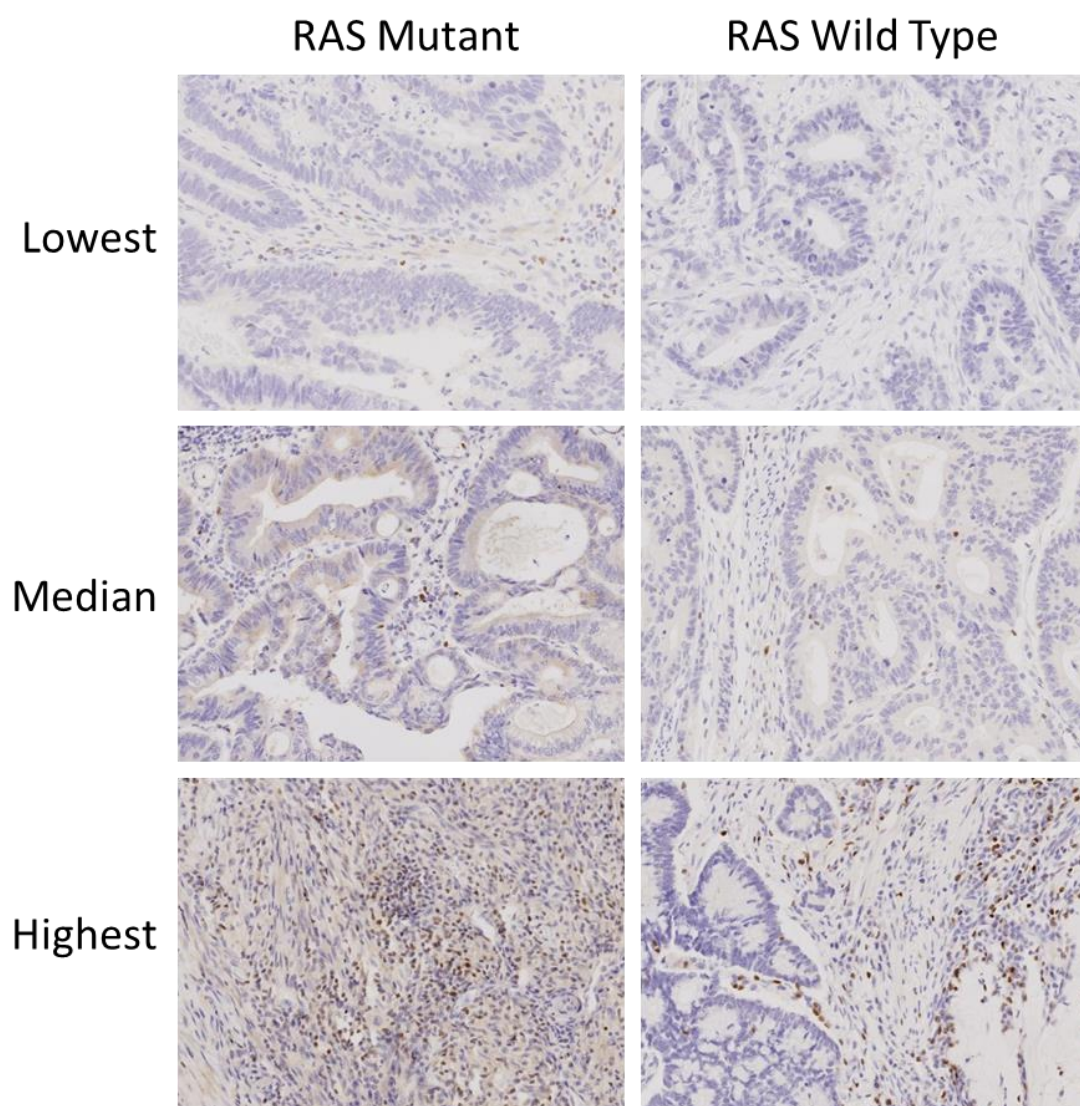


Figure 5.5 Stromal Tbet staining. Representative images (20x magnification) of the tumours with the lowest, median and highest stromal Tbet reactivity in the *RAS* mutant and *RAS* wildtype groups.

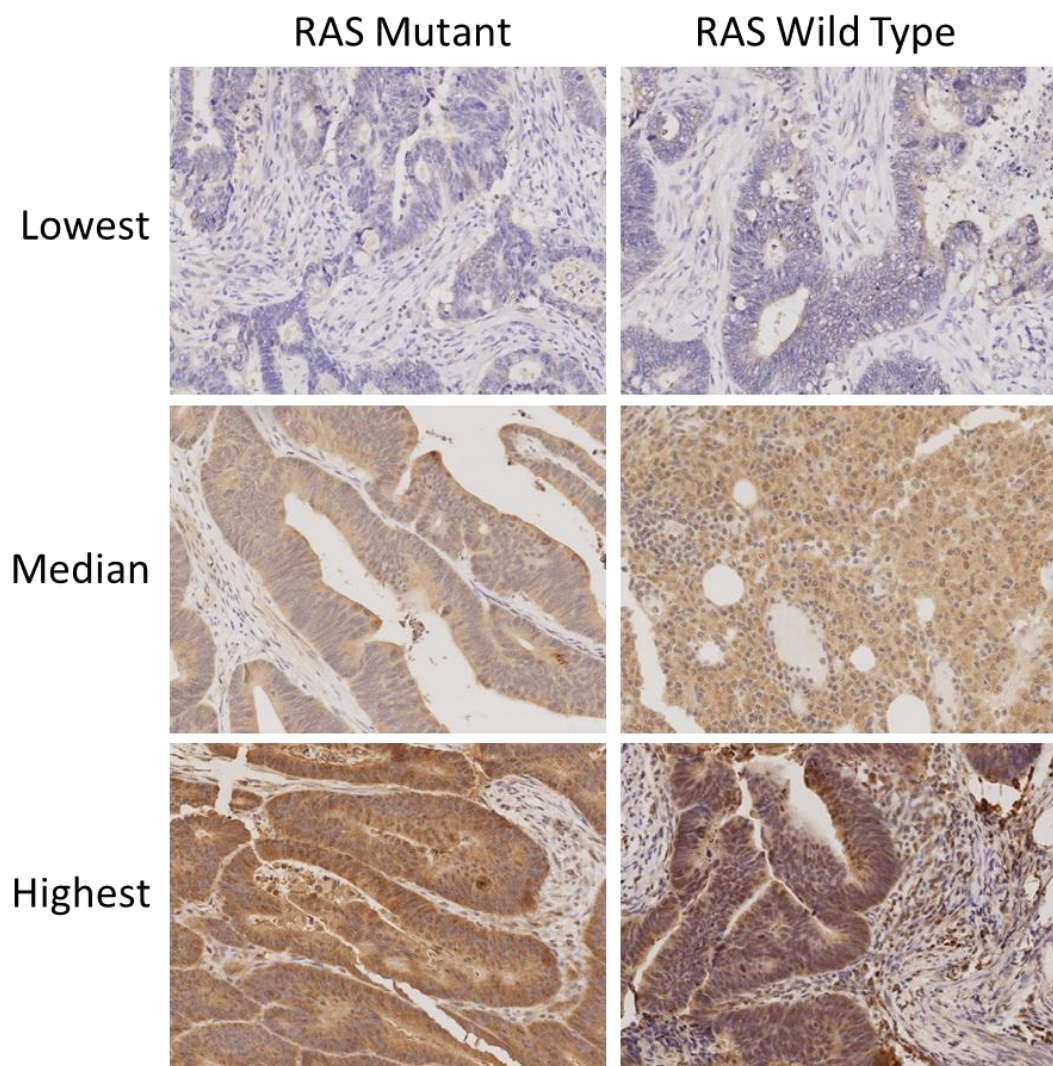


Figure 5.6 Epithelial CXCL10 staining. Representative images (20x magnification) of the tumours with the lowest, median and highest epithelial CXCL10 reactivity in the *RAS* mutant and *RAS* wildtype groups.

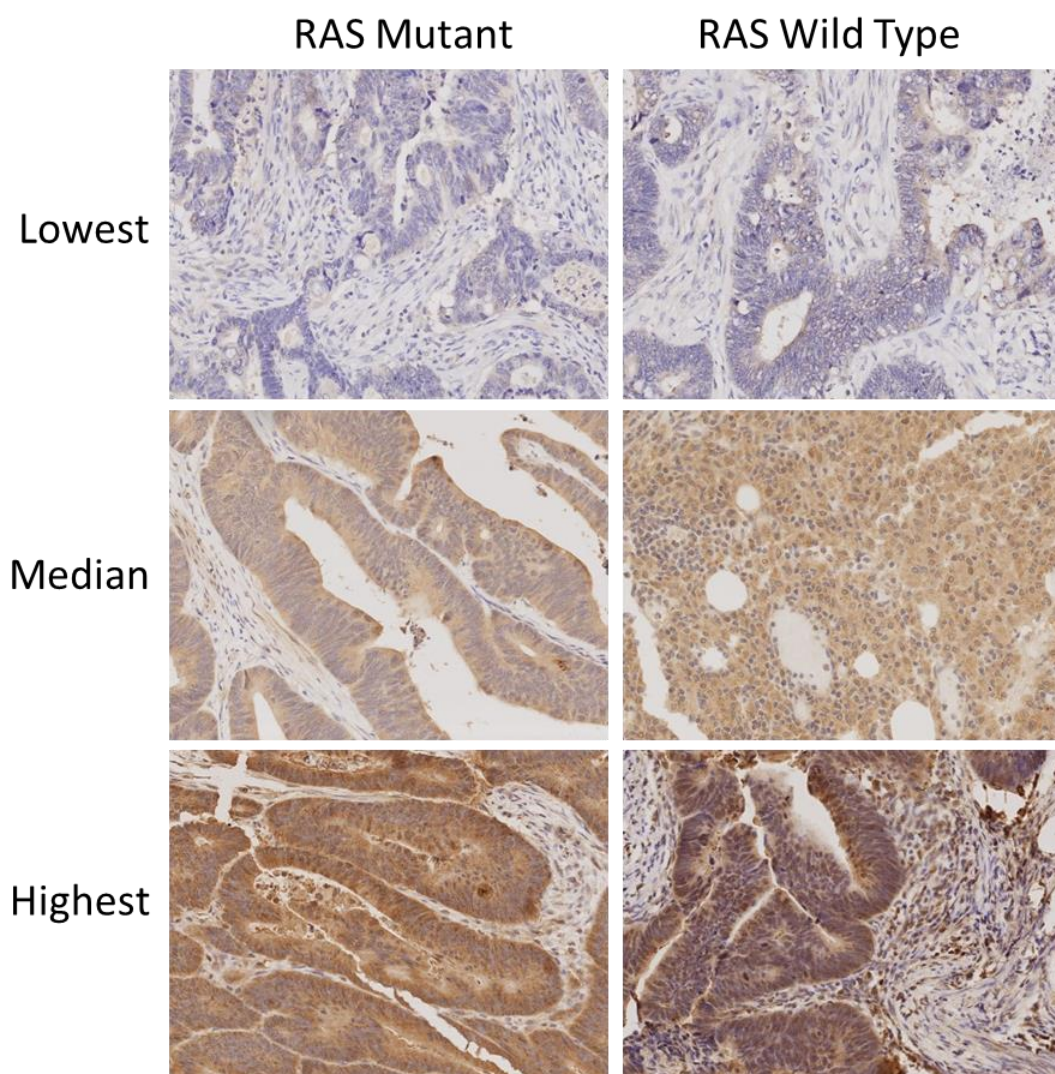


Figure 5.7 Epithelial PD-L1 staining. Representative images (20x magnification) of the tumours with the lowest, median and highest epithelial PD-L1 reactivity in the *RAS* mutant and *RAS* wildtype groups.

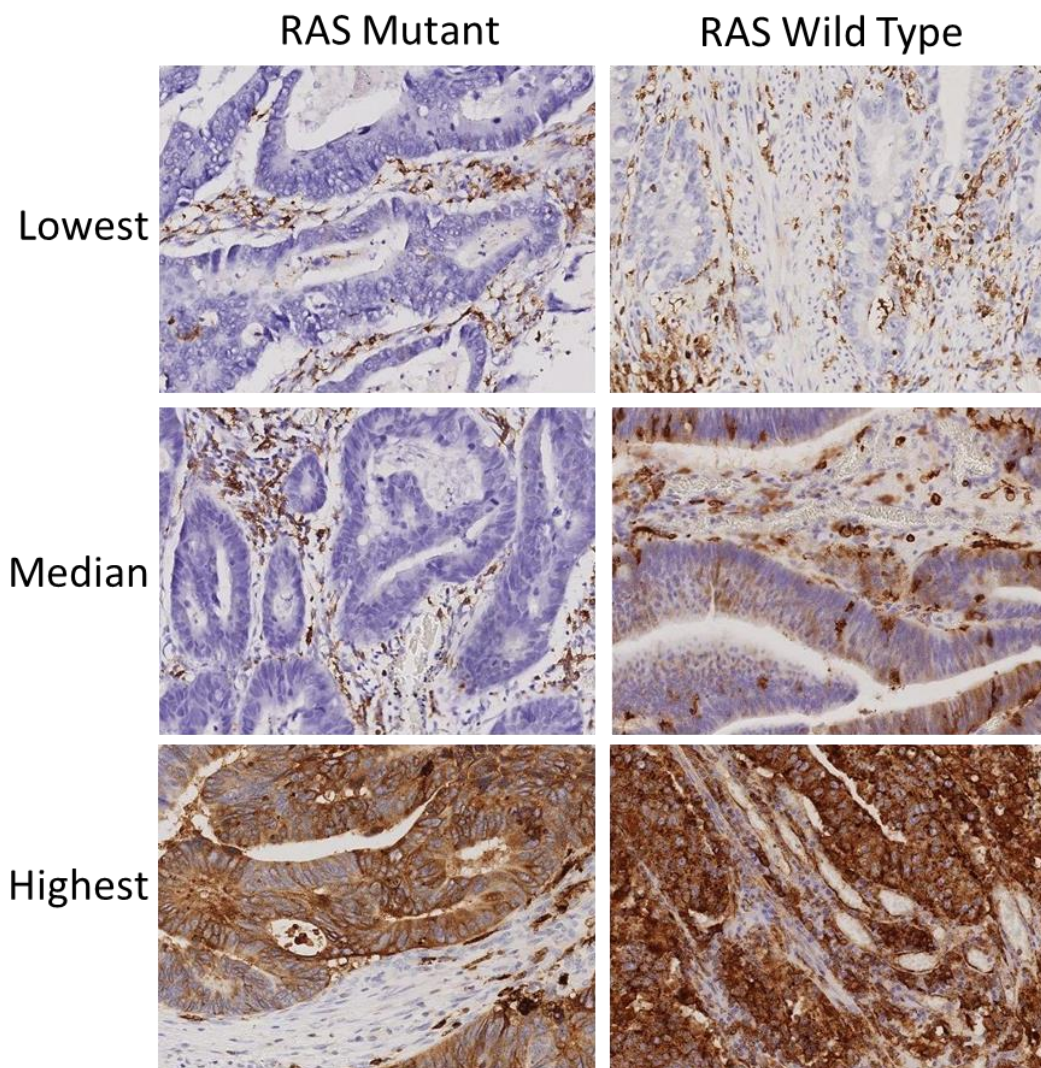


Figure 5.8 Epithelial Class II HLA staining. Representative images (20x magnification) of the tumours with the lowest, median and highest epithelial Class II HLA reactivity in the *RAS* mutant and *RAS* wildtype groups.

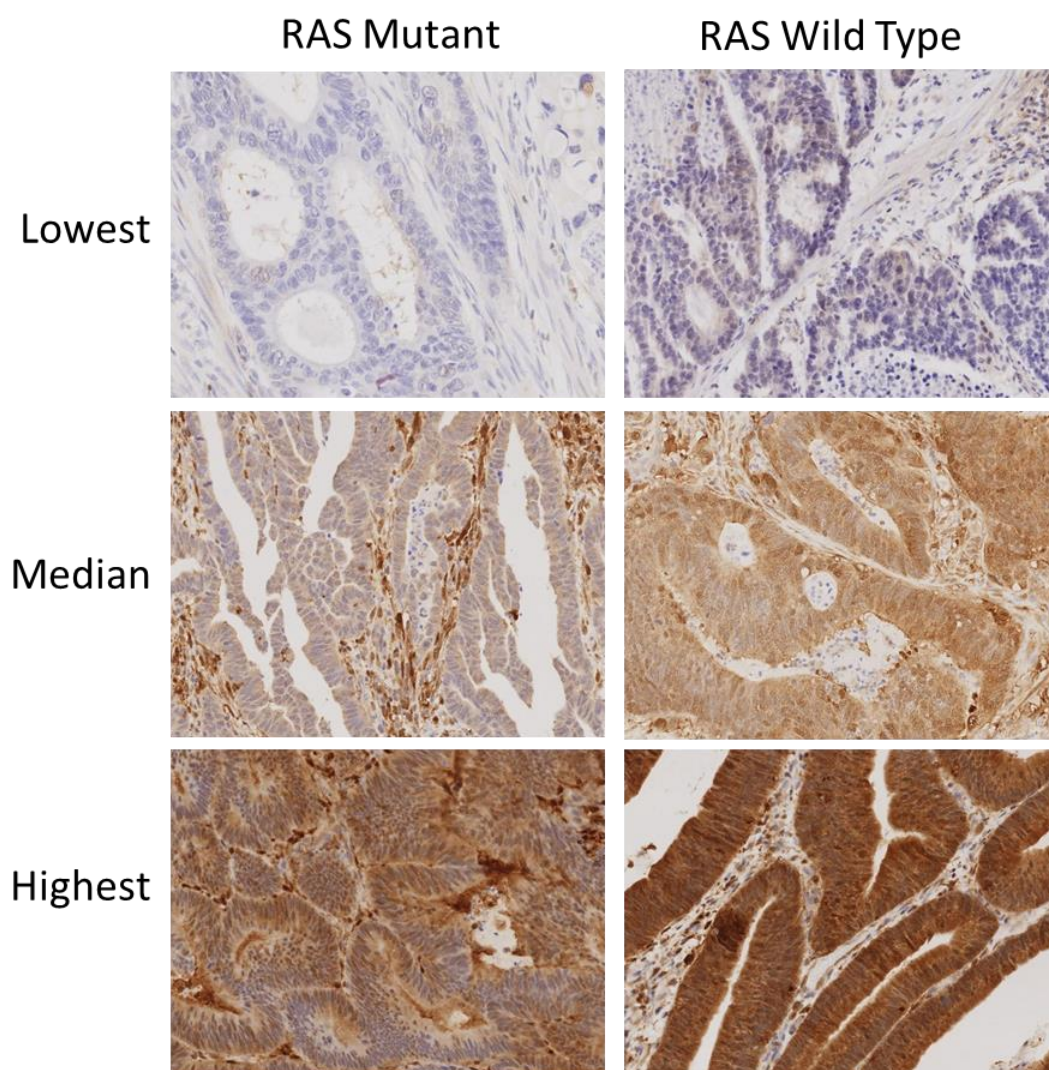


Figure 5.9 Epithelial STAT1 staining. Representative images (20x magnification) of the tumours with the lowest, median and highest epithelial STAT1 reactivity in the *RAS* mutant and *RAS* wildtype groups.

5.2.3 Targeted RNAseq panel

In addition to IHC analysis, targeted RNAseq analysis was performed using RNA extracted from FFPE scrolls from the same patient cohort. This exploratory study took advantage of RNAseq's ability to sequence relatively degraded RNA that would not be suitable for microarray-type analysis, such as the RNA extracted from archived FFPE tissue. If this technique was found to be robust, it would provide the opportunity to perform transcriptomics on samples collected retrospectively, without the need for prospective collection of fresh tissue. Targeted sequencing was performed for all genes from the CIRC signature, as well as several other key immune genes and the Dry et al. CRC-optimised MEK signature as a positive control (full list of genes in methods) [175].

On logistic regression analysis of all included genes against *RAS* mutation status, several genes were significantly differentially expressed in *RAS* mutant versus wild type tumours (Table 5.3). This analysis suggested that the expression of key genes including *CD4*, *CCL5*, *CXC10* and *PDCD1* were significantly lower in *RAS* mutant cases. However, genes including *CD8A*, *GZMH* and *CXCL11* were significantly higher in *RAS* MT cases on multivariate analysis. The CIRC metagene itself was not significantly differentially expressed. The Dry et al. CRC-optimised MEK signature was higher in *RAS* mutant tumours than *RAS* wild type tumours on univariate analysis, but surprisingly this did not reach significance (0.225 vs 0.011, $p=0.285$).

Gene	Odds ratio	Standard error	Z	P value	95% CI Lower limit	95% CI Upper limit
ALOX5AP	9.362858	6.825728	3.07	0.002	2.243179	39.07986
BNIP3	2.130329	0.626167	2.57	0.01	1.197439	3.790007
CCL5	0.331359	0.138236	-2.65	0.008	0.146284	0.750587
CD4	0.560019	0.14001	-2.32	0.02	0.343079	0.914138
CD48	0.654067	0.1225	-2.27	0.023	0.453108	0.944154
CD74	2.449633	0.701375	3.13	0.002	1.397611	4.293541
CD80	1.592431	0.361971	2.05	0.041	1.019944	2.486251
CD8A	1.853823	0.435369	2.63	0.009	1.169942	2.937461
CSF1R	0.562262	0.115678	-2.8	0.005	0.375678	0.841516
CXCL10	0.615967	0.152273	-1.96	0.05	0.379431	0.999959
CXCL11	1.663005	0.408575	2.07	0.038	1.027464	2.691664
DOCK2	0.459117	0.10747	-3.33	0.001	0.290184	0.726394
ETV4	0.571134	0.147028	-2.18	0.03	0.344835	0.945942
EVI2B	2.839844	0.805265	3.68	<0.001	1.629026	4.950636
GNLY	0.158137	0.069613	-4.19	<0.001	0.066731	0.374748
GZMH	4.898323	3.359166	2.32	0.021	1.277354	18.7838
HLAB	2.676023	1.187708	2.22	0.027	1.12124	6.38677
KANK1	3.559602	1.110757	4.07	<0.001	1.931025	6.561678
LAT2	2.231394	0.597264	3	0.003	1.320505	3.770617
LCP2	0.395805	0.101685	-3.61	<0.001	0.239222	0.65488
LGALS3	0.458636	0.124752	-2.87	0.004	0.269115	0.781627
LST1	1.667267	0.415677	2.05	0.04	1.02279	2.71784
PDCD1	0.048851	0.037932	-3.89	<0.001	0.010665	0.223773

Table 5.3 Logistic regression of *RAS* mutant versus wild type tumours using a custom targeted TRUSEQ RNAseq panel. Odds ratio>1 indicates higher expression in *RAS* mutant cases, and odds ratio<1 indicates lower expression in *RAS* mutant cases.

5.2.4 Survival analysis

Clinical data, including age, gender, ethnicity, clinical stage and overall survival were retrieved for all the patients in the clinical cohort. Cox regression analyses were performed to determine which factors associated with survival. On multivariate analysis, patients with higher stage disease had significantly poorer survival, as expected (Figure 5.10A). Furthermore, patients with *NRAS* mutant tumours had significantly poorer survival than *KRAS* or *RAS* wild type patients (Figure 5.10B). In addition, there was a clear trend towards *KRAS* mutant patients having poorer survival than *RAS* WT, consistent with the literature [176]. After 5 years post-resection, 80.7% of *RAS* wild type patients survived, compared to 56.5% of *KRAS* mutant patients. This did not reach statistical significance, most likely because this study was insufficiently powered to detect smaller changes in overall survival. In addition, when *RAS* mutation was categorised according to mutational subtype, there were no significant differences in survival between the *RAS* G12D/G13D mutant group, the non-*RAS* G12D/G13D mutant group and the *RAS* wild type groups. Furthermore, there were no significant differences between MSI-H and MSS patients in survival. Protein expression of CD4, CD8, Tbet, STAT1, Class II HLA, PD-L1 or CXCL10, in epithelium or stroma, as determined by IHC above, were not predictive for survival in this cohort on multivariate analysis.

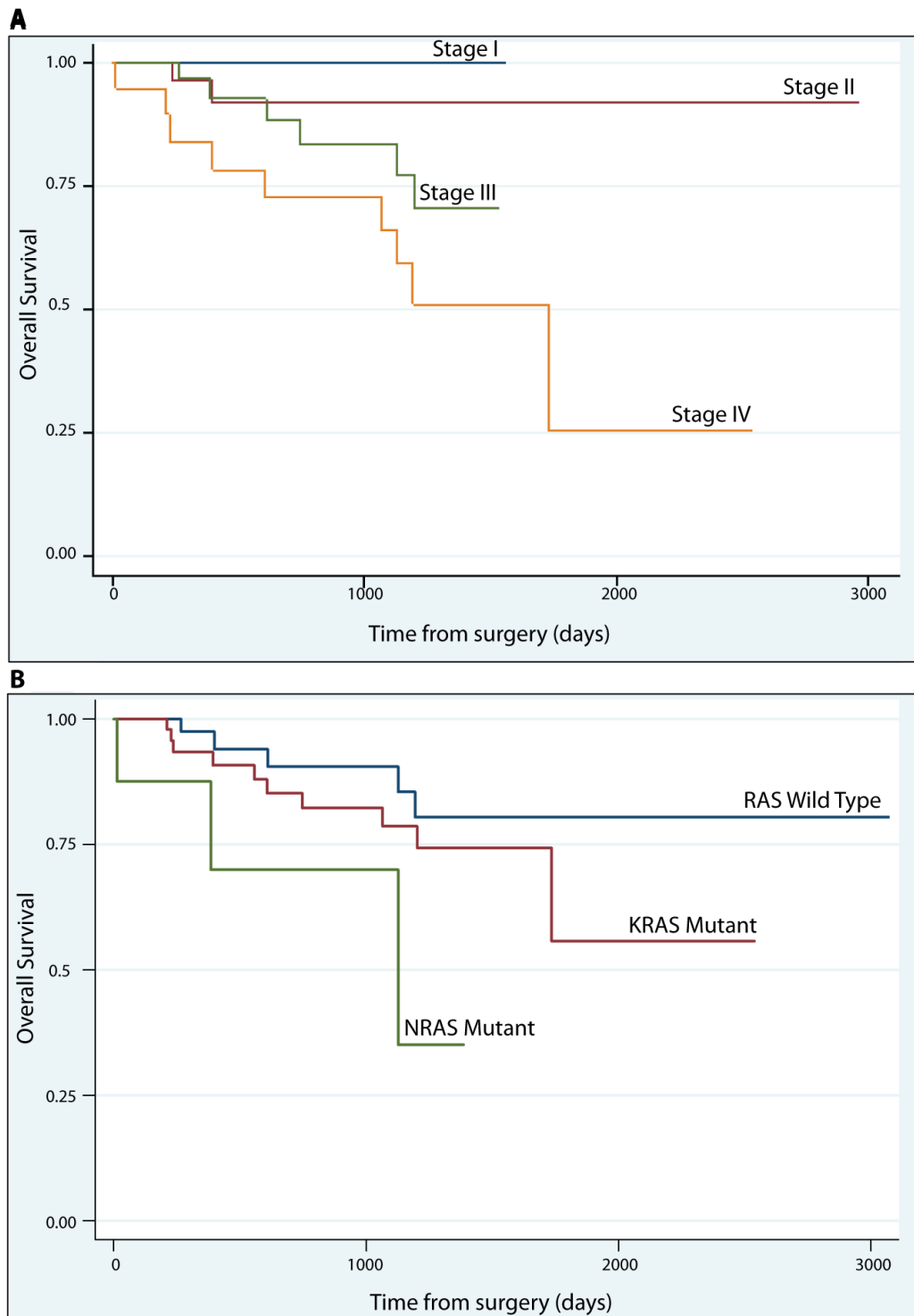


Figure 5.10 Kaplan-Meier curves for overall survival, by:
 d) Clinical Stage
 e) *RAS* mutation status

5.2.5 Cancer Cell Line Encyclopaedia analysis

To investigate whether the effect of *RAS* mutation on immunity was due to a cell autonomous effect, independent of the tumour microenvironment, I calculated expression of key immune genes in multiple CRC cell lines from the Cancer Cell Line Encyclopaedia [93]. This analysis revealed that the MEK signature was significantly higher in *RAS* G12D/G13D mutant (n=11), *RAS* non-G12D/G13D mutant (n=20) and *BRAF* mutant (n=13) cell lines versus the *RAS/RAF/PIK3CA* wildtype (n=7) cell lines ($p < 0.001$ for all). *BRAF* mutants, which are known to have the strongest MEK activation [175], had the highest MEK signature, as expected. However, with only 7 *RAS/RAF/PIK3CA* wildtype lines, this study was insufficiently powered to detect smaller effects. There were no significant differences found in expression of potential mediators of a *RAS*-associated immunological effect such as *STAT1* or *IRF1*, or class I and class II HLA (Table 5.4). *STAT1* expression was lowest in *BRAF* mutants, then non-G12D/G13D *RAS* mutants, then G12D/G13D *RAS* mutants, and was highest in wild type cell lines, as would be expected if its expression was suppressed by MEK signalling. However, this was not statistically significant. In addition, the expression of genes including *CD4* and *PDCD1* (PD1) were not significantly different in *RAS* mutant and wild type cell lines, but this was expected, as the cell line analysis excluded the tumour microenvironment, and therefore microenvironmental cells such as CD4 positive lymphocytes. However, despite the lack of statistical power, this analysis did provide indications that MEK pathway activation may have an impact on *STAT1* expression, though this requires further *in vitro* investigation to confirm causality.

	RAS G12D/G13D MT	Other RAS MT	BRAF MT, RAS WT	BRAF/RAS/ PIK3CA WT
MEK signature	0.347	0.420	0.400	0.026
CCL5	-0.001	-0.102	-0.497	-0.531
CD247	-0.363	-0.325	-0.380	-0.292
CD274	-0.318	-0.208	0.107	0.129
CD276	-0.049	-0.068	-0.056	0.450
CD3E	-0.317	-0.222	-0.102	-0.102
CD4	-0.198	-0.279	-0.314	-0.139
CD8A	-0.166	0.017	-0.131	0.054
CD8B	-0.066	0.047	-0.060	-0.166
CTLA4	-0.239	-0.150	-0.147	-0.201
CXCL10	0.168	0.069	-0.337	-0.009
CXCL9	-0.220	-0.210	-0.094	-0.329
GZMB	-0.104	0.814	0.807	0.206
HAVCR2	-0.235	-0.283	-0.211	-0.223
HLA-A	0.036	0.506	-0.199	0.333
HLA-B	0.257	0.322	-0.288	-0.056
HLA-C	0.353	0.447	-0.165	-0.053
HLA-DMA	-0.179	-0.246	-0.348	-0.446
HLA-DMB	-0.415	-0.362	-0.549	-0.546
HLA-DOA	-0.304	-0.354	-0.328	-0.328
HLA-DPA1	-0.525	-0.575	-0.503	-0.587
HLA-DQA1	-0.278	-0.297	-0.281	-0.345
HLA-DRA	-0.359	-0.473	-0.502	-0.599
IFNG	-0.254	-0.194	-0.182	-0.116
IL12B	-0.173	-0.142	-0.128	-0.127
IL12RB2	-0.076	-0.262	-0.309	-0.187
IL18RAP	-0.220	-0.105	-0.276	-0.339
IRF1	-0.152	0.039	-0.240	-0.125
LAG3	-0.183	-0.176	-0.201	-0.217
PDCD1	-0.186	0.368	0.193	0.020
PDCD1LG2	-0.490	-0.494	-0.267	0.662
STAT1	0.038	-0.105	-0.382	0.373
STAT3	-0.138	-0.272	-1.007	0.209
TBX21	-0.201	-0.304	-0.242	-0.289

Table 5.4 Cell line encyclopaedia analysis. Mean expression of key immune genes in cell lines grouped by *RAS*, *BRAF* and *PIK3CA* status.

5.2.6 Consensus Molecular Subtypes analysis

In view of the lack of significant differences in expression of key proteins including CD4 and CD8 between *RAS* mutant and wild type cancers in my IHC analysis, and the discrepancy with my previous TCGA transcriptomic analysis, I investigated possible reasons for the difference. One hypothesis was that the significance of *RAS* mutation could be related to biological context, and in particular, the transcriptional subgroup of CRC in which the mutation occurs. *RAS* mutation occurs in all of the Consensus Molecular Subtype groupings to varying degrees, even though each of them differ biologically [136]. Therefore the immunosuppressive effect of *RAS* mutation could be restricted to particular CMS groups. When I investigated the impact of *RAS* mutation within specific CMS groupings in the original 195 patient TCGA microarray dataset [91], there was a trend towards CMS2 *RAS* mutants having lower CIRC expression than CMS2 *RAS* wild types (Figure 5.11). To explore this further I collaborated with Justin Guinney, the lead author of the CMS classification.

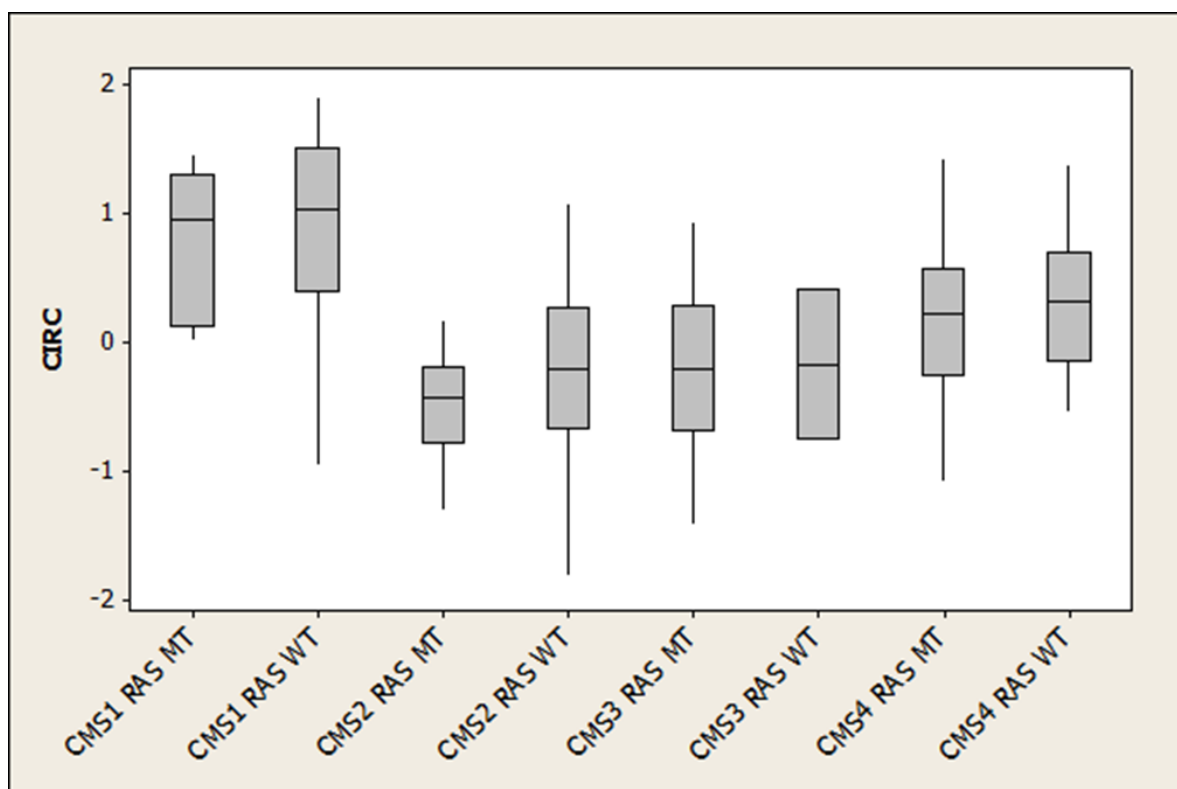


Figure 5.11 CIRC expression z-scores in each CMS grouping by *RAS* mutation status (original 195 patient TCGA microarray dataset).

The boxes represent the 25th and 75th percentile values. The lines within the boxes represent median values. The lines emerging from the boxes represent the maximum and minimum values.

The TCGA dataset was first reinvestigated. 190 of the 195 original TCGA microarray-based samples have subsequently been analysed by RNAseq and were included in an expanded TCGA dataset totalling 291 samples. We initially wanted to confirm whether *RAS* mutation was associated with lower CIRC expression in this expanded dataset. Table 5.5 shows that the expression of the CIRC was indeed significantly lower in *RAS* mutant samples than in wild types ($p=0.0117$). In a separate validation set, using transcriptional data from the Koo Foundation Sun-Yat-Sen Cancer Center dataset ((KFSYSCC), $n=289$) [177], again there was a significantly lower expression of the CIRC in *RAS* mutant cancers ($p=0.00376$). This therefore confirmed that *RAS* mutation is associated with reduced expression of the CIRC metagene in CRC. In both datasets, as expected, expression of the Dry et al. CRC MEK activation signature was significantly higher in *RAS* mutants than wildtypes.

The original CMS study demonstrated that the CMS groups clearly have different levels of the transcription of a range of immune-related genes [136]. The *RAS*-directed IHC analysis had not taken CMS sub-type into account and left open the question as to whether the observed transcriptional differences between *RAS* mutant and *RAS* wild type cancers might be particular to certain molecular sub-types. This could explain why differences were not seen in my local cohort. Thus CIRC expression of *RAS* mutant cancers was compared with *RAS* wild type cancers in each individual CMS in both the extended RNAseq TCGA and the KFSYSCC datasets (Table 5.5).

	All patients	CMS1	CMS2	CMS3	CMS4
Extended TCGA dataset					
CIRC	0.0117	0.816	0.000193	0.268	0.855
MEK	2.20E-13	0.0758	4.03E-09	0.163	0.000235
T cells	0.434	0.792	0.0156	0.563	0.570
Cytotoxic cells	0.162	0.566	0.00227	0.198	0.758
iDC	0.0993	0.699	0.0133	0.930	0.570
KFSYSCC dataset					
CIRC	0.00376	0.494	0.0256	0.696	0.0774
MEK	1.76E-17	0.614	2.59E-08	3.06E-05	1.24E-06
T cells	0.576	0.614	0.582	0.207	0.522
Cytotoxic cells	0.0202	0.243	0.206	0.171	0.241
iDC	0.460	0.157	0.907	0.665	0.141

Table 5.5 CIRC, MEK, T cell, cytotoxic cell and iDC signatures by CMS grouping. p-values of differences in expression of signatures in all patients and by CMS groups between *RAS* mutant and *RAS* wild type colorectal cancers.

Cell colour relates to whether mean expression was higher in *RAS* mutant or wild type cases. Pink cells = higher in *RAS* wild type samples, but not significant. Red cells = significantly higher (p value <0.05) in *RAS* wild type samples. Light blue cells = higher in *RAS* mutant samples, but not significant. Dark blue cells = significantly higher (p value <0.05) in *RAS* mutant samples.

In the TCGA dataset there was no difference in the expression of the CIRC between *RAS* mutant and *RAS* wild type cancers in CMS1, CMS3 and CMS4. CMS2 is the most immune suppressed MSS subtype [136]: expression of the CIRC was significantly lower in *RAS* mutant CMS2 tumours compared with wild type tumours (p=0.000193). This finding was confirmed in the KFSYSCC dataset where expression of the CIRC was lower in *RAS* MT tumours in CMS2 (p=0.0256), but there was no difference in CIRC expression in CMS1, CMS3 and CMS4.

A range of different immune transcriptional signatures, characterised by Bindea et al. [100], were also utilised in this analysis. In the TCGA dataset only, T cell, cytotoxic cell and immature dendritic cell (iDC) signatures, like the CIRC, were significantly lower in CMS2 *RAS* mutant cancers compared with CMS2 *RAS* wild type cancers (Table 5.5). This suggests that at the microenvironmental level the possible Th1 suppressive effect of *RAS* mutation may manifest itself in cancers that are already relatively immune impoverished whereas this effect may be functionally insignificant in CRCs that are more immunologically activated such as CMS4 and especially CMS1. It also suggests that the relationship between *RAS* mutation and immunity in CMS2 patients may go beyond the Th1 phenotype.

To determine the relative impacts of *RAS* mutation and CMS groupings on immunity, the most immunosuppressed MSS CMS subtype (CMS2) and the least suppressed MSS subtype (CMS4) were compared. CIRC expression was significantly lower in CMS2 *RAS* mutants compared to CMS4 *RAS* mutants in the TCGA ($p=2.76 \times 10^{-8}$) and KFSYSCC ($p=0.00042$) datasets. CIRC expression was also significantly lower in CMS2 *RAS* wild type cancers compared with CMS4 *RAS* wild type cancers in both datasets ($p=9.90 \times 10^{-7}$ (TCGA) and 4.30×10^{-6} (KFSYSCC)). Importantly, in the TCGA dataset, CIRC expression was significantly lower in CMS2 *RAS* wild type cancers compared with CMS4 *RAS* mutant cancers ($p=0.000588$) in spite of the fact that, as expected, the Dry et al. 5-gene MEK signature was significantly higher in the CMS4 *RAS* mutants ($p=1.90 \times 10^{-5}$). Finally, in the TCGA dataset, expression of the CIRC was significantly lower in CMS2 *RAS* mutants ($p=0.00832$) compared to *RAS* mutant CMS3 tumours, highlighting that CMS2 is the most immunosuppressed group. Thus,

the CMS status of the patient appears to govern the immunological impact of *RAS* mutation status.

5.3 Discussion

The work in this chapter focussed on the impact of *RAS* mutation on immunity. *RAS* mutant patients form a critical group of unmet need in CRC as well as in a range of other cancer types. Therefore it was important to confirm whether the *RAS*-associated immunosuppressive effect observed in the previous chapter could be replicated in an independent dataset and at a protein level. The TCGA, though large and comprehensive, with 195 patients in the published CRC dataset [91], is nevertheless only a single dataset and it was possible that the *RAS* correlation with immunity was a cohort-specific finding.

A range of experimental approaches were used to attempt to validate this finding. Firstly, digital IHC approaches were used to investigate the effects of *RAS* mutation at a protein level in a local, independent sample set, using digital pathology software. In parallel, exploratory RNAseq was performed from RNA extracted from FFPE tissue. Next, to supplement the *in silico* analyses previously performed, and to investigate the cell autonomous effect of *RAS* mutation, the cell line encyclopaedia was interrogated. Finally, in view of the findings from the above experiments, additional *in silico* analyses were performed on the expanded RNAseq TCGA dataset and the KFSYSCC dataset to investigate the impact of CMS groupings and *RAS* mutation status together on immunity. This variety of approaches proved invaluable in investigating the hypothesis (that *RAS* mutation is associated with immunosuppression) and took advantage of the strengths of different techniques. The advantage of publically available datasets are in accessibility, the speed of analysis, and the data-rich nature of the sample set, with a quantity of data that could

not be easily replicated locally. However, the functional and clinical importance of changes in DNA and RNA depend on the effect on protein expression and structure, as well as changes to the tumour microenvironment. IHC, a clinically aligned approach, is best placed to investigate protein expression and the microenvironment.

The IHC studies analysed the impact of *RAS* mutation on the protein expression of key molecules represented in the CIRC using digital pathology. Digital pathology is emerging as a key technique in the research setting, and attempts are being made to validate and standardise the approach for clinical usage [38]. The advantage that digital pathology provided for this study was that, with pathologist input and validation, I was able to create computer-based learning algorithms that could objectively analyse whole CRC sections, producing data on a per region and cell basis. This approach eliminates the subjective variation that occurs in the scoring of different specimens. However, it cannot completely eliminate user-dependent variation, as the segmentation and setting of scoring thresholds is partly subjective. However, this was minimised by close pathologist collaboration. In addition, any small scoring errors would have been spread equally between the two groups (*RAS* mutant and wildtype), and therefore the effect on the inter-group analysis would have been minimal.

In addition to careful creation of the scoring algorithms, particular efforts were made to ensure that the antibodies used were validated. Antibodies that do not target the correct antigen or that are non-specific are a particular problem in laboratory and clinical studies, leading to a lack of reproducibility of published findings [178]. Therefore, clinically validated antibodies (conforming to *In Vitro* Diagnostics

Conformité Européenne (IVD-CE) standards) were chosen where possible (CD4, CD8). These antibodies go through robust validation before they are approved for clinical use [178]. If these were not available, antibodies that were carefully validated by the vendor, using techniques including gene knockout/knockdown and peptide-blocking, were chosen (Tbet, STAT1, PD-L1). Where these were not available, in-house *in vitro* validation was performed (CXCL10, Class II HLA). This, alongside careful optimisation, ensured robust and reliable staining. In addition, the use of an autostainer and digital analysis reduced the likelihood of inter-batch variability.

The IHC analysis revealed that there were no differences in the expression of class II or PD-L1 between *RAS* mutant and *RAS* wild type cancers in spite of significantly reduced protein expression of STAT1. *CIITA* is the master transcriptional activator of the class II molecules and along with PD-L1 is a STAT1 target gene. Half of all CRC tumours analysed were class II negative in the epithelial component in agreement with previous data [179] and I found that this negative fraction was very similar in *RAS* mutant and *RAS* wild type cancers. Membranous PD-L1 expression was very low again in agreement with recent data [180], using the same antibody as used herein which critically only stains membranous (and hence biologically relevant) PD-L1 [181]. This suggests that CRC at the protein level is generally significantly immunosuppressed independently of *RAS* mutation status. As I was not aware of the possible importance of CMS status on the effect of *RAS* mutation at this stage, I did not perform IHC specifically in CMS2 samples. This would have required full transcriptomics to accurately classify the CMS grouping of each case, but could have

enabled investigation of the CMS-RAS interaction at a protein level. This will be crucial future work.

The targeted RNAseq for transcriptomics from FFPE tissue was a novel approach, and was used as it enabled transcriptomic analysis using degraded RNA from archived FFPE tissue. This work was performed to both investigate this method for transcriptomic analysis, to determine whether it is a robust and feasible approach warranting further evaluation, and to ascertain whether the transcriptomic results observed in the TCGA could be replicated locally, albeit with a different experimental approach (in the TCGA analysis, microarray and RNAseq were performed from fresh-frozen tissue derived RNA). This approach makes the presumption that degraded RNA in FFPE samples is comparable between cases, despite possible differences in the lengths of time samples have been stored prior to RNA extraction. Ideally, a more established method would have been used, such as full transcriptomics (total RNAseq) from fresh resection tissue. However, this approach would take a considerable amount of time to accrue the necessary number of samples. In view of the additional transcriptomics analysis subsequently performed on the extended TCGA and KFSYSCC datasets, this analysis may not have been necessary, as the transcriptomic finding from the previous chapter was validated on the second dataset. The key local analysis was on the protein IHC level. The findings of the targeted panel were not consistent with the TCGA dataset, as univariate analysis failed to show expression differences between key immune genes in *RAS* mutant versus wildtype cases. However, multivariate logistic regression analysis did reveal changes in key genes such as *CD4*. However, though useful to discern biological

interactions, the multivariate analysis in this case does not represent the clinical scenario where absolute differences in gene expression between groups will likely be more meaningful. Regardless of this, the finding that the Dry et al. MEK signature did not differ significantly between *RAS* mutant and *RAS* wildtype cases (though there was a trend in the expected direction) suggests that the data and technique require further evaluation, as this signature has proven robust and has been validated in multiple datasets [175], and by us in the TCGA and KFSYSCC datasets. This lack of significance may have been due to a small number of outliers in the analysis, or a small amount of contamination during tissue processing, rather than an inherent problem with the technique. However, for validation of the FFPE targeted RNAseq approach, firstly a comparative RNAseq analysis of fresh-frozen and FFPE tissue should be performed, to determine whether FFPE analysis aligns with the results obtained from fresh-frozen tissue. This could be performed using fresh tissue obtained at resection, and then with corresponding FFPE tissue blocks at various time points, to determine whether there is a time limit after which the RNA is too highly degraded for analysis. Additionally, established positive and negative controls are needed for the analysis, which may include tissue from tumours with known high and low immune infiltrate in this case. Finally, to ensure that the quantification is accurate, parallel quantitative real-time PCR (qPCR) analysis could be performed for selected targets from the same cDNA.

Despite the lack of significance in the RNAseq analyses, my previous TCGA transcriptomic analyses did reveal a significant association between *RAS* mutation and reduced STAT1 expression, which I have now also observed at the protein level.

There are several potential mechanisms through which *RAS* mutation could potentially suppress immunity, including direct tumour cell mediated effects and indirect effects. In terms of direct effects, a reduction in STAT1 expression in *RAS* mutant tumours may lead to a reduction in IFN γ signalling. CRC cell lines bearing activating mutant *KRAS* have been found to downregulate levels of numerous IFN γ responsive genes, and studies in isogenic HCT116 cell lines have shown that abrogation of *KRAS* signalling pathways elevate basal and IFN γ stimulated expression of IFN-responsive genes, including class II MHC [182]. In addition, such approaches have suggested that mutant *RAS* signalling decreases basal and IFN γ induced expression of STAT1, the critical Th1-associated transcription factor [182], and negatively influences IFN γ induced activation of IRF1 [182, 183], which plays a central role in activating STAT1 expression [184]. Both *IRF1* and *STAT1* are components of the CIRC. CIITA, which is activated by STAT1, is the master transcriptional activator of class II expression, and also effects class I expression [185, 186]. Therefore *RAS* mutation may have an immunosuppressive effect by reducing IRF1 activation and STAT1 expression, which then suppresses IFN γ signalling, thereby reducing CIITA activation and therefore HLA Class I and II expression.

A potential indirect mechanism of *RAS*-associated immunosuppression is a reduction in the cytokine IL-12. IL-12, produced by dendritic cells, is a key cytokine determining Th1 development and IFN γ production by T cells [187, 188]. I have shown in the previous chapter that IL-12 expression is significantly reduced in both *NRAS* ($p=0.016$) and *KRAS* mutant CRC ($p=0.011$) compared with *RAS* wild type. In

the TCGA dataset, there was a negative correlation between IL-12 expression and expression of the Dry et al. MEK signature [175], and a positive correlation between IL-12 expression and expression of the CIRC. The mechanism by which tumour cell *RAS* mutation would decrease IL-12 expression is currently unclear, but a reduction in class II expression and therefore DC activation could be contributory factor.

IL-12 production by DCs could also be suppressed by immunosuppressive cytokines produced by *RAS* mutant tumour cells. *In vitro* expression of *RAS* G12V in HEK cells has been found to result in tumourigenicity that is dependent on production of the immunosuppressive cytokine IL-6 that acts in a paracrine fashion [189]. In addition, melanoma cells bearing mutations in *BRAF*, immediately downstream of *RAS*, have been shown to produce IL-6 [190], which could decrease maturation of, and IL-12 production by, monocytic dendritic cells [190, 191]. MEK inhibition not only decreases immunosuppressive cytokine production by *BRAF* mutant melanoma cells [190], but also affects DCs, abrogating the effects of immunosuppressive cytokines on IL-12 induction, and promoting Th1 induction [191]. Therefore there are multiple potential mechanisms for *RAS*-induced immunosuppression, which warrant further investigation in *in vitro* studies.

The limitations of cancer cell line data from the CCLE were in fact useful for this study as the effects of the microenvironment were excluded, which allowed me to assess if *RAS*-mediated autonomous effects were evident. However, though the MEK signature was associated with *RAS* mutation as expected, there were no significant differences in the expression of key genes such as *STAT1* (though there

was a trend towards lower STAT1 expression in *RAS* mutant lines), perhaps reflecting the small number of wild type cell lines in this analysis. Ideally, to determine the effect of *RAS* mutation in cell lines, local *in vitro* studies would be performed using an isogenic cell line system, with specific *RAS* mutations generated with for example CRISPR-Cas9 gene editing, to determine the effect on possible expression of key genes such as *STAT1* or *CIITA*. This would be important future work for determining whether there is a causal relationship between *RAS* mutation and immune response in CRC.

After finding that *RAS* mutation did not appear to be associated with low immunity at the protein level in the local sample set, I considered possible reasons for the discrepancy with the TCGA data, which included the established variability in the correlation between RNA and protein expression in CRC [171, 172]. However, another possibility was that the relationship between *RAS* and low immunity was restricted to particular subgroups of CRC that only comprised a subset of my cohort. The CMS classification (a consensus classification based on six independent classification systems), introduced in the previous chapter, classifies tumours by their transcriptional profile, and therefore separates them by key biological subtypes rather than by single driver mutations such as *RAS*. As the CMS classification is based upon transcriptomics, rather than genomics, it is more closely linked with tumour phenotype and clinical behaviour, and each CMS group includes tumours with a variety of different mutations that lead to the same phenotype [136]. As key mutations such as *KRAS* are distributed across all four CMS groups, this suggests

that CMS is a more accurate classifier of key biological groupings than a system that uses mutation status alone.

My previous finding that *RAS* mutation was associated with reduced transcription of the CIRC metagene was confirmed in two large bioinformatics datasets. However, in addition, it was found that the net impact of *RAS* mutation was dependent upon the CMS in which the mutation was found. CMS2 is the most immunosuppressed subtype and in this subtype alone the addition of *RAS* mutation appeared to deepen that immunosuppression at least at the level of immune-related gene transcription. As previously discussed, *RAS* mutation is associated with a reduction in STAT1 expression [182], which I have observed at RNA (in the TCGA) and protein level. There are potential reasons why a reduction of STAT1 expression in *RAS* mutant CRC would only cause a concomitant reduction in the transcription of its target genes in CMS2. The previously discussed *in vitro* data demonstrated that *RAS* mutation dampens the STAT1 response to IFN γ and the level of STAT1-driven transcription in CRC cells but does not abolish it [182]. My work has demonstrated that membranous PD-L1 expression is very low across MSS CRC suggesting that MSS CRC in general lacks sufficient immune reactivity to elicit adaptive PD-L1 up-regulation. It is possible that in immune impoverished environments (such as CMS2) *RAS* mutation's effect on STAT1 may be a limiting factor on immunity, whereas in more enriched environments (such as CMS4), *RAS* mutation is not a limiting factor. The analysis showed that the level of STAT1 expression is significantly lower in CMS2 wild type tumours compared with CMS4 wild type tumours in the TCGA. The addition of a possible *RAS* mutation-mediated STAT1 decrease in CMS2 cancers is

likely to have a greater impact on the downstream transcriptional network than that same change in tumours where STAT1 levels are less close to the lower limits of the functional range. Furthermore, in somewhat less immune impoverished tumours such as those in CMS4, there may be at least sufficient microenvironmental IFN γ production to partially abrogate the decrease in STAT1 expression and hence STAT1-driven transcription due to *RAS* mutation.

Thus, prediction of the effect of *RAS* mutation on immune-related microenvironmental gene transcription may be impossible without knowledge of the CMS group in which that mutation is found. Therefore IHC is of limited relevance unless the CMS group can be assigned, which is not currently feasible in the clinical setting, as it requires full transcriptomic analysis. Manipulation of the specific pathways activated by *RAS* mutation in order to up-regulate immune gene transcription may mainly be of therapeutic relevance in CMS2. However, this alone is unlikely to significantly immune activate these cancers. CMS2 *RAS* wild type tumours have lower expression of the CIRC than CMS4 *RAS* mutant tumours. Hence, attention to the underlying biological effect limiting the immune infiltration in each particular CMS group is required.

In view of the potential CMS2-specific effect of *RAS* mutation, it would be interesting to consider the biology of CMS2. This CMS group, termed the 'canonical' CRC subgroup as it is characterised by high activation of the classical CRC-initiating Wnt and MYC pathways, includes 37% of all CRC patients. It includes a substantial proportion of the microsatellite stable group of patients, who I have previously shown

have almost universally low mutation rates, and is associated with a high number of somatic copy number alterations (SCNAs) due to chromosomal instability (CIN) [136]. There is evidence that Wnt signalling is immunosuppressive. Activation of WNT/ β -catenin signalling in melanoma significantly reduces T cell infiltration into the tumour microenvironment as a result of decreased ingress of BATF3-lineage DCs [192]. These findings of the immunosuppressive effect of WNT/ β -catenin signalling have been generalized across other cancer types including CRC [193]. The apparent restriction of *RAS*-associated immunosuppression to CMS2 suggests that the interaction of Wnt signalling and *RAS* mutation warrants further investigation to determine whether there is a synergistic effect. If so, dual inhibition of these pathways could be a route to improving immunity in this CMS grouping.

If *RAS* mutation is causatively linked with immunosuppression, inhibition of downstream signalling may be a route towards increasing intra-tumoural immunity. Recent clinical data have shown that combining the MEK inhibitor cobimetinib with the anti-PD-L1 antibody atezolizumab results in a durable 17% response rate in *KRAS* mutant MSS CRC patients [194]. This approach was based on the pre-clinical observation of synergy between MEK inhibition and PD-L1 blockade, driven by a reduction in antigen-driven T cell exhaustion mediated by activated MEK signalling in microenvironmental T cells and by class I up-regulation [195]. Interestingly therefore, in this case, the MEK activation within T cells, not tumour cells, was crucial. A worthwhile therapeutic avenue to explore may be a combinatory approach utilising the previously discussed neoantigen-directed or tumour antigen-directed vaccination or adoptive cell therapies to increase anti-tumour T cell responses, followed by MEK

inhibition to reduce antigen-driven T cell exhaustion and checkpoint blockade to reduce subsequent inhibition of responses.

Chapter 6

Overall Discussion

This thesis has explored the role of EPCR, a $\gamma\delta$ T cell receptor ligand, in colorectal tumourigenesis and the genetic factors that impact on the immune microenvironment. The EPCR study investigated the mechanism of its upregulation in tumourigenesis and whether this has functional significance in CRC. I found that EPCR is consistently upregulated in CRC and can mediate intra-cellular signalling via ERK phosphorylation. However, the results of my *in vitro* assays did not suggest that EPCR has any positive impact on CRC cell survival. In fact, perturbation of EPCR reduced chemosensitivity and increased cell migration in HCT116 cells. In addition, EPCR expression was not predictive for patient survival, chemotherapy response or cetuximab response. These results, together with my bioinformatics analyses, suggest that EPCR may be upregulated as a bystander gene as a component of chromosome 20q amplification, rather than due to it conferring a selective advantage *per se*. Previous studies have confirmed EPCR expression in breast cancer, ovarian cancer and lung cancer, but these studies have not related EPCR expression to wider abnormalities such as chromosomal amplification or hypomethylation [78, 80, 86]. Significantly, all of these tumour types are associated with 20q amplification to varying degrees. A limitation of previous studies, and of approaches highlighting both beneficial and detrimental effects of EPCR perturbation in animal models [78, 79, 83, 114] is that they primarily address EPCR-intrinsic effects on cancer development, and typically focus on one or a small number of cell lines. This study has utilised a broader approach, and investigated the wider context of EPCR dysregulation, highlighting that the effects of EPCR upregulation on cancer cells may be heavily dependent on biological context, including the exact cancer cell line (even within a single cancer type) and potentially upregulation of functionally important

neighbouring genes co-amplified on chromosome 20q. Therefore studies that focus on cell lines or animal models may not accurately recapitulate the clinical scenario. EPCR expression may be clinically useful as a marker of 20q amplification and as a marker of thrombotic risk in a range of tumours, potentially in serum or urine as sEPCR. This has not been determined in this study and would be important future work. In CRC, amplification of chromosome 20q occurs mostly in the progression of adenomas into carcinomas, and increases further in metastatic disease [117, 125]. EPCR could be a useful biomarker of 20q amplification, and therefore of progression.

EPCR upregulation in CRC may provide an example of how $\gamma\delta$ T cells can recognise cellular stress in different settings. This aspect of EPCR, and whether it is directly recognised by $\gamma\delta$ T cells within colorectal cancer is an area which is being explored by our group and others. $\gamma\delta$ T cells may respond to stress ligands, and as these cells are not MHC restricted, they do not require Class I expression on tumour cells and are traditionally thought to have an innate-like function similar to NK cells, with pre-expanded $\gamma\delta$ T cells able to respond rapidly to stress stimuli in an innate-like fashion [52]. However, work within our group has shown that the $\gamma\delta$ repertoire appears to be private to each individual, and therefore the LES clone, which recognises EPCR, appears unlikely to be present in many patients. The discovery of private repertoires and selective expansions of individual T cell receptors (TCRs) suggest that $\gamma\delta$ T cells may also play an adaptive role. Despite this, a group from Freiburg (Germany) have claimed to have identified the LES clone in breast cancer tissue from multiple patients [196]. If true, this suggests that the $\gamma\delta$ TCR repertoire is larger, or not as private, as it appears and that cells capable of recognising EPCR,

which is upregulated during tumourigenesis, may be enriched in the tumour microenvironment. Importantly, highly sensitive in-situ hybridisation (ISH) approaches have now been developed to explore this issue further. Nevertheless, EPCR may fit a general theme of dysregulation of stress ligands. While LES was expanded in CMV infection [53], the dysregulation of EPCR in cancer suggests the possibility that known $\gamma\delta$ T cell ligands may signal stress in different scenarios, and via different mechanisms. As my work was focussed mainly on the mechanism and biological significance of EPCR expression, the significance of EPCR expression with regards to $\gamma\delta$ T cell recognition in CRC still needs exploration.

In addition to investigating EPCR, I also studied the impact of tumour genetics on the immune microenvironment of CRC. My bioinformatics studies revealed important links between genetics and immunity, allowing the stratification of CRC patients into groups based on immunogenomics. This has implications for patient care, and has led to several predictions which have subsequently been confirmed. Firstly, this work suggested that MSI-H patients, who have high mutation rates and immune infiltration, would likely respond to checkpoint blockade therapy. This has since been confirmed in a clinical study [154].

Secondly, I hypothesised that combination checkpoint blockade would be beneficial, due to the co-ordinate upregulation of several of these molecules as part of the CIRC. Combination therapy has now been shown to be effective in melanoma, where a combination of CTLA4 and PD-1 inhibition had an additive beneficial effect beyond the effect of each individual drug [197]. Currently, it is unclear whether this

benefit is directly related to the co-upregulation of these checkpoint molecules. However, such a combination approach warrants investigation in MSI-H CRC patients, as well as potentially MSS patients with strong immune infiltration, who, like MSI-H patients, have co-upregulation of immune checkpoints. Such combination therapy is associated with higher toxicity. However, the toxicity levels, whilst occasionally severe and life threatening, compare favourably to existing chemotherapeutics regimens [15].

Third, I showed that *POL* mutant cancers are associated with increased immune infiltration. This has recently been confirmed in a retrospective clinical study, which demonstrated that *POLE* mutant CRCs display increased CD8+ lymphocyte infiltration and expression of cytotoxic T cell markers and effector cytokines, similar to the immune infiltration seen in MSI-H cancers [155].

Finally, another key prediction was that MEK inhibition and checkpoint blockade could be synergistic, and this has been partially confirmed in a phase I clinical study that revealed that combining the MEK inhibitor cobimetinib with the anti-PD-L1 antibody atezolizumab resulted in a durable 17% response rate in *KRAS* mutant MSS CRC patients [194]. Clearly, this requires further assessment in phase II and III studies. Interestingly, rather than a tumour cell effect, this response appears to be due to inhibition of MEK signalling within microenvironmental T cells. TCR signalling is known to activate RAS [198, 199].

A further key prediction would be that the subset of MSS patients who have a similar immune profile to MSI-H patients (the 'MSS CIRC-high' group) could also respond to checkpoint blockade therapy. This would be a crucial clinical study, as MSS patients make up over 80% of sporadic CRC cases, and therefore responses in even a small subset of the MSS group could benefit a large number of patients. In such a trial, MSS patients could be treated with checkpoint blockade therapy based on protein expression of key immune markers, such as PD-L1 (the current biomarker for anti-PD-1 mAb therapy), CD4, CD8 or class II HLA. This study would be important even if the mechanism underlying the variability in immunity in MSS is not yet understood, as responses are likely to depend on the immune phenotype itself, rather than the factors determining the immune phenotype.

To investigate one of the factors determining immune phenotype, the relationship between overall tumour mutation rate/number and immunity was explored. This demonstrated that immune infiltration was linked to overall mutation rate and number, but this relationship was weak when intra-group analysis was performed within either MSS or MSI-H patients. This seemed to suggest that the quality and type of mutation may have been the key determinant of immunity, perhaps in relation to certain host genes, e.g. HLA. However, investigation of neoantigen binding using binding prediction algorithms did not show any clear association between the strongest binding neoantigen and CIRC expression, though, as only the strongest single neoantigens was considered, this analysis was superficial. However, it suggests that neoantigen number and strongest binding neoantigen, whilst clearly being highly significant and likely being responsible for the high immune infiltration in

MSI-H tumours, cannot reliably predict the strength of intra-tumoural immunity in MSS *POLE+POLD1* wild type patients, who all have relatively low mutation rates (<15/Mbp). Whilst in MSI-H patients the high neoantigen load may override other factors, in MSS patients other tumour, host and environmental factors may be the key determinants of immune response. As such, there are many possible contributors which are currently poorly understood, including tumour-extrinsic factors such as bowel flora and host genetics, both of which warrant further investigation. Bowel flora have previously been linked to colorectal tumourigenesis [159], with particular microbiota such as *Fusobacterium* enriched in CRC patients [160]. In addition, lipopolysaccharide (LPS), a gram-negative bacterial antigen present in primary CRC due to breakdown of epithelial integrity, has been implicated in cell migration, metastases and immune response in CRC [200-203]. If bowel flora have an immunosuppressive effect in CRC, this could contribute to the low immune infiltration seen in many patients, and modifying the microbiome could conceivably be a route to improving immunity. Host genetics, and in particular expression quantitative trait loci (eQTLs), which are genomic loci which contribute to the expression of nearby (*cis* eQTLs) or distant (*trans* eQTLs) genes [158], have been linked with immune response in the infectious and inflammatory disease settings [204-206]. It is possible that eQTLs could also affect immunity in the tumour setting. It is now established in melanoma that tumour regression after PD-1 blockade depends on pre-existing CD8 T cells and PD-1/PD-L1 expression in the invasive margin and inside the tumour [207]. Understanding the determinants of immunity in MSS patients could lead to the development of new therapeutic approaches that

could 'adjuvantise' the microenvironment, and then enable the effective use of checkpoint blockade in MSS CRC patients.

The frequency and predicted immunogenicity, based on HLA-binding prediction algorithms, of *RAS*-associated neoantigens were also investigated. The prediction of recurrent and commonly presented strong-binding neoantigens derived from mutant *RAS* raises the possibility of utilising these antigens in vaccination or other immunotherapeutic approaches, such as cell-based therapies. In comparison to personalised vaccination, a stratified vaccination approach would have the advantage of being usable at diagnosis, as determination of *RAS* mutation status is now routine in clinical practice, after biopsy or resection. The patient's HLA-type would likely need to be determined, as certain HLA/neoepitope combinations were recurrent. However, full genetic sequencing would not be necessary. In the longer term, it is feasible to envisage an approach where off-the-shelf stratified vaccines are given to patients very soon after biopsy or surgery, potentially bridging the time gap before personalised vaccines can be developed and complementing more costly personalised approaches. Currently, personalised vaccination requires a significant amount of time to be produced following tumour resection [16]. Furthermore, the vaccines could be applicable to multiple patients, and could therefore result in production scale cost savings. As *RAS* is a key cancer driver mutation, it could potentially be less susceptible to immune escape, which is desirable for a potential target antigen [29]. The *RAS*-associated neoantigen could be combined with highly expressed tumour antigens to increase efficacy. MSS *KRAS* mutant patients tend to have low mutation rates, and therefore have a limited number of targetable

neoantigens (a median of 26 class I presented neoepitopes were predicted per patient). *KRAS* mutant as a group makes up 40% of all CRC patients, who do not respond to current targeted therapies such as EGFR monoclonal antibodies [10]. Therefore novel therapeutic options are needed.

The success of such a vaccine programme would have significant implications. It could improve survival in a patient group with very poor survival outcomes. Crucially, the approach focusses the vaccine development on tumour groups with low mutational burden, where immunotherapies such as up-front checkpoint blockade are unlikely to work as single therapies. If successful in generating potent anti-tumour responses, it could increase the effectiveness of such checkpoint blockade approaches, through the ‘adjuvantisation’ of the microenvironment. Such stratified vaccine approaches could be applied to a wide range of tumour types other than CRC, as driver mutations such as *KRAS* are highly commonplace throughout all tumours. Stratified vaccination warrants further study in preclinical and clinical models.

One of the key findings from my *in silico* studies was that *RAS* mutation was correlated with reduced immune infiltration. In the TCGA dataset, this correlation was most marked in those *RAS* mutants with highest MEK signatures – the non-G12D/G13D subtypes. However, my local IHC analysis did not show any association between *RAS* mutation and CD4, CD8, Class II HLA, CXCL10, PD-L1 and Tbet density, although STAT1 was significantly reduced in *RAS* mutants, suggesting that *RAS* signalling may reduce STAT1 expression, which mediates IFN γ receptor

signalling. However, this did not appear to produce significant reductions in immune infiltration in this sample set.

As I did not see a clear association between *RAS* mutation and immune protein expression, the TCGA data was revisited and in collaboration with Justin Guinney, I found that the association of *RAS* mutation with immunity depended on CMS groupings. *RAS* mutation did not have a significant association with immunity within all CMS groupings – only within CMS2. This group is associated with left sided colonic tumours that have high WNT/Myc pathway activation. CMS2 is relatively immunosuppressed, and it is possible that in an immunosuppressed group, *RAS* mutation has a further suppressive effect. However, the fact that CMS2 *RAS* wild types have lower immunity than CMS4 *RAS* mutants suggests that the CMS groupings, which are based on overall transcriptional profiles, are more significant in terms of immune impact than individual driver mutations such as *RAS*. Though understanding the key mechanisms and mutations driving tumours may be of benefit if they can be affected (such as with targeted therapies), targeting of immunotherapies based on transcriptional profiles such as CMS or immune phenotype may be more appropriate than mutational subtype alone, as discussed above. It is possible that different factors limit immunity within the three MSS CMS groups (2, 3 and 4), and specific therapies may need to be developed for each subgroup to improve immune infiltration. For example, dual MEK and Wnt pathway inhibition may be warrant further exploration in CMS2 *RAS* mutant patients, both to reduce signalling through these oncogenic signalling pathways and to reduce immunosuppression.

This study has shown the value of recently developed bioinformatic approaches, which can utilise large publically available databases that include clinical, mutation and transcriptomic data. These large databases could not be produced locally without great time, effort and cost, and enable a 'reverse translation' approach, whereby clinical and scientific hypotheses can be investigated with patients as the starting point. Important findings can then be explored and validated in the laboratory and a local patient set, and then, finally, could impact on patient management.

Overall, this work has increased the understanding of the factors that determine immunity in CRC, and it is a step towards implementing successful immunotherapy in CRC, which though showing great promise in a range of cancer types, has only been successful in MSI-H CRC thus far. CRC is a useful model for the tumour immunology field in general, due to the established relationship between immune infiltration and patient outcome, the variety of well characterised molecular subtypes with both high and low mutational burdens, and the identification of different patient groups in which checkpoint blockade therapy has been both effective and ineffective. It is likely that immunotherapies can succeed in MSS CRC once the factors that determine immune infiltration are better understood, and once existing and emerging immunotherapies are targeted appropriately and are used in intelligent combinations. The future of immunotherapy in CRC and other cancers will likely involve the use of synergistic combinations of immunotherapies, such as combination checkpoint blockade, or vaccination in low immune profile patients followed by checkpoint

blockade. In addition, it is likely that immunotherapies will increasingly be combined with other therapy types including targeted therapies, chemotherapy, radiotherapy and surgery. Several such combinations are being investigated in studies worldwide [20]. My work has provided strong indications as to how immunotherapies should be targeted in CRC.

Appendix. Association of immunity in paired CRC biopsy and resection specimens

In a window-of-opportunity clinical trial, patients may be given a drug or intervention between the time of cancer diagnosis (biopsy) and surgical resection. To investigate how immunity within colonic tumour biopsies (obtained endoscopically) correlates with immunity in full colonic surgical resection samples, resected several days after the biopsy, I collected FFPE sections from 15 CRC biopsy/resection pairs. IHC staining revealed relatively weak correlations between CD8, STAT1 and PD-L1 protein expression in the biopsies and resections. In epithelium, Spearman correlations were 0.369 ($p=0.175$) and -0.322 ($p=0.243$) for CD8 and STAT1 respectively. All biopsy sections were negative for PD-L1 in epithelium. In stroma, Spearman correlations were 0.316 ($p=0.251$), -0.018 ($p=0.952$) and 0.261 ($p=0.347$), for CD8, STAT1 and PD-L1 respectively. Though analysis of further samples is required to confirm this finding, if confirmed this potentially has significant implications for the design of windows-of-opportunity studies. In such a study, biological differences between the biopsy and the resection sample may be measured to determine the effect of a drug or intervention on the tumour microenvironment. However, this result suggests that biopsy and resection sample immune infiltrations are not closely correlated, which would need to be considered in a window study design. It is possible that immune infiltration changes during the period between biopsy and resection, leading to potential bias according to when the resection occurs. Furthermore, heterogeneity in different areas of the tumour could lead to sampling bias depending on which areas of the biopsy and resection are

analysed. However, despite this, this study design may be useful if an intervention or drug has a substantial impact on the microenvironment of the resection sample, overcoming variabilities due to the above factors.

Collaborative Statement

All work in this thesis was my own other than:

Initial confirmation of EPCR expression on CRC cell lines (Carrie Willcox).

Confirmation of EPCR expression on mast cells (Peter Bradding and Aarti Shikotra).

Validation of Class II HLA and CXCL10 antibody (Ghaleb Goussous).

In addition, the neoantigen studies were in collaboration with Sebastian Boegel, who performed the neoantigen prediction. The CMS grouping studies were in collaboration with Justin Guinney, who performed the statistical analyses on the extended TCGA RNAseq and KFSYSCC datasets.

List of References

1. CRUK. *Bowel Cancer Statistics*. 2016 [cited 11/09/2016]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer>.
2. Siegel, R., D. Naishadham, and A. Jemal, *Cancer statistics, 2012*. CA Cancer J Clin, 2012. **62**(1): p. 10-29.
3. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis*. Cell, 1990. **61**(5): p. 759-67.
4. Weinberg, R.A., *The Biology of Cancer*. 2nd ed. 2013: Garland Science.
5. Buckowitz, A., et al., *Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases*. British Journal of Cancer, 2005. **92**(9): p. 1746-1753.
6. Tran, B., et al., *Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer*. Cancer, 2011. **117**(20): p. 4623-32.
7. Gavin, P.G., et al., *Mutation profiling and microsatellite instability in stage II and III colon cancer: an assessment of their prognostic and oxaliplatin predictive value*. Clin Cancer Res, 2012. **18**(23): p. 6531-41.
8. Scaltriti, M. and J. Baselga, *The epidermal growth factor receptor pathway: a model for targeted therapy*. Clin Cancer Res, 2006. **12**(18): p. 5268-72.
9. Seshacharyulu, P., et al., *Targeting the EGFR signaling pathway in cancer therapy*. Expert Opin Ther Targets, 2012. **16**(1): p. 15-31.
10. Maughan, T.S., et al., *Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial*. Lancet, 2011. **377**(9783): p. 2103-14.
11. CRUK. *Treating Bowel Cancer*. 2016; Available from: <http://www.cancerresearchuk.org/about-cancer/type/bowel-cancer/treatment/>.
12. Hodi, F.S., et al., *Improved survival with ipilimumab in patients with metastatic melanoma*. N Engl J Med, 2010. **363**(8): p. 711-23.
13. Brahmer, J.R., et al., *Safety and activity of anti-PD-L1 antibody in patients with advanced cancer*. N Engl J Med, 2012. **366**(26): p. 2455-65.
14. Topalian, S.L., et al., *Safety, activity, and immune correlates of anti-PD-1 antibody in cancer*. N Engl J Med, 2012. **366**(26): p. 2443-54.
15. Brahmer, J., et al., *Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer*. New England Journal of Medicine, 2015. **373**(2): p. 123-135.
16. Kreiter, S., et al., *Mutant MHC class II epitopes drive therapeutic immune responses to cancer*. Nature, 2015. **520**(7549): p. 692-696.
17. Murphy, K., *Janeway's Immunobiology*. Vol. 8th. 2011: Garland Science.
18. Dunn, G.P., et al., *Cancer immunoediting: from immunosurveillance to tumor escape*. Nat Immunol, 2002. **3**(11): p. 991-8.
19. Dobrzanski, M.J., *Expanding Roles for CD4 T Cells and Their Subpopulations in Tumor Immunity and Therapy*. Frontiers in Oncology, 2013. **3**.
20. Farkona, S., E.P. Diamandis, and I.M. Blasutig, *Cancer immunotherapy: the beginning of the end of cancer?* BMC Med, 2016. **14**: p. 73.

21. Khalil, D.N., et al., *The future of cancer treatment: immunomodulation, CARs and combination immunotherapy*. Nat Rev Clin Oncol, 2016. **13**(5): p. 273-90.
22. Kroemer, G., et al., *Immunogenic cell death in cancer therapy*. Annu Rev Immunol, 2013. **31**: p. 51-72.
23. Menard, C., et al., *Cancer chemotherapy: not only a direct cytotoxic effect, but also an adjuvant for antitumor immunity*. Cancer Immunol Immunother, 2008. **57**(11): p. 1579-87.
24. Le, D.T. and E.M. Jaffee, *Regulatory T-cell modulation using cyclophosphamide in vaccine approaches: a current perspective*. Cancer Res, 2012. **72**(14): p. 3439-44.
25. Gattinoni, L., et al., *Removal of homeostatic cytokine sinks by lymphodepletion enhances the efficacy of adoptively transferred tumor-specific CD8+ T cells*. J Exp Med, 2005. **202**(7): p. 907-12.
26. Rosenberg, S.A., et al., *Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy*. Clin Cancer Res, 2011. **17**(13): p. 4550-7.
27. Tran, E., et al., *Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer*. Science, 2014. **344**(6184): p. 641-5.
28. Gross, G., T. Waks, and Z. Eshhar, *Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity*. Proc Natl Acad Sci U S A, 1989. **86**(24): p. 10024-8.
29. Cheever, M.A., et al., *The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research*. Clin Cancer Res, 2009. **15**(17): p. 5323-37.
30. Higano, C.S., et al., *Integrated data from 2 randomized, double-blind, placebo-controlled, phase 3 trials of active cellular immunotherapy with sipuleucel-T in advanced prostate cancer*. Cancer, 2009. **115**(16): p. 3670-9.
31. Kantoff, P.W., et al., *Sipuleucel-T immunotherapy for castration-resistant prostate cancer*. N Engl J Med, 2010. **363**(5): p. 411-22.
32. Ishida, Y., et al., *Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death*. EMBO J, 1992. **11**(11): p. 3887-95.
33. Okazaki, T., et al., *A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application*. Nat Immunol, 2013. **14**(12): p. 1212-8.
34. Robert, C., et al., *Nivolumab in previously untreated melanoma without BRAF mutation*. N Engl J Med, 2015. **372**(4): p. 320-30.
35. Robert, C., et al., *Pembrolizumab versus Ipilimumab in Advanced Melanoma*. N Engl J Med, 2015. **372**(26): p. 2521-32.
36. Chung, K.Y., et al., *Phase II Study of the Anti-Cytotoxic T-Lymphocyte-Associated Antigen 4 Monoclonal Antibody, Tremelimumab, in Patients With Refractory Metastatic Colorectal Cancer*. Journal of Clinical Oncology, 2010. **28**(21): p. 3485-3490.
37. Galon, J., *Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome*. Science, 2006. **313**(5795): p. 1960-1964.
38. Galon, J., et al., *Cancer classification using the Immunoscore: a worldwide task force*. Journal of Translational Medicine, 2012. **10**(1): p. 205.

39. Ropponen, K.M., et al., *Prognostic value of tumour-infiltrating lymphocytes (TILs) in colorectal cancer*. J Pathol, 1997. **182**(3): p. 318-24.
40. Naito, Y., et al., *CD8+ T cells infiltrated within cancer cell nests as a prognostic factor in human colorectal cancer*. Cancer Res, 1998. **58**(16): p. 3491-4.
41. Diederichsen, A.C., et al., *Prognostic value of the CD4+/CD8+ ratio of tumour infiltrating lymphocytes in colorectal cancer and HLA-DR expression on tumour cells*. Cancer Immunol Immunother, 2003. **52**(7): p. 423-8.
42. Mlecnik, B., et al., *Biomolecular Network Reconstruction Identifies T-Cell Homing Factors Associated With Survival in Colorectal Cancer*. Gastroenterology, 2010. **138**(4): p. 1429-1440.
43. Tosolini, M., et al., *Clinical Impact of Different Classes of Infiltrating T Cytotoxic and Helper Cells (Th1, Th2, Treg, Th17) in Patients with Colorectal Cancer*. Cancer Research, 2011. **71**(4): p. 1263-1271.
44. Gajewski, T.F., H. Schreiber, and Y.-X. Fu, *Innate and adaptive immune cells in the tumor microenvironment*. Nature Immunology, 2013. **14**(10): p. 1014-1022.
45. Kim, H., et al., *Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas*. Oncogene, 2004. **23**(37): p. 6218-6225.
46. Phillips, S.M., et al., *Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic*. British Journal of Surgery, 2004. **91**(4): p. 469-475.
47. Banerjea, A., S.A. Bustin, and S. Dorudi, *The immunogenicity of colorectal cancers with high-degree microsatellite instability*. World J Surg Oncol, 2005. **3**: p. 26.
48. Jass, J.R., *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features*. Histopathology, 2007. **50**(1): p. 113-130.
49. Fehlfker, M., et al., *Concerted down-regulation of immune-system related genes predicts metastasis in colorectal carcinoma*. BMC Cancer, 2014. **14**(1): p. 64.
50. Boissiere-Michot, F., et al., *Characterization of an adaptive immune response in microsatellite-unstable colorectal cancer*. OncoImmunology, 2014. **3**: p. e29256.
51. Gentles, A.J., et al., *The prognostic landscape of genes and infiltrating immune cells across human cancers*. Nat Med, 2015. **21**(8): p. 938-45.
52. Hayday, A.C., *$\gamma\delta$ T Cells and the Lymphoid Stress-Surveillance Response*. Immunity, 2009. **31**(2): p. 184-196.
53. Willcox, C.R., et al., *Cytomegalovirus and tumor stress surveillance by binding of a human $\gamma\delta$ T cell antigen receptor to endothelial protein C receptor*. Nature Immunology, 2012. **13**(9): p. 872-879.
54. Wilhelm, M., et al., *Gammadelta T cells for immune therapy of patients with lymphoid malignancies*. Blood, 2003. **102**(1): p. 200-6.
55. Dieli, F., et al., *Targeting Human $\gamma\delta$ T Cells with Zoledronate and Interleukin-2 for Immunotherapy of Hormone-Refractory Prostate Cancer*. Cancer Research, 2007. **67**(15): p. 7450-7457.
56. Kobayashi, H., et al., *Safety profile and anti-tumor effects of adoptive immunotherapy using gamma-delta T cells against advanced renal cell carcinoma: a pilot study*. Cancer Immunol Immunother, 2007. **56**(4): p. 469-76.

57. Kakimi, K., et al., *gammadelta T cell therapy for the treatment of non-small cell lung cancer*. Transl Lung Cancer Res, 2014. **3**(1): p. 23-33.
58. Halary, F., et al., *Shared reactivity of Vδ2 negative γδ T cells against cytomegalovirus-infected cells and tumor intestinal epithelial cells*. J Exp Med, 2005. **201**(10): p. 1567-78.
59. Scheffer, G.L., et al., *Expression of the vascular endothelial cell protein C receptor in epithelial tumour cells*. Eur J Cancer, 2002. **38**(11): p. 1535-42.
60. Esmon, C.T., *The endothelial cell protein C receptor*. Thromb Haemost, 2000. **83**(5): p. 639-43.
61. Fukudome, K. and C.T. Esmon, *Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor*. J Biol Chem, 1994. **269**(42): p. 26486-91.
62. Oganessian, V., et al., *The crystal structure of the endothelial protein C receptor and a bound phospholipid*. J Biol Chem, 2002. **277**(28): p. 24851-4.
63. Koch, M., et al., *The crystal structure of human CD1d with and without alpha-galactosylceramide*. Nat Immunol, 2005. **6**(8): p. 819-26.
64. Mammen Ef Fau - Thomas, W.R., W.H. Thomas Wr Fau - Seegers, and W.H. Seegers, *Activation of purified prothrombin to autoprothrombin I or autoprothrombin II (platelet cofactor II or autoprothrombin II-A)*. (0340-5338 (Print)).
65. Esmon, C.T. and W.G. Owen, *Identification of an endothelial cell cofactor for thrombin-catalyzed activation of protein C*. Proc Natl Acad Sci U S A, 1981. **78**(4): p. 2249-52.
66. Taylor, F.B., Jr., et al., *Protein C prevents the coagulopathic and lethal effects of Escherichia coli infusion in the baboon*. J Clin Invest, 1987. **79**(3): p. 918-25.
67. Macfarlane, S.R., et al., *Proteinase-activated receptors*. Pharmacol Rev, 2001. **53**(2): p. 245-82.
68. Riewald, M., *Activation of Endothelial Cell Protease Activated Receptor 1 by the Protein C Pathway*. Science, 2002. **296**(5574): p. 1880-1882.
69. Joyce, D.E., *Gene Expression Profile of Antithrombotic Protein C Defines New Mechanisms Modulating Inflammation and Apoptosis*. Journal of Biological Chemistry, 2001. **276**(14): p. 11199-11203.
70. Taylor, F.B., Jr., et al., *The endothelial cell protein C receptor aids in host defense against Escherichia coli sepsis*. Blood, 2000. **95**(5): p. 1680-6.
71. Bernard, G.R., et al., *Efficacy and safety of recombinant human activated protein C for severe sepsis*. N Engl J Med, 2001. **344**(10): p. 699-709.
72. Beaulieu, L.M. and F.C. Church, *Activated protein C promotes breast cancer cell migration through interactions with EPCR and PAR-1*. Experimental Cell Research, 2007. **313**(4): p. 677-687.
73. Cheng, T., et al., *Activated protein C blocks p53-mediated apoptosis in ischemic human brain endothelium and is neuroprotective*. Nature Medicine, 2003. **9**(3): p. 338-342.
74. Xue, M., et al., *Endothelial protein C receptor and protease-activated receptor-1 mediate induction of a wound-healing phenotype in human keratinocytes by activated protein C*. J Invest Dermatol, 2005. **125**(6): p. 1279-85.

75. Thiagarajan, M., T. Cheng, and B.V. Zlokovic, *Endothelial Cell Protein C Receptor: Role Beyond Endothelium?* Circulation Research, 2007. **100**(2): p. 155-157.
76. Tsuneyoshi, N., et al., *Expression and anticoagulant function of the endothelial cell protein C receptor (EPCR) in cancer cell lines.* Thromb Haemost, 2001. **85**(2): p. 356-61.
77. Uchiba, M., *Activated Protein C Induces Endothelial Cell Proliferation by Mitogen-Activated Protein Kinase Activation In Vitro and Angiogenesis In Vivo.* Circulation Research, 2004. **95**(1): p. 34-41.
78. Antón, I., et al., *Receptor of Activated Protein C Promotes Metastasis and Correlates with Clinical Outcome in Lung Adenocarcinoma.* American Journal of Respiratory and Critical Care Medicine, 2012. **186**(1): p. 96-105.
79. Bezuhly, M., et al., *Role of activated protein C and its receptor in inhibition of tumor metastasis.* Blood, 2009. **113**(14): p. 3371-3374.
80. Derksen, P., et al., *Endothelial Protein C Receptor Function in Murine and Human Breast Cancer Development.* PLoS ONE, 2013. **8**(4): p. e61071.
81. Park, S.Y., et al., *Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer.* Clin Cancer Res, 2010. **16**(3): p. 876-87.
82. Ranieri, V.M., et al., *Drotrecogin alfa (activated) in adults with septic shock.* N Engl J Med, 2012. **366**(22): p. 2055-64.
83. Keshava, S., et al., *Endothelial Cell Protein C Receptor Opposes Mesothelioma Growth Driven by Tissue Factor.* Cancer Research, 2013.
84. Striggow, F., et al., *The protease thrombin is an endogenous mediator of hippocampal neuroprotection against ischemia at low concentrations but causes degeneration at high concentrations.* Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2264-9.
85. Vaughan, P.J., et al., *Thrombin receptor activation protects neurons and astrocytes from cell death produced by environmental insults.* J Neurosci, 1995. **15**(7 Pt 2): p. 5389-401.
86. Mirshahi, M., *Endothelial protein C receptor expressed by ovarian cancer cells as a possible biomarker of cancer onset.* International Journal of Oncology, 2012.
87. Saposnik, B., et al., *A haplotype of the EPCR gene is associated with increased plasma levels of sEPCR and is a candidate risk factor for thrombosis.* Blood, 2004. **103**(4): p. 1311-8.
88. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.* Science, 2015. **348**(6230): p. 124-8.
89. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
90. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
91. *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
92. *Cbioportal for Cancer Genomics.* Available from: www.cbioportal.org.
93. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-7.
94. Rhodes, D.R., et al., *ONCOMINE: a cancer microarray database and integrated data-mining platform.* Neoplasia, 2004. **6**(1): p. 1-6.

95. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotechnol, 2011. **29**(1): p. 24-6.
96. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types*. Nature, 2013. **502**(7471): p. 333-339.
97. Boegel, S., et al., *HLA typing from RNA-Seq sequence reads*. Genome Med, 2012. **4**(12): p. 102.
98. Boegel, S., et al., *A catalog of HLA type, HLA expression, and neo-epitope candidates in human cancer cell lines*. Oncoimmunology, 2014. **3**(8): p. e954893.
99. Hanzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC Bioinformatics, 2013. **14**: p. 7.
100. Bindea, G., et al., *Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer*. Immunity, 2013. **39**(4): p. 782-795.
101. Butterfield, J.H., et al., *Establishment of an immature mast cell line from a patient with mast cell leukemia*. Leuk Res, 1988. **12**(4): p. 345-55.
102. Sanmugalingam, D., A.J. Wardlaw, and P. Bradding, *Adhesion of human lung mast cells to bronchial epithelium: evidence for a novel carbohydrate-mediated mechanism*. J Leukoc Biol, 2000. **68**(1): p. 38-46.
103. Hollins, F., et al., *Human airway smooth muscle promotes human lung mast cell survival, proliferation, and constitutive activation: cooperative roles for CADM1, stem cell factor, and IL-6*. J Immunol, 2008. **181**(4): p. 2772-80.
104. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Res, 2015. **43**(7): p. e47.
105. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments*. Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
106. Smyth, G.K. and T. Speed, *Normalization of cDNA microarray data*. Methods, 2003. **31**(4): p. 265-73.
107. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc, 2009. **4**(1): p. 44-57.
108. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists*. Nucleic Acids Res, 2009. **37**(1): p. 1-13.
109. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
110. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes*. Nat Genet, 2003. **34**(3): p. 267-73.
111. Buhard, O., et al., *Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors*. J Clin Oncol, 2006. **24**(2): p. 241-51.
112. Esmon, C.T., *Protein C anticoagulant system—anti-inflammatory effects*. Seminars in Immunopathology, 2011. **34**(1): p. 127-132.
113. Boire, A., et al., *PAR1 Is a Matrix Metalloprotease-1 Receptor that Promotes Invasion and Tumorigenesis of Breast Cancer Cells*. Cell, 2005. **120**(3): p. 303-313.

114. Wang, Q., et al., *Endothelial cell protein C receptor promotes MGC803 gastric cancer cells proliferation and migration by activating ERK1/2*. Medical Oncology, 2015. **32**(5).
115. Overton, W.R., *Modified histogram subtraction technique for analysis of flow cytometry data*. Cytometry, 1988. **9**(6): p. 619-26.
116. Tabach, Y., et al., *Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer*. PLoS ONE, 2011. **6**(1): p. e14632.
117. Carvalho, B., et al., *Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression*. Gut, 2008. **58**(1): p. 79-89.
118. Carter, S.L., et al., *A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers*. Nat Genet, 2006. **38**(9): p. 1043-8.
119. Douillard, J.Y., et al., *Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer*. N Engl J Med, 2013. **369**(11): p. 1023-34.
120. Guitton, C., et al., *Protective cross talk between activated protein C and TNF signaling in vascular endothelial cells: implication of EPCR, noncanonical NF-kappaB, and ERK1/2 MAP kinases*. Am J Physiol Cell Physiol, 2011. **300**(4): p. C833-42.
121. Burrell, R.A., et al., *Replication stress links structural and numerical cancer chromosomal instability*. Nature, 2013. **494**(7438): p. 492-496.
122. Hirsch, D., et al., *A new whole genome amplification method for studying clonal evolution patterns in malignant colorectal polyps*. Genes, Chromosomes and Cancer, 2012. **51**(5): p. 490-500.
123. Tanner, M.M., et al., *Frequent amplification of chromosomal region 20q12-q13 in ovarian cancer*. Clin Cancer Res, 2000. **6**(5): p. 1833-9.
124. Hodgson, J.G., et al., *Genome amplification of chromosome 20 in breast cancer*. Breast Cancer Res Treat, 2003. **78**(3): p. 337-45.
125. Tsafir, D., *Relationship of Gene Expression and Chromosomal Abnormalities in Colorectal Cancer*. Cancer Research, 2006. **66**(4): p. 2129-2137.
126. Williams, A.C., S.J. Harper, and C. Paraskeva, *Neoplastic transformation of a human colonic epithelial cell line: in vitro evidence for the adenoma to carcinoma sequence*. Cancer Res, 1990. **50**(15): p. 4724-30.
127. Ruggero, D. and P.P. Pandolfi, *Does the ribosome translate cancer?* Nature Reviews Cancer, 2003. **3**(3): p. 179-192.
128. FDA. *FDA Approval for Cetuximab*. 2016; Available from: <https://www.cancer.gov/about-cancer/treatment/drugs/fda-cetuximab>.
129. *FOCUS4 Homepage*. 2016; Available from: <http://www.focus4trial.org/aboutfocus4/focus4trialschema>.
130. Galon, J., et al., *The immune score as a new possible approach for the classification of cancer*. Journal of Translational Medicine, 2012. **10**(1): p. 1.
131. Lal, N., et al., *An immunogenomic stratification of colorectal cancer: Implications for development of targeted immunotherapy*. Oncoimmunology, 2015. **4**(3): p. e976052.
132. Palles, C., et al., *Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas*. Nat Genet, 2013. **45**(2): p. 136-44.

133. Dierssen, J.W., et al., *HNPCC versus sporadic microsatellite-unstable colon cancers follow different routes toward loss of HLA class I expression*. BMC Cancer, 2007. **7**: p. 33.
134. Hunter, J.C., et al., *Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations*. Molecular Cancer Research, 2015.
135. Cespedes, M.V., et al., *K-ras Asp12 mutant neither interacts with Raf, nor signals through Erk and is less tumorigenic than K-ras Val12*. Carcinogenesis, 2006. **27**(11): p. 2190-2200.
136. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer*. Nat Med, 2015. **21**(11): p. 1350-6.
137. Nagalla, S., et al., *Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis*. Genome Biology, 2013. **14**(4): p. R34.
138. Cheng, W.Y., T.H. Ou Yang, and D. Anastassiou, *Biomolecular events in cancer revealed by attractor metagenes*. PLoS Comput Biol, 2013. **9**(2): p. e1002920.
139. Cheng, W.Y., T.H. Ou Yang, and D. Anastassiou, *Development of a prognostic model for breast cancer survival in an open challenge environment*. Sci Transl Med, 2013. **5**(181): p. 181ra50.
140. Angelova, M., et al., *Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy*. Genome Biology, 2015. **16**(1).
141. Bindea, G., et al., *Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer*. Immunity, 2013. **39**(4): p. 782-95.
142. Bennett, S.R., et al., *Induction of a CD8+ cytotoxic T lymphocyte response by cross-priming requires cognate CD4+ T cell help*. J Exp Med, 1997. **186**(1): p. 65-70.
143. Bennett, S.R., et al., *Help for cytotoxic-T-cell responses is mediated by CD40 signalling*. Nature, 1998. **393**(6684): p. 478-80.
144. Janssen, E.M., et al., *CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes*. Nature, 2003. **421**(6925): p. 852-6.
145. Williams, M.A., A.J. Tyznik, and M.J. Bevan, *Interleukin-2 signals during priming are required for secondary expansion of CD8+ memory T cells*. Nature, 2006. **441**(7095): p. 890-3.
146. Xie, Y., et al., *Naive tumor-specific CD4(+) T cells differentiated in vivo eradicate established melanoma*. J Exp Med, 2010. **207**(3): p. 651-67.
147. Quezada, S.A., et al., *Tumor-reactive CD4(+) T cells develop cytotoxic activity and eradicate large established melanoma after transfer into lymphopenic hosts*. J Exp Med, 2010. **207**(3): p. 637-50.
148. Haabeth, O.A., et al., *How Do CD4(+) T Cells Detect and Eliminate Tumor Cells That Either Lack or Express MHC Class II Molecules?* Front Immunol, 2014. **5**: p. 174.
149. Hunder, N.N., et al., *Treatment of metastatic melanoma with autologous CD4+ T cells against NY-ESO-1*. N Engl J Med, 2008. **358**(25): p. 2698-703.
150. Dong, H., et al., *Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion*. Nat Med, 2002. **8**(8): p. 793-800.
151. Lee, S.J., et al., *Interferon regulatory factor-1 is prerequisite to the constitutive expression and IFN-gamma-induced upregulation of B7-H1 (CD274)*. FEBS Lett, 2006. **580**(3): p. 755-62.

152. Woo, S.R., et al., *Immune Inhibitory Molecules LAG-3 and PD-1 Synergistically Regulate T-cell Function to Promote Tumoral Immune Escape*. Cancer Research, 2011. **72**(4): p. 917-927.
153. Løvig, T., et al., *Strong HLA-DR expression in microsatellite stable carcinomas of the large bowel is associated with good prognosis*. British Journal of Cancer, 2002. **87**(7): p. 756-762.
154. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency*. New England Journal of Medicine, 2015. **372**(26): p. 2509-2520.
155. Domingo, E., et al., *Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study*. The Lancet Gastroenterology & Hepatology, 2016. **In press**.
156. Saeterdal, I., et al., *Frameshift-mutation-derived peptides as tumor-specific antigens in inherited and spontaneous colorectal cancer*. Proc Natl Acad Sci U S A, 2001. **98**(23): p. 13255-60.
157. Gooden, M.J., et al., *The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis*. Br J Cancer, 2011. **105**(1): p. 93-103.
158. Nica, A.C. and E.T. Dermitzakis, *Expression quantitative trait loci: present and future*. Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1620): p. 20120362.
159. Dulal, S. and T.O. Keku, *Gut microbiome and colorectal adenomas*. Cancer J, 2014. **20**(3): p. 225-31.
160. Chen, W., et al., *Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer*. PLoS One, 2012. **7**(6): p. e39743.
161. Nosh, K., et al., *Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review*. J Pathol, 2010. **222**(4): p. 350-66.
162. Maudsley, D.J. and A.G. Morris, *Kirsten murine sarcoma virus abolishes interferon gamma-induced class II but not class I major histocompatibility antigen expression in a murine fibroblast line*. J Exp Med, 1988. **167**(2): p. 706-11.
163. Maudsley, D.J., W.J. Bateman, and A.G. Morris, *Reduced stimulation of helper T cells by Ki-ras transformed cells*. Immunology, 1991. **72**(2): p. 277-81.
164. Atkins, D., et al., *MHC class I antigen processing pathway defects, ras mutations and disease stage in colorectal carcinoma*. International Journal of Cancer, 2004. **109**(2): p. 265-273.
165. Parada, L.F., et al., *Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene*. Nature, 1982. **297**(5866): p. 474-8.
166. Der, C.J., T.G. Krontiris, and G.M. Cooper, *Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses*. Proc Natl Acad Sci U S A, 1982. **79**(11): p. 3637-40.
167. Karapetis, C.S., et al., *K-ras mutations and benefit from cetuximab in advanced colorectal cancer*. N Engl J Med, 2008. **359**(17): p. 1757-65.
168. Paz-Ares, L.G., et al., *PARAMOUNT: Final overall survival results of the phase III study of maintenance pemetrexed versus placebo immediately after induction treatment with pemetrexed plus cisplatin for advanced nonsquamous non-small-cell lung cancer*. J Clin Oncol, 2013. **31**(23): p. 2895-902.

169. Do, K., et al., *Biomarker-driven phase 2 study of MK-2206 and selumetinib (AZD6244, ARRY-142886) in patients with colorectal cancer*. Invest New Drugs, 2015. **33**(3): p. 720-8.
170. Tolcher, A.W., et al., *Antitumor activity in RAS-driven tumors by blocking AKT and MEK*. Clin Cancer Res, 2015. **21**(4): p. 739-48.
171. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control*. Nature, 2011. **473**(7347): p. 337-342.
172. Zhang, B., et al., *Proteogenomic characterization of human colon and rectal cancer*. Nature, 2014. **513**(7518): p. 382-7.
173. Simpson, J.A., et al., *Intratumoral T cell infiltration, MHC class I and STAT1 as biomarkers of good prognosis in colorectal cancer*. Gut, 2010. **59**(7): p. 926-33.
174. Mlecnik, B., et al., *Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability*. Immunity, 2016. **44**(3): p. 698-711.
175. Dry, J.R., et al., *Transcriptional Pathway Signatures Predict MEK Addiction and Response to Selumetinib (AZD6244)*. Cancer Research, 2010. **70**(6): p. 2264-2273.
176. Phipps, A.I., et al., *KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers*. Br J Cancer, 2013. **108**(8): p. 1757-64.
177. Guinney, J., et al., *Modeling RAS phenotype in colorectal cancer uncovers novel molecular traits of RAS dependency and improves prediction of response to targeted agents in patients*. Clin Cancer Res, 2014. **20**(1): p. 265-72.
178. Fitzgibbons, P.L., et al., *Principles of analytic validation of immunohistochemical assays: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center*. Arch Pathol Lab Med, 2014. **138**(11): p. 1432-43.
179. Lovig, T., et al., *Strong HLA-DR expression in microsatellite stable carcinomas of the large bowel is associated with good prognosis*. Br J Cancer, 2002. **87**(7): p. 756-62.
180. Lee, L.H., et al., *Patterns and prognostic relevance of PD-1 and PD-L1 expression in colorectal carcinoma*. Mod Pathol, 2016.
181. Mahoney, K.M., et al., *PD-L1 Antibodies to Its Cytoplasmic Domain Most Clearly Delineate Cell Membranes in Immunohistochemical Staining of Tumor Cells*. Cancer Immunol Res, 2015. **3**(12): p. 1308-15.
182. Klampfer, L., et al., *Oncogenic Ki-ras inhibits the expression of interferon-responsive genes through inhibition of STAT1 and STAT2 expression*. J Biol Chem, 2003. **278**(47): p. 46278-87.
183. Komatsu, Y., et al., *Oncogenic Ras inhibits IRF1 to promote viral oncolysis*. Oncogene, 2015. **34**(30): p. 3985-93.
184. Wong, L.H., et al., *Isolation and characterization of a human STAT1 gene regulatory element. Inducibility by interferon (IFN) types I and II and role of IFN regulatory factor-1*. J Biol Chem, 2002. **277**(22): p. 19408-17.
185. Martin, B.K., et al., *Induction of MHC class I expression by the MHC class II transactivator CIITA*. Immunity, 1997. **6**(5): p. 591-600.
186. Gobin, S.J., et al., *Site alpha is crucial for two routes of IFN gamma-induced MHC class I transactivation: the ISRE-mediated route and a novel pathway involving CIITA*. Immunity, 1997. **6**(5): p. 601-11.

187. Wu, C.Y., et al., *IL-12 induces the production of IFN-gamma by neonatal human CD4 T cells*. J Immunol, 1993. **151**(4): p. 1938-49.
188. Hess, S.D., et al., *Human CD4+ T Cells Present Within the Microenvironment of Human Lung Tumors Are Mobilized by the Local and Sustained Release of IL-12 to Kill Tumors In Situ by Indirect Effects of IFN- γ* . The Journal of Immunology, 2003. **170**(1): p. 400-412.
189. Ancrile, B., K.H. Lim, and C.M. Counter, *Oncogenic Ras-induced secretion of IL6 is required for tumorigenesis*. Genes Dev, 2007. **21**(14): p. 1714-9.
190. Sumimoto, H., et al., *The BRAF-MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells*. J Exp Med, 2006. **203**(7): p. 1651-6.
191. Jackson, A.M., et al., *Tumour-mediated disruption of dendritic cell function: inhibiting the MEK1/2-p44/42 axis restores IL-12 production and Th1-generation*. Int J Cancer, 2008. **123**(3): p. 623-32.
192. Spranger, S., R. Bao, and T.F. Gajewski, *Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity*. Nature, 2015. **523**(7559): p. 231-5.
193. Luke, J.J., et al., *Correlation of WNT/ β -catenin pathway activation with immune exclusion across most human cancers*. J Clin Oncol, 2016. **34**(suppl; abstr 3004).
194. Bendell, J., et al., *Clinical activity and safety of cobimetinib (cobi) and atezolizumab in colorectal cancer (CRC)*. J Clin Oncol, 2016. **34**(suppl; abstr 3502).
195. Ebert, P.J., et al., *MAP Kinase Inhibition Promotes T Cell and Anti-tumor Activity in Combination with PD-L1 Checkpoint Blockade*. Immunity, 2016. **44**(3): p. 609-21.
196. Hidalgo J.V., T.A., Ruggieri M., Heise K., Kock S., Malkovsky M., Kuppers R, Fisch P., *Molecular and immunohistochemical analyses of $\gamma\delta$ T cells infiltrating human breast carcinomas*. Gamma Delta Conference 2016 Abstracts, 2016(Abstract C-03).
197. Larkin, J., et al., *Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma*. New England Journal of Medicine, 2015. **373**(1): p. 23-34.
198. Rubio, I., et al., *TCR-induced activation of Ras proceeds at the plasma membrane and requires palmitoylation of N-Ras*. J Immunol, 2010. **185**(6): p. 3536-43.
199. Ebinu, J.O., et al., *RasGRP links T-cell receptor signaling to Ras*. Blood, 2000. **95**(10): p. 3199-203.
200. Liu, W.-T., et al., *LPS-induced CXCR4-dependent migratory properties and a mesenchymal-like phenotype of colorectal cancer cells*. Cell Adhesion & Migration, 2016: p. 1-11.
201. Hsu, R.Y., et al., *LPS-induced TLR4 signaling in human colorectal cancer cells increases beta1 integrin-mediated cell adhesion and liver metastasis*. Cancer Res, 2011. **71**(5): p. 1989-98.
202. Yao, Y., et al., *ERK and p38 MAPK signaling pathways negatively regulate CIITA gene expression in dendritic cells and macrophages*. J Immunol, 2006. **177**(1): p. 70-6.
203. Wu, P., et al., *gammadeltaT17 cells promote the accumulation and expansion of myeloid-derived suppressor cells in human colorectal cancer*. Immunity, 2014. **40**(5): p. 785-800.
204. Lee, M.N., et al., *Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells*. Science, 2014. **343**(6175): p. 1246980-1246980.

205. Fairfax, B.P., et al., *Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression*. Science, 2014. **343**(6175): p. 1246949-1246949.
206. Barreiro, L.B., et al., *Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection*. Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1204-9.
207. Tumeh, P.C., et al., *PD-1 blockade induces responses by inhibiting adaptive immune resistance*. Nature, 2014. **515**(7528): p. 568-71.