

Characterisation of cancer stem cells and the renin-angiotensin system in colon adenocarcinoma

by

Matthew James Munro

A thesis submitted to the Victoria University of Wellington in fulfilment of the
requirements for the degree of Doctor of Philosophy

Victoria University of Wellington

2020

Abstract

Colorectal cancer (CRC) is the third most common cancer and the second highest cause of cancer deaths globally. More than 70% of CRC-related deaths are due to metastasis to the liver. The cancer stem cell (CSC) concept hypothesises that CSCs drive tumour growth, chemoresistance, recurrence and metastasis. Markers such as CD133, LGR5 and EpCAM, have been used to identify and isolate CSCs in CRC. However, these markers are often expressed by cells with no stem cell properties and are not expressed by all tumour-initiating cells. An improved range of markers to define CSCs is needed. In 2007, adult mouse and human fibroblasts were reprogrammed into a stem cell state and defined as induced pluripotent stem cells (iPSCs) using transcription factors OCT4, SOX2, NANOG, KLF4 and c-MYC. These genes have well-documented roles in embryonic development and the maintenance of pluripotency, and their expression has been investigated in a range of cancers.

The renin-angiotensin system (RAS) physiologically maintains blood pressure and volume and is also acknowledged to play a role in cancer. Over-expression of (pro)renin receptor (PRR), angiotensin II type 1 receptor (AT₁R) and type 2 receptor (AT₂R), and angiotensin-converting enzyme (ACE) have been reported in cancer. Epidemiological studies investigating the effect of RAS inhibitors on cancer outcomes have shown contradictory results.

This thesis investigates the expression of iPSC markers and RAS components in colon adenocarcinoma (CA) with three specific aims: (1) to compare CA-derived primary cell lines to their original CA tissues; (2) to investigate the expression profiles of iPSC markers in CA; and (3) to investigate expression of RAS components by CA CSCs and to determine whether CSCs can be targeted by RAS modulators.

DNA sequencing was carried out to compare the mutational profiles of formalin-fixed paraffin-embedded (FFPE) CA tissues and CA-derived cell lines to confirm whether the cell lines were a suitable *in vitro* model for the parent tumours.

Proteomics was performed to determine proteomic differences between CA tissues and patient-matched normal colon (NC) tissues, CA-derived cell lines and NC-derived cells, and between low grade CA (LGCA) tissues and cell lines and high grade CA (HGCA) tissues and cell lines. Biological processes which may link the RAS and CA were investigated, revealing enrichment of various signalling pathways that may play roles in CA onset and progression directly or via the RAS.

Western blotting and immunohistochemical staining showed elevated protein levels of OCT4, SOX2, NANOG, c-MYC, AT₂R, PRR and cathepsin D in CA tissues relative to their patient-matched NC tissues, with SOX2, ACE and cathepsin B at similar levels and KLF4 less abundant in CA compared with

NC tissues. Co-expression analysis by immunofluorescence staining showed a small number of epithelial cells co-expressed NANOG, SOX2, KLF4, c-MYC and CD133, as well as PRR, ACE2 and AT₂R, while a small number of stromal cells co-expressed OCT4 and AT₂R. This indicates the presence of at least one CSC subpopulation in CA, which expresses RAS components. HGCA tissue-derived cell lines expressed higher levels of OCT4 and SOX2 than LGCA-derived cell lines. The primary cell lines were sorted based on EpCAM expression. These EpCAM^{High} and EpCAM^{Low} cell subpopulations could undergo directed differentiation down the three embryonic lineages. A small number of CA-derived cells, particularly within the HGCA-derived cells, formed tumourspheres. Treatment of HGCA-derived cell lines with RAS modulators revealed that β -blockers and AT₂R antagonists consistently reduced their metabolism, tumoursphere formation and iPSC marker expression.

The findings of this thesis suggest that CA-derived cell lines expressing iPSC markers have stem cell function and express RAS components. Furthermore, RAS modulators may directly influence CSCs in CA by reducing iPSC marker gene expression. This indicates a potential role for RAS modulators in regulating CSCs, which merits further investigation.

Acknowledgements

Firstly, thank you to my supervisors, Dr Lifeng Peng, Dr Swee Tan, and Dr Susrutha Wickremesekera, for your dedication, for your belief, for sharing your scientific and clinical knowledge, for critiquing and proof-reading hundreds of documents, and for guiding me into a position where I could take ownership of my project. Your valuable experience and infinite patience have been an inspiration.

Thank you to Mr John Groom and associates, who performed the surgeries and supplied tissue samples to the GMRITB. Also, thank you to the patients who donated their tissues to the GMRITB.

Special thanks to Ms Liz Jones, the GMRI histologist, for all her time and effort in preparing tissue blocks and sections, performing IHC and IF staining and assisting with ordering consumables. You really keep the lab functioning!

I would also like to acknowledge the hours which Dr Helen Brasch, GMRI and DHB pathologist, devoted to selecting the cases which fit our parameters, checking the staining, teaching me with patience and insight what I needed to know to get by, and advocating for my cause.

This project would not have gone ahead without the funding of the NZ Community Trust; thank you for your support through a PhD scholarship presented to the GMRI. A special thank you as well to the Yellow Chili Fund, which contributed towards my attendance at the Royal Australasian College of Surgeons' 88th Annual Scientific Congress in Bangkok and helped bridge the gap between the end of the NZCT scholarship and my submission.

Thank you to Victoria University of Wellington and the School of Biological Sciences for your support during my study, including the PhD Doctoral Submission Scholarship which allowed me time to focus on writing my thesis. In particular, thank you to Lifeng and the rest of the proteomics lab for your assistance running the mass spectrometer.

Thank you to the various GMRI technicians whom I have worked with through the years, including Nick Bockett, Erin Paterson, Josie Patel, Bede van Schaijik, Jen de Jongh and Susana Enriquez. Your collective experience was the most helpful thing to me in the lab, and you all had unique perspectives from which to address a scientific question. You have helped open my mind to the many ways in which we can think about science. I am also grateful that I have had the chance to form friendships with some of you outside of the lab as well.

The GMRI Research Nurses, Carolyn Croasdale and Merie Claridge, were super helpful when it came to collating anonymised data from the patient cohort, as well as having great lunchtime chat.

Dr Sean Hall arrived at the GMRI from Switzerland to take up the position of Chief Scientist just in time for us to have valuable conversations regarding appropriate experimental controls, the importance of a good post-doctoral project, and the strangeness of what we might think of as quintessential NZ foods.

To my cousin, Dr Kelly Paton, for reassuring me that I was on track during the last hectic months and that all the little annoying details tend to smooth themselves out, especially when there are other people around you who are going through, or have recently been through, the same thing. Your experience and your calmness helped me realise that things were not usually as big of a deal as they seemed in my head.

For their help in the early days, either in establishing the project direction or helping me find my feet, I would like to acknowledge Dr Tinte Itinteang, Dr Paul Davis, Dr Reginald Marsh and Dr Melanie McConnell. Also, thank you to GMRI EA Cindy Naresh, for keeping us all in line, keeping the tea-and-biscuit cupboard fully stocked, and orchestrating everything that goes on in the GMRI offices while hardly breaking a sweat.

Thank you, Dr Sean Mackay, for your donation of the proprietary compound SMM02 to the GMRI, which was a valuable inclusion to my experiments.

Finally, thank you to my wife, Ingrid; my parents, Paul and Liz; and my brother and sister-in-law, David and Jodie; for all the behind-the-scenes work you have done to make sure I relax and laugh and play games and watch rugby, and most importantly to think about something other than stem cells or the renin-angiotensin system for 5 minutes; for being patient and curious, and for supporting me in many different ways.

Table of Contents

Abstract	i
Acknowledgements	iii
List of Figures	x
List of Tables	xiii
Abbreviations	xv
Chapter 1. Introduction	1
1.1 The Colon	1
1.2 Colorectal Cancer	5
1.2.1 Colorectal Cancer Incidence, Treatment and Mortality	5
1.2.2 Staging of CRC	6
1.2.3 Grading of CRC	6
1.3 Genetic Causes of CRC	8
1.3.1 Hereditary Disorders	8
1.3.2 Microsatellite Instability	9
1.3.3 CpG Island Methylator Phenotype	10
1.3.4 Chromosomal Instability	10
1.3.5 Common Mutations in CRC	10
1.3.6 Consensus Molecular Subtypes	12
1.4 Cancer Stem Cells	13
1.4.1 The Cancer Stem Cell Concept	13
1.4.2 Stem Cell Markers	13
1.4.3 Issues of current CSC markers	19
1.4.4 Pluripotency Markers	20
1.4.5 Induced Pluripotent Stem Cells	23
1.5 The Renin-Angiotensin System in Cancer	23
1.5.1 Renin-Angiotensin System	23
1.5.2 (Pro)renin Receptor	24
1.5.3 PRR and Wnt Signalling	25
1.5.4 The RAS in Cancer	28
1.5.5 Cathepsins as Bypass Loops of the RAS	31
1.6 Epithelial-to-Mesenchymal Transition	33
1.7 Aims and Hypotheses	34
Chapter 2: Methods and materials	37
2.1 Tissue Samples	37

2.2 Immunohistochemistry	37
2.2.1 DAB Immunohistochemical staining	37
2.2.2 Positive controls	37
2.2.3 Primary antibodies and staining protocols	37
2.2.4 Immunofluorescence staining	38
2.2.5 <i>In situ</i> Hybridisation	39
2.2.6 Image capture and analysis	39
2.3 Western blotting	39
2.3.1 Protein extraction from tissue	39
2.3.2 Protein extraction from cells	40
2.3.3 Protein quantification	40
2.3.4 Gel electrophoresis	41
2.3.5 Western blotting	41
2.4 Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR)	42
2.4.1 RNA Extraction	42
2.4.2 Rotor Gene PCR Protocol	43
2.5 Cell Culture	44
2.5.1 Cell culture	44
2.5.2 Cell sorting	44
2.5.3 Tumoursphere formation assays	45
2.5.4 Immunocytochemistry	46
2.5.5 Mesoderm differentiation assay	47
2.5.6 Endoderm differentiation assay	47
2.5.7 Ectoderm differentiation assay	48
2.5.8 Cathepsin activity assays	48
2.5.9 RAS modulation drug assays	49
2.6 Proteomics	50
2.6.1 Sample definition	50
2.6.2 Protein extraction	50
2.6.3 Protein precipitation and digestion	51
2.6.4 Liquid Chromatography Tandem Mass Spectrometry	51
2.6.5 Protein identification	52
2.6.6 Label-free quantitation	53
2.6.7 Functional analysis	54
2.7 DNA Sequencing	54
2.7.1 Sample preparation	54

2.7.2 DNA quality validation	55
2.7.3 DNA sequencing	56
2.7.4 Data analysis	56
Chapter 3: Validating Colon Adenocarcinoma Tissue-derived Primary Cell Lines by DNA Sequencing	57
3.1 Introduction	57
3.2 Results	58
3.2.1 DNA Quality Assurance	58
3.2.2 Summary of Oncomine™ TML Assay data	58
3.2.3 Mutational signatures of CA tissues and CA-derived cells.....	60
3.2.4 Types of nucleotide substitutions	63
3.3 Discussion	64
3.4 Conclusion	66
Chapter 4: Cancer Stem Cell Subpopulations in Primary Colon Adenocarcinoma.....	67
4.1 Introduction	67
4.2 Results	68
4.2.1 EpCAM expression in the normal colon and CA	68
4.2.2 OCT4 is expressed in the CA stroma	69
4.2.3 SOX2 is expressed in CA	70
4.2.4 The NANOG protein is expressed in CA but not NC.....	70
4.2.5 KLF4 abundance is inversely proportional to CA grade	73
4.2.6 NC and CA express high levels of c-MYC.....	73
4.2.7 CD133 and LGR5 IHC staining	73
4.2.8 IF staining	74
4.3 Discussion	77
4.4 Conclusion	80
Chapter 5: Colon Adenocarcinoma-derived Cells that Express Induced-Pluripotent Stem Cell Markers Possess Stem Cell Function	82
5.1 Introduction	82
5.2 Results	82
5.2.1 Co-localisation of pluripotency markers by Immunocytochemistry.....	82
5.2.2 HGCA-derived cells form tumourspheres	86
5.2.3 CA-derived cells display multilineage differentiation capacity.....	89
5.2.4 RT-qPCR.....	92
5.2.5 Western blotting	93
5.3 Discussion	97

5.4 Conclusion	101
Chapter 6: Changes in the Proteomes of Colon Adenocarcinoma Tissues and Cell Lines Compared to the Normal Colon	102
6.1 Introduction	102
6.2 Results	103
6.2.1 Protein identification	103
6.2.2 Label-free quantification	109
6.2.3 Functional analysis of differentially expressed proteins	110
6.2.4 Comparison of protein changes seen in CA tissues and CA-derived cells	127
6.2.5 Validation of mass spectrometry results by western blotting	128
6.3 Discussion	133
6.3.1 Comparing CA-derived cell lines to CA tissues	133
6.3.2 Similarities with other publications	135
6.3.3 Detection of RAS components by LC-MS/MS	139
6.3.4 Integration of the RAS with other signalling pathways	140
6.3.5 Convergence of the RAS and Wnt signalling	148
6.3.6 The RAS and pluripotency	149
6.4 Conclusions	151
Chapter 7: Primary Colon Adenocarcinoma Stem Cells Express Components of the Renin-Angiotensin System	153
7.1 Introduction	153
7.2 Results	156
7.2.1 ACE and ACE2 are expressed variably by CA tissues and CA-derived cell lines	156
7.2.2 AT ₂ R expression in CA tissues is proportional to grade	161
7.2.3 PRR is upregulated in CA tissues and CA-derived cells	161
7.2.4 Cathepsins are expressed in the CA epithelium and stroma	164
7.2.5 Cells expressing OCT4 and NANOG co-express RAS components	168
7.2.6 Cathepsin activity assays	170
7.3 Discussion	171
7.4 Conclusions	174
Chapter 8: The Effects of Renin-Angiotensin System Modulation on Metabolism and Stem Cell Phenotype of Colon Adenocarcinoma derived Primary Cell Lines	176
8.1 Introduction	176
8.1.1 β -Blockers	176
8.1.2 ACE Inhibitors	178
8.1.3 Antagonists of AT ₁ R and AT ₂ R	179

8.2 Results	180
8.2.1 Cell metabolism assays	180
8.2.2 Effect of RAS modulation on tumoursphere formation	187
8.2.3 RT-qPCR on treated cells	189
8.3 Discussion	190
8.4 Conclusions	193
Chapter 9: General Discussion	195
9.1 Key findings	195
9.2 Clinical Significance	200
9.3 Limitations and Future Directions	201
9.4 Concluding Remarks	205
References	206
Appendices	226
Appendix A: Supplementary figures and tables	226
Appendix B: List of Publications	266

List of Figures

Figure 1.1: Cellular organisation of the intestinal epithelium.....	3
Figure 1.2: Cells of the colonic epithelium.....	4
Figure 1.3: Crypts of the normal colon and colon adenocarcinoma.....	7
Figure 1.4: Classical RAS and contemporary additions.....	27
Figure 1.5: PRR and Wnt signalling.....	28
Figure 4.1: IHC staining for EpCAM.....	68
Figure 4.2: iPSC marker IHC staining.....	69
Figure 4.3: RT-qPCR data from tissues.....	71
Figure 4.4 <i>In situ</i> hybridisation.....	72
Figure 4.5 Stem cell marker IHC staining.....	74
Figure 4.6 IF staining.....	76
Figure 5.1: Representative SSEA4/OCT4 ICC images.....	84
Figure 5.2: Representative SOX2/TRA-1-60 ICC images.....	85
Figure 5.3: Representative images of tumoursphere formation assays.....	88
Figure 5.4: Representative images of Mesoderm differentiation.....	89
Figure 5.5: Representative images of Endoderm differentiation.....	90
Figure 5.6: Representative images of Ectoderm differentiation.....	91
Figure 5.7: RT-qPCR data from EpCAM ^{Low} and EpCAM ^{High} cells.....	93
Figure 5.8: Western blotting for iPSC and stem cell markers in cells.....	95
Figure 5.9: Densitometry performed on western blot.....	96
Figure 6.1 The distribution of proteins identified in each LC-MS/MS technical replicate run for each LGCA-derived cell line.....	106
Figure 6.2 The distribution of proteins identified in each LC-MS/MS technical replicate run for each HGCA-derived cell line.....	107
Figure 6.3 The distribution of proteins identified in each tissue biological replicate.....	108
Figure 6.4: Heat map displaying the relative abundances of all quantified proteins in NC, HGCA and LGCA tissues.....	111
Figure 6.5: Cluster 1 from the tissue samples heat map.....	112
Figure 6.6: Cluster 2 from the tissue samples heat map.....	113
Figure 6.7: Cluster 3 from the tissue samples heat map.....	114

Figure 6.8: Cluster 4 from the tissue samples heat map.....	115
Figure 6.9: Cluster 5 from the tissue samples heat map.....	116
Figure 6.10: Heat map displaying the relative abundances of all quantified proteins in NC, HGCA and LGCA-derived cell lines.....	118
Figure 6.11: Cluster 1 from the cell lines heat map.....	119
Figure 6.12: Cluster 2 from the cell lines heat map.....	120
Figure 6.13: Cluster 3 from the cell lines heat map.....	121
Figure 6.14: Cluster 4 from the cell lines heat map.....	122
Figure 6.15: Cluster 5 from the cell lines heat map.....	123
Figure 6.16: CD44 Densitometry for NC and CA-derived cell lines.....	129
Figure 6.17: FN1 densitometry for NC and CA-derived cell lines.....	130
Figure 6.18: S100A8 and S100A9 Densitometry for tissues.....	132
Figure 6.19: AT ₁ R-PI3K/AKT signalling.....	142
Figure 6.20: AT ₁ R-MAP kinase signalling.....	145
Figure 6.21: AT ₁ R-c-Src-JAK/STAT signalling.....	146
Figure 6.22: Synthesis of relevant pathways.....	150
Figure 7.1: IHC staining for ACE.....	157
Figure 7.2: IHC staining for ACE2.....	158
Figure 7.3: Western blot for RAS components.....	159
Figure 7.4: RT-qPCR for RAS components in CA tissues.....	160
Figure 7.5: RT-qPCR for RAS components in CA-derived cells.....	160
Figure 7.6: IHC staining for AT ₂ R.....	162
Figure 7.7: IHC staining for PRR.....	163
Figure 7.8: IHC staining for cathepsin B.....	165
Figure 7.9: IHC staining for cathepsin D.....	166
Figure 7.10: IHC staining for cathepsin G.....	167
Figure 7.11: Co-localisation of iPSC markers with RAS components by IF staining.....	169
Figure 7.12: Cathepsin activity assays.....	171
Figure 8.1: Effect of propranolol on the metabolism of HGCA-derived cell lines.....	182
Figure 8.2: Effect of timolol on the metabolism of HGCA-derived cell lines.....	183
Figure 8.3: Effect of ACEIs on the metabolism of HGCA-derived cell lines.....	184

Figure 8.4: Effect of ARBs on the metabolism of HGCA-derived cell lines.....	185
Figure 8.5: Effect of AT ₂ R antagonists on the metabolism of HGCA-derived cell lines.....	186
Figure 8.6: Effect of RAS modulators on tumoursphere formation.....	188
Figure 8.7: mRNA levels of iPSC markers in HGCA-derived cells treated with RAS modulators.....	189
Figure 9.1: Representative images of CA cases used for DNA sequencing, mass spectrometry and cell cultures.....	199
Figure A4.1: Supplementary iPSC marker IHC stained images.....	230
Figure A4.2: iPSC marker IHC staining positive controls.....	231
Figure A4.3: iPSC marker IHC staining negative controls.....	232
Figure A4.4: iPSC marker IF staining negative controls.....	233
Figure A4.5: <i>In situ</i> hybridisation controls.....	234
Figure A5.1: Representative images of ICC negative controls.....	239
Figure A5.2: Representative images of ICC staining for sample LGCA1.....	240
Figure A5.3: Representative images of ICC staining for sample LGCA2.....	241
Figure A5.4: Representative images of ICC staining for sample LGCA3.....	242
Figure A5.5: Representative images of ICC staining for sample HGCA1.....	243
Figure A5.6: Representative images of ICC staining for sample HGCA2.....	244
Figure A5.7: Representative images of ICC staining for sample HGCA3.....	245
Figure A5.8: Representative images of tumoursphere formation assay positive controls.....	246
Figure A5.9: Mesodermal differentiation positive controls.....	246
Figure A5.10: Mesodermal differentiation negative controls.....	247
Figure A5.11: Endodermal differentiation negative controls.....	248
Figure A5.12: Endodermal differentiation negative controls.....	249
Figure A5.13: Endodermal differentiation control cell line.....	250
Figure A5.14: Ectodermal differentiation negative controls.....	251
Figure A5.15: Ectodermal differentiation negative controls.....	252
Figure A5.16: Ectodermal differentiation control cell line.....	252
Figure A5.17: Unmodified western blot images.....	253
Figure A6.1: Heat map for individual tissue samples.....	254
Figure A6.2: Heat map for individual cell lines.....	255
Figure A7.1: IHC staining positive and negative controls.....	256

Figure A7.2: Unmodified western blot images.....	257
Figure A7.3: iPSC marker IF staining negative controls.....	258

List of Tables

Table 2.1: Primary antibodies used for immunohistochemical and immunofluorescence staining.....	37
Table 2.2: Primary antibodies used for western blotting.....	41
Table 3.1: Summary of TML assay results for CA tissues and CA-derived primary cell lines.....	59
Table 3.2: Mutations to CA-related genes in FFPE CA tissues and CA-derived primary cell lines....	60-62
Table 3.3: Summary of coverage depth and DNA substitutions.....	64
Table 5.1. Analysis of tumoursphere formation assay data.....	86
Table 5.2. Data from passaged tumourspheres.....	87
Table 6.1: Number of proteins identified in each technical replicate LC-MS/MS run for each CA-derived cell line.....	103
Table 6.2: Total number of proteins identified in each CA-derived cell line biological replicate across three LC-MS/MS technical replicate runs.....	104
Table 6.3: Total number of proteins identified in each CA tissue biological replicate across three LC-MS/MS technical replicate runs.....	104
Table 6.4: Mass spectrometry summary data.....	109
Table 6.5: Functions of proteins significantly downregulated in LGCA and HGCA tissues compared to NC tissues.....	125
Table 6.6: Functions of proteins significantly upregulated in HGCA tissues compared to LGCA tissues.....	125
Table 6.7: Functions of proteins significantly downregulated in HGCA tissues compared to LGCA tissues.....	125
Table 6.8: Functions of proteins significantly upregulated in LGCA and HGCA-derived cell lines compared to the NC cell line.....	126
Table 6.9: Functions of proteins significantly downregulated in LGCA and HGCA-derived cell lines compared to the NC cell line.....	126
Table 6.10: Functions of proteins significantly upregulated in HGCA-derived cell lines compared to the LGCA-derived cell lines.....	127
Table 6.11: Functions of proteins significantly downregulated in HGCA-derived cell lines compared to the LGCA-derived cell lines.....	127
Table 6.12: Peptidase inhibitors with differential expression in CA-derived cell lines compared to the NC cell line.....	134
Table 6.13: Comparison with TCGA study.....	136

Table 6.14: Comparison with Vasaikar <i>et al</i>	138
Table 6.15: RAS components detected in tissues by LC-MS/MS.....	139
Table 6.16: RAS components detected in cell lines by LC-MS/MS.....	140
Table 6.17: Caveolae-related proteins detected in NC and CA tissues and tissue-derived cell lines..	141
Table 6.18: PI3K signalling-related proteins detected in NC and CA tissues and tissue-derived cell lines.....	143
Table 8.1 Analysis of tumoursphere formation from treated cells.....	187
Table A1.1: Patient data.....	227-229
Table A4.1: Supplementary data for iPSC marker RT-qPCR in tissues.....	235-238
Table A7.1: Supplementary data for RAS components RT-qPCR in CA tissues and CA-derived cells.....	259-264

Abbreviations

ACE: angiotensin-converting enzyme
ACE2: angiotensin-converting enzyme 2
ACEI: angiotensin-converting enzyme inhibitor
ARB: angiotensin receptor blocker
ATG: angiotensinogen
ATI: angiotensin I
ATII: angiotensin II
AT₁R: ATII type 1 receptor
AT₂R: ATII type 2 receptor
CA: colon adenocarcinoma
CIN: chromosome instability
CRC: colorectal cancer
CSC: cancer stem cell
CTSB: cathepsin B
CTSD: cathepsin D
CTSG: cathepsin G
DFS: disease-free survival
DMEM: Dulbecco's modified eagle medium
FCS: fetal calf serum
FFPE: formalin-fixed paraffin embedded
HGCA: high-grade colon adenocarcinoma
HPLC: high performance liquid chromatography
ICC: immunocytochemistry
IF: immunofluorescence
IHC: immunohistochemical
INDEL: insertion-deletion mutation
iPSC: induced pluripotent stem cell
ISH: *in situ* hybridisation
LGCA: low-grade colon adenocarcinoma

MAP: mitogen-activated protein

Mb: megabases

MMP: matrix metalloproteinase

MMR: mismatch repair

LC-MS/MS: liquid chromatography tandem mass spectrometry

MSI: microsatellite instability

MSS: microsatellite-stable

NC: normal colon

NGS: next-generation sequencing

OS: overall survival

PCR: polymerase chain reaction

PFS: progression-free survival

PRR: (pro)renin receptor

PtdIns: phosphatidylinositol

PTM: post-translational modification

RTK: receptor tyrosine kinase

RT-qPCR: reverse transcription quantitative PCR

SNV: single nucleotide variant

TCGA: the cancer genome atlas

TIMP: tissue inhibitor of metalloproteinases

TMB: tumour mutational burden

WB: western blot

WXS: whole exome sequencing

Chapter 1. Introduction

1.1 The Colon

The luminal surface of the gut consists of a single layer of columnar epithelial cells which form folds called the crypts of Lieberkühn – the functional unit of the intestine (Khalek et al., 2010) (Figure 1.1). The epithelium of the gut contains cells of four distinct lineages: enterocytes, goblet cells, endocrine cells and Paneth cells (Takashima et al., 2013) (Fig 1.1 and Fig 1.2).

Crypt base columnar cells (CBCs) are small undifferentiated cells thought to be the true intestinal stem cells (ISCs) which give rise to the epithelial lineages (Humphries et al., 2008; Khalek et al., 2010; Takashima et al., 2013; van der Wath et al., 2013) (Fig 1.1). These stem cells possess the ability to divide asymmetrically giving rise to identical ISCs, and transit amplifying cells which proliferate and differentiate into enterocytes, goblet cells and endocrine cells during their upward movement through the crypt (Humphries et al., 2008; Khalek et al., 2010; Takashima et al., 2013) (Fig 1.1 and Fig 1.2). These three epithelial cell types occupy the top half of the crypt as well as the luminal epithelial surface. Paneth cells differentiate from transit amplifying cells as they move back downwards to the crypt base, where they are eventually found amongst the stem cell/CBC cell population (Khalek et al., 2010). Paneth cells may contribute to maintaining the stem cell niche by producing mucosal defence barriers, modulating intestinal microflora and producing growth factors and other regulatory molecules (Khalek et al., 2010; Sato et al., 2011; van der Wath et al., 2013). Paneth cells are only found in the right side (proximal) of the colon, but are seen in the left side (distal) of the colon in states of disease or inflammation (such as cancer, ulcerative colitis or Crohn's disease) and due to certain diets associated with Western lifestyle (Boman et al., 2008). However, the normal colon does contain a population of CD24⁺ cells thought to play a role equivalent to that of Paneth cells (Boman et al., 2008). The lack of Paneth cells, and hence reduced support for ISCs, in the distal colon may contribute to the observation that the prevalence of cancer is four times greater in the colon relative to the small intestine, despite the epithelium of both renewing at a similar rate (Boman et al., 2008).

The epithelium of the gut is replaced every 3-5 days in humans due to the harsh, acidic environment (Boman et al., 2008; van der Wath et al., 2013). This requires tight regulation of stem cell proliferation and differentiation, and defects in this process are responsible for inflammation and cancer in the intestine (Takashima et al., 2013; van der Wath et al., 2013). The lamina propria harbours a large

number of immune cells to prevent infection or contamination introduced by ingestion (van der Wath et al., 2013). Beneath the gut epithelium lies the basement membrane, separating the epithelium from the lamina propria mucosae (Fleming et al., 2012). The epithelium and the lamina propria are collectively called the mucosa. A layer of smooth muscle called the muscularis mucosae sits between the mucosa and submucosa (Fleming et al., 2012). The submucosa contains arteries and veins that supply the gut with blood. Underneath the submucosa is the muscularis externa or muscularis propria, consisting of the inner circular muscle layer and the outer longitudinal muscle layer, which is responsible for peristalsis (Fleming et al., 2012). The outer layer of the gut is surrounded by the adventitia and another epithelial layer. The lamina propria, submucosa and adventitia are all comprised of loose areolar connective tissue (Fleming et al., 2012).

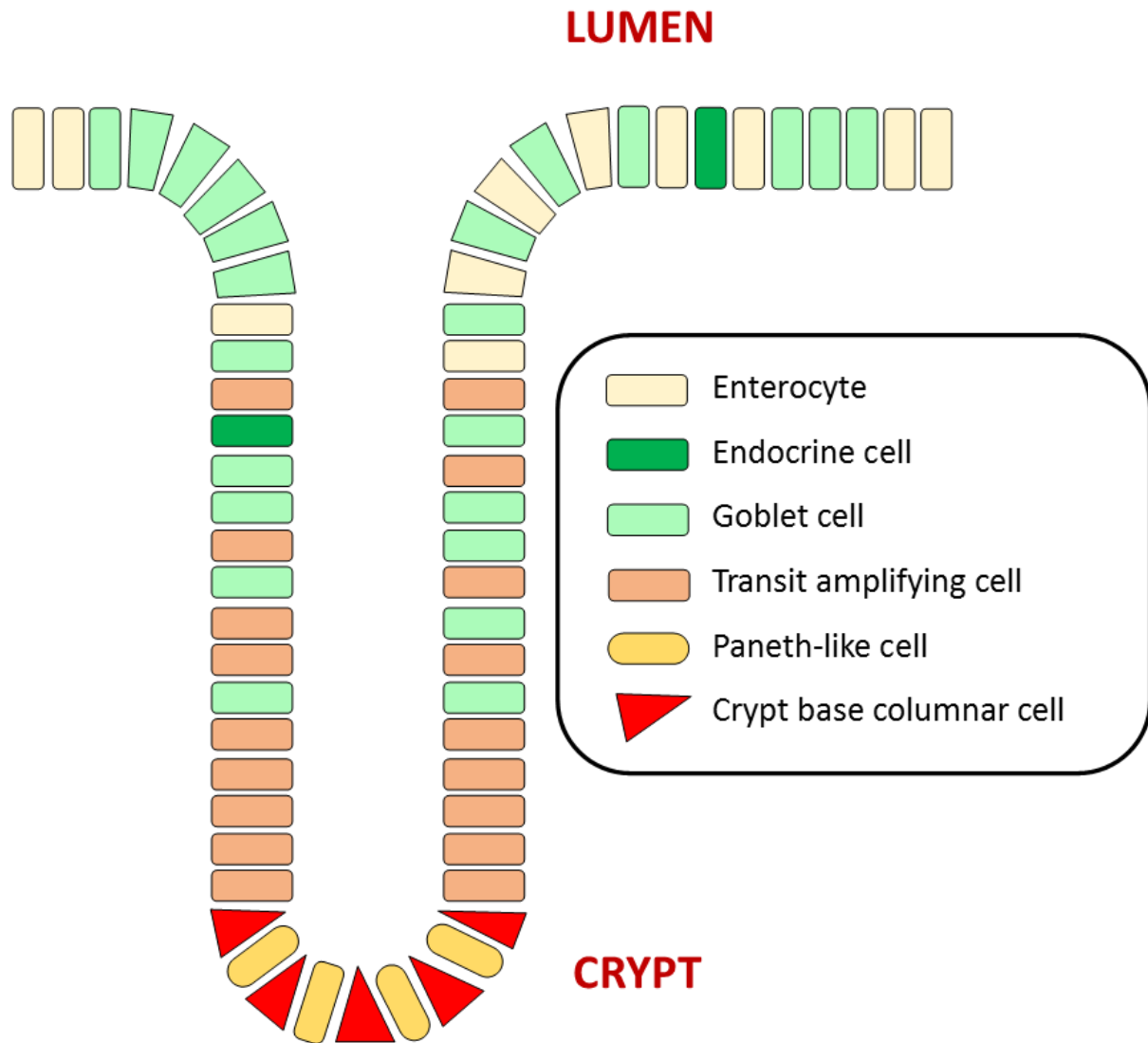


Figure 1.1: Cellular organisation of the intestinal epithelium. Crypt base columnar cells, the proposed intestinal stem cell, divide to produce an identical daughter stem cell and a transit amplifying cell, which differentiate into the four epithelial cell types as they migrate upwards through the crypt. Enterocytes, goblet cells and endocrine cells remain in the luminal epithelium, whereas Paneth cells return to the crypt base.

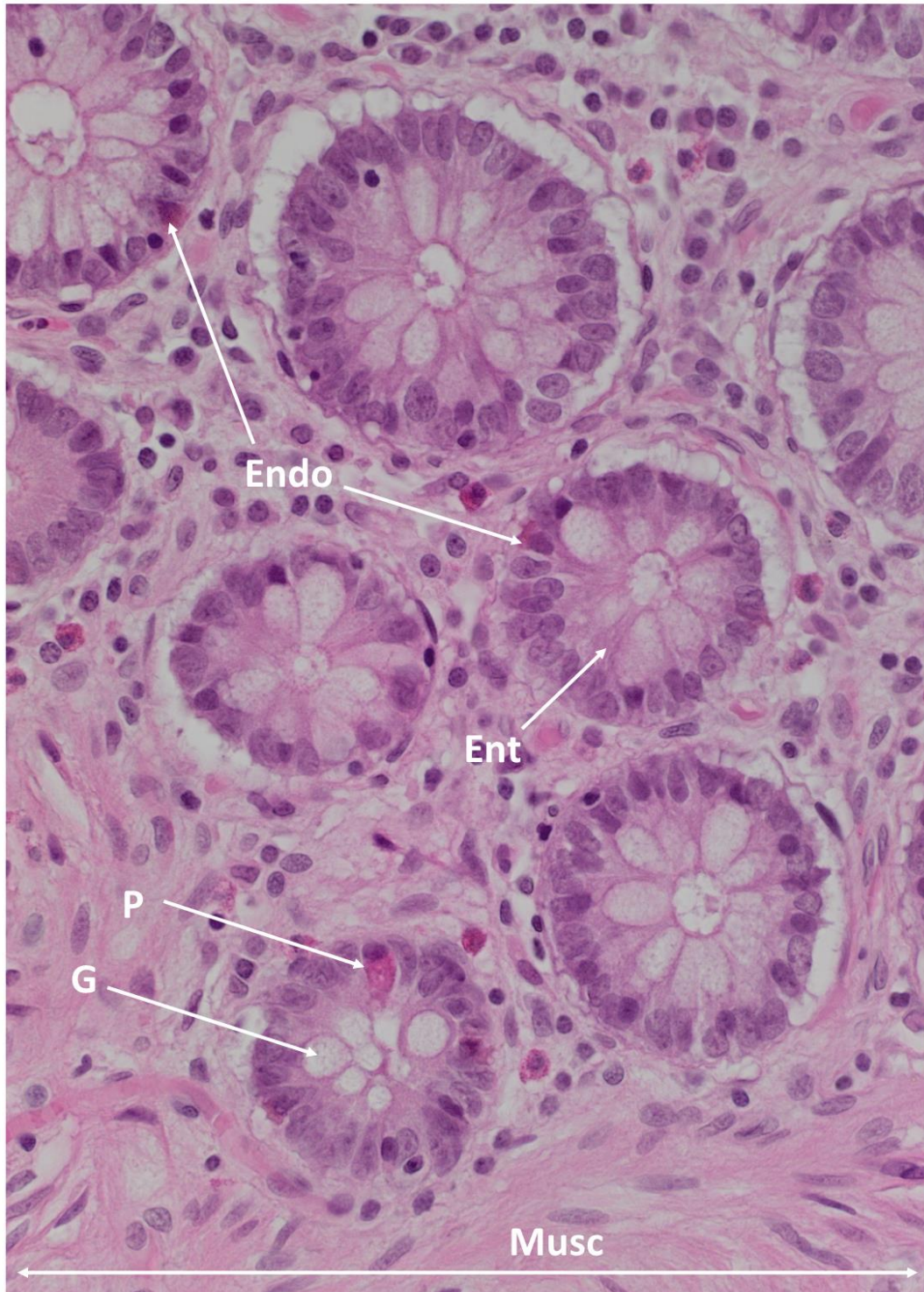


Figure 1.2: Cells of the colonic epithelium. Endo = Endocrine cell, Ent = Enterocyte, P = Paneth cell, G = Goblet cell, Musc = Muscularis mucosae. Original magnification = 400x.

1.2 Colorectal Cancer

1.2.1 Colorectal Cancer Incidence, Treatment and Mortality

The World Health Organisation (WHO) ranks colorectal cancer (CRC) as the third most common cancer type, accounting for 10.2% of new cancer cases in 2018, and is the second highest cause of cancer deaths. Adenocarcinoma arises from the colonic epithelium and accounts for 90% of CRC cases (Fleming et al., 2012). More than 70% of CRC-related deaths are caused by metastases to the liver (Neo et al., 2010). The first line treatment for early-stage colon cancer is surgical resection, which may be preceded (neoadjuvant) or followed by (adjuvant) chemotherapy for later stage disease (Braun et al., 2011; Werner et al., 2016). Chemotherapeutic options usually involve combinations of oxaliplatin (OX), 5-fluorouracil (5-FU), leucovorin/folinic acid (FA), and irinotecan (IR) – these combinations are offered as FOLFOX (OX, 5-FU and FA), FOLFIRI (IR, 5-FU and FA), or FOLFOXIRI (all four) (Braun et al., 2011; Goldberg et al., 2004; A. Rahman, 2014). While surgery can potentially be curative, recurrence occurs in up to 70% of cases (Neo et al., 2010). Recurrent or metastatic CRC can sometimes be operated on, but inoperable cases are generally treated by palliative chemotherapy (Adam et al., 2004; Neo et al., 2010).

Western countries account for almost two-thirds of CRC cases globally, predominantly in Australia, New Zealand, Canada, United States and Europe, although it is becoming more common in China, India, Eastern Europe, South America and Africa (Ghoncheh et al.; Hagggar et al., 2009). The increased incidence of CRC has been linked with increased developmental status of a country, although a decrease in both the incidence and mortality occurs once a country attains a “highly developed” status (Fidler et al., 2017; Ghoncheh et al.; Siegel et al., 2017). These statistics may be influenced by the presentation and reporting rates in developing countries (Hagggar et al., 2009).

The incidence of CRC increases dramatically after age 50, with 90% of cases falling into this age bracket (Hagggar et al., 2009; Siegel et al., 2017). However, 65% of male CRC deaths occur between the age of 50 and 80, compared to 53% of female deaths (Siegel et al., 2017). Furthermore, 40% of CRC-related deaths in women occur after age 80, whereas in males this age bracket accounts for 27% of deaths (Siegel et al., 2017).

1.2.2 Staging of CRC

CRC confined to the wall of the colon (stages I and II) is regarded as early-stage and is potentially curable (Markowitz et al., 2009). It has a 5-year survival rate of 70-90%, however most countries do not have a screening programme to detect the disease at an early stage of its development (Ghoncheh et al.; Siegel et al., 2017). In contrast, the average 5-year survival for regional-stage (stage III) and distant-stage (metastatic; stage IV) CRC is approximately 50-70% and 10-14%, respectively (Haggar et al., 2009; Markowitz et al., 2009; Siegel et al., 2017). This is mainly attributed to the tumour breaching the intestinal wall and spreading via the lymphatics and bloodstream to lymph nodes and distant organs, respectively (Fleming et al., 2012). For CRC to be classed as invasive it must penetrate the muscularis mucosa; this is a point of difference from other cancers, which may be termed invasive following their infiltration of the basement membrane (Fleming et al., 2012).

TNM pathological staging is currently the best predictor of clinical outcome for CRC patients. It determines the depth of invasion of the primary tumour (T-stage) and metastasis to lymph nodes (N-stage) and distant organs (M-stage) (Fleming et al., 2012). The first three degrees of tumour invasion are invasion into the submucosa (T1), into the muscularis externa/propria (T2), and through the muscularis into pericorectal tissue (T3), respectively (Fleming et al., 2012). T4a describes a CRC in which tumour cells have reached the visceral peritoneum at the serosal surface, with an inflammatory reaction, mesothelial hyperplasia and/or erosion of tissue, as well as ulceration of the visceral peritoneum (Fleming et al., 2012). A T4b tumour displays invasion or adherence to other organs or structures (Fleming et al., 2012). CRC is also staged based on its spread to the lymph nodes, designated as N1a (1 node), N1b (2-3 nodes), N2a (4-6 nodes) or N2b (7+ nodes) (Fleming et al., 2012). Tumour deposits are small clusters of tumour cells which may represent discontinuous extension of the tumour (T3), venous invasion, or a lymph node which is totally replaced by tumour cells, and they are considered to be nodal metastases associated with higher relapse rates and poorer survival (Fleming et al., 2012).

1.2.3 Grading of CRC

Adenocarcinomas comprise more than 90% of CRCs (Fleming et al., 2012) and fall into three categories of differentiation based on gland formation (Fig 1.3). A tumour is well differentiated if >95% of the tumour is comprised of glands and crypts; moderately differentiated if it displays 50-95% gland and crypt composition; and poorly differentiated when the tumour is comprised of densely packed tumour cells and has <50% gland or crypt composition (Fleming et al., 2012). Due to the subjectivity of grading CRCs, colon

adenocarcinoma (CA) grading has generally used a 2-tier system with low-grade CA (LGCA) comprising both well and moderately differentiated tumours of more than 50% gland formation (Fig 1.3B), and high-grade CA (HGCA) being all poorly differentiated tumours of <50% gland formation (Fig 1.3C) (Fleming et al., 2012). Approximately 70% of CAs are moderately differentiated when histological analysis is undertaken (Fleming et al., 2012). This grading system does not apply to histological variants or rare cases, such as squamous cell carcinoma, adenosquamous carcinoma, spindle cell carcinoma, neuroendocrine tumours and undifferentiated carcinoma, which can be influenced by the microsatellite-instability (MSI) status of the tumour (Fleming et al., 2012).

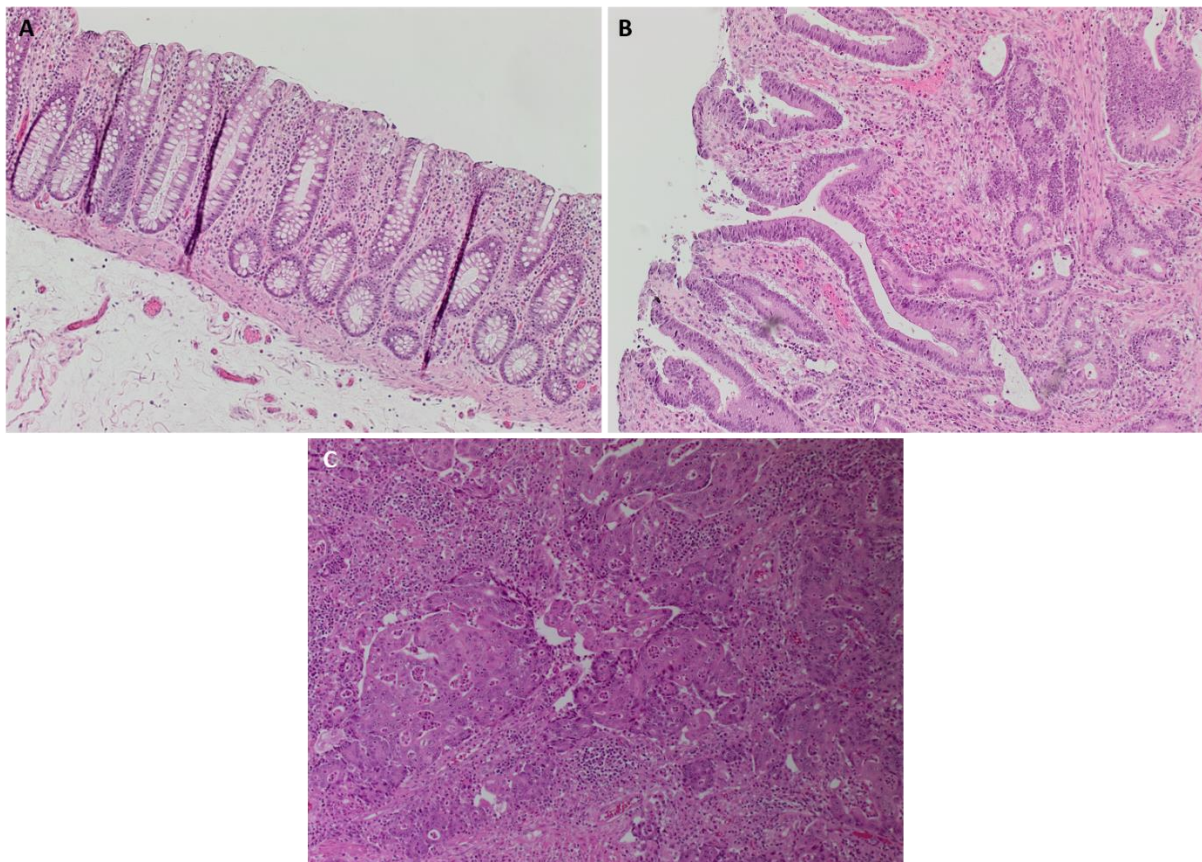


Figure 1.3: Crypts of the normal colon and colon adenocarcinoma. H&E slides showing normal colon (A), low-grade adenocarcinoma (B) and high-grade adenocarcinoma (C). Original magnification = 100x.

Mucinous adenocarcinoma is a histological variant of CA in which more than half of the tumour volume is composed of extracellular mucin (Fleming et al., 2012). It has large glandular structures and is

associated with Lynch syndrome which is characterised by MSI; however, if the mucinous adenocarcinoma is microsatellite-stable (MSS) it behaves more aggressively and has a worse prognosis (Fleming et al., 2012).

High-grade dysplasia refers to an immediate precursor to invasive CRC (Fleming et al., 2012). It usually describes a tumour which has penetrated the basement membrane but not the muscularis mucosa, yet it possesses greater architectural complexity and atypical cellular features compared to adenomas (Fleming et al., 2012).

1.3 Genetic Causes of CRC

1.3.1 Hereditary Disorders

Lynch syndrome and Familial Adenomatous Polyposis (FAP) are two hereditary disorders known to predispose the affected individuals to the development of CRC (Strate et al., 2005).

Lynch syndrome is an autosomal dominant hereditary disorder and the most common precursor to hereditary CRC, with varying reports of 40-80% likelihood of progressing to malignancy (Fleming et al., 2012; Jang et al., 2010; Strate et al., 2005). Lynch syndrome is thought to have a prevalence of between 1:300 and 1:2000 in the general population (Hampel et al., 2011; Haraldsdottir et al., 2017), and accounts for about 3% of all CRC cases, with the affected individuals developing CRC from their 20's or 30's (Boland et al., 2010; Fleming et al., 2012). It is caused by germline mutations that cause inactivation of any of four DNA mismatch repair (MMR) genes; *MSH2*, *MSH6*, *MLH1* and *PMS2* (Boland et al., 2010; Fleming et al., 2012; Haraldsdottir et al., 2017). MMR activity is impaired by an initial mutation, but total loss of MMR gene function requires a second somatic mutation in the remaining allele (Hemminki et al., 1994). The epithelial cell adhesion molecule (EPCAM) gene is upstream of *MSH2*, and deletions of *EPCAM* cause *MSH2* silencing due to transcriptional read-through (Huth et al., 2012). This is the cause of approximately 25% of the Lynch syndrome cases in which the MSH2 protein is inactive despite a lack of detectable mutations, and this accounts for 2-3% of all Lynch syndrome cases (Huth et al., 2012). Impairment of the MMR mechanism leads to microsatellite instability (MSI), a feature of up to 90% of Lynch syndrome colorectal tumours and up to 20% of sporadic CRCs (Sanchez et al., 2008), which will be discussed in greater detail in section 1.3.2.

FAP is thought to progress to CRC in almost 100% of cases, but this can be prevented by early diagnosis and total colectomy (Jang et al., 2010; Strate et al., 2005). It is primarily caused by germline

mutations of the Adenomatous Polyposis Coli (APC) tumour suppressor gene, usually a nonsense or frameshift mutation resulting in a truncated protein (Bolocan et al., 2011; Fleming et al., 2012; Jang et al., 2010; Markowitz et al., 2009; Strate et al., 2005).

Inactivation of the *MYH* base excision repair gene leads to MYH-associated polyposis (MAP syndrome) (Bolocan et al., 2011). This syndrome is caused by the bi-allelic mutation of *MYH* and is inherited in an autosomal recessive manner leading to seemingly random development in descendants of those affected, in contrast to the direct autosomal dominant inheritance pattern of FAP (Bolocan et al., 2011; Fleming et al., 2012; Strate et al., 2005). In addition, the APC gene is not mutated in MAP (Fleming et al., 2012). It leads to CRC in approximately 80% of cases (Jang et al., 2010).

1.3.2 Microsatellite Instability

Microsatellites are also known as short tandem repeats and are units of between 2 and 5 base pairs which are repeated up to 50 times consecutively (Ellegren, 2004). Because of their repetitive nature and tendency to be located in non-coding regions, they have a high mutation rate which leads to genetic diversity between individual genomes, but they can also result in phenotypic change and disease.

Microsatellite instability (MSI) arises from defective DNA MMR mechanisms, either due to a germline mutation or methylation of the promoter region of an MMR gene (Boland et al., 2010; Haydon et al., 2002; Sanchez et al., 2008). DNA methylation occurs at sites where a cytosine is followed by a guanine, called a CpG site. When the cytosine is methylated it tends to spontaneously deaminate into a thymine. Thus, CpG islands form where methylation of DNA is rare, such as in a promoter region. About 70% of mammalian promoters have a high CpG content because the spontaneous deamination into a thymine would affect the promoters' functionality, and methylation of a promoter prevents transcription of its gene.

MSI is an important factor in the development of CRC, with about 15% having MSI (Boland et al., 2010; Fleming et al., 2012; Haydon et al., 2002; Markowitz et al., 2009; Popat et al., 2005). This unique pathway for the development of CRC is shared by a small subset of sporadic cases (12% of CRC), usually due to *MLH1* promoter methylation, and those with a hereditary disease that has undergone malignant transformation (3% of CRC), usually due to MMR gene mutations (Boland et al., 2010; Fleming et al., 2012; Sanchez et al., 2008; Strate et al., 2005). Therefore, 85% of CRC cases are microsatellite stable (MSS).

1.3.3 CpG Island Methylator Phenotype

CpG island methylator phenotype (CIMP) is the main factor that distinguishes sporadic MSI CRC from hereditary CRC of Lynch syndrome origin (Boland et al., 2010). CIMP can be involved in both MSS and MSI sporadic CRC (Fleming et al., 2012).

The CIMP phenotype typically leads to the bi-allelic methylation of *MLH1* (Boland et al., 2010; Markowitz et al., 2009). Sporadic cases of CRC with MSI also have the *BRAF* V600E mutation 50% of the time, which lowers mortality (Boland et al., 2010). This contrasts with CRC that develops in patients with Lynch syndrome, who have mutations in the *KRAS* gene but never in the *BRAF* gene, and arises from germline mutations in MMR genes as opposed to CIMP-mediated gene silencing (Boland et al., 2010; Fleming et al., 2012). Overall, the prognosis is better for those with a hereditary form of CRC (Boland et al., 2010). Numerous studies have found that methylation patterns differ between colon tumours and normal mucosa (Beggs et al., 2013; Illingworth et al., 2010; Wei et al., 2016).

1.3.4 Chromosomal Instability

The most common cause of genetic instability in CRC is chromosomal instability (CIN) rather than MSI (Markowitz et al., 2009). Like MSI, it is implicated in both sporadic and hereditary CRC, however it appears that the two are mutually exclusive (Fleming et al., 2012). Hereditary CRC caused by FAP and MAP develop due to CIN (Fleming et al., 2012).

CIN involves changes in chromosome copy number or structure, including abnormal karyotype and loss or gain of entire chromosomes; often the wild-type copy of a tumour-suppressor gene, most commonly *TP53*, *APC*, *KRAS* or *SMAD4*, is lost completely (Fleming et al., 2012; Markowitz et al., 2009; Vogelstein et al., 1989).

1.3.5 Common Mutations in CRC

APC is the most common mutation of CRC and is considered to be the initiating mutation of CRC due to its influence in Wnt signalling (Markowitz et al., 2009). It is a key component of the β -catenin destruction complex, so mutation or knockout of the *APC* gene causes canonical Wnt signalling to be constitutively active (Markowitz et al., 2009).

The TGF- β signalling pathway is involved in proliferation, differentiation, migration and apoptosis, as well as stem cell maintenance and function (Khalek et al., 2010). It is one of the most commonly altered pathways across all cancer types due to its inherent tumour-suppressive actions, such as c-MYC suppression, being altered or lost in cancer (Khalek et al., 2010; Miyazono, 2009). TGF- β signalling occurs when TGF- β binds to the type 2 TGF- β receptor (TGF β R2), which then recruits the type 1 receptor (TGF β R1) (Wolf, 2006). Once activated, TGF β R1 sends signals to intracellular mediators called SMAD proteins (Khalek et al., 2010; Wolf, 2006). The phosphorylated SMAD2/SMAD3 complex recruits SMAD4 and crosses the nuclear membrane to act as a transcription factor. This leads to transcription of genes involved in cell survival, proliferation and tumourigenesis (Khalek et al., 2010). TGF β R2 is localised to differentiated cells of the villus and undifferentiated cells at the crypt base during gut development (Khalek et al., 2010). The *TBR2* gene which codes for TGF β R2 is mutated in one third of CRC cases (Markowitz et al., 2009) and is prone to errors in replication, especially when there is inactivation of MMR genes (Khalek et al., 2010). In fact, it is affected by frameshift mutations in up to 80% of MSI CRCs (Khalek et al., 2010). SMAD genes are mutated or deleted in up to 50% of CRC cases (Markowitz et al., 2009). In general, the downstream targets of TGF- β signalling are cell cycle checkpoint genes which mediate growth arrest (Khalek et al., 2010).

The p53 protein, coded by the *TP53* gene, mediates cell cycle arrest at a cell death checkpoint (Markowitz et al., 2009). However, in most CRC tumours both *TP53* alleles are inactive, and this is usually due to a combination of missense mutations inactivating the transcriptional activity of one allele and a deletion from chromosome 17q which causes the loss of the other allele (Markowitz et al., 2009). It has been shown that the loss of p53 predisposes mammary stem cells to symmetric division rather than asymmetric division, increasing the likelihood of tumour development due to an increase in total stem cell numbers (Tang, 2012). Inactive p53 is associated with adenoma to carcinoma sequence transition (Markowitz et al., 2009).

The *BRAF* gene codes for B-raf, a serine-threonine MAP3K (Markowitz et al., 2009). Oncogenic mutations causing over-expression of *BRAF* leads to overproduction of B-raf. This is then activated when phosphorylated by ras, a GTP-ase encoded by the *KRAS* gene. These mutations increase MAP-kinase cascade signalling: ras activates B-raf (MAP3K), which then phosphorylates MEK (MAP2K), and MEK subsequently activates MAP-kinases (Markowitz et al., 2009). As mentioned previously, 50% of CIMP-related sporadic CRC cases have *BRAF* mutations, and hereditary CRC cases can have a *KRAS* mutation but never *BRAF*, as these two mutations are mutually exclusive (Boland et al., 2010).

KRAS is constitutively active in 37% of CRC and its protein product, ras, is a critical downstream molecule in EGFR signalling (Fleming et al., 2012). The *KRAS* mutant has reduced GTPase activity, meaning that the ras protein accumulates in an active GTP-bound state (Le Rolle et al., 2016). Tumours with the *KRAS* mutation are resistant to EGFR therapy so it is necessary to test for a *KRAS* mutation before administering anti-EGFR therapies (Fleming et al., 2012; Markowitz et al., 2009; Tack et al., 2018). *KRAS* mutations can be detected by PCR or DNA sequencing (Fleming et al., 2012; Tack et al., 2018). While *APC* or *CTNNB1* (β -catenin) mutations initiate adenomas via constitutive Wnt signalling, the activating mutation in *KRAS* is thought to drive the malignant transformation of adenomas to carcinomas (Le Rolle et al., 2016).

Approximately 13% of CRCs have the *BRAF* V600E point mutation, which is mutually exclusive with *KRAS* mutations (Fleming et al., 2012; Markowitz et al., 2009). Therefore, if *KRAS* tests do not identify a mutation, *BRAF* can be investigated. Point mutations like that seen in *BRAF* are easy to detect with a commercially available PCR-based assay (Fleming et al., 2012). *BRAF* testing is highly valuable in determining whether a patient with MSI or CIMP CRC has developed it sporadically or has inherited it, because the *BRAF* mutation has never been reported in Lynch syndrome (Fleming et al., 2012).

1.3.6 Consensus Molecular Subtypes

It was recently shown that despite the high degree of heterogeneity between cases, CRC can be distilled down into 4 consensus molecular subtypes (CMS) (Guinney et al., 2015). Most CRC cases (~87%) are able to be grouped in this way, based on: hypermutation and hypermethylation with MSI, *BRAF* mutations and immune activation (CMS1), a differentiated epithelial phenotype with either abnormal Wnt and c-MYC signalling (CMS2) or metabolic dysregulation and *KRAS* mutations (CMS3), or a mesenchymal phenotype with high levels of angiogenesis, stromal invasion, EMT signatures, complement pathway components and TGF- β signalling (CMS4). CMS4 tumours tend to present at more advanced stages and have the poorest outcomes, while recurrent CMS1 tumours also lead to poor survival. Tumours of the classical CMS2 subtype have the best outcomes. This is reflected in CRC survival based on the location of presentation, with CMS1 tumours more prevalent in the proximal colon and CMS2 in the distal colon (Guinney et al., 2015). The CMS classification system is expected to be more biologically relevant and more suitable for translational therapies due to the range of genomic, epigenetic and proteomic factors taken into consideration (Guinney et al., 2015).

1.4 Cancer Stem Cells

1.4.1 The Cancer Stem Cell Concept

The cancer stem cell (CSC) concept hypothesises that tumour growth is driven by CSCs, a small subpopulation of cancer cells with stem cell characteristics (Khalek et al., 2010; Kreso et al., 2014; Shimokawa et al., 2017; Zheng et al., 2013). These cells divide asymmetrically to produce identical daughter pluripotent cells as well as progenitor cells which are more committed and sit on a hierarchy between CSCs and terminally differentiated cancer cells (Gage, 2000; Seaberg et al., 2003; Tang, 2012). CSCs have lost control of cellular homeostasis leading to dysregulated cellular replication and differentiation, resulting in uncontrolled growth and tumour formation (Humphries et al., 2008; Khalek et al., 2010; Kreso et al., 2014; Zheng et al., 2013). Despite this, CSCs have a slow cellular turnover cycle bordering on quiescence, rendering them resistant to standard therapies which target rapidly dividing cells (Khalek et al., 2010; Kreso et al., 2014; Zheng et al., 2013). Rather than being a direct result of CSCs, Tang (Tang, 2012) has postulated that uncontrolled growth is caused by progenitor cells which sit in the hierarchy between stem cells and differentiated cells. The bulk of the tumour results from aberrant progenitor cell proliferation producing differentiated cancer cells, which are unable to initiate tumour formation or self-renew (Khalek et al., 2010; Tang, 2012).

Two theories have been proposed to explain how CSCs arise: oncogenic mutations accumulating within normal adult stem cells, leading to uncontrolled proliferation (Khalek et al., 2010; Munro, Wickremesekera, Davis, et al., 2017); or cellular de-differentiation of a tumour cell into a stem-like state (Khalek et al., 2010; Kreso et al., 2014; Munro, Wickremesekera, Davis, et al., 2017). In both scenarios, it is thought that the progenitor cells produced by CSCs during asymmetrical division inherit mutations that allow them to proliferate in an unregulated manner and produce the differentiated cancer cells of the tumour bulk. CSCs and their progenies can harbour a surprising amount of variability, possibly due to mutations in DNA repair genes that can contribute to carcinogenesis, which in turn produces heterogenous tumours (Zheng et al., 2013).

1.4.2 Stem Cell Markers

As CSCs have become a more prominent area of research, there has been a focus on finding markers that can be used to identify them. One of the earliest CSC markers was CD133, which was first discovered on

the surface of hematopoietic stem cells (Miraglia et al., 1997; Yin et al., 1997). Similarly, leucine-rich repeat-containing G-protein coupled-receptor 5 (LGR5) is expressed by normal colonic stem cells where its expression decreases as cells differentiate, and it is elevated in CRC (Khalek et al., 2010). Other markers hypothesised to identify CSCs include EpCAM, which is expressed by cancers of epithelial origin (Dalerba et al., 2007; Trzpis et al., 2007; van der Gun et al., 2010), aldehyde dehydrogenase 1 (ALDH1), which enables chemoresistance (Y. Deng et al., 2014), and CD44, which is involved in extracellular interactions (Basakran, 2015; Dalerba et al., 2007; Wielenga et al., 1999). These markers tend to be overexpressed in cancer, and their expression has been reported in tumour-initiating cells considered to be CSCs.

CD133

CD133 is one of the more well-documented stem cell markers (Z. Li, 2013; Schmohl et al., 2016; Wu et al., 2009; X. Yu et al., 2011). It is also known as the prominin-1 glycoprotein and is thought to organise cell membrane topology (Dalerba et al., 2007). Initial studies aiming to characterise CRC CSCs began by isolating CD133⁺ cells (Z. Li, 2013). When isolated from primary colon cancer samples, these cells are capable of forming tumours in mice, remain undifferentiated when cultured in serum-free media, and become more aggressive over the span of generations (Ricci-Vitiani et al., 2007). Furthermore, some studies have found that the majority of tumours are comprised of CD133⁺ cells which are incapable of tumour initiation (O'Brien et al., 2007). CD133⁺ cells produce IL-4 and use it to evade apoptosis, as confirmed by the observation that an anti-IL-4 neutralising antibody or treatment with IL-4R α will greatly enhance the sensitivity of CD133⁺ cells to chemotherapeutic drugs and increase their anti-tumour efficacy (Todaro et al., 2007).

LGR5

LGR5 has emerged as a unique marker for cycling CBCs and therefore colon stem cells. LGR5⁺ cells have been proven to be ISCs by being able to differentiate into all colonic epithelium lineages within 60 days *in vitro* (Barker et al., 2007; Khalek et al., 2010). The ISC signature was further clarified as being positive for LGR5 and the EphB2 receptor, both of which become less abundant as the crypt epithelial cells differentiate (Le Rolle et al., 2016; Merlos-Suarez et al., 2011). The erythropoietin-producing human hepatoma amplified sequence (Eph) family is the largest subfamily of receptor tyrosine kinases (RTKs), and their ligands are known as ephrins (W. Liu et al., 2002). EphB2 and EphB3 receptors are found at the crypt base, with decreasing expression further up the crypt (Clevers et al., 2006; Merlos-Suarez et al., 2011). Ephs are target genes of Wnt signalling (Clevers et al., 2006) and blocking these receptors reduces

the number of proliferating cells at the crypt base without causing a reduction in total cell number, suggesting a role in Wnt-mediated proliferation and differentiation of ISCs (Khalek et al., 2010).

These LGR5⁺/EphB2⁺ crypt base cells are controlled by Wnt signalling, of which the *LGR5* gene is a target (Le Rolle et al., 2016). They are found amongst Paneth cells at the crypt base (Schepers et al., 2012). Paneth cells produce mucosal defence components such as lysozymes and α -defensins (cryptdins in mice), as well as growth factors and other regulatory molecules essential to stem cell maintenance, namely epidermal growth factor (EGF), TGF- α , WNT3 and Notch ligand D114 (Sato et al., 2011). For this reason, Paneth cells are hypothesised to be essential in maintaining the intestinal stem cell niche, and so it is logical that proposed epithelial stem cells are located amongst them.

LGR5 was found to be required, but not sufficient, for malignant transformation, and LGR5⁺ cells can form tumours and enhance their proliferative capacity (Khalek et al., 2010; Le Rolle et al., 2016; Sato et al., 2011; Shimokawa et al., 2017). After using LGR5 to identify colon stem cells, its co-expression with other stem cell markers during the investigation of CSCs has been confirmed. The cell of origin of CRC is proposed to be the LGR5⁺/EphB2⁺ crypt base cells, under the following model: an *APC/CTNNB1* mutation in LGR5⁺ cells causes adenoma formation, and a subsequent *KRAS* mutation initiates an ESC-like expression signature, forming tumour-fuelling CSCs (Merlos-Suarez et al., 2011). Further evidence that CRC CSCs may be LGR5⁺ has been shown through LGR5 knock-down causing tumour regression, and subsequent re-emergence of the LGR5⁺ cell population coinciding with tumour growth and recurrence (Shimokawa et al., 2017). Additionally, Wnt signalling is downregulated when LGR5 expression is lowered, and this causes premature Paneth cell differentiation and reduced tumour production (Khalek et al., 2010; Merlos-Suarez et al., 2011).

EpCAM

The epithelial cell adhesion molecule (EpCAM) is a marker expressed solely by epithelial cells and cancers of epithelial origin (van der Gun et al., 2010). It was discovered first in colon carcinoma, and subsequently seen in other cancers (Dalerba et al., 2007; Trzpis et al., 2007). Its native functions include homotypic Ca²⁺-independent cell-cell adhesion in the epithelium, cell signalling, proliferation and differentiation, and it also influences metastasis and migration, tumourigenesis and upregulation of Wnt targets c-MYC and cyclins A and E (Dalerba et al., 2007; Patriarca et al., 2012; Schnell, Cirulli, et al., 2013; Trzpis et al., 2007; van der Gun et al., 2010). EpCAM has a basal or basolateral cell membrane expression pattern in normal

epithelium, solely basolateral in well-differentiated CA, and can be found in the membranes, cytoplasm and lumen of moderately-differentiated CA (Trzpis et al., 2007).

EpCAM undergoes an evolutionarily conserved two-step proteolytic processing mechanism called regulated intramembrane proteolysis (RIP). First, TNF- α converting enzyme (TACE/ADAM17) releases the N-terminal extracellular domain (EpEX); second, the γ -secretase presenilin 2 (PS2) cleaves the C-terminal intracellular domain (EpICD) from the transmembrane region (Maetzel et al., 2009; T. Yu et al., 2017). There have been observations of increased RIP in tumours, resulting in higher levels of EpEX and EpICD, leading to progression of epithelial cancers, invasion, metastasis, and epithelial-to-mesenchymal transition (EMT) (Martowicz et al., 2016; Munz et al., 2009).

EpEX functions as a paracrine and autocrine signalling molecule (Schnell, Kuipers, et al., 2013). It is an EGFR ligand which initiates proliferation via downstream ERK1/2 and AKT signalling, but counteracts EGF-mediated EMT (Liang et al., 2018; M. Pan et al., 2018).

EpICD is translocated to the nucleus and carries transcriptional co-factors in with it, where it stimulates gene expression in conjunction with β -catenin (Schnell, Kuipers, et al., 2013; T. Yu et al., 2017). Nuclear accumulation of EpICD is associated with metastasis and a worse prognosis in CRC and does not occur in the normal colon (Liang et al., 2018; Maetzel et al., 2009). Furthermore, EpICD has been shown to upregulate the expression and stability of pluripotency markers *OCT4*, *SOX2*, *NANOG*, *c-MYC* and *KLF4* by associating with Lef-1 and β -catenin (L. Huang et al., 2018; Lu et al., 2010; Maetzel et al., 2009; Munz et al., 2009).

A recent study demonstrated that induced pluripotent stem cells (iPSCs) can be generated using EpCAM, EpEX, and either OCT4 or KLF4 (Kuan et al., 2017). Reprogramming using the Yamanaka factors requires STAT3 and HIF-2 α expression (Kuan et al., 2017). They found that EpEX increases phosphorylation of STAT3 and β -catenin, while EpCAM allows HIF-2 α expression which in turn increases the levels of *OCT4*, *SOX2* and *NANOG*. It was concluded that EpEX and EpCAM can functionally replace SOX2 and OCT4/KLF4 in reprogramming via the same signalling pathways, and that this may occur through a mesenchymal-to-epithelial transition (MET) function (Kuan et al., 2017).

EpCAM is a target of β -catenin/Wnt signalling. The EpCAM promoter region contains two Tcf4 binding elements, where the β -catenin/Tcf4 transcription factor complex binds to upregulate EpCAM expression (Yamashita et al., 2007). It is a potential diagnostic marker in some epithelial cancers, where high expression correlates with poor survival and an advanced disease state (Patriarca et al., 2012).

Furthermore, due to its overexpression in adenocarcinomas, metastases and CSCs, it has been used as a target for detecting circulating or metastasising cancer cells (Patriarca et al., 2012).

ALDH

Aldehyde dehydrogenase (ALDH) is widely considered to be a marker of both normal stem cells and CSCs (Khalek et al., 2010). ALDH is known to be important in protecting haematopoietic stem cells and may play a similar role in CSCs, which exhibit raised levels of ALDH activity (Hadjimichael et al., 2015; I. Ma et al., 2011; X. Xu et al., 2015). There is weak staining for ALDH1 in nearly all crypt bases, with a small number of cells that stain strongly positive for ALDH1 (S. Deng et al., 2010). These ALDH1^{High} cells are found in low abundance at crypt bases, but increase during progression from normal to APC-mutant to adenoma (E. H. Huang et al., 2009). A high percentage of tumour cells expressing ALDH1 correlates with poor prognoses in a range of cancers (S. Deng et al., 2010; Tomita et al., 2016). ALDH1^{High} cells display properties of CSCs, including self-renewal, *in vitro* and *in vivo* tumour growth potential, and chemotherapy resistance (S. Deng et al., 2010). Because the normal colon has low ALDH1 expression relative to CRC tumours, ALDH1 has been proposed as a promising marker for identifying CRC CSCs and as a potential target for CSC-directed therapy in CRC (S. Deng et al., 2010; Tomita et al., 2016).

A clinical study investigating CSC markers in stage II and III rectal cancer showed that CD44, LGR5 and CD166 are all significantly upregulated following chemoradiation, suggesting that they are markers of therapy-resistant CRC CSC populations (Y. Deng et al., 2014). Patients who exhibit low ALDH prior to chemoradiation do not recur following surgical resection, however patients with high ALDH levels before treatment have a recurrence rate of 40% within 36 months post-resection. This highlights the importance of ALDH in predicting the prognosis of patients receiving pre- or post-operative chemoradiation, and a possible correlation between ALDH and the abundance of CSCs (Y. Deng et al., 2014). Cells expressing both CD44 and EpCAM generate heterogeneous tumours, are enriched following chemotherapy, and have elevated expression and activity of ALDH1, which has been proven to mediate cyclophosphamide chemotherapy resistance in these CSCs (Dylla et al., 2008). ALDH has been used to further clarify the role of CD133 as a stem cell marker. A study by Ma *et al.* (S. Ma et al., 2008) showed that in hepatocellular carcinoma most ALDH⁺ cells are also CD133⁺, but there are many CD133⁺/ALDH⁻ cells. Furthermore, CD133⁺/ALDH⁺ cells had the highest tumour-forming potential, followed by CD133⁺/ALDH⁻ and then CD133⁻/ALDH⁻ (S. Ma et al., 2008).

CD44

CD44 has emerged as a potential marker for CRC CSCs, and is considered by some to be the most consistent marker for isolating such cells (Basakran, 2015; Morath et al., 2016; Tang, 2012). However, other reports suggest that it is expressed by the majority of colon tumour cells (Jaggupilli et al., 2012). It is a cell surface glycoprotein which functions in cell-cell interactions, as a receptor for hyaluronic acid to allow adhesion to the extracellular matrix (ECM), and in cell migration (Basakran, 2015; Dalerba et al., 2007; Jaggupilli et al., 2012; Wielenga et al., 1999). Transcription of CD44 is at least in part activated by canonical β -catenin/Wnt signalling, and overexpression is an early event in the transformation of colorectal adenoma to carcinoma (Dalerba et al., 2007; Wielenga et al., 1999). In the colonic mucosa it is used as a marker of immature differentiation (Wielenga et al., 1999). Knock-down of CD44 prevents tumourigenesis and clonal formation (L. Du et al., 2008). Furthermore, injection of only 100 CD44⁺ cells is sufficient to initiate tumour formation in nude mice, and single CD44⁺ cells form tumourspheres with stem cell characteristics which develop into tumours when xenografted into nude mice (L. Du et al., 2008). CD44⁺ cells also possess the stem cell characteristic of multilineage differentiation, with the potential to form enterocyte, enteroendocrine and goblet cell lineages *in vitro* (Yeung et al., 2010).

In contrast to earlier studies, Shmelkov *et al.* (Shmelkov et al., 2008) show that CD44⁺/CD24⁻/CD133⁻ cells form the most aggressive colon tumours, removing the necessity for CD133 in tumour initiation. Furthermore, they observed that CD133⁺ cell populations are generally CD44^{low}/CD24⁺, and CD133 is occasionally present on normal differentiated colonic epithelium (Shmelkov et al., 2008). They hypothesise that the CD133⁺ population may give rise to a more aggressive CD133⁻ subset that retains the ability to initiate tumours (Shmelkov et al., 2008). It has also been suggested that CD133 expression is correlated with cell culture conditions and hypoxia, making it a less reliable marker for identifying isolated CSCs (Hadjimichael et al., 2015).

STAT3

Signal-transducer and activator of transcription 3 (STAT3) is a transcription factor that is translocated to the nucleus when it becomes phosphorylated (pSTAT3). When constitutively active, STAT3 leads to proliferation, angiogenesis and invasion of cancer cells (Corvinus et al., 2005; Lassmann et al., 2007; H. Yu et al., 2009). IL-6 and OCT4 induce STAT3-mediated proliferation by causing STAT3 to be phosphorylated (Corvinus et al., 2005; Z. Du et al., 2009; H. Yu et al., 2009). This has been observed in intestinal tumour-initiating cells (Grivennikov et al., 2009). Furthermore, IL-6 produced by the lamina propria plays a role in preventing apoptosis of intestinal epithelial cells via STAT3 activity (Grivennikov et al., 2009). There is also evidence that STAT3 promotes forms of inflammation which are beneficial to the tumour and prevents

anti-tumour immune responses (H. Yu et al., 2009). In CRC, the presence of STAT3 has been observed in the cytoplasm of crypt epithelial cells. Active pSTAT3 has been localised to both the cytoplasm and the nucleus (Lassmann et al., 2007).

STAT3 represents a promising target in therapy-resistant CRCs. Expression of STAT3 has been related to chemotherapy and radiotherapy resistance in CRC, and inhibiting STAT3 causes previously resistant cells to become sensitive to 5-FU (Spitzner et al., 2014). Furthermore, the inhibition of pSTAT3 can cause apoptosis and cell cycle arrest in CRC cells (Q. Lin et al., 2005). A link between pSTAT3 and both the EMT marker vimentin and vascular mimicry-related protein VE-cadherin has also been discovered (Han et al., 2017), suggesting roles for pSTAT3 in metastasis and angiogenesis. STAT3 involvement specifically in CRC CSCs has also been investigated. Tumour-initiating ALDH⁺/CD133⁺ CRC cells contain more pSTAT3 than ALDH⁻/CD133⁻ cells, and inhibition of STAT3 prevents tumour initiation and CSC survival (L. Lin et al., 2011).

1.4.3 Issues of current CSC markers

A major issue with the current markers used to identify colon CSCs is the paucity of markers that are expressed exclusively by CSCs, either individually or in combination. This prevents markers from being therapeutic targets without having side-effects on ISCs and homeostasis. Furthermore, it has been well documented that the markers commonly used to identify colon CSCs, CD133, LGR5 and EpCAM, are not expressed by all colon CSCs (Guo et al., 2011). One solution to address the variability in CSC marker expression could be to identify panels of markers from which the expression of specific combinations could allow CSCs to be identified. This technique has been used when characterising embryonic stem cells (ESCs). Various independent gene expression signatures have been created for human ESCs, and the common genes identified by multiple studies are a proposed “consensus hESC gene list” (Assou et al., 2007). As mentioned above, LGR5 is associated with the signature for ISCs, whereas the *KRAS* mutation is believed to initiate an ESC signature (Le Rolle et al., 2016). *KRAS* mutants induced an ESC signature which included upregulation of the pluripotency gene *SOX2*, and suppression of the colonic goblet-lineage differentiation marker *KLF4* (Le Rolle et al., 2016; I. C. Li et al., 2011). From these results, an example of a hypothetical panel for colon CSCs could be a combination of the expression of LGR5 and *SOX2*, reduction in *KLF4* levels, and a mutation in the *KRAS* gene. Some genes, such as *MYC* and *CTNNB1*, are associated with both proliferation and differentiation depending on the context of their expression, but their co-expression with other markers as part of a consensus panel might help to narrow down their role in a

tumour context (Ben-Porath et al., 2008). Additionally, the level of expression in the tumour relative to normal tissue can be as informative as presence-absence, especially with ubiquitously expressed genes such as *MYC* and *CTNNB1*. Consensus expression panels such as these are an emerging tool in cancer research, and could become useful for distinguishing cancer from normal (Cancer Genome Atlas, 2012), grading and subtyping of cancers (Guinney et al., 2015), and identifying pluripotent stem-like cells in cancer.

1.4.4 Pluripotency Markers

Octamer-binding transcription factor 4 (*OCT4/POU5F1*), Sex-determining region Y-box 2 (*SOX2*) and *NANOG* are the three genes which play a dominant role in regulating the pluripotency of cells such as ESCs (Hadjimichael et al., 2015; Takahashi et al., 2006). They are transcription factors which were originally found to be key regulators of embryogenesis (Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998), but recently their expression has been documented in a range of cancers, including glioblastoma (A. Bradshaw et al., 2016), oral cavity squamous cell carcinoma of different subsites (Baillie et al., 2016; R. Ram et al., 2017; H. H. Yu et al., 2016), metastatic SCC (Kilmister et al., 2020) and breast tumours (Al-Hajj et al., 2003).

OCT4

OCT4 is known to influence embryogenesis, stem cell maintenance, tumour growth and metastasis (Amini et al., 2014; Hadjimichael et al., 2015; Shi et al., 2010). OCT4 expression has been noted in normal colonic stem cells (Amini et al., 2014). Studies have shown OCT4 expression in CRC, often in the cytoplasm of epithelial cells (Amini et al., 2014; J. Hu et al., 2017; Wen et al., 2013). This expression has been suggested to drive recurrence, leading to poorer clinical outcomes, presumably by preventing apoptosis (J. Hu et al., 2017; Wen et al., 2013). The expression of OCT4 in CRC has been reported in cells which are undergoing EMT, a key step in cancer progression and metastasis, which has been shown to increase the stem-like phenotype of cells (Dai et al., 2013; Mani et al., 2008). The OCT4A isoform is 38.6 kDa and is responsible for functions related to pluripotency and stem cell maintenance (Atlasi et al., 2008). OCT4B, a 30 kDa isoform, has predominantly been found in the cytoplasm but its function is unclear; it does not sustain pluripotency or transcribe genes which OCT4 is traditionally known to regulate (Atlasi et al., 2008). Most antibodies and PCR probes do not discriminate between the two isoforms (Atlasi et al., 2008).

The OCT4B1 splice variant of the B isoform is particularly overexpressed in CRC and generally localised to the cytoplasm (Atlasi et al., 2008). Like OCT4A and unlike OCT4B, it is highly expressed in pluripotent cells and downregulated after differentiation; however, there is currently no antibody specific to the OCT4B1 splice variant (Atlasi et al., 2008; Gazouli et al., 2012). Additionally, OCT4 has 6 pseudogenes, 2 of which are transcribed and act as antisense regulators of OCT4 predominantly in cancers (Suo et al., 2005).

SOX2

SOX2 maintains pluripotency of ESCs and neural progenitor cells and is critical for early embryogenesis (Amini et al., 2014; S. Zhang et al., 2014). SOX2 is involved in regulating *OCT4* expression by binding to its promoter region, and pluripotency is usually maintained by this combination of transcription factors (Avery et al., 2006; Hadjimichael et al., 2015; Masui et al., 2007). OCT4 and SOX2 commonly work together as a transcription complex, with around half of the promoters bound by OCT4 also co-occupied by SOX2 (Boyer et al., 2005). This complex is able to induce transcription of the *OCT4/POU5F1*, *SOX2* and *NANOG* genes (Boyer et al., 2005). Studies have shown localisation of SOX2 to the cytoplasm and nuclei of both normal and cancerous crypt epithelial cells (Talebi et al., 2015). SOX2 expression is associated with lymph node infiltration and metastasis in CRC (J. Neumann et al., 2011).

NANOG

NANOG also influences pluripotency via transcriptional control. Although it is vital in establishing embryonic pluripotency, it has been suggested that NANOG is not necessary for maintaining undifferentiated ESCs in culture (Hadjimichael et al., 2015; Takahashi et al., 2006; J. Zhang et al., 2013). NANOG also forms complexes with OCT4 and SOX2, with more than 90% of promoters occupied by the OCT4/SOX2 complex also co-occupied by NANOG (Boyer et al., 2005). These three factors work together to regulate their own expression, enhance the transcription of other pluripotency genes, and repress the expression of transcription factors involved in differentiation (Boyer et al., 2005). Transcription of the *NANOG* gene is also induced by the Wnt signalling pathway via Lf/STAT3, by transcription factor FoxD3, and by AKT signalling (G. Pan et al., 2007). Conversely, *NANOG* is suppressed by Tcf3 and by p53 phosphorylated at Ser315, to allow differentiation to occur (G. Pan et al., 2007). NANOG has ten pseudogenes, but only the retrogene *NANOGP8* is translated into a functional protein (Booth et al., 2004). *NANOGP8* expression has been confirmed in sphere-forming gastric cancer cells, which also overexpress LGR5, OCT4, SOX2 and the EMT gene *TWIST1* (B. Wang et al., 2016). Furthermore, *NANOGP8* can replace

NANOG when producing iPSCs, and so it may also function in the same way as NANOG in CSCs (W. Zhang et al., 2016).

NANOG has been detected in colon tumours and dysplastic polyps, often exhibiting strong nuclear staining in a subpopulation of crypt epithelial cells (Amini et al., 2014; Ibrahim et al., 2012). High expression of NANOG is correlated with a poor prognosis in CRC patients (Hadjimichael et al., 2015; Muller et al., 2016).

c-MYC

The proto-oncogene c-MYC plays a key role as a transcription factor in proliferation, apoptosis, cell growth and differentiation in normal cells, as well as stem cell renewal (Takahashi et al., 2006). It binds to E-box sequences (CACGTG) in the promoter regions of genes (Fernandez et al., 2003). c-MYC is often upregulated in cancers and is a downstream target of the Wnt and MAP kinase signalling pathways, which are frequently altered in CRC (Markowitz et al., 2009; Takahashi et al., 2006). Duplication of the *c-MYC* gene is associated with a worse prognosis in CRC (Lee et al., 2015). Therapy-naïve CRC cells with high c-MYC expression progress more quickly, and CRC metastases exhibit greater c-MYC expression than the primary tumour (M. Martini et al., 2016).

KLF4

Krüppel-like factor 4 (KLF4) is highly expressed in undifferentiated cells, regulating both the cell cycle and pluripotency, and is necessary for maintaining self-renewal capacity (Hadjimichael et al., 2015; Leng et al., 2013). KLF4 has been shown to be overexpressed in CRC and it enables tumoursphere formation (Hadjimichael et al., 2015; Leng et al., 2013). KLF4 enhances cell movement in the form of migration, invasion and EMT, suggesting a possible role in metastasis (Hadjimichael et al., 2015). Knock-down of KLF4 expression inhibits these capabilities, as well as reducing chemoresistance, tumour initiation and expression of CSC markers (Leng et al., 2013). However, research into the role of KLF4 in cancer often yields conflicting results, and its role is likely affected by other factors (Muller et al., 2016). For instance, KLF4 expression seems to be inversely correlated with tumour grade, with the highest expression seen in the normal colon adjacent to tumours, and the lowest expression seen within high-grade, poorly differentiated tumours (R. Hu et al., 2011). During differentiation of the intestinal epithelium from LGR5⁺ intestinal stem cells, KLF4 is a marker specific to the goblet cell lineage, which is the most abundant epithelial cell type in colonic crypts (May et al., 2010).

1.4.5 Induced Pluripotent Stem Cells

OCT4, *SOX2*, *KLF4* and *c-MYC* are the genes which Takahashi and Yamanaka used to reprogramme somatic mouse fibroblast cells and produce the first iPSCs in 2006 (Takahashi et al., 2006). Interestingly, *NANOG* was not required for reprogramming in this case (Hadjimichael et al., 2015; Takahashi et al., 2006). Noting that *c-MYC* expression causes death in human ESCs, a similar experiment in 2007 by the Thomson lab produced iPSCs from human primary fibroblasts using *OCT4*, *SOX2*, *NANOG* and *LIN28* (J. Yu et al., 2007). *OCT4* and *SOX2* were essential, while *NANOG* and *LIN28* were not absolutely necessary for clone formation, although the efficiency of reprogramming was drastically increased by *NANOG* and modestly increased by *LIN28* (J. Yu et al., 2007).

The ways in which these genes interact in the initiation and maintenance of gastrointestinal tumours were recently summarised in a review which seeks to clarify the role of each Yamanaka factor and place them relative to each other in the context of signalling pathways involved in tumour initiation, maintenance and metastasis (Muller et al., 2016). However, it is yet to be established whether the iPSC markers could be utilised as part of a panel to identify CSCs in CRC.

1.5 The Renin-Angiotensin System in Cancer

1.5.1 Renin-Angiotensin System

The role of the renin-angiotensin system (RAS) in maintenance of blood pressure and body fluid is well known, although its involvement in cancer is an emerging field of research (Munro, Wickremesekera, Davis, et al., 2017). Of particular interest here is its possible paracrine role in regulating CSCs (Munro, Wickremesekera, Davis, et al., 2017).

The classical RAS (Fig 1.4) begins with renin, which is produced as a pro-enzyme called pro-renin by the juxtaglomerular cells of the kidney (Sparks et al., 2014). The juxtaglomerular cells are activated when blood pressure falls or when macula densa cells detect a decrease in sodium, and this leads to cleavage and activation of pro-renin (Fountain et al., 2019). Active renin converts angiotensinogen (AGT) to angiotensin I (ATI), which is a substrate of the angiotensin-converting enzyme (ACE) in a reaction that produces angiotensin II (ATII) (Fyhrquist et al., 2008; Sparks et al., 2014). ACE is expressed on the endothelium of blood vessels throughout the body, as well as circulating in plasma (Sparks et al., 2014). ATII is the major effector molecule of the RAS, interacting with its two receptors, the angiotensin II type 1 receptor (AT₁R) and angiotensin II type 2 receptor (AT₂R), to initiate downstream signalling (Fyhrquist et

al., 2008; Sparks et al., 2014). Its actions include stimulating sodium reabsorption, release of aldosterone from the adrenal cortex and anti-diuretic hormone from the posterior pituitary gland, vasoconstriction of arterioles, and stimulation of thirst by binding the hypothalamus (Fountain et al., 2019).

In the peripheral blood, ATII has a half-life of 1-2 minutes as it is quickly degraded by peptidases into ATIII and ATIV (Fountain et al., 2019).

ACE2 converts ATI into Ang1-9 and Ang1-7, and ATII into Ang1-7, which bind the Mas receptor and play a protective role, similar to that of the ATII-AT₂R axis by antagonising the effects of AT₁R-mediated signalling (Fountain et al., 2019).

The highest ACE2 expression is seen in the kidney, heart and testis, and its primary role is to degrade ATII (Sparks et al., 2014). Its activity plays a beneficial role by reducing inflammation and oxidative stress via Ang1-7 and the Mas receptor, and by reducing susceptibility to cardiovascular diseases (Sparks et al., 2014). Accordingly, ACE2 loss is thought to be a marker of poor prognosis (Duan et al., 2018). In the GI tract, ACE2 plays a role in the transport of amino acids (Sparks et al., 2014).

1.5.2 (Pro)renin Receptor

The (pro)renin receptor (PRR) is a 35 kDa receptor with no intrinsic kinase ability (G. Nguyen et al., 2010). It is a single pass transmembrane protein with a short intracellular domain (ICD); the transmembrane and ICD together are 8.9 kDa, and the extracellular domain (ECD) can be cleaved off to form soluble PRR (sPRR) with a molecular weight of 28 kDa (G. Nguyen et al., 2010). Both full length PRR and sPRR can bind pro-renin and mature renin. Pro-renin levels are 7-9 times higher in the plasma than renin, and it is usually activated by irreversible proteolytic removal of a 43 amino acid pro-segment (G. Nguyen et al., 2010). The binding of pro-renin to PRR allows it to become reversibly active without proteolysis via a conformational change, causing the pro-segment to be folded away (G. Nguyen et al., 2010). While bound to PRR, pro-renin can cleave AGT to produce ATI. Furthermore, while mature renin is bound to PRR it has four times greater activity at cleaving AGT (G. Nguyen et al., 2010).

The binding of pro-renin or renin to PRR induces ERK1/2 and p38 MAP kinase signalling, independent of the RAS (G. Nguyen, 2011; G. Nguyen et al., 2010; Sakoda et al., 2007). ERK1/2 signalling has downstream profibrotic effects acting via TGF- β 1, plasminogen activator inhibitor 2 (PAI-2/SERPINB2), collagen, fibronectin (FN1) and COX2 (G. Nguyen et al., 2010; Ramkumar et al., 2019).

1.5.3 PRR and Wnt Signalling

PRR is a vital component of the Wnt signalling receptor complex (Fig 1.5). This discovery was made during a study investigating which proteins co-purify with the hydrogen ion pump V-ATPase (G. Nguyen et al., 2010). The PRR transmembrane-ICD fragment and full length PRR provide a vital link between the Wnt receptor Frizzled (FZD) and low-density lipoprotein receptor-related protein 5/6 (LRP5/6) (G. Nguyen, 2011; Nusse, 2005). Wnt proteins bind the FZD receptor, causing Dishevelled (Dvl) to be recruited and allowing LRP5/6 to polymerise with FZD (C. Gao et al., 2010; Khalek et al., 2010). PRR then provides a vital link between LRP5/6 and V-ATPase: to become active via phosphorylation, LRP5/6 must be in an acidic environment, and this is achieved through its close proximity to the proton pump V-ATPase as is facilitated by PRR (G. Nguyen, 2011; Nusse, 2005; Ray, 2010). Once active, LRP5/6 binds Axin to prevent it from participating in the β -catenin destruction complex (G. Nguyen, 2011; Nusse, 2005). Additionally, Dickkopf-1 (DKK1) and Dickkopf-2 (DKK2) are inhibitors of LRP5/6 but can be bound by EpCAM to allow LRP5/6 to be activated (L. Huang et al., 2018). Importantly, PRR enables Wnt/ β -catenin signalling to occur via a mechanism which is independent of pro-renin or renin binding to PRR (L. Zhou et al., 2015).

APC is a member of the destruction complex (Markowitz et al., 2009; Valenta et al., 2012). The APC gene mutation is the most common mutation in CRC and is also the cause of FAP (Markowitz et al., 2009; Strate et al., 2005). Glycogen-synthase kinase 3 β (GSK3 β) forms a complex with APC and Axin, which allows GSK3 β to phosphorylate β -catenin and thereby target it for ubiquitination and subsequent degradation (Markowitz et al., 2009; Nusse, 2005; Valenta et al., 2012). This prevents translocation of β -catenin to the nucleus, where it would enable Wnt-mediated gene transcription to occur (Khalek et al., 2010). Therefore, the combination of PRR enabling LRP5/6 to be active and degrade Axin, and the APC gene mutation which is common in CRC, lead to the accumulation of β -catenin and canonical Wnt signalling capability.

β -catenin is an oncoprotein which recruits T-cell factors (Tcf) and lymphoid enhancer factors (Lef) to activate transcription of Wnt-targeted genes (Khalek et al., 2010; Yamashita et al., 2007). Wnt signalling controls cell fate and polarity, axis formation, cell proliferation and migration, as well as adult stem cell self-renewal and differentiation (Khalek et al., 2010; Markowitz et al., 2009; G. Nguyen, 2011; Ramkumar et al., 2019). The activation of Wnt signalling is considered to be the initiating event of CRC and it is constitutively active in CRC and FAP (Markowitz et al., 2009; G. Nguyen, 2011; Ray, 2010). Normal crypt cells of the gut accumulate β -catenin and differentiated villus cells present β -catenin on their basolateral

membranes where it functions in cell adhesion (Khalek et al., 2010). The interaction between β -catenin and Tcf4 has been investigated by deletion of Tcf4, which causes a complete absence of intestinal crypts and proliferative cells, illustrating the importance of the β -catenin-Tcf4 complex in transcribing genes responsible for crypt cell characteristics (Khalek et al., 2010).

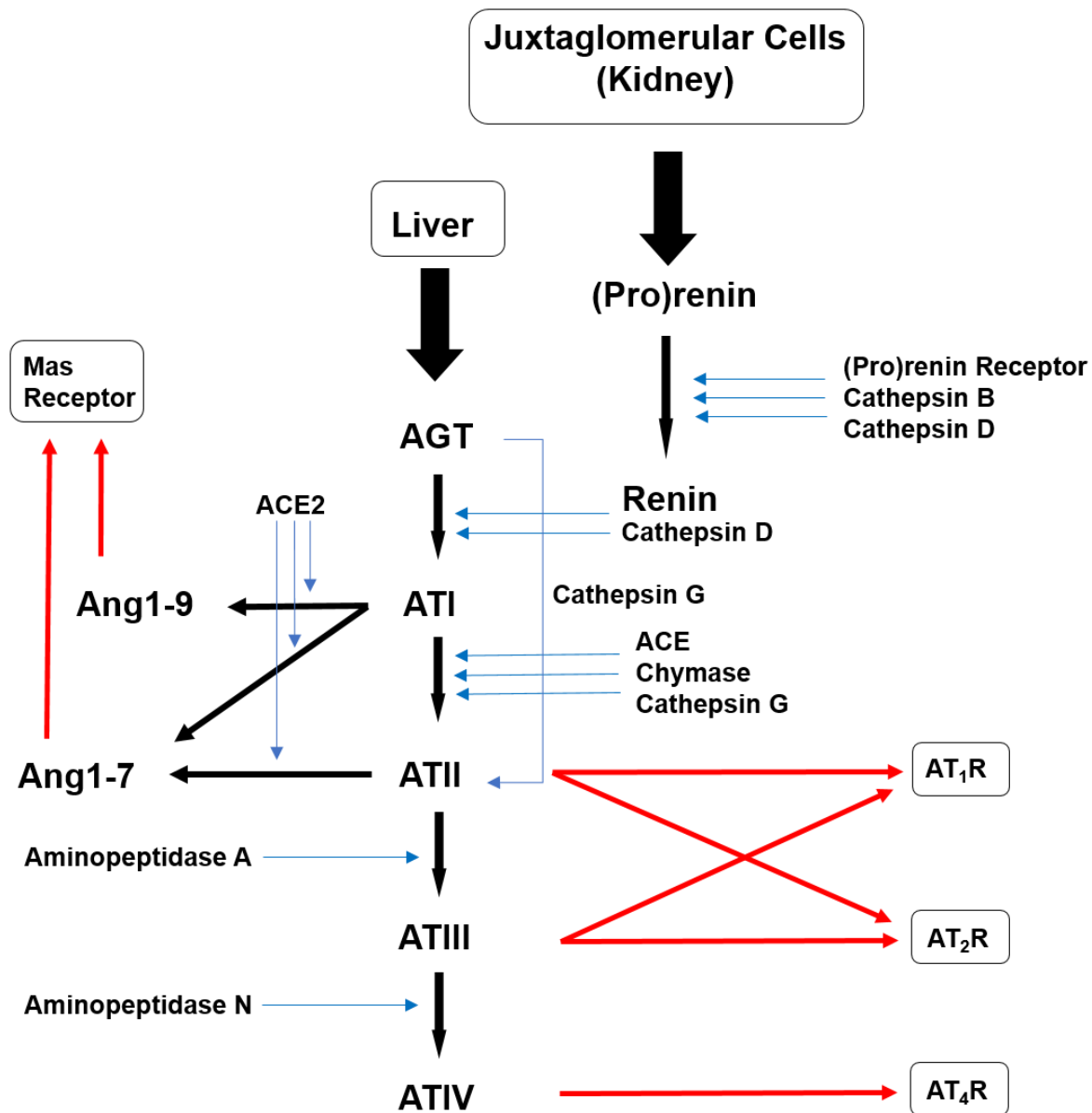


Figure 1.4: Classical RAS and Contemporary additions (modified from Munro MJ, Wickremesekera AC, Davis PF, Marsh R, Tan ST, Itinteang T. (2017) Renin-angiotensin system and cancer: A review. *Integr Cancer Sci Therap.* 4: DOI: 10.15761/ICST.1000231). Once secreted from the kidneys, (pro)renin becomes active upon binding to the (pro)renin receptor. Cathepsins B and D also activate renin. Renin and cathepsin D convert angiotensinogen (AGT) to angiotensin I (ATI). Angiotensin converting enzyme (ACE), chymase and cathepsin G act on ATI to produce angiotensin II (ATII). AGT is also directly converted into ATII by cathepsin G. Angiotensin III (ATIII) is the result of further cleavage of ATII by Aminopeptidase A. Both ATII and ATIII act on angiotensin II receptor 1 (AT1R) and angiotensin II receptor 2 (AT2R). ACE2 converts ATI and ATII into Ang1-9 and Ang1-7, which both bind the Mas receptor. The downstream effects of AT2R and Mas receptor agonism counteract those of AT1R.

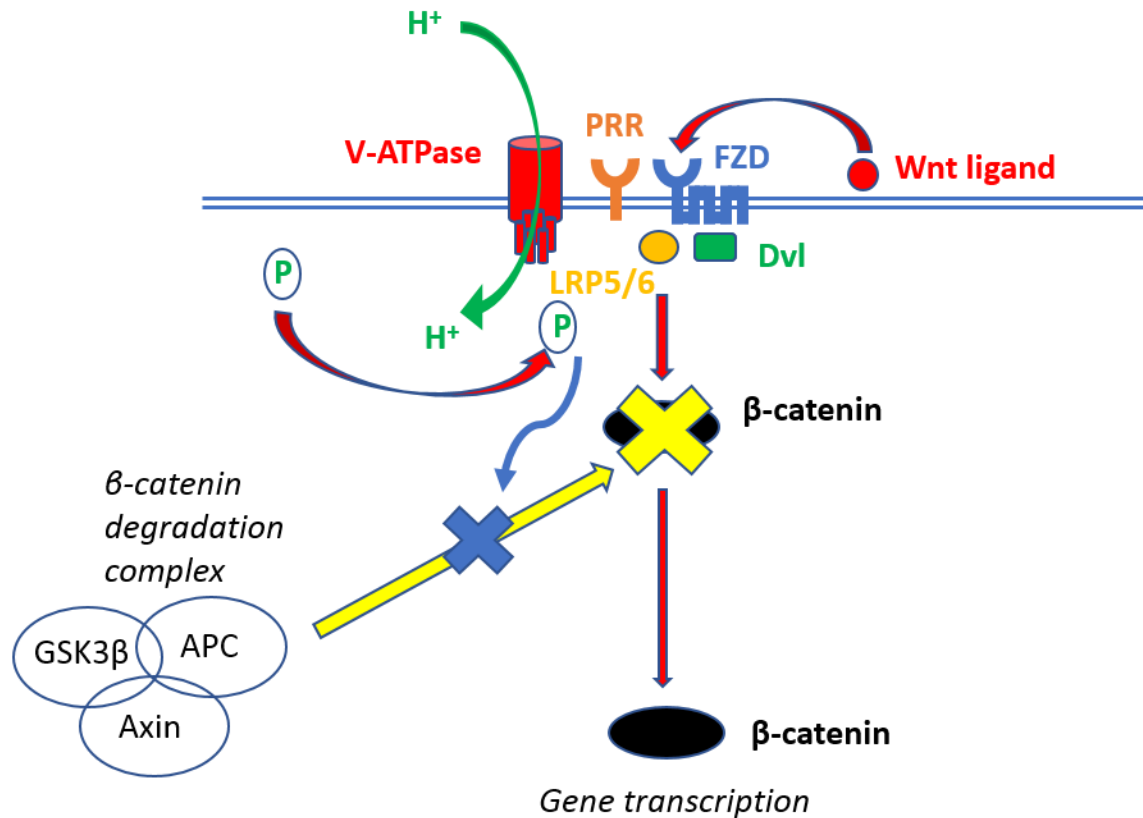


Figure 1.5: PRR and Wnt signaling. PRR links the Frizzled (FZD) receptor and V-ATPase proton pump. V-ATPase pumps protons into the cells, causing the region around the complex to become acidic and therefore allowing LRP5/6 to become phosphorylated. Once phosphorylated, LRP5/6 binds Axin to prevent it from participating in the β-catenin destruction complex. This allows β-catenin signalling to occur via the FZD receptor.

Interestingly, c-MYC, EpCAM, and RAS components including ACE, AT₁R, AT₂R and PRR are among the Wnt-targeted genes which are transcribed by the β-catenin/Tcf/Lef complex (Khalek et al., 2010; Yamashita et al., 2007; L. Zhou et al., 2015). Therefore, although PRR is vital for Wnt signalling in a manner which is renin-independent and not directly attributable to its RAS activity, the involvement of PRR in the Wnt receptor complex enables the downstream transcription of RAS components.

1.5.4 The RAS in Cancer

ATII is the main driver of cancer-related functions of the RAS. However, the downstream effects of the two ATII receptors, AT₁R and AT₂R, seem to play antagonistic roles. AT₁R is typically associated with

adverse or cancer-related outcomes via VEGF, PDGF and FGF, such as angiogenesis, proliferation, inflammation and fibrosis; while AT₂R antagonises these effects to reduce or prevent carcinogenesis (Beitia et al., 2019; Childers, 2015; Munro, Wickremesekera, Davis, et al., 2017).

AT₁R overexpression is seen in some breast cancers and SCCs (Ager et al., 2008; George et al., 2010). It is associated with malignant transformation in breast, ovarian and gastric cancers, poor differentiation of SCCs, and invasiveness of ovarian and cervical cancers (Deshayes et al., 2005; Munro, Wickremesekera, Davis, et al., 2017). AT₁R activation by ATII has been shown to increase global DNA and protein synthesis by around 50% (Sayeski et al., 2001). Signalling downstream of AT₁R can occur via tyrosine kinases in a manner similar to growth factor receptors, however, AT₁R does not have intrinsic tyrosine kinase activity, and instead recruits cytoplasmic “non-receptor” tyrosine kinases such as c-Src and JAK2 (Sayeski et al., 2001). This causes direct activation of the JAK/STAT and ERK/MAPK pathways. Alternatively, AT₁R can act as a G-protein-coupled receptor to induce PI3K signalling and prevent apoptosis by suppressing caspases (Wegman-Ostrosky et al., 2015). There is a large degree of crosstalk between tyrosine kinase and G-protein signalling systems to produce a wide range of downstream effects (Sayeski et al., 2001; Wegman-Ostrosky et al., 2015). AT₁R also upregulates mediators of inflammation, including IL-1 β , TNF- α , TGF- β , STATs, and HIF-1 α (Wegman-Ostrosky et al., 2015).

Conversely, AT₂R has been shown to induce apoptosis in cancer cells (H. Li et al., 2009), and AT₂R knock-out mice display increased growth and vascularisation of xenotransplanted tumours (Doi et al., 2010). AT₂R causes phosphotyrosine phosphatase to become active and inhibit tyrosine kinase signalling (Wegman-Ostrosky et al., 2015). Furthermore, it works with p53 and MAPK signalling to induce apoptosis (Wegman-Ostrosky et al., 2015). Expression levels of AT₂R are highest in the developing embryo and decline continuously from the time of birth, with residual low expression in the adult kidney, adrenal gland and brain (Sparks et al., 2014). Despite its beneficial opposition to AT₁R signalling, overexpression of AT₂R has been documented in various disease states (Sparks et al., 2014).

There is evidence of AT₁R and AT₂R in the plasma membrane and AT₂R in the nuclei of CRC cells, in which ATII binding to AT₁R leads to tumour growth, invasion and VEGF-A secretion, and AT₂R opposes all actions of AT₁R at high levels of ATII (Zhou et al., 2014).

The importance of ATII is further highlighted by the finding that reduced activity of aminopeptidases, which convert ATII to ATIII, correlates with a poorer prognosis in CRC patients, suggesting that accumulation of ATII leads to adverse outcomes (Beitia et al., 2019). Furthermore, the

Mas receptor is overexpressed in CRC and its abundance correlates with grade, suggesting that the ATII cleavage product Ang1-7 may play a role in progression (Beitia et al., 2019).

Retrospective studies have shown that patients taking RAS inhibitors have reduced incidence of CRC, and reduced risk of tumour recurrence and metastasis (Childers, 2015). Specifically, AT₁R inhibition has the greatest effect in reducing the development and metastasis of solid tumours (Childers, 2015).

The liver produces AGT and naturally has high levels of ATII and AT₁R. The liver is the most common site of CRC metastasis, and these secondary tumours contain higher levels of ACE and Mas receptor compared with the primary tumour (Childers, 2015). The liver may be a lucrative location for metastasis due to its production of AGT, which CRC metastases could utilise via elevated levels of ACE to produce more ATII and drive cancer growth via AT₁R-mediated mechanisms. Evidence for this hypothesis has been provided with the observation that RAS inhibitors reduce the spread of CRC to the liver (Y. Luo et al., 2011; Shimomoto et al., 2012).

A large-scale meta-analysis by Sun *et al.* (Sun et al., 2017) sought to clarify whether outcomes were influenced by the type of cancer or by the RAS inhibitor class administered. The authors show that administration of AT₁R blockers (ARBs) or ACE inhibitors (ACEIs) is associated with lower risk of tumour recurrence and increased overall survival (OS), progression-free survival (PFS) and disease-free survival (DFS) compared with patients not using RAS modulating medication (Sun et al., 2017). ARBs seemed to have a larger positive effect than ACEIs. However, patients with breast, oesophageal and biliary tract cancers do not benefit from ARBs or ACEIs, and patients with acute myelocytic leukaemia and multiple myeloma who are administered RAS inhibitors have poorer survival outcomes (Sun et al., 2017). CRC is amongst the cancer types which display a positive trend in survival associated with ARB and ACEI administration, with up to 20% increases in OS, PFS and DFS rates. This study highlights that the RAS is relevant to cancer and underscores the need for further study of the mechanisms which may be involved.

Despite the fact that the RAS has been investigated in a range of cancers, including co-localisation of iPSC markers and RAS components in glioblastoma (A. R. Bradshaw et al., 2016) and oral cavity cancers (R. S. Ram et al., 2017) by our laboratory, it is yet to be elucidated whether colon CSCs express RAS components and, if so, whether the RAS plays a functional role in CSC behaviour.

1.5.5 Cathepsins as Bypass Loops of the RAS

Cathepsins are lysosomal peptidases which belong to the papain family (S. Chen et al., 2017; Turk et al., 2012). They are translated by ribosomes into a zymogen (pre-pro-cathepsin) and transported to the endoplasmic reticulum (ER) (Liaudet-Coopman et al., 2006; Soond et al., 2019). The signal peptide is then removed, producing the pro-enzyme form (S. Chen et al., 2017; Liaudet-Coopman et al., 2006). In the ER and Golgi apparatus, cathepsins undergo a range of PTMs, including phosphorylation and glycosylation (S. Chen et al., 2017; Liaudet-Coopman et al., 2006; Soond et al., 2019). The modified pro-enzyme is labelled with mannose-6-phosphate to direct its transport into lysosomes (Liaudet-Coopman et al., 2006), where the acidic pH allows for the removal of the pro-domain either by a conformational change or through the activity of other proteases (S. Chen et al., 2017; Soond et al., 2019). Once activated most cathepsins require the acidic environment within the lysosome to operate optimally (Turk et al., 2012). In cancer, it is common for cathepsins to be secreted and then act within the cytoplasm or extracellularly (Soond et al., 2019).

Cathepsins are involved in cell cycle regulation and proliferation by processing cytokines and histones (Turk et al., 2012). They carry out normal protein degradation and turnover, including a role in immune responses involving MCH proteins and antigen processing (S. Chen et al., 2017; Garcia et al., 1996; Kuester et al., 2008; Turk et al., 2012). Cathepsins have a wide range of roles related to apoptosis. When grown in low-serum conditions, cathepsin D overexpression facilitates cells to undergo autophagy-induced apoptosis more quickly, whereas cathepsin B overexpression slows this process (Uchiyama, 2001). Furthermore, cathepsin B is involved in a type of autophagy called tumour cannibalism, which aids immune avoidance and nutrient recycling in tumour cells (Gondi et al., 2013). Cathepsins B, D and G are of particular interest in this study due to their roles in the RAS and in cancer.

Cathepsins B, D and G have been shown to function in the RAS (Munro, Wickremesekera, Davis, et al., 2017). Cathepsin B co-localises with renin in the kidney (Neves et al., 1996), and when isolated from kidney tissue it specifically removes the pro-segment from renin *in vitro* without causing further degradation (P. H. Wang et al., 1991). Cathepsin D is also capable of activating pro-renin under physiological conditions, though at a much lower rate than pepsin because the activity of cathepsins is more general and broad (Morris, 1978). Cathepsin D also acts in a manner redundant to ACE by converting ATI into ATII (Ferrario et al., 1989). Cathepsin G is involved in another RAS bypass reaction through its ability to convert AGT directly into ATII or via initial conversion to ATI (Ferrario et al., 1989; Tonnesen et

al., 1982). In fact, cathepsin G-mediated ATII production is in the same order of magnitude as that of ACE, suggesting that it plays a larger role in the RAS than previously suspected (Rykl et al., 2006).

Cathepsin B is a cysteine protease (Turk et al., 2012) which acts primarily in lysosomes, where it is cleaved and activated by cathepsin D (Bian et al., 2016; S. Chen et al., 2017). Pro-cathepsin B is 43-46 kDa, and its active form is a 31 kDa protein comprising a 25 kDa heavy chain and a 5 kDa light chain (Aggarwal et al., 2014). Cathepsin B has been found in most human tissues (Turk et al., 2012), including the thyroid gland where it processes and activates thyroid hormones (S. Chen et al., 2017), and in macrophages where it digests proteins from phagocytosed pathogens for antigen presentation to other immune cells (Gondi et al., 2013). It can function at acidic pH as a carboxydipeptidase, and neutral pH as an endopeptidase, meaning that it can digest proteins in the lysosome, the cytosol, or extracellularly (Aggarwal et al., 2014). Cathepsin B enables tissue remodelling and cell migration by cleaving components of the ECM, including laminin, collagen IV, elastin, fibronectin (FN1) and E-cadherin (S. Chen et al., 2017; Turk et al., 2012). In this capacity, it initiates a cascade which ultimately leads to matrix metalloproteinase (MMP) activation, while also degrading MMP inhibitors (TIMPs) (Gondi et al., 2013; Ruan et al., 2015). However, its localisation affects its function: the MMP-activating cascade is only initiated when cathepsin B is localised to caveolin-1 at the cell membrane, which occurs most commonly in tumour cells (Ruan et al., 2015). The role of cathepsin B in physiological homeostasis is disrupted by aberrant regulation, often due to splice variants, abnormal localisation or increased expression (Gondi et al., 2013).

Cathepsin D is an aspartic protease consisting of a 34 kDa heavy chain and a 14 kDa light chain which are non-covalently linked (Turk et al., 2012). When exposed to an acidic pH, such as in the lysosome, pro-cathepsin D undergoes a conformational change which removes the pro-segment from the active site (Benes et al., 2008). Its primary functions are degradation of proteins within the lysosome following autophagy, and activation of other proteases (Benes et al., 2008; Turk et al., 2012). For example, pro-cathepsin B is cleaved by cathepsin D to become active (S. Chen et al., 2017); cathepsin D is also responsible for degrading and therefore deactivating cathepsin B (Kuester et al., 2008). It plays a vital role in tissue homeostasis, especially in the intestinal epithelium where it activates growth factors, however, it is not necessary for embryonic development (Benes et al., 2008; Liaudet-Coopman et al., 2006). Cathepsin D can integrate with traditional apoptosis pathways to trigger cell death in response to chemotherapeutic compounds (Benes et al., 2008; Liaudet-Coopman et al., 2006).

Cathepsin G is a serine endopeptidase with an optimal pH range of 7-8 (Burster et al., 2010; Kudo et al., 2009; Turk et al., 2012) which is expressed by cells of the myeloid lineage (Burster et al., 2010). The

2-amino acid pro-segment of pro-cathepsin G is cleaved by cathepsin C to produce its active 26 kDa form (Burster et al., 2010; Kudo et al., 2009), and mature cathepsin G is capable of activating cathepsin B (Mason et al., 2011). Cathepsin G is most commonly associated with neutrophils, which express cathepsin G, neutrophil elastase, and proteinase 3 (Burster et al., 2010). In this capacity, cathepsin G helps neutrophils to balance tissue protection and destruction during inflammation, destroy internalised pathogens, modify cytokines and cell surface receptors, and potentiate chemotactic factors CCL15 and CXCL5 to attract monocytes and more neutrophils, respectively (Burster et al., 2010). It can act proteolytically by cleaving proteins, or non-proteolytically by using cationic residues to interfere with the negatively charged surface membranes of bacteria (Burster et al., 2010). Inhibition of cathepsin G reduces the levels of VEGF and MCP-1, leading to reduced TGF- β signalling and impaired angiogenesis (S. Gao et al., 2018; Wilson et al., 2010).

Our laboratory has previously co-localised cathepsins with iPSC markers in CA metastases to the liver (Mehrotra et al., 2018), but this has not been investigated in primary CA.

1.6 Epithelial-to-Mesenchymal Transition

EMT occurs during embryogenesis but is also a vital process in the initiation of metastasis (Gawlik-Rzemieniewska et al., 2016; C. W. Liu et al., 2014; Mani et al., 2008; L. Nguyen et al., 2016). Cancer cells which have undergone EMT tend to possess CSC properties and the capacity for migration and invasion, and CSCs in a range of cancers have a mesenchymal phenotype (Mani et al., 2008). OCT4 and NANOG are strongly associated with EMT; knock-down of OCT4 reduces cancer cell migration, and EMT can be induced by overexpression of NANOG causing increased migration (Dai et al., 2013; Gawlik-Rzemieniewska et al., 2016). The EMT gene *Snai1* increases NANOG expression via SMAD1/AKT signalling, and by phosphorylating GSK3 β to deactivate it and allow for NANOG transcription via Wnt signalling (C. W. Liu et al., 2014).

A study of the relationship between the RAS and EMT in CRC showed that cellular migration increases in response to ATII and is mediated by both AT₁R and AT₂R (L. Nguyen et al., 2016). Interestingly, the downstream effects of AT₁R interacting with ATII are consistent with the changes expected in EMT (L. Nguyen et al., 2016). This suggests that the RAS may play a key role in CRC metastasis.

CRC metastases in the liver express high levels of ACE (Neo et al., 2010). In combination with the large amount of AGT produced by the healthy liver, this leads to high levels of ATII and increases

downstream RAS activity (Neo et al., 2010). Experiments in a mouse model showed that the ACEI captopril markedly decreases metastasis of CRC to the liver, as well as decreasing the levels of AGT produced by the liver which therefore deprives the metastases of ATII (Neo et al., 2010). IHC staining shows that liver-resident macrophages are responsible for some of the expression of ACE within liver metastases and AT₁R within stromal intrusions, as these cells are known to express both of these RAS components (Neo et al., 2010). Macrophages may in fact assist tumour growth via growth factors and promotion of angiogenesis, as well as carrying out the tumour-reducing effects of RAS modulation (Neo et al., 2010).

The proven involvement of the RAS in EMT, the knowledge that stem cells can both undergo EMT and be generated through EMT, and the prevalence and mortality of CRC metastasis to the liver, underscore the need for research into the roles of the RAS in CSCs.

TGF- β also has a well-established association with EMT. TGF- β signalling via receptor tyrosine kinases, such as FGF, EGF and PDGF, induces EMT (J. Xu et al., 2009). The SMAD transcription factor complex regulates the transcription of three key EMT families: snail, ZEB and bHLH (J. Xu et al., 2009). The actions of these families lead to reduced tight junctions and E-cadherin, as well as increased N-cadherin and MMPs (J. Xu et al., 2009). SMADs upregulate GSK3 β expression, which goes on to activate Snai1. Snai1 has been identified in all cases of EMT; it directly binds the promoter region of E-cadherin to suppress its transcription (Miyazono, 2009; J. Xu et al., 2009). EMT induced by TGF- β has been linked to de-differentiation and the acquisition of stem-like characteristics (Katsuno et al., 2013), as well as producing cancer-associated fibroblasts and other stromal cells which support malignant transformation and cancer progression (Katsuno et al., 2013; Miyazono, 2009).

1.7 Aims and Hypotheses

This project focuses on primary colon adenocarcinoma (CA), which is the most common type of CRC. CSCs are considered to be a pluripotent subpopulation of tumour cells. The iPSC markers, which drive pluripotency and are capable of reprogramming differentiated cells to a pluripotent state, were chosen to be investigated as potential markers of CSCs. The role of the RAS in malignancy is an emerging focus of cancer research. This thesis assesses whether RAS components are present within the CSCs of CA, which is yet to be established in the literature.

Hypothesis 1: CA contains CSCs, tumour-initiating cells which can be identified by their expression of a range of iPSC markers. It may be that iPSC markers are expressed at a higher level in CA relative to normal tissue, as opposed to being exclusively expressed by the cancer. Some markers are known to be expressed by normal tissues but have an association with cancer, such as c-MYC, and so their expression with other members of a panel will give them greater value. Marker expression levels suggest a distinct proteomic pattern, so it is thought that this may hint at molecular networks that drive cancer growth and metastasis.

Hypothesis 2: CSCs in CA express components of the RAS, including cathepsins that constitute bypass loops of the RAS, which may play a role in CSC regulation. If RAS components are expressed in CA and localised to tumour cells and CSCs, then they are likely to reflect those found in the normal colon, though their abundance will presumably be altered.

Aim 1 (Chapters 3 and 6): To compare CA-derived primary cell lines to their parent tissues and identify key signalling pathways and how they relate to CSCs and the RAS.

In order to compare CA-derived primary cell lines to their parent CA tissues, next-generation sequencing will be performed using the OncoPrint® Tumor Mutation Load Assay, which sequences 409 cancer-related genes. Mass spectrometry will be carried out to compare patient matched normal colon (NC) to CA tissue samples and the CA-derived primary cell lines, allowing for analysis of upregulated proteins and pathways in these CA samples relative to the patient-matched NC samples.

Aim 2 (Chapters 4 and 5): To identify and characterise the CSC subpopulations within primary CA tumours.

First, the number of cells expressing iPSC markers in LGCA and HGCA tissue samples will be compared with patient-matched NC samples using IHC staining of tissue sections. The co-expression of multiple markers will be interrogated using immunofluorescence (IF) staining.

Gene expression of these markers will be investigated and quantified by *in situ* hybridisation (ISH) and RT-qPCR.

Furthermore, the stem cell properties of CA-derived cell lines will be investigated. The formation and maintenance of tumourspheres in culture and the ability of these cells to differentiate down three embryonic lineages would suggest that cells with pluripotent characteristics are present. The gene and

protein expression of pluripotency markers by tumoursphere-forming cells will be analysed using immunocytochemistry (ICC), RT-qPCR and western blot (WB).

Aim 3 (Chapters 7 and 8): To determine the expression of components of the RAS by CSCs in primary CA.

To determine the presence of RAS markers, including cathepsins, the same tests will be implemented as for iPSC markers. IHC staining will reveal the locations within the tissue of cells expressing RAS components, as well as their sub-cellular localisation. Confirmation of whether these markers are transcribed and translated by the tissue-derived primary cell lines will be undertaken using RT-qPCR and WB, also allowing for a quantitative comparison of CA tissues and cell lines to patient matched NC samples.

Further to this, co-staining tumour sections with combinations of iPSC markers and RAS components by IF staining will reveal whether the CSCs express RAS components.

Finally, RAS modulation will be carried out on cultured cells to investigate whether targeting the RAS has any effect on colon CSCs.

Chapter 2: Methods and materials

2.1 Tissue Samples

Formalin-fixed paraffin-embedded (FFPE) and snap-frozen tissue samples of low-grade CA (LGCA) from 12 patients and high-grade CA (HGCA) from 8 patients, with patient-matched normal colon (NC) from 19 of the 20 patients, were provided by the Gillies McIndoe Research Institute Tissue Bank (GMRITB) for this study, which was approved by the Central Health and Disability Ethics Committee (Ref. 15/CEN/106).

2.2 Immunohistochemistry

2.2.1 DAB Immunohistochemical staining

3,3-diaminobenzidine (DAB) immunohistochemical (IHC) staining was carried out on all 20 cases of CA, as well as patient-matched NC tissue samples from 19 of these patients. These FFPE tissue samples were sectioned into 4 µm slices and mounted on slides. Staining was carried out for the iPSC markers OCT4, SOX2, NANOG, KLF4 and c-MYC, CSC markers CD133 and LGR5, epithelial marker EpCAM, and RAS components PRR, ACE, ACE2 and AT₂R.

2.2.2 Positive controls

Positive control tissues were included in each run to validate staining success. These tissues were seminoma (OCT4, NANOG), skin (SOX2), prostate (c-MYC), breast (KLF4), kidney (ACE, ACE2 and AT₂R), and placenta/tonsil (PRR). Each IHC staining procedure also included isotype control antibodies as a negative control.

2.2.3 Primary antibodies and staining protocols

Pre-defined automated staining protocols, which are optimised for each antibody, were carried out using the Leica BOND™ RX Research Auto-stainer (Leica, Nussloch, Germany). Primary antibodies are outlined in Table 2.1.

Table 2.1: Primary antibodies used for immunohistochemical and immunofluorescence staining

Marker	Species/clonality	Dilution	Catalogue number
OCT4	Mouse monoclonal	1:30	MRQ-10 (Cell Marque)

SOX2	Rabbit polyclonal	1:200	PA1-094 (ThermoFisher Scientific)
SOX2	Rabbit polyclonal	1:200	ab97959 (Abcam)
NANOG	Rabbit monoclonal	1:200	443R-16 (Cell Marque)
NANOG	Mouse monoclonal	1:100	ab62734 (Abcam)
KLF4	Rabbit polyclonal	1:200	NBP2-24749SS (Novus)
c-MYC	Mouse monoclonal	1:1000	ab32 (Abcam)
c-MYC	Rabbit monoclonal	1:100	ab32072 (Abcam)
EpCAM	Rabbit monoclonal	1:200	ab124825 (Abcam)
CD133	Rabbit polyclonal	1:200	ab19898 (Abcam)
LGR5	Mouse monoclonal	1:1000	MA5-25644 (ThermoFisher Scientific)
PRR	Rabbit polyclonal	1:200	ab264763 (Abcam)
ACE	Mouse monoclonal	1:40	3C5 (Serotec)
ACE	Rabbit polyclonal	1:50	PA5-83080 (ThermoFisher Scientific)
ACE2	Mouse monoclonal	1:1000	MAB933 (ThermoFisher Scientific)
AT ₂ R	Rabbit polyclonal	1:2000	NBP1-77368 (Novus)

The predefined protocols, designated as F, TF and double-TF, were used for IHC staining on the BOND™ RX, which only differed in terms of exposure time to the primary antibody. All three protocols had the following base template: peroxide block (5 min), wash solution (3x washes), primary antibody (15 min for protocol F, 45 min for protocol TF, and 2x 45 min for protocol double-TF), wash solution (3x), post-primary (8 min), wash solution (3x 2 min), polymer (8 min), wash solution (2x 2 min), dH₂O wash (1x), DAB Refine (1x wash followed by 1x 10 min), dH₂O wash (3x), hematoxylin (5 min), dH₂O wash (1x), wash solution (1x), dH₂O wash (1x). Protocol F was run for NANOG, KLF4, c-MYC, ACE and AT₂R. Protocol TF was run for SOX2. Protocol double-TF was run for PRR and OCT4.

2.2.4 Immunofluorescence staining

Protein co-localisation was carried out using dual immunofluorescence (IF) staining. Primary antibodies were the same used for IHC staining. Secondary antibodies used were Vectafluor Excel goat anti-mouse 488 (ready-to-use; cat # DK2488, Vector Laboratories, Burlingame, CA, USA) and Alexa Fluor donkey anti-rabbit 594 (1:500; cat # ab150076, Life Technologies, Carlsbad, CA, USA). All stained slides were mounted using Vecta Shield Hardset mounting medium with 4',6'-diamino-2-phenylindole (DAPI) nuclear stain (cat

H-1200, Vector, Abacus DX, Auckland, NZ). Negative controls were performed using matched isotype controls for both mouse (ready-to-use; cat#IR750, Dako, Copenhagen, Denmark) and rabbit (ready-to-use; cat#IR600, Dako).

2.2.5 *In situ* Hybridisation

RNA *in situ* hybridisation (ISH) was performed on FFPE tissues sectioned into 4 µm slices and mounted on slides. Staining was carried out on the BOND™ RX Research Auto-stainer. The ViewRNA eZ Detection Kit (Affymetrix, Santa Clara, CA, USA) was used with probes for OCT4 (NM_002701), SOX2 (NM_003106), NANOG (NM_024865), KLF4 (NM_004235) and c-MYC (NM_002467) to detect the presence of mRNA. Positive control tissues were seminoma (OCT4, NANOG, KLF4), normal skin (SOX2) and normal colon (c-MYC). To determine the specificity of probes, negative controls were created using a probe for *Bacillus* (NM_L38424).

2.2.6 Image capture and analysis

Images of IHC and ISH slides were captured using an Olympus BX53 light microscope, Olympus SC100 digital camera and cellSens 2.0 software (Olympus, Tokyo, Japan). Images were captured using ImageJ software (National Institutes of Health (NIH), Bethesda, MD, USA). The location of expression of iPSC and RAS markers was recorded. This included commenting on whether staining was present in the epithelium, stroma, blood vessel endothelium and/or smooth muscle, and whether it was nuclear, perinuclear, cytoplasmic and/or membranous, or any combination thereof.

IF-stained slides were visualised and imaged using an Olympus FV1200 biological confocal laser-scanning microscope (Olympus) and processed using cellSens 2.0 software (Olympus).

2.3 Western blotting

2.3.1 Protein extraction from tissue

Protein extractions from pieces of snap-frozen tissue were performed using a volume of RIPA buffer equal to ten times the volume of tissue used (e.g. 500 µL to 50 mg). Solution was prepared by mixing Pierce™ RIPA buffer (cat # 89901, ThermoFisher Scientific, Waltham, MA, USA) and Halt™ Protease and Phosphatase inhibitor cocktail (cat # 78442, ThermoFisher Scientific) at a ratio of 99:1. RIPA solution was

added to the tissue piece and ground down in a glass homogeniser or in an Eppendorf tube with an Axygen™ plastic pestle (cat # PSE-15-B-SI; ThermoFisher Scientific) until dissociated. Solution was briefly vortexed and agitated on ice for 1 h. After 30 min, the sample was re-homogenised and returned to the agitator. Following agitation, the tube was centrifuged at 17,000 g for 10 min at 4°C. The supernatant, containing isolated proteins, was retained in a fresh tube.

2.3.2 Protein extraction from cells

Protein extractions from cell pellets washed with PBS were performed using 200 µL of RIPA buffer per million cells. RIPA solution was added and the cell pellet was resuspended by pipetting up and down until completely dissociated. Solution was briefly vortexed and agitated on ice for 30 min, followed by centrifugation at 17,000 g for 10 min at 4°C. The supernatant, containing isolated proteins, was retained in a fresh tube.

2.3.3 Protein quantification

Protein was quantified either by Qubit or BCA. Qubit was carried out as per the manufacturer's instructions (cat # Q33211, ThermoFisher Scientific). Briefly, the three supplied Qubit protein standards were made by diluting the stock 10µL:190µL in Qubit buffer. Protein extracts were diluted 1µL:199µL in Qubit buffer and left to stand for 15 min. The Qubit standards were used to create a standard curve before quantifying each sample.

For BCA quantification (cat # 23227, ThermoFisher Scientific), an eight-point serial dilution of BSA protein standards was used to create a standard curve. BCA Reagents A and B were combined at a 50:1 ratio, and 200 µL of this working reagent was added to each well of a black-walled clear-bottom 96-well plate. BSA standards were run in duplicate, with 10 µL of each added to the appropriate wells. Samples to be quantified were also run in duplicate, but only 5 µL of each sample was added to each appropriate well to prevent the readings from being above the range of the standard curve. The plate was incubated at 37°C for 30 min before being read in a plate reader.

2.3.4 Gel electrophoresis

Protein samples were diluted in 1x Bolt LDS sample buffer (cat # B0007, ThermoFisher Scientific) and heated at 85°C for 5 min. Samples were run on Bolt 4-12% Bis-Tris gels (cat # NW04125BOX, ThermoFisher Scientific) with 20 µg of sample protein per lane, and 1-2 µL of Precision Plus Protein Kaleidoscope MW ladder (cat # 1610375, Biorad, Rosedale, Auckland, NZ) run in the first lane. Gels were run in Bolt MES SDS Running Buffer (cat # B0002, ThermoFisher Scientific) for 50 min at 150V, 3.00A and 300W.

2.3.5 Western blotting

Electrophoresed proteins were transferred to PVDF membranes using an iBlot 2 apparatus (cat # IB21001, ThermoFisher Scientific) with the following predefined program: 20V for 1 min, 23V for 4 min, and 25V for 2 min. Once transferred, the membrane was briefly washed in water and blocked using iBind Flex FD solution (cat # SLF2019, ThermoFisher Scientific) for 5 min. An iBind card (cat # SLF1010 or SLF2010, ThermoFisher Scientific) was placed in the iBind apparatus (cat # SLF1000 or SLF2000, ThermoFisher Scientific) and soaked with iBind Flex FD solution. The membrane was then placed protein-side down on the card and a roller was used to remove any bubbles.

Primary antibodies are outlined in Table 2.2. Secondary antibodies included: HRP-linked goat anti-rabbit (1:1000; cat # ab6721, Abcam), HRP-linked goat anti-rabbit (1:1000; cat # 111-035-045, Jackson Immunology), and Alexa Fluor® 488 donkey anti-mouse (1:1000; cat # A-21202, ThermoFisher Scientific).

Table 2.2: Primary antibodies used for western blotting

Marker	Species/clonality	Dilution	Catalogue number
OCT4	Rabbit monoclonal	1:500	ab109183 (Abcam)
SOX2	Rabbit polyclonal	1:1000	48-1400 (ThermoFisher Scientific)
NANOG	Rabbit monoclonal	1:1000	ab109250 (Abcam)
NANOG	Mouse monoclonal	1:1000	ab62734 (Abcam)
KLF4	Rabbit polyclonal	1:1000	NBP2-24749 (Novus)
c-MYC	Rabbit monoclonal	1:1000	ab32072 (Abcam)
EpCAM	Rabbit polyclonal	1:1000	ab71916 (Abcam)
α-SMA	Rabbit polyclonal	1:2000	ab5694 (Abcam)
CD133	Rabbit polyclonal	1:2000	ab19898 (Abcam)
LGR5	Mouse monoclonal	1:2000	MA5-25644 (ThermoFisher Scientific)
PRR	Rabbit polyclonal	1:250	ab40790 (Abcam)

ACE	Goat polyclonal	1:200	sc12184 (Santa Cruz)
ACE2	Mouse monoclonal	1:500	MAB933 (R&D Systems)
AT ₂ R	Rabbit monoclonal	1:500	ab92445 (Abcam)
Cathepsin B	Mouse monoclonal	1:1000	ab58802 (Abcam)
Cathepsin D	Rabbit monoclonal	1:1000	ab75852 (Abcam)
α -tubulin	Mouse monoclonal	1:2000	ab7291 (Abcam)
CD44	Rabbit polyclonal	1:5000	ab157107 (Abcam)
FN1	Rabbit polyclonal	1:500	ab2413 (Abcam)
S100A8	Rabbit monoclonal	1:2000	ab92331 (Abcam)
S100A9	Rabbit monoclonal	1:500	ab92507 (Abcam)

Membranes were incubated in the iBind apparatus for 2.5 h or overnight. Following incubation, the membranes were briefly washed in water, developed using Clarity Western ECL Substrate (cat # 170-5061, Biorad) and imaged using a ChemiDoc MP Imaging System (Biorad) and ImageLab 6.0 software (Biorad). Densitometry was performed using ImageLab 6.0, with the intensity values for the protein-of-interest normalised against α -tubulin. Densitometry data were analysed using GraphPad Prism 8 (San Diego, CA, USA).

2.4 Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR)

2.4.1 RNA Extraction

RNA was extracted using a QIAcube (Qiagen) by running their preset protocols for extraction of RNA from tissue and cells.

For tissue extractions, tissue pieces were cut, placed in a vial and weighed to ensure they were less than 20 mg. 350 μ L of RLT lysis buffer and 14 μ L DTT were added to each tissue vial. Tissues were homogenised using a tissue homogeniser (OMNI International, Kennesaw, USA) until entirely dissolved, and then centrifuged at 13,300 RPM for 3 min at 17°C. The supernatant was retained in a 2 mL round-bottom Eppendorf tube.

For extraction from cells, a thawed cell pellet of a maximum of 5×10^5 cells was centrifuged at 17,000 g for 3 min at 17°C. Supernatant was discarded and 350 μ L of RLT lysis buffer with DTT (40 μ L/mL) was added to each pellet. Pellets were resuspended by pipetting.

DNase was prepared by diluting in RDD buffer as per Qiagen instructions, in a 2 mL round-bottom Eppendorf tube. The DNase tube was placed in “slot A” within the QIAcube (Qiagen).

In the QIAcube Rotor Adapters, a pink spin column was placed in position L1, and a labelled 1.5 mL Eppendorf collection tube was placed in position L3 for RNA collection. Rotor adapters were placed in the centrifuge component of the QIAcube in numbered positions, and the lysed sample tubes placed in adjacent slots numbered corresponding to centrifuge positions. Lids were removed from the QIAcube buffer bottles and buffers were topped up as required. On the QIAcube display, “Mini kit” was selected for tissue extraction and “Micro kit” was selected for cell extraction. DNA digest option was selected. Following the run, only the 1.5 mL Eppendorf collection tube was retained, containing extracted RNA in 15 µL of buffer (micro kit) or 45 µL of buffer (mini kit). RNA was quantified using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific) before being diluted with RNase-free water and aliquoted to give a volume containing enough RNA for triplicates to be run during PCR.

2.4.2 Rotor Gene PCR Protocol

A Rotor-Gene Q (Qiagen) was used for RT-qPCR reactions. For each sample, three 0.2 µL PCR tubes were labelled, plus one for positive control and one for no-template control. A mastermix for each gene was prepared immediately prior to each run, containing Master Mix, gene-of-interest primer/probe, Reverse Transcriptase and RNase-free water. Mastermix was aliquoted to each PCR tube before adding 40 ng RNA. Tubes containing mastermix or RNA were kept on ice at all times. PCR tubes were placed in the Rotor Gene apparatus and the standard PCR run template was selected. This protocol ran as follows: 1. Reverse transcription at 50°C for 15 min; 2. Taq Polymerase activation at 95°C for 5 min; 3. Cycling: denature at 95°C for 15 sec, anneal and extend at either 60°C (AT₂R, PRR) or 62°C (OCT4, SOX2, NANOG, KLF4, c-MYC, ACE, CTSB, CTSD) for 15 sec. Primer optimisation was carried out prior to use to determine a suitable annealing temperature and whether any additives were required in the mastermix. 5% DMSO was added to the mastermix for OCT4, SOX2, NANOG, KLF4 and c-MYC, and 1 M Betaine was added to the mastermix for ACE, cathepsin B and cathepsin D. Due to availability and difficulties in culturing patient-matched normal colon cell lines, RNA abundance was measured relative to pooled data from 4 NC tissues which were run in triplicate in each PCR experiment.

Following the run, tubes were stored at -20°C until being run on an E-Gel™ EX 2% Agarose Gel (cat # G401001; ThermoFisher Scientific) to confirm amplicon size and negative control. Results were exported from the software as a Word document for analysis.

2.5 Cell Culture

2.5.1 Cell culture

Primary cell lines were provided by the GMRITB with approval by the Central Health and Disability Ethics Committee (Ref. 15/CEN/106). Primary cell lines were derived from CA tissue samples at the GMRITB by the explant method. This involves embedding small tissue pieces (<10mm³) in Matrigel, which allows cells to migrate out from the tissue piece. Once there is considerable outgrowth, the matrix is dissociated using dispase and diluted with PBS. The cells are pelleted and plated to cell culture flasks to grow as a monolayer in DMEM media supplemented with 10% FCS and 5% mTeSR. Commercial cell lines CaCo2 (cat # HTB-37, ATCC, In Vitro Technologies, Auckland, New Zealand), 3T3 (cat # CRL-1658, ATCC) and NTERA-2 (cat # CRL-1973, ATCC) were used as positive controls for tumoursphere formation assays, differentiation assays and expression of iPSC markers, respectively. Cells were cultured in Nunc™ EasYFlasks™ (ThermoFisher Scientific) using DMEM media with high glucose and containing pyruvate (cat # 10569010, ThermoFisher Scientific) and supplemented with 10% fetal calf serum (FCS; cat # 10091148, ThermoFisher Scientific), 5% mTeSR Complete (cat # 85850, STEMCELL Technologies, Tullamarine, Victoria, Australia), 1% penicillin-streptomycin (cat # 15140122, ThermoFisher Scientific) and 0.2% gentamicin/amphotericin B (cat # R01510 ThermoFisher Scientific). Cells were passaged upon reaching 75-95% confluency using 1x PBS (cat # 70013032, ThermoFisher Scientific) to wash the cells and TrypLE Express Enzyme (cat # 12605093, ThermoFisher Scientific) to detach them from the flask.

2.5.2 Cell sorting

CA-derived primary cell lines were sorted into EpCAM^{High} and EpCAM^{Low} subpopulations using the CELlection™ Epithelial Enrich Dynabeads kit (cat # 16203, ThermoFisher Scientific). Cells were lifted from their culture flask using TrypLE and a cell count was performed to ensure there were between 1x10⁶ and 2x10⁷ live cells. Cell pellet was resuspended in a 15 mL Falcon tube using 1 mL of PBS with 0.1% FCS. 50 µL of washed Dynabeads were added, and the tube was incubated for 30 min in the fridge with gentle tilting and rotation. Following incubation, the tube was placed in a DynaMag magnet (ThermoFisher Scientific)

for 2 min, and the supernatant containing unbound EpCAM^{Low} cells was transferred to a new tube. The incubation tube was then removed from the magnet, the beads washed gently with 1 mL of PSB with 0.1% FCS, and then returned to the magnet for 2 min. This supernatant was then pooled with the first, and a total of 3 washes were performed in this manner. After 3 washes, the beads were resuspended in 200 µL of DMEM with 1% FCS and 4 µL of Release Enzyme Buffer and incubated at room temperature for 15 min with gentle tilting and rotation. Meanwhile, a cell count was performed on the collected supernatants containing EpCAM^{Low} cells, and these cells were plated to an appropriately sized culture flask.

After incubation for 15 min, the bead-bound EpCAM^{High} cells should have been released. The tube was placed in the DynaMag for 2 min and the supernatant containing unbound EpCAM^{High} cells was transferred to a new tube. Wash steps as above were performed but using 200 µL of DMEM with 1% FCS, and supernatants were pooled for a cell count. These cells were then plated into an appropriately sized culture flask.

2.5.3 Tumoursphere formation assays

Cells were lifted from their culture flask using TrypLE and a cell count was performed. Sphere formation assays were carried out in low-adherence plates or flasks. Plates were Corning Costar 6-well ultra-low attachment plates (cat # 3471, In Vitro Technologies) and flasks were T25 Nunclon Sphera EasyFlasks (cat # 174951, ThermoFisher Scientific). Cell pellets were resuspended in StemXVivo Serum-free Tumoursphere media (cat # CCM012, R&D Systems, In Vitro Technologies) and seeded at a density of 1×10^4 live cells per mL, with 4 mL per well in a 6-well plate or 10 mL per T25. Progress was checked daily using an inverted microscope, and fresh media was added every 3-4 d; no media was removed due to the cells being non-adherent. Spheres were expected to form within 7-10 d. They were harvested at a point where spheres had formed but had not yet developed dark centres which indicate necrosis. If spheres had not formed by 14 d, the cells were harvested or discarded. A cell line was considered to be positive in the sphere forming assay if more than half of measured spheres reached a diameter of at least 50 µm. Spheres were measured using an inverted microscope and cellSens 2.0 software (Olympus).

To harvest spheres, the tumoursphere media was aspirated into a 50 mL falcon tube. Each plate or flask was rinsed with PBS and this was collected into the same tube to maximise yield. More PBS was added to the tube to allow for a 1:3 dilution of the viscous tumoursphere media in PBS. Tubes were

centrifuged and supernatant carefully removed. Pellets were resuspended in PBS and transferred to a 1.5 mL Eppendorf tube for pelleting and cryopreservation at -80°C.

2.5.4 Immunocytochemistry

Immunocytochemistry assays were performed using the PSC 4-marker Immunocytochemistry Kit (cat # A24881, ThermoFisher Scientific). Cells were seeded onto 8-chamber culture slides (cat # 354118, Corning, In Vitro Technologies) at a density of 5000 cells per well. After allowing for the confluency to reach 75-95%, growth medium was removed and each well was washed once with PBS. Fixative Solution (cat # A24344) was added for 15 min, before being removed and replaced with Permeabilization Solution S (cat # A24878) for 15 min, both at room temperature. Finally, a Blocking Solution was added (cat # A24353) for 30 min at room temperature.

Primary antibodies included rabbit anti-OCT4 (cat # A24867), rat anti-SOX2 (cat # A24759), mouse IgG3 anti-SSEA4 (cat # A24866) and mouse IgM anti-TRA-1-60 (cat # A24868). Primary antibodies were diluted 1:150 in Blocking Solution and cells were incubated with these in the fridge overnight. After exposure to the primary antibodies, the cells were washed with Wash Buffer (cat # A24348) 3 times for 2-3 min each. Secondary antibodies included Alexa Fluor® 555 donkey anti-rabbit (cat # A24869) and goat anti-mouse IgM (cat # A24871), Alexa Fluor® 594 donkey anti-rabbit (cat # A24870) and goat anti-mouse IgM (cat # A24872), and Alexa Fluor® 488 goat anti-mouse IgG3 (cat # A24877) and donkey anti-rat (cat # A24872).

The required combinations of secondary antibodies were diluted 1:250 in Blocking Solution and cells were incubated with these for 1 h at room temperature in the dark.

Following exposure to secondary antibodies, the cells were washed 3 times with Wash Buffer as above. NucBlue™ Fixed Cell nuclear stain (cat # R37606, ThermoFisher Scientific) was added to the third wash and left on the cells for 5 min. Finally, the chambers were removed and a coverslip was mounted using Histomount (cat # 008030, ThermoFisher Scientific). Fluorescence was visualised using the FV1200 Laser Scanning Microscope (Olympus).

2.5.5 Mesoderm differentiation assay

Cells were induced to undergo osteogenic differentiation using the StemPro® Osteogenesis Differentiation Kit (ThermoFisher Scientific). Cells were seeded onto 8-chamber culture slides (cat# 354118, Corning, In Vitro Technologies, Auckland, NZ) at a density of 5000 cells per well. After allowing 2-3 d to adhere, regular DMEM media was replaced with Osteogenesis Differentiation Medium. This was replaced every 2-3 d for a total of 10-14 d. After the differentiation period, the media was removed and cells were fixed in 5% formalin for 5 min and then thoroughly washed with distilled water. Fixed cells were then stained with 2% Alizarin Red solution (pH 4.2) for 5 min. Once the dye had been removed, cells were washed 5 times with distilled water and visualised under the inverted microscope.

2.5.6 Endoderm differentiation assay

Endoderm differentiation assays were performed using the StemXVivo® Endoderm Kit (cat # SC019B, R&D Systems, In Vitro Technologies). Cells were seeded onto 8-chamber culture slides (cat # 354118, Corning, In Vitro Technologies) at a density of 5000 cells per well. After allowing 2-3 d to adhere, regular DMEM media was refreshed with the addition of *b*FGF and left for 4 h. Following this, the cells were washed with PBS and cultured in Differentiation Media I overnight. This was then replaced with Differentiation Media II, which was refreshed twice daily for a further 2 d. After 2 d in Media II, the cells were washed with PBS and fixed using 10% formalin for 20 min at room temperature. Cells were washed 3 times using 1% BSA in PBS, and permeabilized using 5% BSA in PBS with 0.1% Tween-20 for 45 min at room temperature. Following this, anti-human SOX17 primary antibody was added to the permeabilisation buffer (final concentration 10 µg/mL) and left in the fridge overnight. Cells were then washed 3 times using 1% BSA in PBS before being exposed to the NorthernLights™ secondary antibody (cat # NL001, R&D Systems, In Vitro Technologies), at a 1:200 dilution in permeabilisation buffer, for 1 hour in the dark at room temperature. Secondary antibody was washed off 3 times using 1% BSA in PBS. NucBlue™ Fixed Cell nuclear stain (cat # R37606, ThermoFisher Scientific) was added to the third wash and left on the cells for 5 min. Finally, the chambers were removed and a coverslip was mounted using Histomount (cat # 008030, ThermoFisher Scientific). Fluorescence was visualised using the FV1200 Laser Scanning Microscope (Olympus).

2.5.7 Ectoderm differentiation assay

Ectoderm differentiation assays were performed using the StemXVivo® Ectoderm Kit (cat # SC031B, R&D Systems, In Vitro Technologies). Cells were seeded onto 8-chamber culture slides (cat # 354118, Corning, In Vitro Technologies) at a density of 5000 cells per well. After allowing 2-3 d to adhere, regular DMEM media was replaced with Ectoderm Differentiation Media. This was refreshed once daily for a further 2 d. After a total of 3 d in differentiation media, the cells were washed with PBS and fixed using 10% formalin for 20 min at room temperature. Cells were washed 3 times using 1% BSA in PBS and permeabilised using 5% BSA in PBS with 0.1% Tween-20, for 45 min at room temperature. Following this, anti-human Otx2 primary antibody was added to the permeabilisation buffer (final concentration 10 µg/mL) and left in the fridge overnight. Cells were then washed 3 times using 1% BSA in PBS before being exposed to the NorthernLights™ secondary antibody (cat # NL001, R&D Systems, In Vitro Technologies), at a 1:200 dilution in permeabilisation buffer, for 1 h in the dark at room temperature. Secondary antibody was washed off 3 times using 1% BSA in PBS. NucBlue™ Fixed Cell nuclear stain (cat # R37606, ThermoFisher Scientific) was added to the third wash and left on the cells for 5 min. Finally, the chambers were removed and a coverslip was mounted using Histomount (cat # 008030, ThermoFisher Scientific). Fluorescence was visualised using the FV1200 Laser Scanning Microscope (Olympus).

2.5.8 Cathepsin activity assays

Cathepsin activity assays were performed using the kit from Abcam (cat # ab65300 and cat # 65302). Total protein was extracted from tissue and cell samples using lysis buffers provided in each kit, which were specific to the cathepsin B and D assays. The samples were washed briefly with cold PBS and homogenised in lysis buffer using an Axygen™ plastic pestle (cat # PSE-15-B-SI; ThermoFisher Scientific), then incubated shaking on ice for 20 min. For the cathepsin B assay, 100 µL of lysis buffer was added to 10 mg of tissue or 1×10^6 cells. For the cathepsin D assay, 200 µL of lysis buffer was added to 100 mg of tissue or 1×10^6 cells, and then 800 µL extra lysis buffer was added to the tissues after homogenisation. Supernatant was collected after centrifugation for 5 min at 17,000 g, and the protein concentration was measured by BCA. For measurement of fluorescence, a black-walled clear-bottom 96-well plate was set up with 50 µL of protein extract from each sample, in duplicate. Blank wells contained 50 µL of buffer without protein. To each well, 2 µL of substrate was added and the plates were incubated at 37°C for 1 h. Fluorescence was measured in a plate reader with excitation and emission wavelengths of 400 nm and 505 nm for cathepsin B and 328 nm and 460 nm for cathepsin D. Raw fluorescence values were adjusted by subtracting the

background fluorescence from blank wells, and then normalised by dividing the adjusted value by the μg of protein added to give the enzyme activity in relative fluorescent units per μg of protein. Tonsil tissue was used as a positive control for both assays.

2.5.9 RAS modulation drug assays

The effect of RAS modulators on the metabolism of CA-derived primary cell lines was investigated using the RealTime-Glo™ Cell Viability Assay (cat # Q9712, Promega, In Vitro Technologies). The assay does not require cell lysis, and so the same plate can be repeatedly read for at least 72 h. The substrate is reduced within metabolically active cells, and then diffuses into the culture medium where the NanoLuc® luciferase enzyme catalyses a reaction which produces luminescence. Cells with impeded metabolism will reduce less of the substrate and produce less luminescence.

A seeding assay was performed to determine that 1000 cells per well was the optimal seeding density for a 96-well plate for the CA-derived cells, with control cells reaching approximately 90% confluency after 4 days. Initial drug doses used for test plates were 100 μM , 50 μM , 10 μM , 5 μM and 1 μM . Final doses were 50 μM , 10 μM and 1 μM for R-, S- and R/S-propranolol, and 100 μM , 50 μM and 10 μM for all other drugs.

Cells were seeded on day 1 at 1000 cells per well in white-walled clear-bottom 96-well plates (cat # FAL353377, In Vitro Technologies), with triplicate wells per dose. Control wells were seeded in triplicate, and included cells grown in adjuvant + media and media alone to assess possible effects of the adjuvants on cell metabolism, as well as controls of adjuvant + media and media without cells to measure background luminescence. On day 2, NanoLuc® enzyme and substrate were added to each well to a final concentration of 1x from a 1000x stock. Drug doses were formulated from stocks to give the final desired concentrations. The first luminescence reading was performed 1 h after the initial dose. On days 3 and 4, cells were given further doses of each drug and the plates were read again, at time points of 24 h and 48 h after the initial dose. The final reading was performed on day 5, 72 h after the initial dose, and the media was collected and stored.

To study the effects of RAS modulation on tumoursphere formation and iPSC gene transcription, 6000 cells per well were seeded in 24-well tissue culture plates, after a test plate was used to determine the optimal seeding density. Cell doses were selected based on the RealTime-Glo™ Cell Viability Assay results to ensure that metabolism was affected but with minimal cell death. R-propranolol was dosed at

30 μ M and 10 μ M, R-Timolol at 100 μ M and 50 μ M, EMA401 at 50 μ M and 10 μ M, and losartan and SMM02 at 100 μ M. RNA was extracted and pooled from 9 wells per dose for all drugs. Tumoursphere assays were only performed on cells dosed with R-propranolol, EMA401 and losartan, with 6 wells harvested and pooled for this. Cells were dosed as outlined above. Harvested cells were seeded for tumoursphere assays and RNA processed for RT-qPCR both as described earlier.

Statistical analysis and graphing were performed on GraphPad Prism 8.

2.6 Proteomics

2.6.1 Sample definition

The proteomics included various sample types. There were 4 HGCA and 4 LGCA tissues from individual patients, and each included a patient-matched normal colon (NC) tissue sample. Therefore, there were 16 tissue samples analysed. Each of the CA tissues had a tissue-derived primary cell line, and there was one NC-derived cell line used as a control which was not matched to any of the patients. Therefore, there were 9 cell lines analysed.

2.6.2 Protein extraction

Tissues and cells were lysed in lysis buffer containing 30 mM Tris-HCl, pH 7.4; 7 M urea, 2 M thiourea, and 4% CHAPS plus Halt Protease and Phosphatase inhibitor cocktail (cat # 78442, ThermoFisher Scientific) at a ratio of 99:1, with a 1 mL glass Dounce homogeniser (Corning Inc, Corning, NY, USA) on ice followed by agitation for 45 min in the cold room at 4°C. The lysate was checked under a microscope to ensure that the lysis was efficient. If not, the samples were homogenised again and agitated for a further 30 min.

After the extraction steps, the samples were centrifuged at 17,000 g for 20 min at 4°C. The protein concentration was measured by Bradford assay (Bio-Rad). Samples were added at 5x, 10x and 20x dilutions in Bradford reagent, and a standard curve was produced using BSA. The plate was incubated at room temperature for 15 min before reading in a plate reader. The samples were stored at -20 °C until further being processed.

2.6.3 Protein precipitation and digestion

Extracted proteins were precipitated using a Calbiochem ProteoExtract® Protein Precipitation Kit (cat # 539180, Merck, North Shore City, NZ). Precipitant 1 was added to each sample at a ratio of 4:1, vortexed briefly, and placed in the freezer overnight. Following this, each tube was briefly vortexed and centrifuged at 17,000 g for 10 min at 4°C. Supernatant was discarded and the pellet was washed twice using Wash Buffer by centrifuging at 17,000 g for 5 min. The protein pellet was dissolved in a digestion buffer of 8 M urea and 100 mM Tris-HCl, pH 8.5. A second Bradford assay was performed to measure the protein concentration after precipitation, and 20 µg of protein was taken from each sample and the volume brought to 25 µL using digestion buffer. To reduce the proteins, DTT was added to a final concentration of 5 mM and incubated at 56°C for 30 min. The tube was then briefly centrifuged and iodoacetamide was added to a final concentration of 10 mM and incubated in the dark at room temperature for 40 min to alkylate the proteins. The sample was then diluted 3-fold using 100 mM Tris-HCl to bring the urea concentration to 2 M. To digest the proteins, CaCl₂ to a final concentration of 1 mM and 0.67 µg trypsin were added and incubated at 37°C for 16 h. The reaction was quenched with formic acid to a final concentration of 4%.

The resulting tryptic peptides were then purified by using 100 µL OMIX C-18 zip tips (cat # A57003100K, Agilent Technologies, USA) to remove salts, using a pre-wet buffer of 0.4% TFA and 100% ACN at 1:1, a rinse buffer of 0.4% TFA and 100% ACN at 1:1, and elution buffers of 0.1% formic acid with 70% ACN (elution A) and 50% ACN (elution B) at 1:1. Zip tip eluates were pooled and dried down to 2-3 µL using a vacuum centrifuge, then brought to 100 µL using 0.1% formic acid in HPLC water (so the final constitution of the sample solution was 0.1% FA and ~ 2% ACN) .

2.6.4 Liquid Chromatography Tandem Mass Spectrometry

Proteomic analysis of the prepared samples was performed by liquid chromatography tandem mass spectrometry (LC-MS/MS). The LC-MS/MS included an UltiMate 3000 RSLCnano system and an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer that are coupled via a Nanospray Flex ion source (ThermoFisher Scientific).

The sample vials containing tryptic peptides in 0.1% FA and 2% ACN were placed in the autosampler of the HPLC unit for injection, maintained at 10°C. Xcalibur™ software (Version 2.1.0, ThermoFisher Scientific) was used to define the method and acquire LC-MS/MS data. Peptides were first

loaded onto an Acclaim™ PepMap™ 100 C18, 5 µm 0.3 x 5 mm trap column (cat # 160454, ThermoFisher Scientific) with the loading pump at a flow rate of 8.0 µL/min of 2% ACN and 0.05% TFA and then separated on an Acclaim™ PepMap™ 100 C18, 2 µm, 100 Å, 75 µm x 15 cm analytical column (cat # 164941, ThermoFisher Scientific) with the Nano/Cap pump running at 0.3 µL/min with an organic solvent gradient constructed from buffer A (0.1% FA) and buffer B (0.1% formic acid in 80% ACN). The gradient was programmed as follows: 3% from 0 to 5 min, 30% from 5 – 70 min, 50% from 70 – 82 min, 95% from 82 – 88 min, and finally 3% from 88 – 99 min. Elution was based on reverse-phase liquid chromatography, whereby the more hydrophobic the peptide is the slower it will pass through the column. Peptide mass also affects transit time; larger peptides tend to retain for longer.

The peptides eluted in solution were ionised by nanoelectrospray ionisation (Nanospray Flex, ThermoFisher Scientific) with the 25 µM Ion Transfer capillary tube set to 275°C and voltage set at 1.8 kV. MS scans were acquired in the Orbitrap (OT) with the following settings: detector type OT, resolution 120,000, scan range 375 - 1500 m/z, AGC target 5.0e3, Maximum Injection Time 50 ms, charge state 2 – 7, and data type Profile. Data-dependent MS/MS (ddMS/MS) scans were acquired in Ion trap (IT) having the following settings: detector type IT, scan range mode Auto: m/z Normal, IT scan rate Rapid, AGC target 5.0e3, Maximum Injection Time 300 ms, and data type Centroid. For MS/MS, high-energy collision-induced dissociation (HCD) fragmentation was performed in the linear Quadrupole ion trap (isolation window 1.6 m/z, HCD collision energy 30%). The “Top 20” highest-intensity ions from each MS scan were selected for the subsequent MS/MS scans. Dynamic exclusion was included having the following settings: mass tolerance 10 ppm, exclusion duration 60 s. Each sample was run with LC-MS/MS at least 3 times.

2.6.5 Protein identification

The LC-MS/MS spectra were exported as .raw files and searched against SwissProt human protein sequence database (TaxID=9606 and Subtaxonomies, version 2017-10-25, downloaded on 24-10-2019, 42,253 sequences) using the SequestHT search engine in Proteome Discoverer 2.4 (ThermoFisher Scientific). Protein identification settings were as follows: a peptide length range of 6-144 and allowing for 2 missed trypsin cleavages, with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.5 Da; the carboxyamidomethylation of cysteine was set as a static modification (+57.021); dynamic modifications included: oxidation (+15.995 Da) at M, carbamylation (+43.066 Da) at K, acetylation (+42.011 Da) at K, deamidation (+0.984 Da) at N and Q, and peptide terminus modification of carbamylation (+43.066 Da) at the N-terminus. The files were searched against the protein sequence

database and the decoy database (Percolator node) with the false discovery rate (FDR) filter set to 0.01, plus a relaxed FDR of 0.05. Proteins were identified with high peptide confidence and a minimum peptide number of 1. Proteins identified sharing the common peptides were grouped into protein groups for export.

2.6.6 Label-free quantitation

Label-free quantitation (LFQ) was carried out using Proteome Discoverer 2.4 (ThermoFisher Scientific). Study factors were defined as “patients”, “grade” and “technical replicates”. The .raw files were imported, and each file was assigned to one category from each study factor. First, each file was assigned a “grade”, either NC, LG or HG. Next, the biological replicates from each of the samples were assigned a “patient” number (from 1 to 4 for CA tissues and CA-derived cell lines, and from 1 to 8 for NC tissues). Finally, the technical replicates from each samples were numbered from 1 to 3. Before running the analysis, comparisons are defined manually. The 3 comparisons were LGCA/NC, HGCA/LGCA and HGCA/NC.

Protein quantitation data are displayed as a grouped abundance, which is a measure of the normalised intensity of all the peptide ions assigned to the protein across all input files for each group being compared. The grouped abundance is the average of normalised spectral counts from all technical and biological replicates for each grade. Proteome Discoverer used a t-test to determine the statistical significance for each comparison. Grouped abundances are used to calculate the abundance ratio for any given comparison. The selection criteria for significantly differentially expressed proteins were set at a fold change of 2 or greater (\log_2 fold change ≥ 1), and a p-value of 0.05 or smaller ($-\log_{10}$ p-value ≥ 1.30103).

The output from the LFQ analysis was visualised as volcano plots and heat maps in Proteome Discoverer. Volcano plots showed fold change on the x-axis and the p-value on the y-axis. When comparing the abundance of all proteins in two conditions (e.g. HG/LG), the distance that a protein fell from the origin ($x=0$, $y=0$) was related to the abundance fold change and the p-value for that change. Therefore, the selection criteria selected proteins that were further from the origin. These proteins were exported as the significantly differentially expressed proteins.

The input for heat maps were the complete lists of all quantified proteins. The heat map displayed differential expression as green (low expression), black (similar expression) and red (high expression) when comparing the three conditions. All the quantified proteins were clustered by Proteome Discoverer

based on similarity of protein sequences, and each cluster showed a distinct pattern of change in abundance between the conditions.

The most highly differentially expressed proteins from the volcano plots and the clusters from the heat maps which saw significant differences between each condition were exported to elucidate protein function.

2.6.7 Functional analysis

The protein interaction database called STRING (string-db.org) (Szklarczyk et al., 2019) was used to identify the pathways that are enriched amongst the proteins significantly upregulated or downregulated in one condition (NC, LGCA, HGCA) compared to another. The Uniprot accession numbers for proteins with significant abundance changes (fold change >2; $p < 0.05$) and proteins from each cluster of the heat maps were uploaded to STRING. A network map was created based on the input list of accession numbers using a database of known protein interactions in STRING. Each node in the network represents an individual protein, and the connections between them indicated both functional and physical protein interactions. The thickness of the connection was relative to the strength of the interaction based on evidence from the database. The minimum interaction score required for an interaction was set to medium. Nodes with no connections to other nodes under these criteria were removed from the map. Furthermore, the optional setting of adding in an additional shell of proteins, which were not present in the input list but are known to have strong interactions with other members, was not selected. Once these settings had been defined, the functional enrichments from the network were displayed, which included the Gene Ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that were enriched within the input list of proteins. For a category or pathway to be considered significant, at least 2 protein members were required to be present within the network map, with a false discovery rate (Benjamini-Hochberg adjusted p-value) for the category/pathway of less than 0.05.

2.7 DNA Sequencing

2.7.1 Sample preparation

DNA was extracted from FFPE tissue samples and cells using the PureLink Genomic DNA Mini Kit (cat # 1820-00, Invitrogen, ThermoFisher Scientific).

For cells, a pellet of 3×10^6 cells was resuspended in 200 μ L PBS with 20 μ L proteinase K and 20 μ L RNase A and incubated at room temperature for 2 min. Next, 200 μ L PureLink Genomic Lysis/Binding Buffer was added, mixed by vortex and incubated at 55°C for 10 min. Following this, 200 μ L of 100% ethanol was added, mixed by vortex and the solution was transferred to a PureLink Spin Column in a collection tube and centrifuged at 17,000 g for 1 min at room temperature. The flow-through was discarded, 500 μ L Wash Buffer 1 was added to the column and it was centrifuged as above. The flow-through was again discarded, 500 μ L Wash Buffer 2 was added to the column and it was centrifuged as above. Columns were transferred to a clean 1.5 mL Eppendorf tube and 25 μ L of elution buffer was added and incubated for 1 min at room temperature before centrifugation as above. The elution step was repeated to give a final elution volume of 50 μ L containing DNA.

For FFPE tissues, 5 sections of 5 μ m were cut from the tissue block and placed in a 1.5 mL Eppendorf tube, to which 1 mL of xylene was added and vortexed. The tube was incubated at 50°C until the paraffin had melted (~20 min) and centrifuged at 17,000 g for 2 min at room temperature. The supernatant was discarded, and 1 mL of 100% ethanol added and vortexed before centrifuging as above. This ethanol wash step was then repeated. The tubes were left at room temperature with the lid open to allow the ethanol to evaporate. Protease (110 μ L) was added and incubated at 55°C for 2 h, and then at 90°C for 1 h. The tubes were then briefly centrifuged and left to cool. Next, 155 μ L DNA Binding Buffer containing magnetic beads was added, the tubes were shaken at 1000 RPM using an orbital shaker for 5 min and then placed in a magnet for 2 min. The supernatant was discarded and then 200 μ L DNA Wash Buffer was added to the beads. Tubes were shaken at 1150 RPM for 2 min, placed in a magnet for 2 min and the supernatant discarded. This wash step was repeated before another wash step using 200 μ L Wash Solution 2. After removing the supernatant, the tubes were shaken at 1150 RPM for 2 min with the lids open to dry the beads before adding 50 μ L elution buffer. The tubes were shaken at 1150 RPM for 5 min and placed in a magnet for 2 min. The supernatant containing DNA was collected.

2.7.2 DNA quality validation

The quality of extracted DNA was assessed using the TaqMan™ RNase P Detection Reagents Kit (cat # 4316831, ThermoFisher Scientific), to ensure the DNA was of high enough quality for sequencing. A serial dilution was created from a DNA control provided with the kit ranging from 5 ng/ μ L to 0.078125 ng/ μ L. Extracted DNA was diluted 1:500 and 1:1000 in nuclease-free water. Controls and samples were run in triplicate, with 2.5 μ L added to 17.5 μ L of a mastermix containing RNase P primers. A PCR protocol was

performed in the Rotor Gene Q as follows: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. The CT values for samples were used to measure the concentration of sequencable DNA by using the standard curve generated from the control DNA serial dilution.

2.7.3 DNA sequencing

DNA sequencing was performed by ThermoFisher Scientific (Life Technologies Australia, 5 Caribbean Drive, Scoresby, Melbourne, Australia).

Ion AmpliSeq™ Library Preparation was carried out by ThermoFisher Scientific staff as follows. DNA samples were added to a mastermix and split in two. Each pool received a different Tumor Mutation Load Assay primer pool. DNA amplification was performed using the following PCR protocol: 99°C for 2 min to activate the enzyme, and 15 cycles of 99°C for 15 s to denature and 60°C for 16 min to anneal and extend. Reactions were then combined and partially digested before adapters were ligated to the amplicons. Each resulting barcoded library was created from a different sample. The libraries were then purified by washing twice with 70% ethanol and quantified using the following PCR protocol: 50°C for 2 min to allow Uracil-DNA Glycosylase (UDG) to remove deaminated cytosines, 95°C for 2 min to activate the polymerase, and 40 cycles of 95°C for 15 s and 60°C for 1 min to amplify the DNA. Pools containing 4 barcoded libraries were created and prepared for sequencing using the Ion Chef template system. DNA was sequenced on Ion 540 chips using the Ion GeneStudio™ S5 Series sequencer.

2.7.4 Data analysis

DNA sequencing data were analysed using the ThermoFisher Scientific cloud-based Ion Reporter system. Single-nucleotide variants (SNVs), variant impacts and tumour mutational burden (TMB) were compared between FFPE tissue and cells derived from the same patient. Mutations to genes important in CA were searched for within the variants table and were compared between samples.

Chapter 3: Validating Colon Adenocarcinoma Tissue-derived Primary Cell Lines by DNA Sequencing

3.1 Introduction

DNA sequencing is used to detect genomic differences between a cancer tissue sample and its patient-matched normal tissue, and to determine the significance of such changes (Cancer Genome Atlas, 2012). Often, this will focus on regions of the genome that code for proteins and therefore provides a more functional basis. The Cancer Genome Network (TCGA) has examined the genomes and transcriptomes of 276 CRC samples (Cancer Genome Atlas, 2012). Whole exome sequencing (WXS) revealed the mutation rates for each sample, which varied greatly between tumours, and identified the most highly mutated genes. Patient-matched NC tissue samples were analysed to allow comparisons between the normal colon tissues and tumour tissues from the same patient. Vasaikar et al. (Vasaikar et al., 2019) also used WXS to identify a core set of 17 genes most commonly mutated across their 106 CRC tumour tissue samples, 13 of which were also reported by TCGA.

The OncoPrint™ Tumor Mutation Load (TML) Assay (ThermoFisher Scientific) uses targeted next-generation sequencing (NGS) to sequence 409 cancer-related genes in order to detect low-frequency single nucleotide variants (SNVs) as well as insertion/deletion (INDEL) mutations. The impact of each variant is annotated as being either synonymous, missense, nonsense, frameshift, splice variant or of unknown effect. Furthermore, it provides a tumour mutational burden (TMB) score that shows the number of mutations per megabase (Mb) of sequenced DNA. The TML assay is a PCR-based method which employs primer pairs for each of the 409 genes to produce an amplicon library by amplifying the genetic regions of interest. Each library is produced from an individual sample, and barcodes which are unique to each library are ligated to the amplicons. In this way, up to 8 libraries can be combined and sequenced in a single reaction to ensure equal amplification, reduce reagents and save time. This targeted approach allows greater sequencing depth than whole genome sequencing and higher confidence in low-frequency variants. Overall, the TML assay covers ~1.7 Mb of the genome, with about 70% of this comprising exonic sequences (1.2 Mb). A variant calling algorithm removes germline mutations before data analysis and does not require the matched NC tissue in order to do so, leaving only the somatic cancer-related mutations to be analysed.

The primary tissue-derived cell lines are not characterised before being banked in the GMRITB. Therefore, it was important to determine whether the CA tissue-derived primary cell lines for this project reflected the original CA tissue from which they were derived. The aim of this chapter was to sequence the DNA of the matched CA tissues and CA-derived primary cell lines using the Oncomine™ TML assay to assess whether the mutational signatures for key CA-related genes were shared between the two sample types, and therefore determine the suitability of the cells as an *in vitro* model of the tumour.

3.2 Results

3.2.1 DNA Quality Assurance

For this project, a total of 4 LGCA-derived and 4 HGCA-derived primary cell lines were used. DNA was extracted from all 8 cell lines and patient-matched FFPE CA tissue samples. PCR was carried out using the TaqMan™ RNase P Detection Reagents Kit (cat # 4316831, ThermoFisher Scientific) to ensure the DNA quality was sufficient for sequencing, which found that all the samples except for LGCA4 were above the threshold for sequencing. The DNA samples from all 8 CA tissues and 8 CA-derived cell lines were sent to ThermoFisher Scientific (Australia) for DNA sequencing to be performed. Unfortunately, no reliable data was produced for LGCA4, so analysis was conducted on the tissues and cells from the remaining 7 (3 LGCA and 4 HGCA) patients.

3.2.2 Summary of Oncomine™ TML Assay data

ThermoFisher provided the DNA sequencing data, which was uploaded to the cloud-based analysis platform, Ion Reporter, where the data could be accessed and analysed. The pre-defined Oncomine Tumor Mutation Load w3.2 analysis workflow was launched for each individual tissue and cell sample, which calculated the TMB score for each sample, identified and annotated variants within the 409 genes, and assigned an impact level to each variant. Following this, paired analyses were performed to compare each tissue with its matched cell line.

The TMB scores were higher in the FFPE tissues than their patient-matched cell lines for 6 of the 7 patients, with the exception of HGCA4 (Table 3.1). CA-derived cells had more consistent TMB scores, with 6 out of 7 having a score below 4 and no significant difference between LGCA and HGCA-derived cells, whereas the FFPE tissues had TMB scores ranging from 4.63 to 38.65. Unsurprisingly, HGCA tissues had higher TMB scores than LGCA tissues.

Similarly, 5 out of 7 FFPE CA tissue samples had more unique variants within the 409 genes than the cell lines derived from them, but LGCA1 and LGCA3 cell lines had more unique variants than their matched FFPE tissues. The total number of variants across both sample types was relatively similar for 6 of the 7 patients (2023-2182), however HGCA4 displayed a lower variant count within the TML gene panel (1643) despite having the highest TMB scores. This suggests that HGCA4 contains a larger proportion of non-synonymous variants than the other samples, as synonymous variants are not considered when calculating the TMB score, or that it contains more mutations outside of the exons. When considering the impact of these variants, all 7 samples were relatively consistent, with the number of total impactful variants per patient ranging from 266 to 319 within these 409 genes. The number of unique impactful variants was proportional to the total number of unique variants for each sample, with the HGCA tissues and cells containing a lower proportion of shared impactful variants than the LGCA samples.

Table 3.1: Summary of TML assay results for CA tissues and CA-derived primary cell lines

Samples	TMB (mutations/Mb)		Total variants			Impactful variants		
	Cells	Tissues	Cells	Tissues	Shared	Cells	Tissues	Shared
LGCA1	3.36	4.63	154	40	1856	23	6	263
			2050			292		
LGCA2	1.67	5.92	44	63	1986	4	9	291
			2093			304		
LGCA3	1.67	6.01	58	40	1925	5	4	272
			2023			281		
HGCA1	2.51	7.75	64	163	1932	7	27	278
			2159			312		
HGCA2	2.51	10.49	95	519	1445	17	97	201
			2059			315		
HGCA3	1.68	38.65	165	477	1540	24	65	230
			2182			319		
HGCA4	20.69	17.67	83	124	1436	24	53	189
			1643			266		

Tumour mutational burden (TMB) score, the total number of variants within the 409 TML assay genes and the number of variants that impact the gene product (i.e. missense, nonsense, etc.) are displayed. For each CA sample, the number of variants unique to the cells and to the tissues and those shared by both are reported, with the total number of variants listed directly below.

Overall, these data suggest that the CA-derived cell lines contain a subset of the mutations present within their matched FFPE tissues, given that the cells tended to have fewer total and impactful variants and lower TMB scores.

3.2.3 Mutational signatures of CA tissues and CA-derived cells

To confirm whether the mutations detected in the cells were the same as those in the tissues and thereby determine the similarity of the cells to their parent tissues, several CA-related genes were selected and the specific mutations within these genes in the cells and tissues were explored (Table 3.2). Mutations with low frequency (<6%) and/or coverage depth (<300 reads), which were unique to 1 sample type (either the FFPE tissues or cells from a patient, but not both) and/or unique to 1 patient, were considered to be read errors and were disregarded. Table 3.3 details the average coverage depth for each sample. The *APC*, *TP53*, *KRAS*, *BRAF*, *PIK3CA* and *FBXW7* genes were identified by other studies as being relevant to CRC (Cancer Genome Atlas, 2012; Dos Santos et al., 2019; Vasaikar et al., 2019), and the *MSH2*, *MSH6* and *MLH1* genes are vital to DNA MMR and MSI.

Table 3.2: Mutations to CA-related genes in FFPE CA tissues and CA-derived primary cell lines

Gene	Mutation		LGCA1	LGCA2	LGCA3	HGCA1	HGCA2	HGCA3	HGCA4
APC	c.646 C>T	Tissue	-	-	✓	-	-	-	-
		Cells	-	-	-	-	-	-	-
	c.688_689 ins ACTTC	Tissue	-	-	✓	-	-	-	-
		Cells	-	-	-	-	-	-	-
	c.1458 T>C	Tissue	✓	✓	✓	✓	-	✓	✓
		Cells	✓	✓	✓	✓	✓	✓	-
	c.1743+19 A>G	Tissue	-	✓	-	-	-	-	-
		Cells	-	✓	-	-	-	-	-
	c.1476 C>G	Tissue	-	-	-	-	-	-	-
		Cells	-	-	-	✓	-	-	-
	c.3871 C>T	Tissue	-	✓	-	-	-	-	-
		Cells	-	-	-	-	-	-	-
	c.4118 del C	Tissue	-	-	-	✓	-	-	-
		Cells	-	-	-	-	-	-	-
	c.4479 G>A	Tissue	✓	✓	✓	✓	-	-	-
		Cells	✓	✓	✓	✓	✓	✓	-
	c.5034 G>A	Tissue	✓	✓	✓	✓	✓	✓	✓
		Cells	✓	✓	✓	✓	✓	✓	✓
	c.5268 T>G	Tissue	-	✓	✓	✓	-	✓	✓
		Cells	✓	✓	✓	✓	✓	✓	-
	c.5880 G>A	Tissue	-	✓	✓	✓	-	✓	✓
		Cells	✓	✓	✓	✓	✓	✓	-
	c.7201 C>T	Tissue	-	✓	-	-	-	-	-

		Cells	-	✓	-	-	-	-	-
TP53	c.215 C>A	Tissue	✓	✓	✓	✓	-	-	-
		Cells	✓	✓	✓	✓	-	-	-
	c.391 A>T	Tissue	✓	-	-	-	-	-	-
		Cells	-	-	-	-	-	-	-
	c.524 C>T	Tissue	-	-	-	✓	-	-	-
		Cells	-	-	-	-	-	-	-
	c.538 G>T	Tissue	-	✓	-	-	-	-	-
		Cells	-	-	-	-	-	-	-
	c.754 del C	Tissue	-	-	✓	-	-	-	-
		Cells	-	-	-	-	-	-	-
	c.844 G>A	Tissue	-	-	-	-	✓	-	-
		Cells	-	-	-	-	-	-	-
	c.919+1 G>A	Tissue	-	-	-	-	-	-	✓
		Cells	-	-	-	-	-	-	✓
KRAS	c.483 C>T	Tissue	✓	✓	✓	✓	-	-	-
		Cells	✓	✓	✓	✓	✓	✓	-
	c.*5598 A>G	Tissue	-	-	-	-	-	-	-
		Cells	-	-	-	-	✓	-	-
BRAF	c.980+27 C>T	Tissue	✓	✓	-	-	-	-	-
		Cells	✓	✓	-	-	-	-	-
	c.1518-48 G>A	Tissue	✓	✓	-	-	-	-	-
		Cells	✓	✓	-	-	-	-	-
MSH2	c.211+9 C>G	Tissue	✓	-	✓	-	-	-	-
		Cells	✓	-	✓	-	-	-	-
	c.2006-6 T>C	Tissue	-	-	✓	-	-	-	-
		Cells	-	-	✓	-	-	-	-
	c.186 C>A	Tissue	-	-	-	-	-	-	✓
		Cells	-	-	-	-	-	-	-
	(with FBXO11:) c.3647- 70_3647-54 del TTTTTGTTTAA ATTCCT, c.*1985 AGGAATTAAA ACAAAAAT>T	Tissue	-	✓	-	-	-	✓	-
		Cells	-	✓	-	-	-	✓	-

MSH6	c.116 G>A	Tissue	-	✓	✓	-	-	-	-
		Cells	-	✓	✓	-	-	-	-
	c.540 T>C	Tissue	-	✓	-	-	-	-	-
		Cells	-	✓	-	-	-	✓	-
	c.642 C>T	Tissue	-	✓	-	-	-	-	-
		Cells	-	✓	-	-	-	✓	-
	c.1186 C>G	Tissue	-	-	-	✓	-	-	-
		Cells	-	-	-	✓	-	-	-
	c.3306 T>A	Tissue	-	-	✓	-	-	-	-
		Cells	-	-	✓	-	-	-	-
	c.3438+14 A>T	Tissue	✓	-	-	✓	-	-	-
		Cells	✓	-	-	✓	-	-	-
	c.3646+29_36 46+32 del CTAT	Tissue	✓	✓	✓	✓	-	-	-
		Cells	✓	✓	✓	✓	✓	✓	-
	c.3646+35_36 46+40 indel T	Tissue	-	-	-	✓	-	-	-
		Cells	-	-	✓	✓	-	-	-
MLH1	c.655 A>G	Tissue	✓	✓	✓	✓	-	-	-
		Cells	✓	✓	✓	✓	✓	-	-
	c.1558+14 G>A	Tissue	-	✓	-	-	✓	-	✓
		Cells	-	✓	-	-	✓	-	✓
	c.1668-19 A>G	Tissue	✓	✓	-	-	-	-	-
		Cells	✓	✓	-	-	-	✓	-
PIK3CA	c.1060-17 C>A	Tissue	-	✓	-	-	-	✓	✓
		Cells	✓	✓	-	-	✓	✓	-
	c.1173 A>G	Tissue	-	-	-	-	-	-	-
		Cells	-	-	-	-	-	✓	-
	c.2295-57 C>G	Tissue	✓	✓	-	-	-	-	-
		Cells	✓	✓	-	-	✓	✓	-
	c.2416+67 A>G	Tissue	-	-	✓	-	-	-	-
		Cells	-	-	✓	-	-	-	-
FBXW7	c.1855+67 G>A	Tissue	✓	-	-	-	-	-	-
		Cells	✓	-	-	-	-	-	-
	c.1972 G>A	Tissue	-	-	-	-	-	-	✓
		Cells	-	-	-	-	-	-	✓

% mutations shared by tissues and cells)	78.9%	91.3%	77.8%	80.0%	15.4%	46.7%	44.4%
--	-------	-------	-------	-------	-------	-------	-------

Mutations detected in the FFPE CA tissues and CA-derived cells for 9 CA-related genes are displayed. Presence (✓) or absence (-) of mutations are indicated. Green font is used for mutations that were detected in both the CA tissue and CA-derived cells from one patient, and red font for mutations only found in one sample type from each patient. Mutation types are substitutions (>), deletions (del), insertions (ins) and insertion + deletion mutations (indel). The percentage of mutations shared by the CA tissue and the CA-derived cell line from each patient were calculated by dividing the number of mutations shared by tissues and cells (green ticks) by the total number of mutations across the tissues and/or the cells (green and red ticks), and are displayed at the bottom of each column.

The APC gene was the most commonly mutated, with the c.5034 G>A mutation present in the tissue and cells for all 7 samples. Other common mutations to the APC gene were c.1458 T>C, c.4479 G>A, c.5268 T>G and c.5880 G>A.

Other frequent mutations were in the TP53 (c.215 C>A), KRAS (c.483 C>T) and MSH6 (c.3646+91 T>C SNV and c.3646+29_3646+32 deletion) genes. Overall, the mutational patterns seemed random, although TP53 mutations in FFPE were most likely to be absent in the tissue-derived cell lines, whereas some PIK3CA mutations in the tissue-derived cell lines were absent in their parent FFPE tissue.

The number of mutations present in the 9 CA-related genes and shared by both sample types (the FFPE CA tissues and the CA-derived cell lines) was calculated manually from the raw sequencing data (BAM files). These are displayed in Table 3.2 above. The green ticks signify that an identical mutation was present in both sample types, whereas a red tick identifies a mutation seen only in the tissue or the cell line. Overall, 78 out of 115 (67.8%) mutations were shared by the FFPE CA tissue samples and the matched CA-derived cell lines based on the analysis of the 9 genes. There were 71 occasions of a shared SNV, plus 6 instances of shared MSH6 deletions and 1 instance of a shared MSH6 INDEL (Table 3.2, all shown in green). Of the 37 mutations present in 1 only sample type per patient (either the tissue or the cells) and not the other (32.2%), 13 were only in FFPE CA tissue samples and 24 only in cell lines (Table 3.2, red). As expected, 13 of these 37 mutations were only observed in 1 out of 7 patients, representing very low-frequency mutations and possibly present through sequencing errors.

3.2.4 Types of nucleotide substitutions

Next, the types of substitutions were analysed (Table 3.3). Transition substitutions were predominant in HGCA tissue samples, with the most common being C>T and T>C transitions. However, transversions ranked amongst the most common substitutions in tissue samples and cell lines from all 3 LGCA patients and HGCA1, predominantly C>G and C>A transversions.

Table 3.3: Summary of coverage depth and DNA substitutions

		Average coverage depth	Substitutions (frequency)
LGCA1	Tissue	1407	C>T (50%), T>A (25%), T>C (25%)
	Cells	1549	C>G (75%), C>T (25%)
LGCA2	Tissue	888	C>T (71.4%), C>G (14.3%), C>A (14.3%)
	Cells	1837	C>A (50%), C>G (50%)
LGCA3	Tissue	1399	C>T (60%), C>A (40%)
	Cells	2255	C>A (100%)
HGCA1	Tissue	1177	C>T (57.1%), C>G (14.3%), T>C (14.3%), T>A (14.3%)
	Cells	1549	T>C (33.3%), C>G (33.3%), C>T (33.3%)
HGCA2	Tissue	925	C>T (100%)
	Cells	1631	C>T (33.3%), C>A (33.3%), T>C (33.3%)
HGCA3	Tissue	2655	C>T (79.2%), T>C (16.7%)
	Cells	2002	T>C (100%)
HGCA4	Tissue	2247	C>T (83.3%), T>C (16.7%)
	Cells	398	C>T (88.9%), T>C (11.1%)

Average depth of sequencing coverage and the predominant substitution types for each sample.

3.3 Discussion

The data summarising the overall Oncomine™ TML assay suggests that the tissues are highly variable, reflecting the heterogenous nature of CA tumours. Evidence for this is the high proportion of unique mutations and the wide range of TMB scores across the FFPE CA tissue samples. This was also observed in the study by TGCA (Cancer Genome Atlas, 2012), which found that mutation rates were highly variable between samples. In contrast, the CA-derived primary cell lines contained fewer unique mutations and had similar TMB scores, therefore appearing more homogenous. The high TMB scores in tissues may also reflect that the *in vivo* tumour contains necrotic regions and cells with lower fitness, some of which may die due to their mutational burden. Tissue-derived cells need to be more robust in order to survive the transition to culture conditions and therefore might be expected to carry fewer mutations. Alternatively, some cells with high mutation frequencies may contain mutations beneficial to growing in culture leading to their dominance *in vitro*.

When the analysis was narrowed to just 9 CA-relevant genes, common mutations present in the CA tissue samples were usually also present in the CA-derived cell lines. Furthermore, there were more mutations within these 9 genes which were unique to cells (24) than to tissues (13), suggesting that further mutations may have arisen during cell culture. Alternatively, a low-frequency mutation in the tissues may have been filtered out by the algorithm, but cells derived from this tissue which contain this mutation may

be enriched in the culture, increasing the prevalence of the mutation in the population and preventing it from being filtered out by the algorithm. This also suggests that many of the mutations detected in the tissues are passenger mutations which may not be directly relevant to the initiation or progression of the cancer.

The FFPE tissues and cell lines from all 3 LGCA patients plus HGCA1 had a higher frequency of mutations to the *MSH2* and *MSH6* genes, and more mutations overall within the 9 genes analysed. However, IHC staining performed as part of the diagnosis at the hospitals found that LGCA1 and LGCA3, as well as HGCA2, had normal staining patterns for MSH2, MSH6 and MLH1, while the other 4 samples were not tested for MMR/MSI during diagnosis (Appendix Table A1.1). This suggests that though these samples contain similar *MSH6* gene mutations, the function of the MSH6 protein may still be intact. This is exemplified by the low TMB scores for these 4 CA tissue samples. *MSH6* mutations are common in CRC, but there are redundant mechanisms to allow DNA mismatches to be repaired, provided the mutations to MMR genes are not germline (Salem et al., 2020).

The relevance of specific substitutions to CMS and MSI status have previously been reported (Vasaikar et al., 2019), which associated transition mutations with MSI-H and transversion mutations with MSS tumours. CMS1 is characterised by hypermutation, hypermethylation and MSI-H. The most likely cause of MSI-H has been reported to be *MLH1* promoter methylation (Salem et al., 2020). Furthermore, MSI-H tumours are more likely to have a high TMB score (Salem et al., 2018). Collectively, this provides a link between hypermethylation, MSI-H and a high mutation rate. HGCA3 and HGCA4 had the highest TMB scores and contained transition substitutions almost exclusively (Table 3.3), which supports this assertion and aligns them most closely with the CMS1 category, despite a lack of *BRAF* mutations (Guinney et al., 2015). The 3 LGCA and the HGCA1 tissue samples, which all had low TMB scores, exhibited predominantly transversion substitutions (Table 3.3). They also had a higher prevalence of *KRAS* mutations, which seems to place them closest to the CMS3 category (Guinney et al., 2015). Furthermore, along with the substitution types displayed, somatic MMR mutations occur in less than half of MSI-H tumours, further evidence that these samples do not fall into the MSI-H/CMS1 category. Despite the tissue-derived cells having lower TMB scores than their patient-matched FFPE tissue samples, they are still very similar in the sense that they contain the same mutations to key genes.

There have been reports of mutations that are an artefact of the formalin fixation process. These are typically deamination mutations that cause substitutions (C>T and G>A) (Prentice et al., 2018). Deamination events typically occur when tissue is left in formalin for more than 48 h, and the ideal fixation

protocol has been suggested as 24 h in 10% formalin (Einaga et al., 2017; Prentice et al., 2018). However, the frequency of formalin-induced deamination mutations is generally less than 2%, so they should be filtered out by the variant calling algorithm (Prentice et al., 2018). Deamination events at methylated cytosines, such as in CpG sites, produce uracils that are either repaired or spontaneously transition to thymines (Do et al., 2015). One method of reducing these artefactual mutations is to treat the DNA with uracil-DNA glycosylase (UDG) before amplification, which removes the uracil to leave an abasic site or a 1-3 base deletion (Do et al., 2015). Treatment of DNA with UDG before sequencing has been shown to greatly reduce the presence of deamination artefact (Do et al., 2015). All DNA samples were treated using UDG before being sequenced for the OncoPrint™ TML assay, so there can be a high degree of confidence that the C>T and G>A transition mutations in these FFPE tissue samples are not artefacts. Furthermore, this is validated by the presence of the same mutations in the patient-matched cell lines which were derived from small samples of the CA tissues before fixation.

To ensure that the data for identified mutations were reliable, the BAM files containing aligned sequences were opened using the Integrative Genomics Viewer (IGV) tool on the Ion Reporter server to assess the coverage depth and frequency of each mutation. In this way, mutations with low frequency (<6%) or low read depth (<300 reads) could be excluded. The sequencing depth of flanking regions was also checked to ensure that the chromosomal region being sequenced was reliable overall.

3.4 Conclusion

While the FFPE CA tissue samples were variable in terms of their unique mutations and TMB scores, the cell lines derived from them shared 78 of the 115 SNVs identified within 9 key CA-related genes. The HGCA tissue samples with high TMB scores have a mutational signature that aligns with the CMS1 category, whereas the LGCA tissue samples have fewer mutations and a mutational signature more similar to the CMS3 category. Based on the sequencing data, it can be concluded that although they do not fully reflect the heterogeneity of the tumour tissues, the cells have a high degree of similarity to the parent tissues in terms of mutations to CA-relevant genes (67.8%), and therefore they are likely to be useful as an *in vitro* model for functional studies.

Chapter 4: Cancer Stem Cell Subpopulations in Primary Colon

Adenocarcinoma

4.1 Introduction

The cancer stem cell (CSC) concept hypothesises that tumour growth is driven by CSCs, a small subpopulation of cancer cells with stem cell characteristics (Khalek et al., 2010; Kreso et al., 2014; Shimokawa et al., 2017; Zheng et al., 2013). CSCs produce identical daughter cells which are pluripotent and can self-renew, as well as progenitor cells which are more committed and sit on a hierarchy between CSCs and terminally differentiated cancer cells (Gage, 2000; Seaberg et al., 2003; Tang, 2012). Cells within this hierarchy can be identified by their expression of different combinations of markers (A. Bradshaw et al., 2016; Seaberg et al., 2003). Tang (Tang, 2012) postulates that progenitor cells are responsible for uncontrolled growth.

The most common markers used to identify colon CSCs are CD133, LGR5 and EpCAM, though it is now widely accepted that they are not expressed by all colon CSCs, and are present in some non-stem cells (Guo et al., 2011). One way to enhance the identification of CSCs may be to introduce panels of markers which are more likely to be expressed, or at a higher level, in CSCs than in non-CSC tumour cells. CSCs are considered pluripotent, and so the aberrant expression of markers associated with pluripotency and development might be good candidates for a CSC identification panel.

Takahashi and Yamanaka first used *OCT4*, *SOX2*, *KLF4* and *c-MYC* to produce induced-pluripotent stem cells (iPSCs) from mouse fibroblasts (Takahashi et al., 2006) and adult human fibroblasts (Takahashi et al., 2007). The Thomson laboratory also successfully produced iPSCs from human fibroblasts using *OCT4*, *SOX2*, *NANOG* and *LIN28* (J. Yu et al., 2007).

Primary CA, the most common type of CRC, is categorised as low-grade CA (LGCA; well and moderately differentiated tumours with greater than 50% crypt and gland composition) or high-grade CA (HGCA; poorly differentiated tumours with densely packed tumour cells) (Fleming et al., 2012). Although CSCs have been previously studied in CRC, the putative subpopulations of CSCs are yet to be fully characterised.

The aims of this chapter were to investigate the level of iPSC marker gene transcription in CA tissues (n=12) using RT-qPCR and *in situ* hybridisation (ISH), and to determine their distribution within CA using IHC (n=18) and immunofluorescence (IF) staining (n=6). It was hypothesised that *c-MYC* and *KLF4* expression would be relatively widespread due to their function in the NC, but that *OCT4*, *SOX2* and

NANOG would be present in a limited number of cells – the CSCs or early progenitors. The presence of established colon CSC markers CD133, LGR5 and EpCAM was investigated to see whether the inclusion of these markers could help to verify the utility of iPSC markers in a CSC panel. Using these data, the expression levels and localisation of iPSC and CSC markers were assessed in LGCA and HGCA tissue samples compared to their patient-matched NC tissues.

4.2 Results

4.2.1 EpCAM expression in the normal colon and CA

EpCAM is a cell surface protein expressed exclusively by epithelial cells and overexpressed in cancers of epithelial origin (Dalerba et al., 2007; van der Gun et al., 2010). CA tissues and their patient-matched NC samples were probed with an antibody against EpCAM to distinguish between epithelial cells and stromal cells. It was found that EpCAM expression was restricted to epithelial cells in all NC, LGCA and HGCA tissues (Fig 4.1). This specificity was utilised when attempting to distinguish tumour epithelium from stroma in tumours with highly chaotic architecture. Overall, EpCAM expression was weak to moderate in NC, and moderate to strong in LGCA. Some HGCA samples exhibited very strong EpCAM staining, but others were weak to negative and presumably represent poorly differentiated tumours.

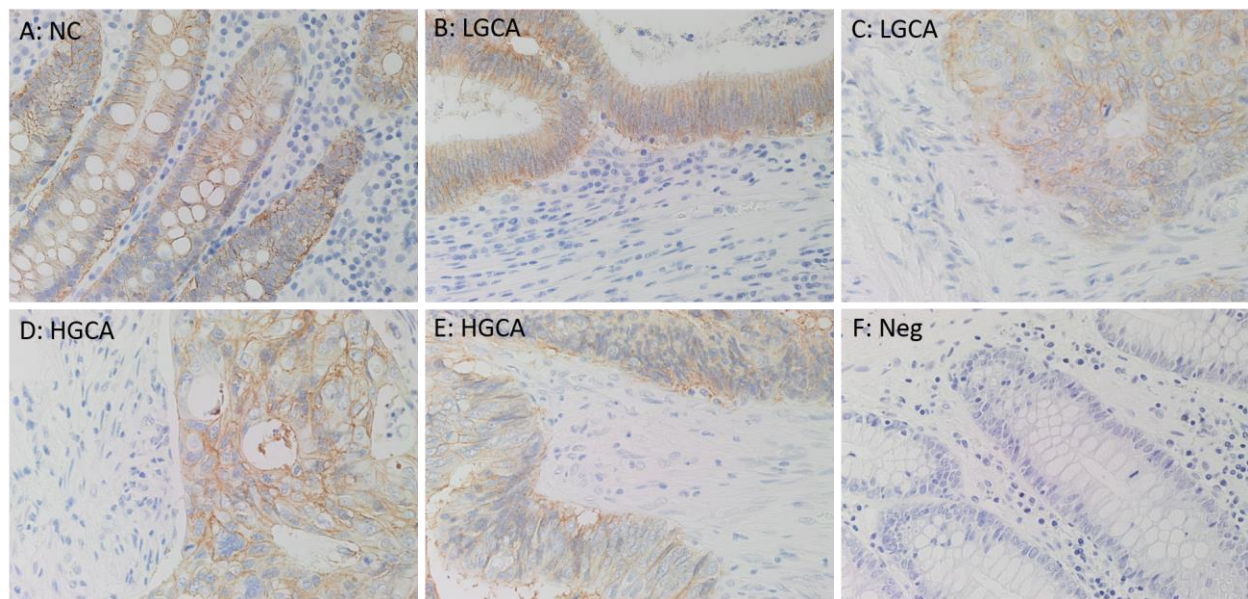


Figure 4.1: IHC staining for EpCAM. Representative IHC staining images showing protein expression of EpCAM (brown) in NC (A), LGCA (B&C), and HGCA (D&E). In all normal and tumor samples, EpCAM was expressed only by the epithelial cells and not by stromal cells. Negative control = F. Nuclei were counterstained with hematoxylin (blue). Original magnification: 400x.

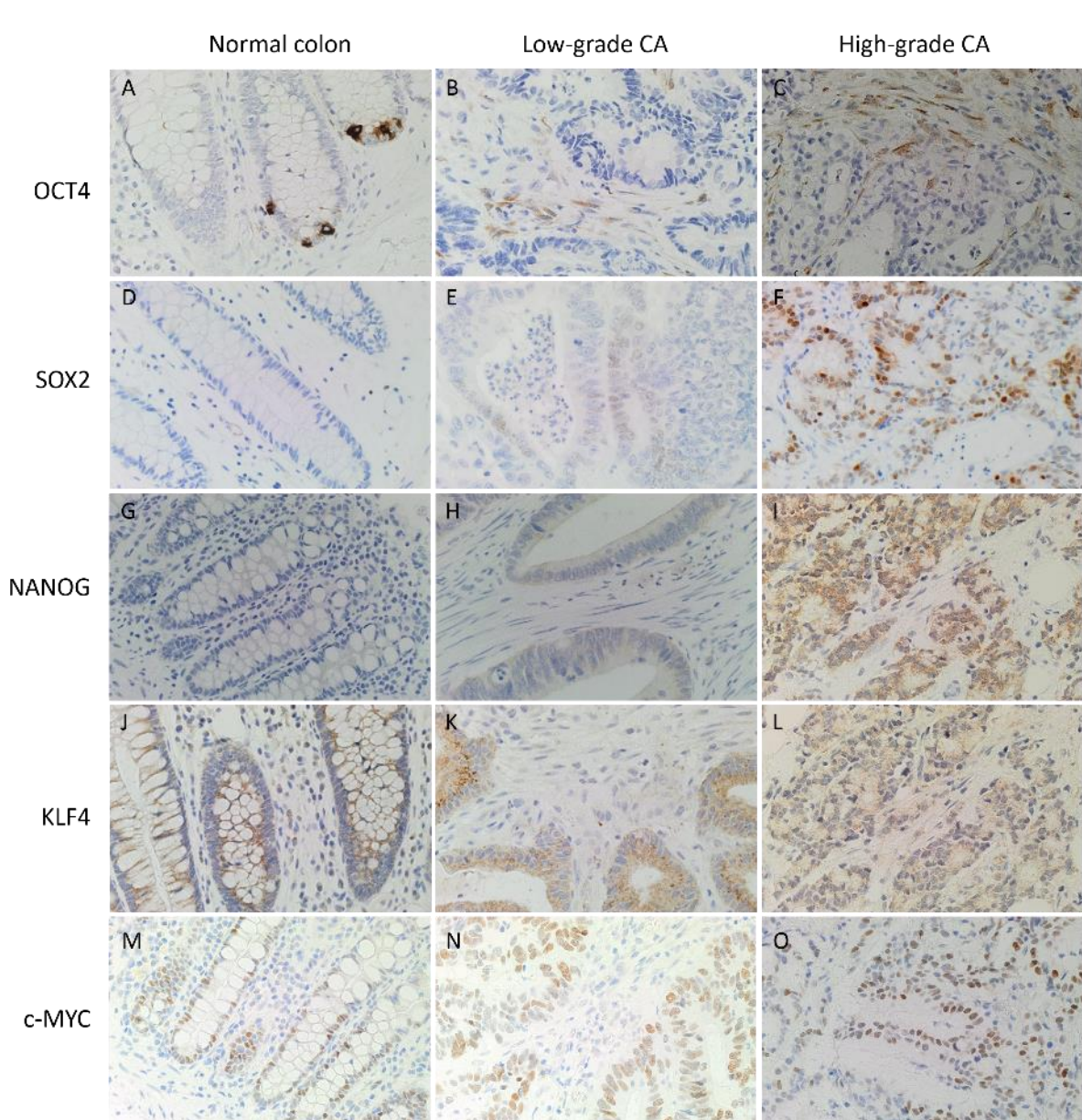


Figure 4.2: iPSC marker IHC staining. Representative IHC staining images showing protein expression of iPSC markers OCT4 (A-C, brown), SOX2 (D-F, brown), NANOG (G-I, brown), KLF4 (J-L, brown) and c-MYC (M-O, brown) in NC (A,D,G,J,M), LGCA (B,F,H,K,N) HGCA (C,F,I,L,O) tissue samples. Nuclei were counter-stained with hematoxylin (blue). Original magnification: 400x.

4.2.2 OCT4 is expressed in the CA stroma

Very strong staining for OCT4 (Fig 4.2A-C) was detected in the cytoplasm of a small proportion of epithelial cells in the NC tissues (Fig 4.2A). The staining pattern was almost identical to that seen by Alexander *et al.* (Alexander et al., 2014), who attributed this to OCT4 expression by enterochromaffin cells, a type of neuroendocrine cell. Other than neuroendocrine cell staining, the NC tissues did not express OCT4. In

contrast, OCT4 was found in the cytoplasm of some elongated stromal cells in LGCA (Fig 4.2B) and HGCA (Fig 4.2C) tissues, with little or no expression in the tumour epithelium and an absence of obvious neuroendocrine cell staining.

OCT4 mRNA was detected by RT-qPCR in all CA tissues and all but one NC tissue sample, with significant upregulation seen in two HGCA cases and significant down regulation in one LGCA and one HGCA case (Fig 4.3A). When assessed by ISH, there was a noticeable stepwise increase in *OCT4* mRNA abundance from NC to LGCA and from LGCA to HGCA (Fig 4.4A-C, brown).

4.2.3 SOX2 is expressed in CA

The expression of SOX2 was investigated in 3 LGCA and 3 HGCA tissues and their patient-matched NC tissues by IHC staining. SOX2 (Fig 4.2D-F) was generally not expressed in NC tissues, but in 2 out of 6 samples there were occasional crypts with positive staining in the nuclei (Appendix Figure A4.1A). Only 1 LGCA sample displayed weak nuclear staining (Fig 4.2E), however all 3 HGCA cases showed SOX2 nuclear staining which was weak and scattered in 2 cases and strong in the other (Fig 4.2F). Furthermore, 1 of the 3 HGCA tissues showed weak to moderate cytoplasmic staining of epithelial cells.

SOX2 mRNA abundance was measured in 6 LGCA and 6 HGCA tissues. When measured by RT-qPCR, SOX2 mRNA was below the detection threshold in the NC and CA tissue samples from 3 HGCA and 3 LGCA cases, with significant upregulation in 2 HGCA cases (Fig 4.3B). ISH demonstrated the presence of SOX2 mRNA (Fig 4.4D-F, brown) at very low levels in the epithelial cells, though it was more abundant in CA tissues than NC tissues.

4.2.4 The NANOG protein is expressed in CA but not NC

NANOG (Fig 4.2G-I) was not detected by IHC staining in NC tissues (Fig 4.2G) but was present in 3 out of 10 LGCA cases (Fig 4.2H) and 5 out of 8 HGCA cases (Fig 4.2I). Staining was predominantly seen in the cytoplasm, and the intensity was greater in HGCA tissues than LGCA tissues. These results were verified using an alternative antibody (ab62734) on 2 LGCA and 2 HGCA samples which had positive staining with the first antibody (Appendix Figure A4.1B-D). The new antibody resulted in slightly weaker staining in the positive control tissue (seminoma) despite using double the concentration. Furthermore, while both LGCA tissues stained positively, albeit with a greater presence of nuclear staining, only 1 of the 2 HGCA samples stained positive.

NANOG mRNA was detected in all 12 NC tissue samples; 3 out of 6 LGCA cases had significant upregulation of *NANOG* relative to their matched NC samples, with 2 LGCA and 2 out of 6 HGCA showing

significant downregulation (Fig 4.3C). ISH revealed that *NANOG* mRNA was predominantly expressed by epithelial cells, but occasionally seen in CA stromal cells (Fig 4.4G-I, brown).

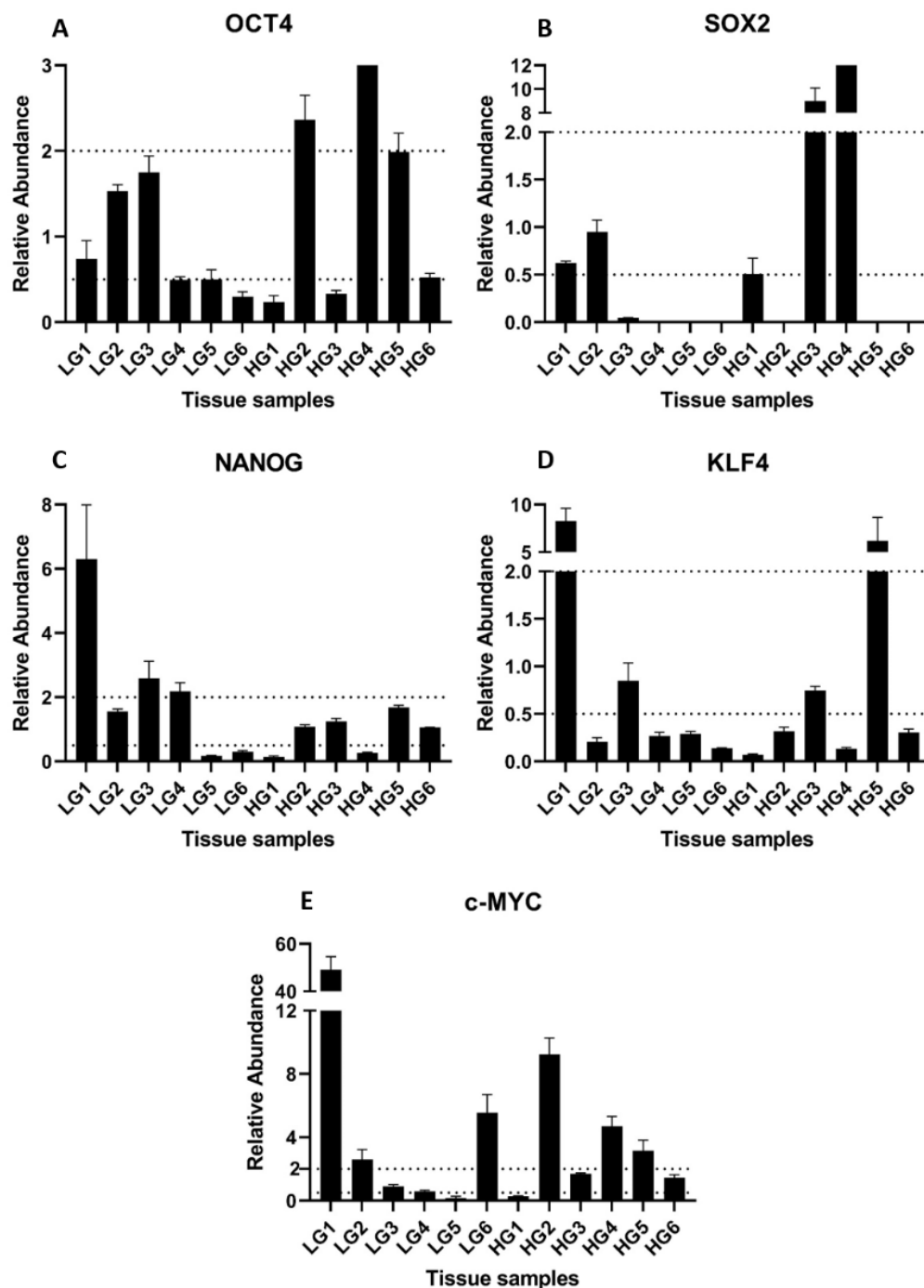


Figure 4.3: RT-qPCR data from tissues. Expression of iPSC genes *OCT4* (A), *SOX2* (B), *NANOG* (C), *KLF4* (D) and *c-MYC* (E) detected by RT-qPCR. Data displayed as the relative abundance of mRNA in tumour samples relative to their patient-matched NC sample (Y-axis). A cohort of six LGCA tissue samples and six HGCA tissue were analysed (X-axis). Δ CT and fold-change data are displayed in Appendix Table A4.1.

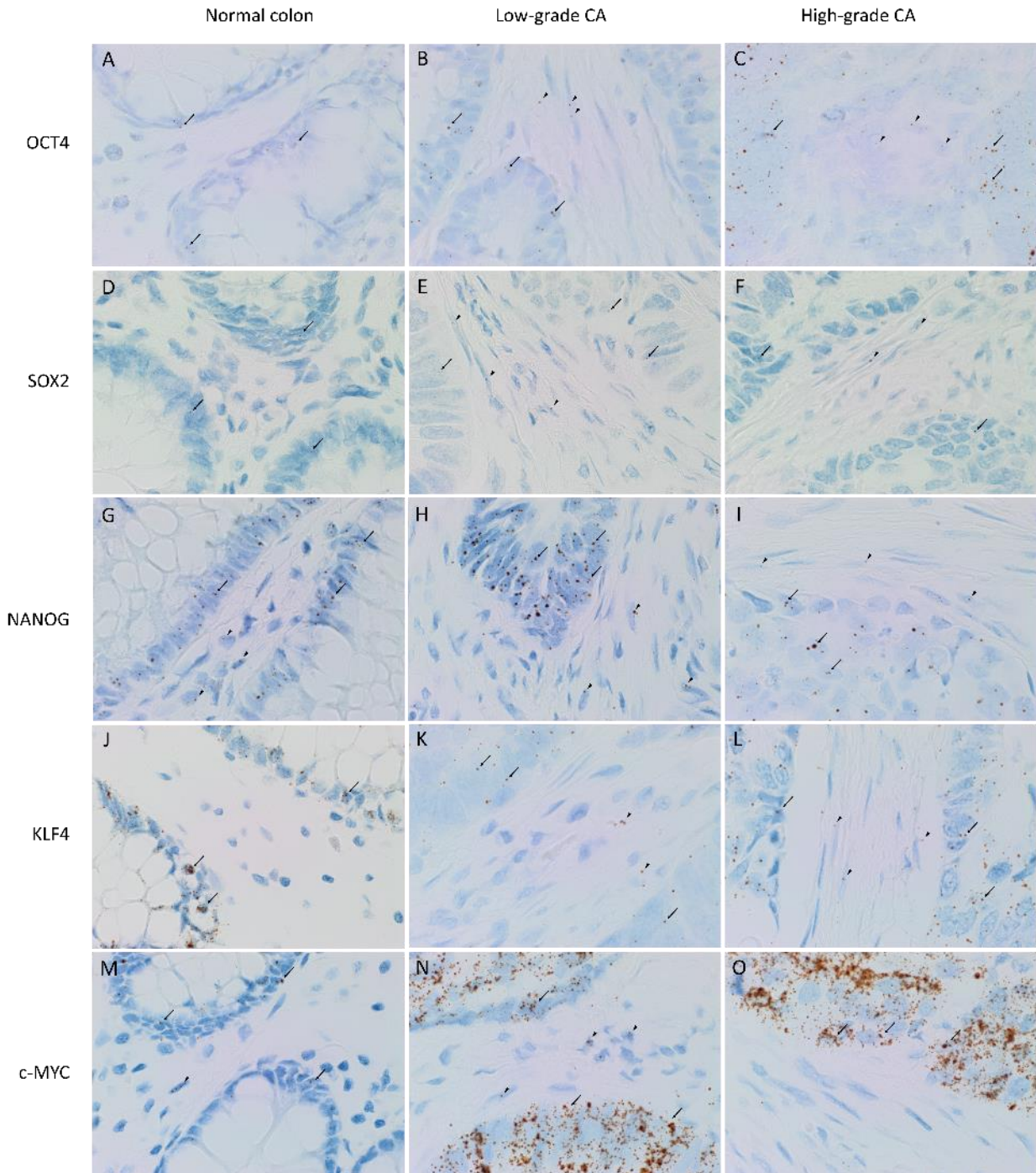


Figure 4.4: *In situ* hybridization. Representative images of *in situ* hybridization, showing mRNA expression of iPSC genes *OCT4* (A-C, brown), *SOX2* (D-F, brown), *NANOG* (G-I, brown), *KLF4* (J-L, brown) and *c-MYC* (M-O, brown) in epithelial cells (*arrows*) and stromal cells (*arrowheads*) of NC (A,D,G,J,M), LGCA (B,E,H,K,N) and HGCA (C,F,I,L,O) tissue samples. Nuclei were counter-stained with hematoxylin (blue). Positive and negative controls are shown in Appendix Figure A4.5. Original magnification: 1000x.

4.2.5 KLF4 abundance is inversely proportional to CA grade

KLF4 (Fig 4.2J-L) showed perinuclear expression in NC epithelial cells (Fig 4.2J). Nuclear staining was more prevalent in the epithelial cells of HGCA (Fig 4.2L) than in those of LGCA (Fig 4.2K) and NC (Fig 4.2J). Stromal KLF4 staining was more abundant in HGCA-matched NC tissues than it was in LGCA-matched NC tissues. Similarly, the stroma of HGCA tissues displayed more widespread KLF4 staining than the stroma of LGCA tissues.

As expected, *KLF4* mRNA was detected in all 12 NC samples. Furthermore, while it was significantly upregulated in 1 LGCA and 1 HGCA sample, it was significantly downregulated in 4 LGCA and 4 HGCA samples (Fig 4.3D). This was validated using ISH, which confirmed that *KLF4* mRNA was more highly expressed in the epithelium of NC tissues than that of CA tissues, though LGCA appeared to have the lowest levels (Fig 4.4J-L, brown).

4.2.6 NC and CA express high levels of c-MYC

c-MYC (Fig 4.2M-O) generally demonstrated moderate to strong nuclear staining in the epithelial cells of all NC and CA tissues. Some NC tissues (Fig 4.2M) stained weakly, and most LGCA tissues (Fig 4.2N) had a similar staining intensity to their matched NC. Staining intensity in HGCA tissues (Fig 4.2O) was variable, but generally stronger than LGCA, and most HGCA tissues stained more strongly than their matched NC tissues.

There was significant upregulation of c-MYC mRNA in 3 LGCA and 3 HGCA tissues and downregulation in 1 LGCA and 1 HGCA tissue (Fig 4.3E). ISH revealed that the NC contained relatively low levels of c-MYC mRNA, with significantly more detected in LGCA and HGCA tissues (Fig 4.4M-O, brown).

4.2.7 CD133 and LGR5 IHC staining

IHC staining for CD133 (Fig 4.5A-C) was stronger in tumours than their matched NC tissues, with expression seeming to be localised to the luminal surface of epithelial cell membranes. The CD133 antibody seemed to be taken up non-specifically in some areas of mucin and necrosis in HGCA tissues, but also stained with specificity in the cytoplasm of tumour epithelial cells.

LGR5 expression was localised to pockets within the tumour, but specifically around the edge of tumour nests (Fig 4.5D-F). In most cases, expression was noticeably higher in CA tissues relative to their matched NC samples.

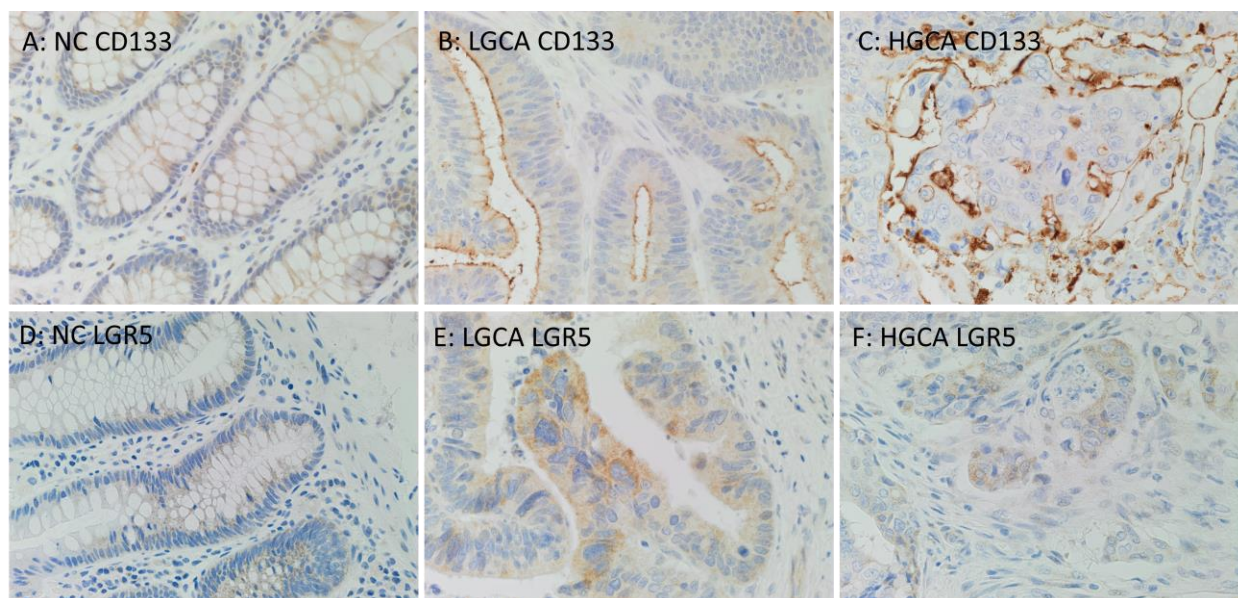


Figure 4.5: Stem cell marker IHC staining. Representative IHC staining images showing protein expression of CD133 (A-C, brown) and LGR5 (D-F, brown) in NC (A,D), LGCA (B,E) and HGCA (C,F) tissue samples. Nuclei were counter-stained with hematoxylin (blue). Positive and negative controls are shown in Appendix Figures A4.2 and A4.3, respectively. Original magnification: 400x.

4.2.8 IF staining

Based on IHC staining for iPSC markers, two potential CSC subpopulations were identified: one within the CA epithelium, with 3 of 10 of LGCA and 5 of 8 of HGCA cases expressing NANOG; and the other within the CA stroma expressing OCT4 in both LGCA and HGCA tissue samples, but not in the stroma of NC tissues. IF staining expanded on the findings of IHC staining by interrogating the localisation of two iPSC markers simultaneously.

OCT4 (Fig 4.6, green) was expressed in the cytoplasm of cells within the stroma of LGCA (Fig 4.6B,E,H,K,N) and HGCA (Fig 4.6C,F,I,L,O) tissue samples, and by neuroendocrine cells in NC crypts (Fig 4.6A,D,G,J,M).

KLF4 (Fig 4.6A-C, red) stained positively in the cytoplasm of epithelial cells in NC (Fig 4.6A), LGCA (Fig 4.6B) and HGCA (Fig 4.6C), and a few stromal cells in LGCA and HGCA tissues. However, these KLF4⁺ stromal cells did not express OCT4.

NANOG (Fig 4.6D-F, red) was absent in the NC (Fig 4.6D), and despite staining positively in IHC staining it was difficult to detect in the cytoplasm of epithelial cells in LGCA (Fig 4.6E) and HGCA (Fig 4.6F) samples by IF staining.

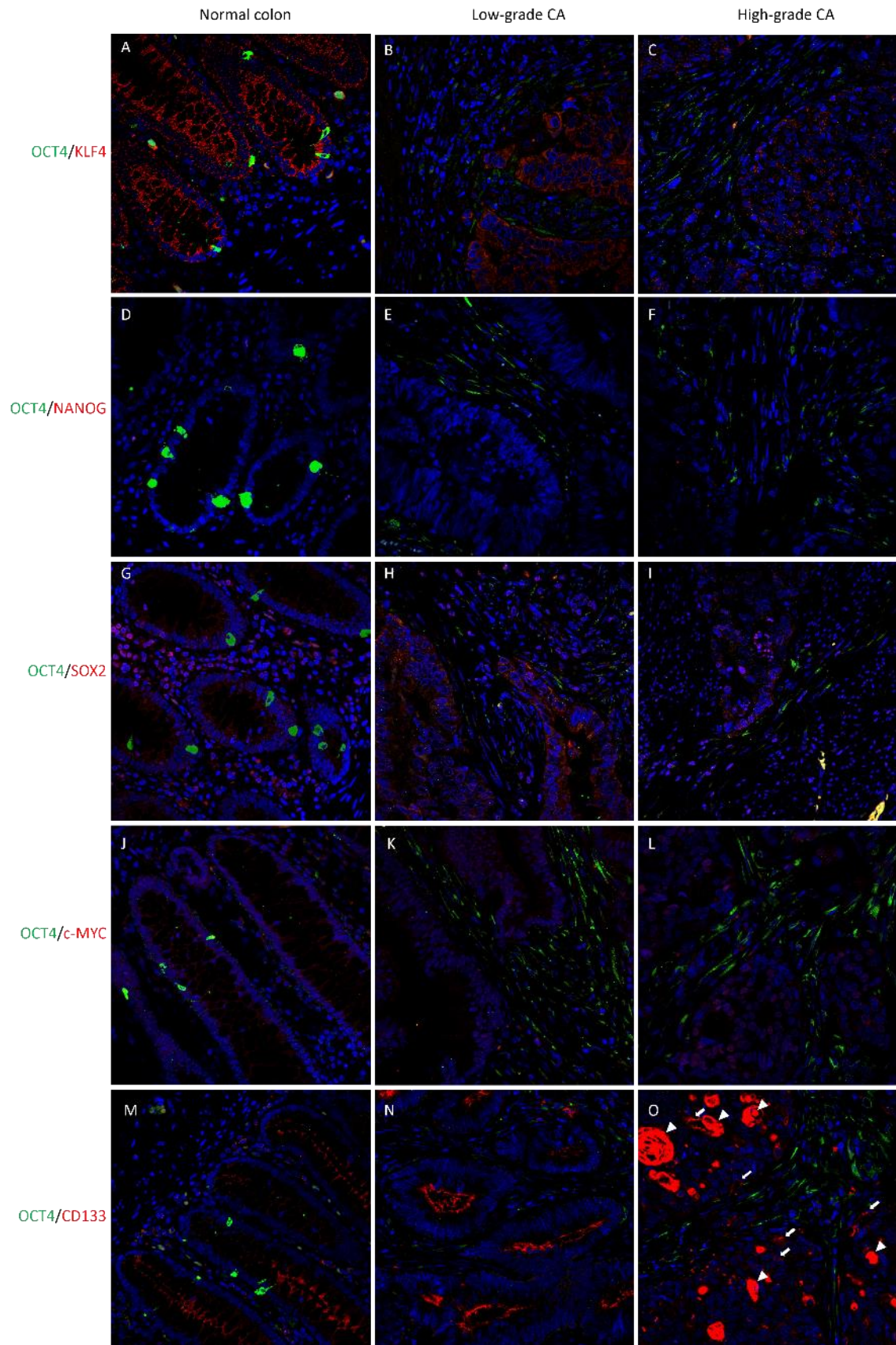
IHC staining showed that SOX2 was only present in the nuclei of NC epithelial cells, and so the red staining seen in the cytoplasm in NC by IF staining was not considered to be true SOX2 staining (Fig 4.6G). SOX2 (Fig 4.6G-I, red) was widely expressed in the nuclei of epithelial cells in LGCA (Fig 4.6H) and HGCA (Fig 4.6I) tissue samples by IF staining. However, OCT4 was not expressed by SOX2⁺ stromal cells.

Nuclear c-MYC staining (Fig 4.6J-L, red) was seen in the epithelial cells of NC (Fig 4.6J), LGCA (Fig 4.6K) and HGCA (Fig 4.6L).

CD133, a potential CSC marker, was restricted to the luminal membrane surface of many epithelial cells, but it did not co-localise with OCT4 (Fig 4.6M-O). As with the IHC staining, the CD133 antibody seemed to non-specifically bind mucin in one HGCA case (Fig 4.6O, arrowheads) but was also expressed on the luminal membrane and cytoplasm of epithelial cells (Fig 4.6O, arrows).

From the above IHC and IF staining data, it was inferred that there was one possible CSC subpopulation in the epithelium which co-expressed NANOG, SOX2, KLF4, c-MYC, CD133 and EpCAM. Another subpopulation of cells in the tumour stroma expressed only OCT4 and it is therefore unclear whether this is a CSC subpopulation.

Figure 4.6: IF staining (page 76). Representative immunofluorescence images showing protein expression of iPSC markers OCT4 (A-O, green), KLF4 (A-C, red), NANOG (D-F, red), SOX2 (G-I, red), c-MYC (J-L, red) and CD133 (M-O, red) in NC (A,D,G,J,M), LGCA (B,E,H,K,N) and HGCA (C,F,I,L,O) tissue samples. Cell nuclei were counterstained with 4', 6'-diamidino-2-phenylindole (DAPI; blue). Negative controls are shown in Appendix Figure A4.4. Original magnification: 400x.



4.3 Discussion

This chapter investigated the mRNA and protein expression of *OCT4*, *SOX2*, *NANOG*, *KLF4* and *c-MYC* to identify their presence in CA and, by inference from their localisation, the expression phenotypes of possible CSC subpopulations.

It has become clear that finding single markers to identify CSCs from different cancers is not a feasible endeavour due to the heterogeneity inherent to cancers, and the plasticity which allows them to change their phenotype in response to treatment or other stresses (Guo et al., 2011). Therefore, panels of markers could be employed to better differentiate between normal cells, non-tumourigenic cancer cells and functional CSCs.

LGR5 is one of the current candidates for identifying colon CSCs. However, it is expressed by normal adult stem cells as well as progenitor or transit amplifying cells in a wide range of tissues (Kemper et al., 2012). Nonetheless, it may be useful as a functional marker of colon CSCs because the binding of R-spondin to LGR5 enhances the phosphorylation of LRP5/6 and therefore contributes to Wnt signalling (Kemper et al., 2012). When LGR5⁺ cells are targeted for ablation, LGR5⁺/KRT20⁺ cells revert to an LGR5⁺ state to replace the ablated cells and allow tumour growth to continue (Shimokawa et al., 2017). This reveals a high degree of plasticity in the tumour and suggests that differentiated tumour cells can revert to a more stem-like LGR5⁺ state to enhance growth. Furthermore, the de-differentiation or plasticity of cancer cells provides an explanation for observations of cells initially negative for LGR5 having CSC capabilities (Guo et al., 2011). The IHC data presented above revealed low to moderate LGR5 expression in NC tissues and moderate to strong levels in pockets of the tumour, often at the edge of tumour nests where other CSC markers, including NANOG, have been documented (W. Luo et al., 2013). Accordingly, if a tumour is poorly differentiated then it might be expected that a majority of tumour cells express LGR5 and possibly the iPSC markers.

Similarly, CD133 was identified as an early CSC candidate which has subsequently been questioned as a definitive CSC marker. It was seen on the luminal membrane of epithelial cells in LGCA and HGCA. However, it was also relatively widespread in NC epithelial cell cytoplasm, lending weight to the suggestion that it is not selective enough as a CSC marker on its own.

RT-qPCR and ISH data for *SOX2* corroborated each other, with RT-qPCR failing to detect *SOX2* in three NC samples and two CA samples, and ISH showing *SOX2* to be the least abundant in terms of the number of cells containing mRNA. However, the *SOX2* protein seemed to be more abundant than the mRNA levels would suggest. Other studies have also shown an abundance of *SOX2* protein in both the nuclei and cytoplasm of CRC tumour cells even when mRNA levels are low (Amini et al., 2014; Talebi et

al., 2015). ISH showed an abundance of c-MYC mRNA, predominantly in the epithelium, which mirrored the results of IHC staining which showed moderate to strong nuclear staining.

KLF4 has been previously studied in CRC and shown to be associated with EMT, cell migration and metastasis (Hadjimichael et al., 2015). However, studies on the role of KLF4 in cancer often yield conflicting results (Muller et al., 2016). In the normal colon, KLF4 helps direct epithelial progenitor cells down the goblet cell lineage, the most abundant epithelial cell type in colonic crypts (May et al., 2010). As the grade of CA increases, tumours become less differentiated, and this may explain the observation of decreased KLF4 in both LGCA and HGCA tumours relative to NC tissues. This has been reinforced by other reports of lower KLF4 levels in cancer, and with low KLF4 expression conferring a poor prognosis (Ghaleb et al., 2016; Hashimoto et al., 2017; Hsu et al., 2014). Decreased cytoplasmic KLF4 specifically correlates with worse outcomes, further evidence that differential sub-cellular localisation is responsible for aspects of the disease state (Z. Y. Chen et al., 2000; Z. Liu et al., 2013; Y. Ma et al., 2017). CRM1 facilitates the translocation of proteins including KLF4 from the nucleus to the cytoplasm in a nuclear export signal-dependent manner (Y. Liu et al., 2013). In the cytoplasm of vascular smooth muscle cells, KLF4 has been shown to stabilise the cytoskeleton and promote stress fibre formation by interacting with actin (Y. Liu et al., 2013). The spliced variant KLF4 α is associated with promotion of tumourigenesis, and it has lost its nuclear localisation signal causing accumulation in the cytoplasm (Le Magnen et al., 2013). However, it is unclear whether the interaction of cytoplasmic KLF4 or its spliced variant KLF4 α with actin promotes tumorigenesis, and the antibodies and PCR probes used in this study are unable to distinguish between the two.

As highlighted in this chapter, elevated KLF4 and OCT4 protein expression in stromal cells of HGCA may reflect the migration of cancer cells away from the epithelium via EMT, which has been postulated as a major factor in CRC progression (Loboda et al., 2011). Furthermore, when applying the concept of a stem cell hierarchy in cancer (A. Bradshaw et al., 2016; Gage, 2000; Seaberg et al., 2003), it may be that cells at different levels of this hierarchy will express different combinations of these markers. For instance, OCT4 is known to be expressed by primitive stem cells such as ESCs (Hadjimichael et al., 2015; Shi et al., 2010) as well as by migratory cells, whereas KLF4 is associated with a more differentiated, non-proliferative phenotype (R. Hu et al., 2011; May et al., 2010), and so a cell co-expressing both of these markers may be in an intermediate state on the hierarchy between primitive and differentiated. Furthermore, it may be that cancer cells can flow up and down the hierarchy depending on the cues they receive.

In this study, IF staining identified two distinct subpopulations of cells. The first was a NANOG⁺/OCT4⁻ subpopulation localised to the epithelium and thought to be a CSC subpopulation. The second was an OCT4⁺/NANOG⁻ subpopulation within the stroma. The stromal OCT4⁺ subpopulation did not co-express other iPSC markers. Similarly, in the epithelium, SOX2, KLF4 and CD133 staining was widespread but comparatively few of these cells were also NANOG⁺. Based on the staining patterns of these markers, the presence of one predominant yet low-abundance CSC subpopulation in the epithelium of CA could be inferred, with a co-expression phenotype of NANOG⁺/SOX2⁺/KLF4⁺/c-MYC⁺/CD133⁺/EpCAM⁺. This indicates that the inclusion of CD133 and EpCAM along with iPSC markers may help to increase the accuracy of a CSC identification panel.

The literature correlating OCT4 with EMT and metastasis provides evidence supporting a stromal subpopulation expressing OCT4 that migrates away from the tumour (Dai et al., 2013; Muller et al., 2016; J. Neumann et al., 2011). Furthermore, NANOG is associated with maintenance of the stem-like phenotype of CSCs within the tumour, consistent with the observation of NANOG expression by what may be considered resident CSC-like cells within the CA tissues (J. Zhang et al., 2013). KLF4 and c-MYC are associated with proliferation and differentiation and it is therefore not unexpected that these two markers were co-expressed by epithelial cells within CA (Halim et al., 2018; R. Hu et al., 2011; M. Martini et al., 2016).

Some stromal cells within CA that stained positively for OCT4 did not express SOX2, KLF4 or c-MYC. It is possible that the OCT4⁺ stromal cells are cancer-associated fibroblasts recruited by the tumour and induced to express OCT4 (Som et al., 2016).

One intriguing hypothesis to explain the widespread expression of various iPSC markers, specifically in the context of tumour cell plasticity, is that pluripotency is better described as a function of a population rather than a phenotype of individual cells. It has been shown that cancer cells display a large degree of plasticity, including the replenishment of LGR5⁺/KRT20⁻ colon CSCs from LGR5⁻/KRT20⁺ differentiated cells (Shimokawa et al., 2017). This model of pluripotency suggests that the expression of key markers such as NANOG are stochastically dynamic, with fluctuations between periods of expression and of non-expression, allowing cells to respond to extracellular stimuli and retain balance within the cell population as a whole (MacArthur et al., 2013). One function attributed to CSCs is the ability to respond to stimuli such as chemotherapy by replenishing the sensitive cells which are ablated by the treatment. This model goes further by suggesting that there is no dedicated subpopulation of CSCs, but rather that at any point in time some of the cancer cells within the tumour are primed to replenish the depleted cells due to being in a suitable phase of gene expression. Accordingly, the expression patterns seen in the CA

cases explored in this study are representative of a single point in time, influenced by the specific extracellular stimuli and environment at that time. This model does not diminish the importance of targeting cells with CSC characteristics to treat cancer, but could prompt a shift in perspective from thinking of CSCs as a discrete subpopulation to considering them to be a fluid subset of cells which happen to have the expression profile at that time to enable survival, resistance, recurrence and migration. Furthermore, it emphasises the need for combination treatments which target different cell states, such as differentiation-primed and stem-like cells.

There are various limitations to this body of work. First, it must be kept in mind that the patient-matched ‘normal colon’ samples used as a control may not represent true normal colon; although NC samples were excised from a region of colon away from the tumour (i.e. “non-involved” and/or “non-adjacent”), it is feasible that despite the regular appearance of crypt architecture the cells may be abnormal in other ways. Secondly, the issue of antibody specificity and suitability for IHC staining can influence the results of a study. Originally, SOX2 IHC staining was carried out using the PA1-094 antibody from ThermoFisher Scientific. However, due to incorrect localisation, including the degree of cytoplasmic staining seen in this and other studies within our lab, these results were deemed to be unreliable; it is suspected that the antibody had expired. A new antibody from Abcam (cat # ab97959) was obtained, and a sub-cohort of 3 LGCA and 3 HGCA cases were stained using the new antibody. Similarly, after staining for c-MYC had been carried out, Abcam updated the product data sheet for our original c-MYC antibody (cat # ab32; Abcam) to reflect that it was no longer recommended for the detection of full-length c-MYC protein in human samples. Once it became clear that it was unsuitable, it was replaced with an extensively validated alternative (cat # ab32072), which, along with the antibodies used for OCT4 (cat # MRQ-10) and NANOG (cat # 443R-16), is of diagnostic grade suitable for clinical use. Finally, in-house validation of antibodies was ongoing throughout the time of this project and some antibodies were not able to be optimised for the IF workflow despite being used reliably for IHC staining. This, as well as species cross-reactivity, limited the number of combinations possible for IF staining. Originally, the co-expression of NANOG and c-MYC had been interrogated by using rabbit anti-NANOG and mouse anti-c-MYC primary antibodies, but the replacement c-MYC antibody was raised in rabbit, and a mouse anti-NANOG primary was not able to be optimised for IF staining at the time of submission.

4.4 Conclusion

The aim of this chapter was to examine the expression of iPSC markers and other stem cell markers in CA tissues and patient-matched NC samples. While most markers were present in the NC, the extent and

localisation of expression differed between CA and NC tissues. Notably, all iPSC markers besides OCT4 were co-expressed by a small subpopulation of cells within the epithelium, which also expressed CD133 and LGR5, while OCT4 was expressed by stromal cells within CA but not in NC tissue. The significance of these expression patterns will be addressed by using functional assays involving cells derived from these tissues.

Once validated, interpretation of localisation and expression levels of novel combinations of iPSC markers with established CSC markers may provide a valuable tool to help guide patient management by further stratifying tumour grade, identifying cases with higher potential for metastasis or relapse, or tracking response to therapy.

Chapter 5: Colon Adenocarcinoma-derived Cells that Express Induced-Pluripotent Stem Cell Markers Possess Stem Cell Function

5.1 Introduction

It has been hypothesised that markers of pluripotency may be used to identify subpopulations of CSCs, due to their documented expression in a range of cancers and their capacity to cause de-differentiation. However, there is a need to demonstrate that tumour cells expressing these markers exhibit functional characteristics of stem cells, and to determine whether they are suitable markers for identifying CSCs. There are now a range of validated *in vitro* tests for pluripotency which mitigate the reliance on both animal testing, such as patient-derived xenografts with their inherent ethical considerations, and teratoma assays, which are not standardised and are therefore inconsistent. In place of *in vivo* work, stem cell function is typically demonstrated by confirmation of the expression of pluripotency markers, tumoursphere formation assays and multilineage differentiation capability.

The aim of this chapter was to assess the expression of the iPSC markers and test the stem cell functionality of primary cell lines derived from CA tissue samples, with the hypothesis that a small number of CA-derived cells would express the combinations of iPSC markers seen in the CA tissues and would demonstrate *in vitro* stem cell functionality. The pluripotent stem cell 4-marker Immunocytochemistry (ICC) Kit (cat # A24881, ThermoFisher Scientific), an established method for identifying pluripotent cells, was used to investigate the expression of verified pluripotency markers TRA-1-60 and SSEA-4 and iPSC markers OCT4 and SOX2 in primary CA-derived cells. Cells were also grown in suspension culture using StemXVivo media (R&D Systems) to assess their tumoursphere-forming capacity. Furthermore, directed differentiation assays were employed to determine whether they had the potential to develop ectoderm, endoderm and mesoderm lineage phenotypes.

5.2 Results

5.2.1 Co-localisation of pluripotency markers by Immunocytochemistry

CA-derived primary cell lines were seeded and fixed on 8 chamber cell culture slides to assess their expression of pluripotency markers by using the PSC 4-marker ICC kit, which includes two of the iPSC

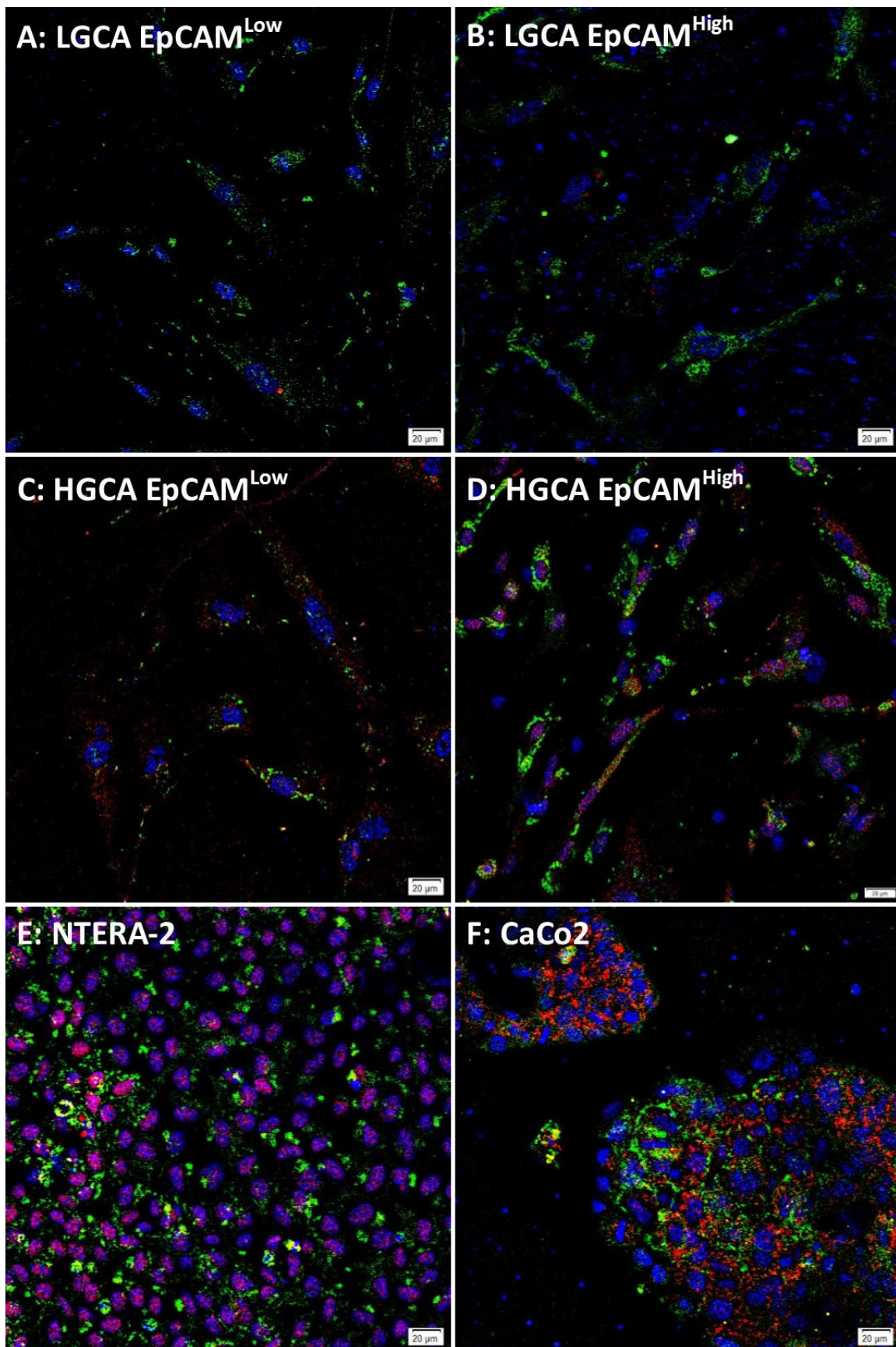
markers of interest, OCT4 and SOX2, as well as two validated pluripotency markers, SSEA4 and TRA-1-60, to verify whether the cells expressing OCT4 and SOX2 were pluripotent.

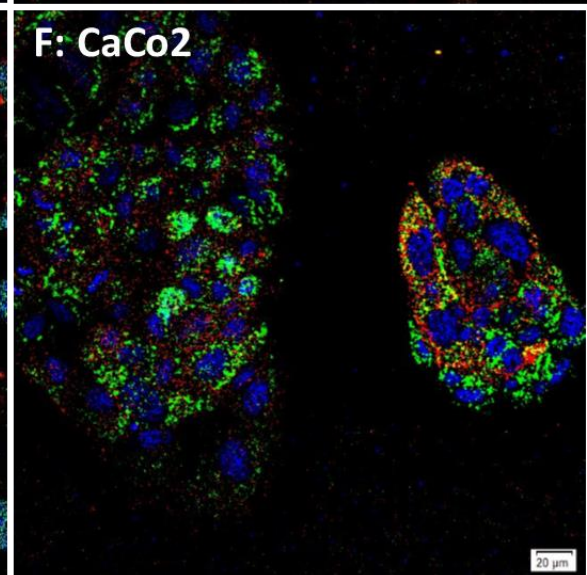
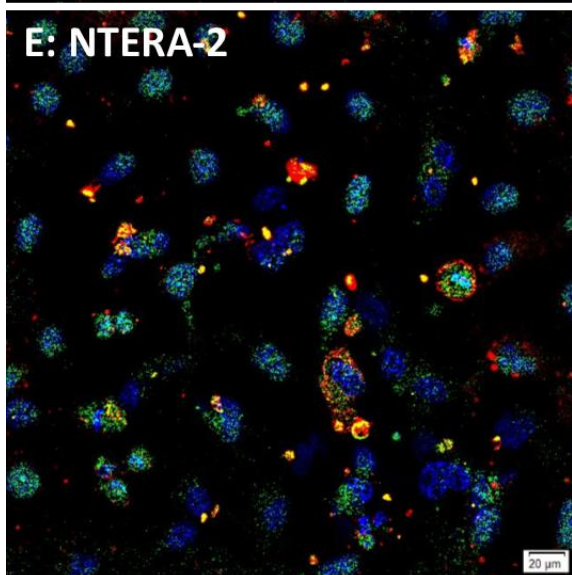
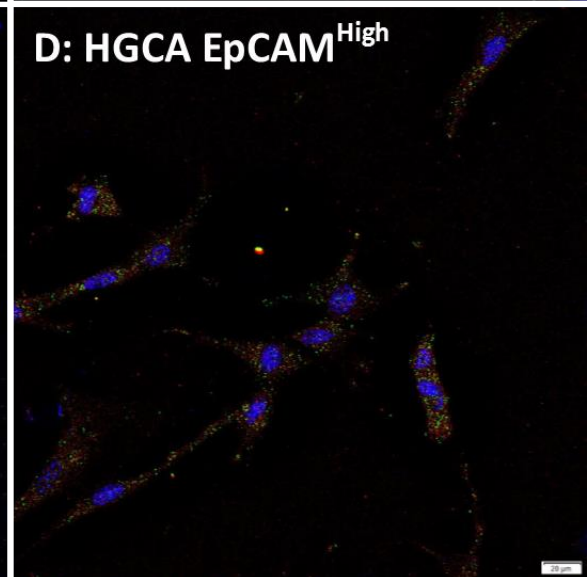
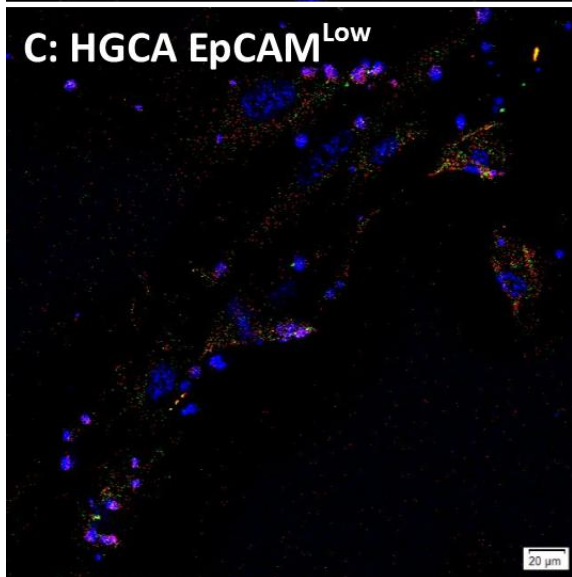
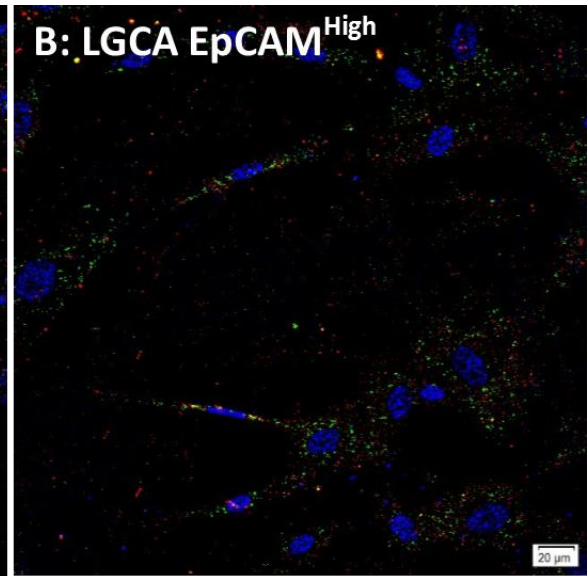
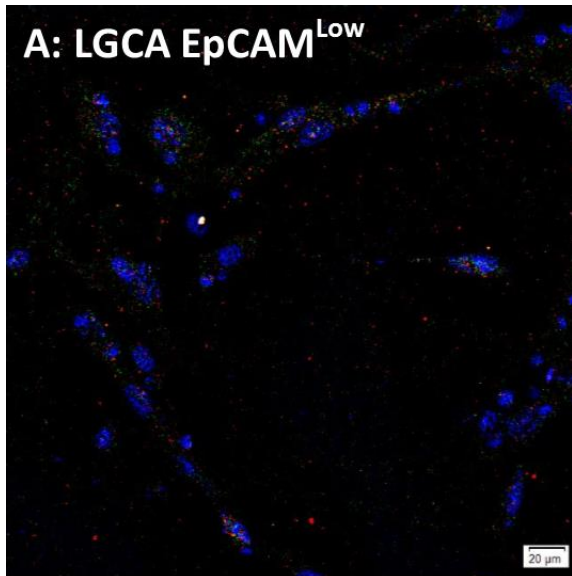
SSEA4 (Fig 5.1A-D) was expressed by all cell lines and co-expressed with OCT4 in 2 of the 6 EpCAM^{Low} and 4 of the 6 EpCAM^{High} cell lines. Similarly, TRA-1-60 (Fig 5.2A-D) was expressed in all cell lines and co-expressed with SOX2 in 4 of the 6 EpCAM^{Low} and 5 of the 6 EpCAM^{High} cell lines.

Commercial cell lines NTERA-2 and CaCo2 were used as the positive controls for RT-qPCR and tumoursphere formation assays, respectively, and so their expression of key pluripotency markers was also assessed by ICC. Both cell lines contained cells that co-expressed each of the four PSC markers: OCT4 and SSEA4 (Fig 5.1E,F), and SOX2 and TRA-1-60 (Fig 5.2E,F).

Figure 5.1: Representative SSEA4/OCT4 ICC images (page 84). EpCAM^{Low} (A) and EpCAM^{High} (B) cells from LGCA-derived primary cell lines, and EpCAM^{Low} (C) and EpCAM^{High} (D) cells from HGCA-derived primary cell lines, showing expression of SSEA4 (green) and OCT4 (red). LGCA (n=3); HGCA (n=3). Positive control NTERA-2 (E) and CaCo2 (F) cells were stained for SSEA (green) and OCT4 (red). Original magnification: 400x; scale bar = 20 µm.

Figure 5.2: Representative SOX2/TRA-1-60 ICC images (page 85). EpCAM^{Low} (A) and EpCAM^{High} (B) cells from LGCA-derived primary cell lines, and EpCAM^{Low} (C) and EpCAM^{High} (D) cells from HGCA-derived primary cell lines, showing expression of SOX2 (green) and TRA-1-60 (red). LGCA (n=3); HGCA (n=3). Positive control NTERA-2 (E) and CaCo2 (F) cells were stained for SOX2 (green) and TRA-1-60 (red). Original magnification: 400x; scale bar = 20 µm.





5.2.2 HGCA-derived cells form tumourspheres

EpCAM^{High} and EpCAM^{Low} cells derived from 3 LGCA and 3 HGCA tissue samples were cultured in ultra-low adherence plates with StemXVivo tumoursphere media. The threshold for positive tumoursphere formation was chosen to be an average diameter of 50 µm across all measured spheres per field of view (Hou et al., 2017; Pizon et al., 2016; X. Zhou et al., 2015).

The tumoursphere forming assay was carried out with three concurrent technical triplicates for each biological replicate. These results are displayed in Table 5.1, which shows that there was considerable variation between each of the biological replicates.

Table 5.1. Analysis of tumoursphere formation assay data

Sample ID	Average maximum diameter, μm (days to maximum diameter) \pm SD	Average maximum diameter, μm (days) \pm SD	Average maximum diameter, μm (days) \pm SD
HGCA1 EpCAM ^{Low}	102.6 (1) \pm 26.93	HGCA EpCAM ^{Low} 70.79 (3.0) \pm 28.26	EpCAM ^{Low} 54.74 (5.7) \pm 23.32
HGCA2 EpCAM ^{Low}	50.82 (3) \pm 9.08		
HGCA3 EpCAM ^{Low}	64.54 (5) \pm 13.25		
LGCA1 EpCAM ^{Low}	60.25 (6) \pm 10.96	LGCA EpCAM ^{Low} 48.10 (8.3) \pm 17.32	
LGCA2 EpCAM ^{Low}	32.35 (10) \pm 9.71		
LGCA3 EpCAM ^{Low}	32.31 (9) \pm 9.65		
HGCA1 EpCAM ^{High}	99.46 (2) \pm 11.48	HGCA EpCAM ^{High} 73.06 (4.7) \pm 18.65	EpCAM ^{High} 80.86 (5.5) \pm 38.27
HGCA2 EpCAM ^{High}	66.06 (6) \pm 14.28		
HGCA3 EpCAM ^{High}	65.85 (6) \pm 6.21		
LGCA1 EpCAM ^{High}	114.69 (4) \pm 42.55	LGCA EpCAM ^{High} 88.10 (6.3) \pm 49.22	
LGCA2 EpCAM ^{High}	33.20 (7) \pm 5.93		
LGCA3 EpCAM ^{High}	47.88 (8) \pm 7.35		

Primary cell lines derived from 3 LGCA and 3 HGCA tissues samples were sorted into EpCAM^{High} and EpCAM^{Low} fractions. Tumoursphere diameter was measured in µm. Diameter values in column 1 represent the average maximum diameter of all measured tumourspheres across technical replicates for each biological replicate. Column 2 shows the average maximum sphere size across the 4 conditions (HGCA EpCAM^{Low}, LGCA EpCAM^{Low}, HGCA EpCAM^{High} and LGCA EpCAM^{High}). Column 3 displays the average sphere diameters for all EpCAM^{Low} cells and EpCAM^{High} cells.

The EpCAM^{High} and EpCAM^{Low} cells derived from the LGCA1 sample (Fig 5.3A,B) formed tumourspheres (114.69 µm and 60.25 µm, respectively), but the EpCAM^{High} cells derived from LGCA2 and the EpCAM^{Low} cells derived from LGCA2 and LGCA3 did not reach the size threshold (33.32 µm, 32.35 µm and 32.31 µm, respectively). Tumourspheres formed by LGCA3-derived EpCAM^{High} cells were just below the diameter threshold (47.88 µm), and this was regarded as a negative result. Both the EpCAM^{High} and EpCAM^{Low} cells from all three HGCA-derived primary cell lines attained a positive result for tumoursphere formation (Fig 5.3C,D).

Interestingly, the EpCAM^{High} and the EpCAM^{Low} cells that were capable of forming tumourspheres reached the size threshold after a comparable number of days. However, the average size of the tumourspheres produced by EpCAM^{High} cells tended to be larger (80.86 µm after 5.5 d) than those produced by EpCAM^{Low} cells (54.74 µm after 5.7 d).

HGCA-derived cells produced tumourspheres that reached an average maximum diameter of 71.42 µm after 4.7 d for EpCAM^{High} cells and 60.11 µm after 3.0 d for EpCAM^{Low} cells. In comparison, when LGCA cells were capable of producing tumourspheres, they reached an average maximum diameter of 60.11 µm after 6.3 d for EpCAM^{High} cells and 45.92 µm after 8.3 d for EpCAM^{Low} cells. Overall, both the EpCAM^{High} and EpCAM^{Low} cells from 1 of 3 LGCA and all 3 HGCA were capable for forming tumourspheres that reached the size threshold.

Tumourspheres derived from 2 cell lines, 1 from a LGCA and the other from a HGCA, were passaged to confirm formation of true tumourspheres as opposed to cell clusters (Table 5.2). The EpCAM^{High} and EpCAM^{Low} cells derived from the LGCA sample both produced tumourspheres in their second passage. This took longer for the LGCA-derived EpCAM^{Low} cells in the second passage (53.29 µm after 7 d) than in their first passage (60.25 µm after 6 d).

For the LGCA-derived EpCAM^{High} cells, the difference in time taken to reach the maximum tumoursphere diameter was even more pronounced. On average, tumourspheres reached a maximum diameter of 114.69 µm after 4 d in the first passage versus 81.09 µm after 9 d in the second passage.

The HGCA sample formed tumourspheres which reached the diameter threshold for both the EpCAM^{High} and EpCAM^{Low} cells (53.02 µm and 55.49 µm, respectively). However, this took considerably longer in the second passage (5 d and 6 d, respectively) than the first passage (2 d and 1 d, respectively).

Table 5.2. Data from passaged tumourspheres

Sample ID	Average diameter at maximum, μm (Days to maximum diameter)
LGCA1 EpCAM ^{High}	81.09 (9) \pm 21.65
LGCA1 EpCAM ^{Low}	53.29 (7) \pm 4.46
HGCA1 EpCAM ^{High}	53.02 (5) \pm 7.44
HGCA1 EpCAM ^{Low}	55.49 (6) \pm 10.90

Tumourspheres from 1 LGCA and 1 HGCA-derived primary cell line were passaged. Tumourspheres were isolated and separated into single cells to reform tumourspheres. Tumoursphere diameter was measured in μm . Diameter values represent the average diameter of all measured tumourspheres across three technical replicates for each biological replicate.

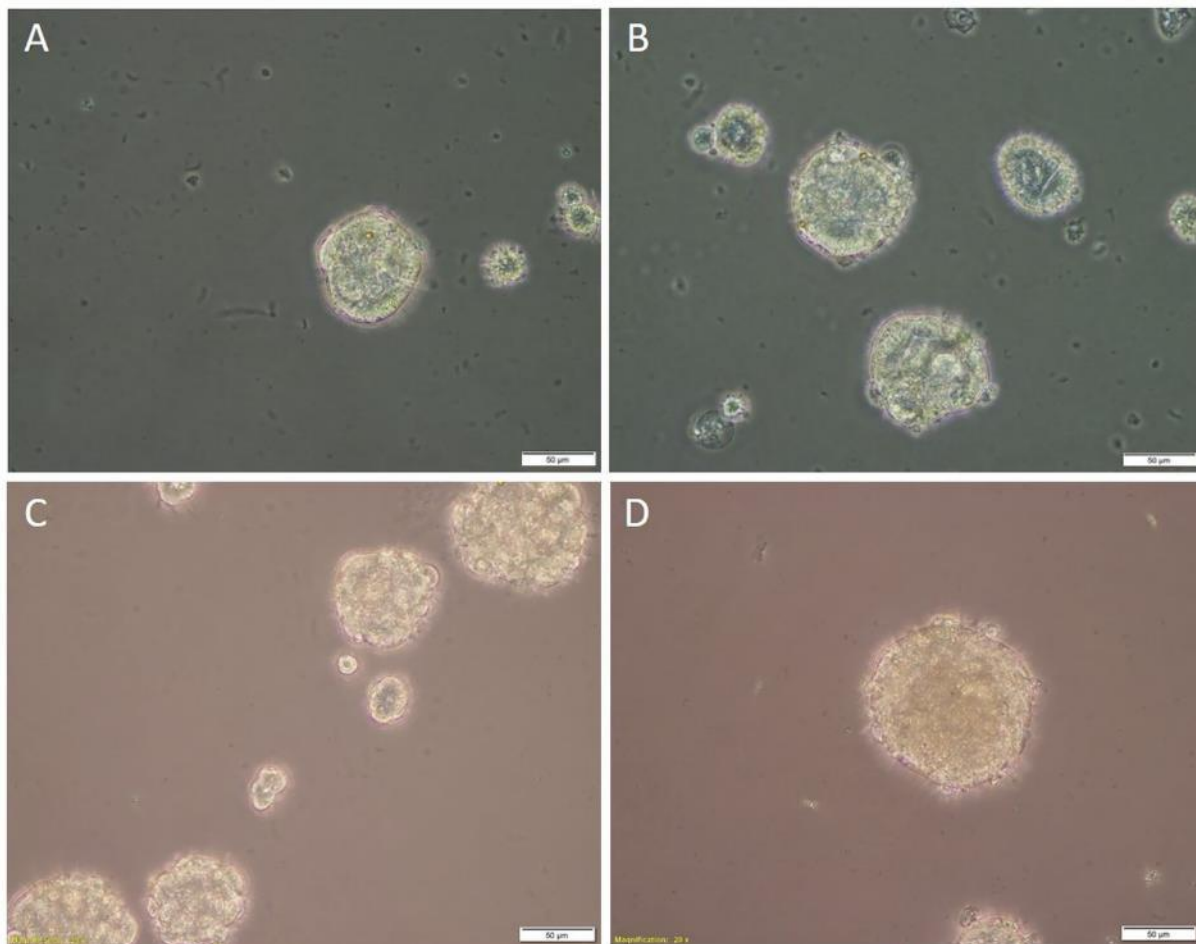


Figure 5.3: Representative images of tumoursphere formation assays. Tumoursphere formation in EpCAM^{Low} (A) and EpCAM^{High} (B) cells from LGCA-derived primary cell lines, and EpCAM^{Low} (C) and EpCAM^{High} (D) cells from HGCA-derived primary cell lines. LGCA (n=3); HGCA (n=3). Original magnification: 400x; scale bar = 50 μm .

5.2.3 CA-derived cells display multilineage differentiation capacity

To further explore the stem cell functionality of these CA-derived cells, differentiation down the three embryonic germ lineages – mesoderm, endoderm and ectoderm – was induced.

For mesodermal differentiation, cells were incubated in StemPro® Osteogenesis media, and after up to 14 d in culture they were fixed and stained with Alizarin Red dye which binds specifically to calcium at pH 4.2. All 3 LGCA and 3 HGCA cell lines showed positive staining to varying degrees (Fig 5.4), suggesting the capacity to differentiate down the mesodermal lineage.

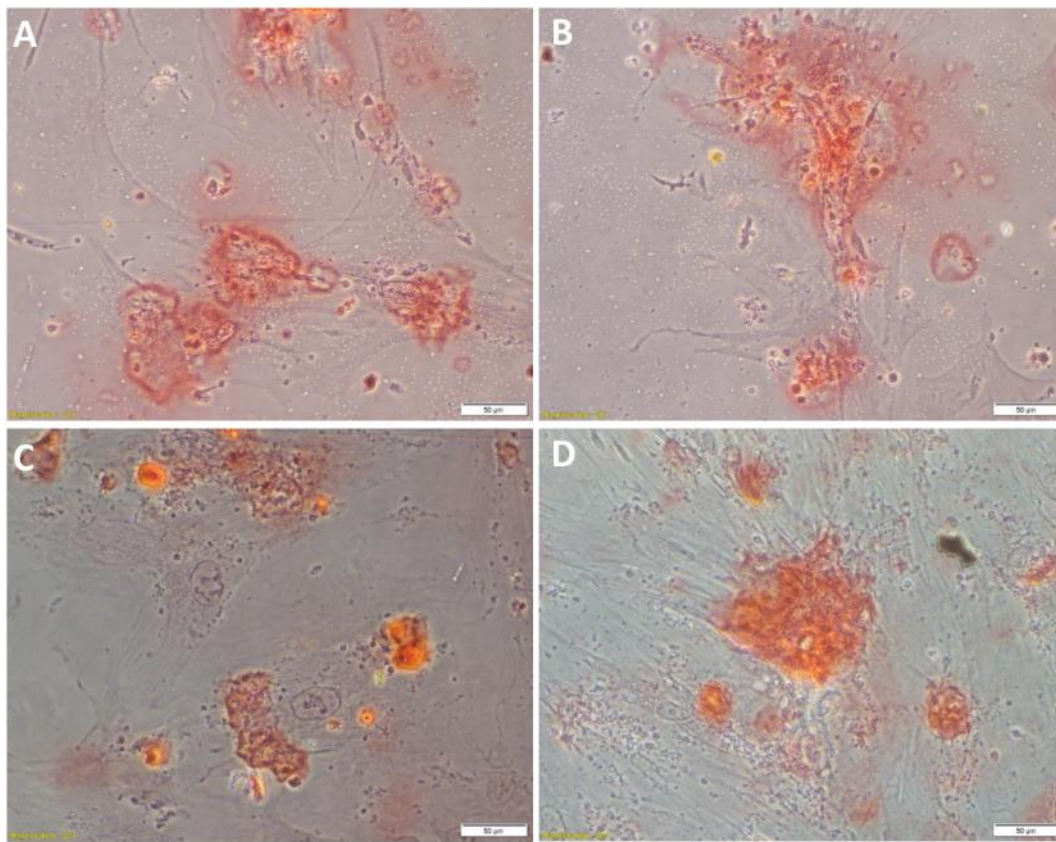


Figure 5.4: Representative images of Mesoderm differentiation. Primary tumour-derived EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA cells, and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA cells, were grown in chambered slides for differentiation assays. Alizarin Red stain (pH 4.2) was used to detect calcium deposits. LGCA (n=3); HGCA (n=3). Original magnification: 400x; scale bar = 50 µm.

Endodermal differentiation was induced using StemXVivo® Endoderm media and confirmed using an antibody against SOX17. When imaged by confocal microscopy, all 3 LGCA and 3 HGCA cell lines

expressed SOX17 following incubation with the differentiation media, confirming the ability of these cells to differentiate down the endodermal lineage (Fig 5.5). These cells also expressed SOX17 when grown in regular culture media, but at much lower levels than when in differentiation media (Appendix Fig A5.12). This is consistent with the endodermal origins of colon.

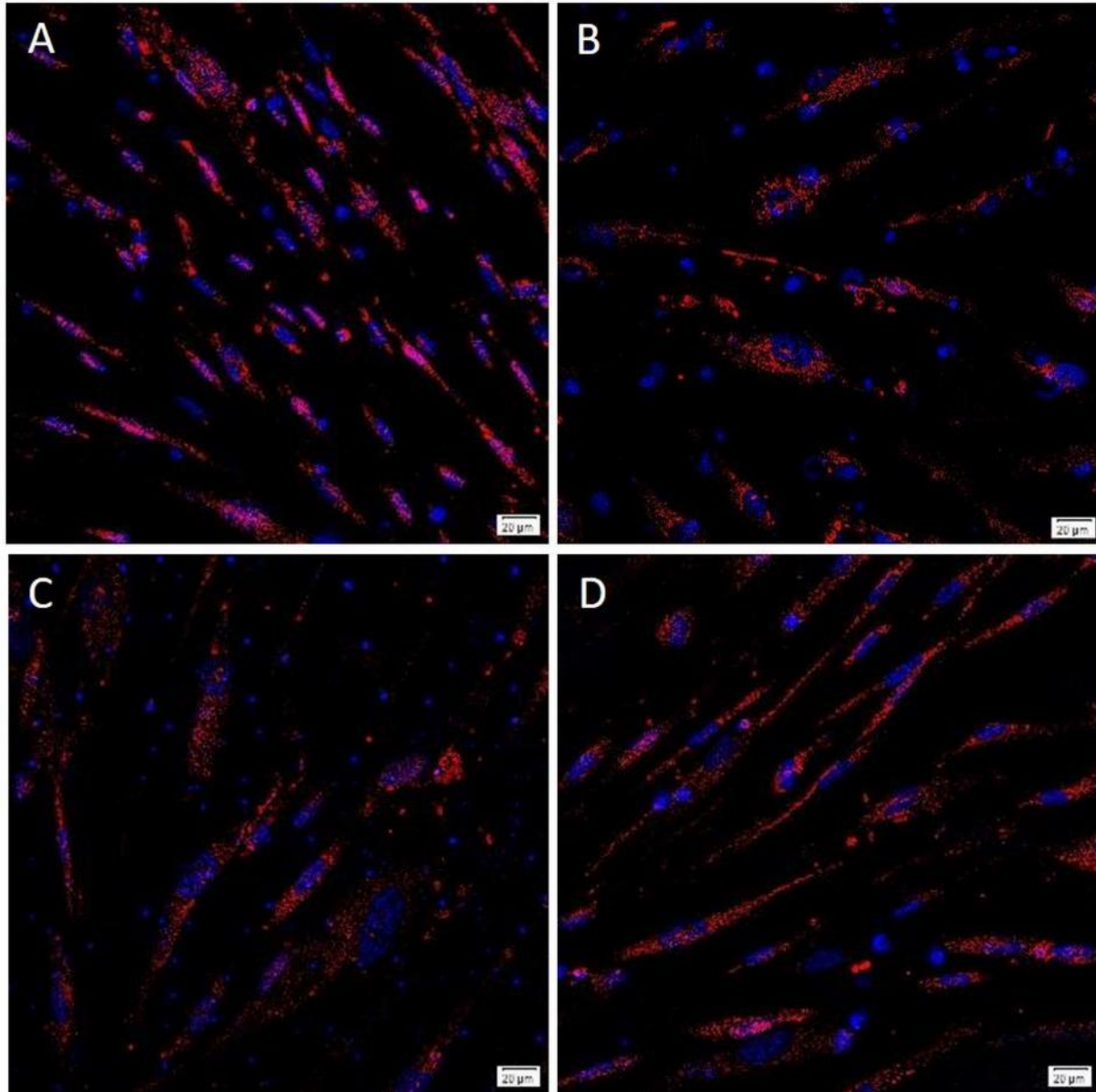


Figure 5.5: Representative images of Endoderm differentiation. Primary tumour-derived EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA cells, and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA cells, were grown in chambered slides for differentiation assays. NorthernLights™ fluorescent secondary antibody (red) detected the rabbit anti-SOX17 primary antibody. LGCA (n=3); HGCA (n=3). Original magnification: 400x; scale bar = 20 µm.

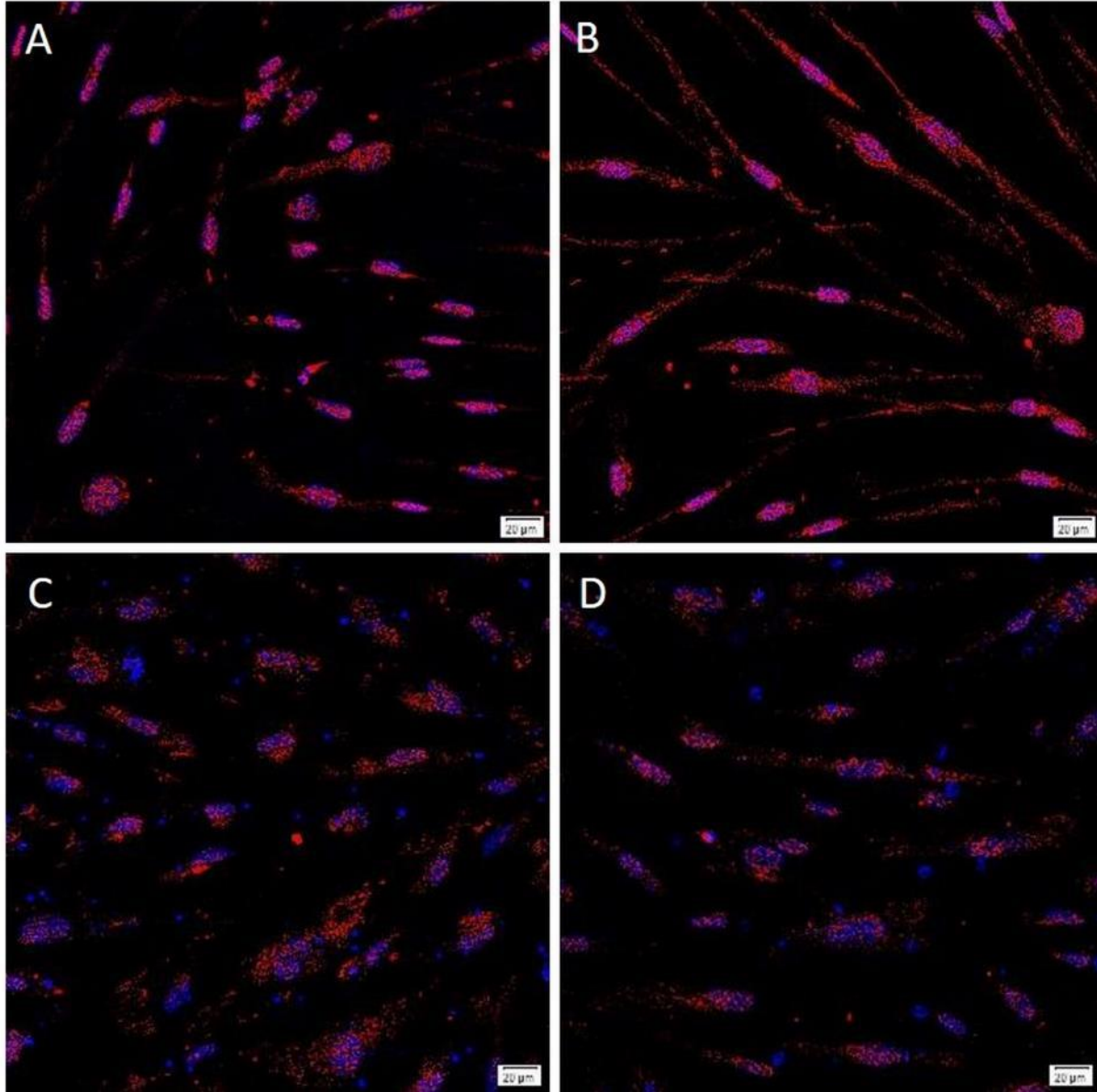


Figure 5.6: Representative images of Ectoderm differentiation. Primary tumour-derived EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA cells, and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA cells, were grown in chambered slides for differentiation assays. NorthernLights™ fluorescent secondary antibody (red) detected the rabbit anti-Otx2 primary antibody. LGCA (n=3); HGCA (n=3). Original magnification: 400x; scale bar = 20 μm.

The StemXVivo® Ectoderm kit was used to induce ectodermal differentiation, using Otx2 as an ectodermal marker. Surprisingly, all control cells grown in regular media (Appendix Fig A5.15) expressed Otx2 at similar levels to the cells incubated in the differentiation media (Fig 5.6).

5.2.4 RT-qPCR

To investigate the mRNA expression of iPSC genes by the 2 sorted cell subpopulations, RNA was extracted from cells and analysed by RT-qPCR, using the pooled results from 4 NC tissue samples as a reference for relative abundance (Fig 5.7).

OCT4 mRNA was detected in all CA-derived primary cell lines and was found to be more abundant in the EpCAM^{Low} cells than the EpCAM^{High} cells derived from all 3 LGCA samples. However, it was more abundant in the EpCAM^{High} cells than the EpCAM^{Low} cells derived from 2 of the 3 HGCA samples (Fig 5.7A). Surprisingly, all cells had lower levels of *OCT4* expression than the NC reference tissues.

SOX2 mRNA was present in the EpCAM^{Low} and EpCAM^{High} cells from 1 LGCA-derived cell line at higher levels than the NC reference. However, besides low levels being detected in 1 HGCA-derived EpCAM^{Low} cell line, *SOX2* was not detected in any other cells (Fig 5.7B).

NANOG was below the detection threshold in all cell lines but was detected at low levels in all 4 NC tissue samples (Fig 5.7C).

KLF4 mRNA was more abundant in LGCA-derived cells than HGCA-derived cells, with detection above the threshold in all EpCAM^{Low} cells but only in the EpCAM^{Low} and EpCAM^{High} cells from 1 HGCA cell line (Fig 5.7D). This was in line with data from other studies, with the highest levels of *KLF4* expected in NC and the lowest in HGCA.

Of the 5 iPSC genes, *c-MYC* was the most highly expressed and was seen in all EpCAM^{Low} and EpCAM^{High} cells (Fig 5.7E). The EpCAM^{Low} and EpCAM^{High} cells from all 3 LGCA-derived and 1 HGCA-derived cell lines displayed a significant increase in expression relative to the NC reference, while the remaining 2 HGCA-derived EpCAM^{Low} and EpCAM^{High} cells showed no significant change.

Overall, HGCA-derived cells had higher expression of *OCT4* and lower expression of *KLF4* compared with LGCA-derived cells, however, LGCA-derived cells had higher expression of *c-MYC*. *NANOG* was not detected in any of the cells, and *SOX2* expression was very low in all but 1 cell line.

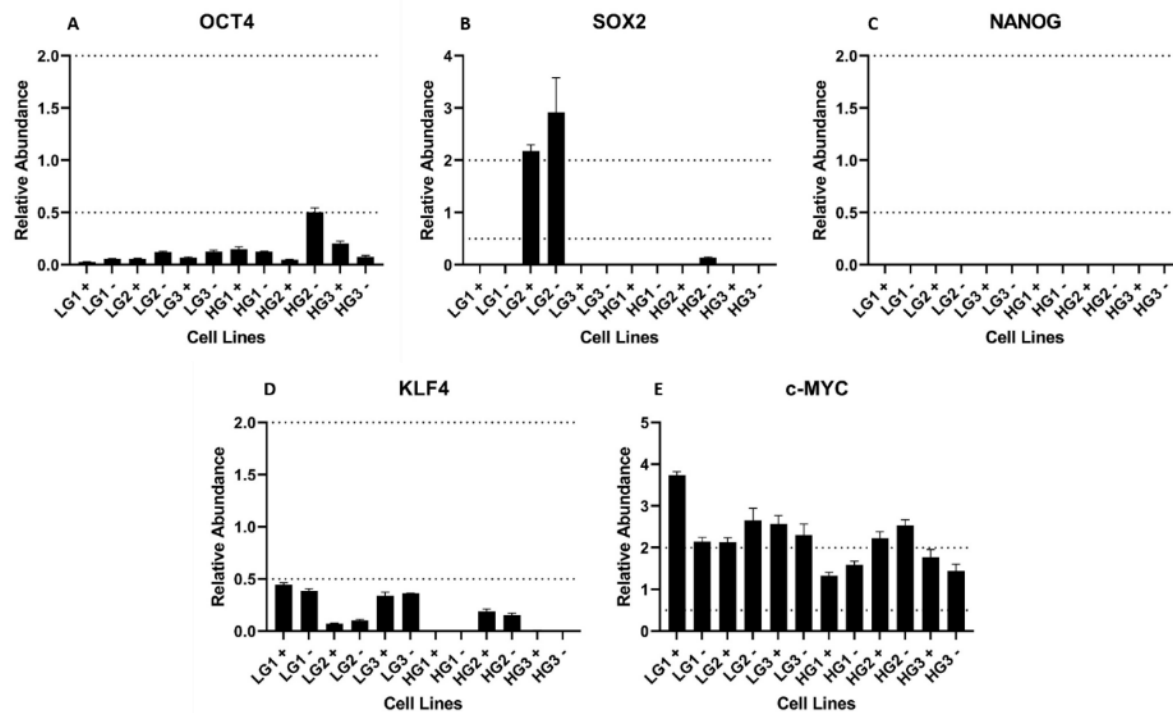


Figure 5.7: RT-qPCR data from EpCAM^{Low} and EpCAM^{High} cells. RNA was extracted from EpCAM^{Low} and EpCAM^{High} cells derived from 3 LGCA and 3 HGCA cases, and RT-qPCR was performed to measure the mRNA levels of *OCT4* (A), *SOX2* (B), *NANOG* (C), *KLF4* (D) and *c-MYC* (E). The abundance of each marker was measured relative to the pooled data generated from 4 NC tissues, all run separately as triplicates in each PCR reaction. The average relative abundance from triplicates are displayed with error bars representing standard deviation (SD). LGCA (n=3); HGCA (n=3).

5.2.5 Western blotting

The protein products of the 5 iPSC genes were investigated by WB (Fig 5.8A-E) and analysed semi-quantitatively using densitometry (Fig 5.9).

OCT4, SOX2, KLF4 and c-MYC proteins were detected in both the EpCAM^{High} and EpCAM^{Low} cells from all 3 HGCA-derived primary cell lines.

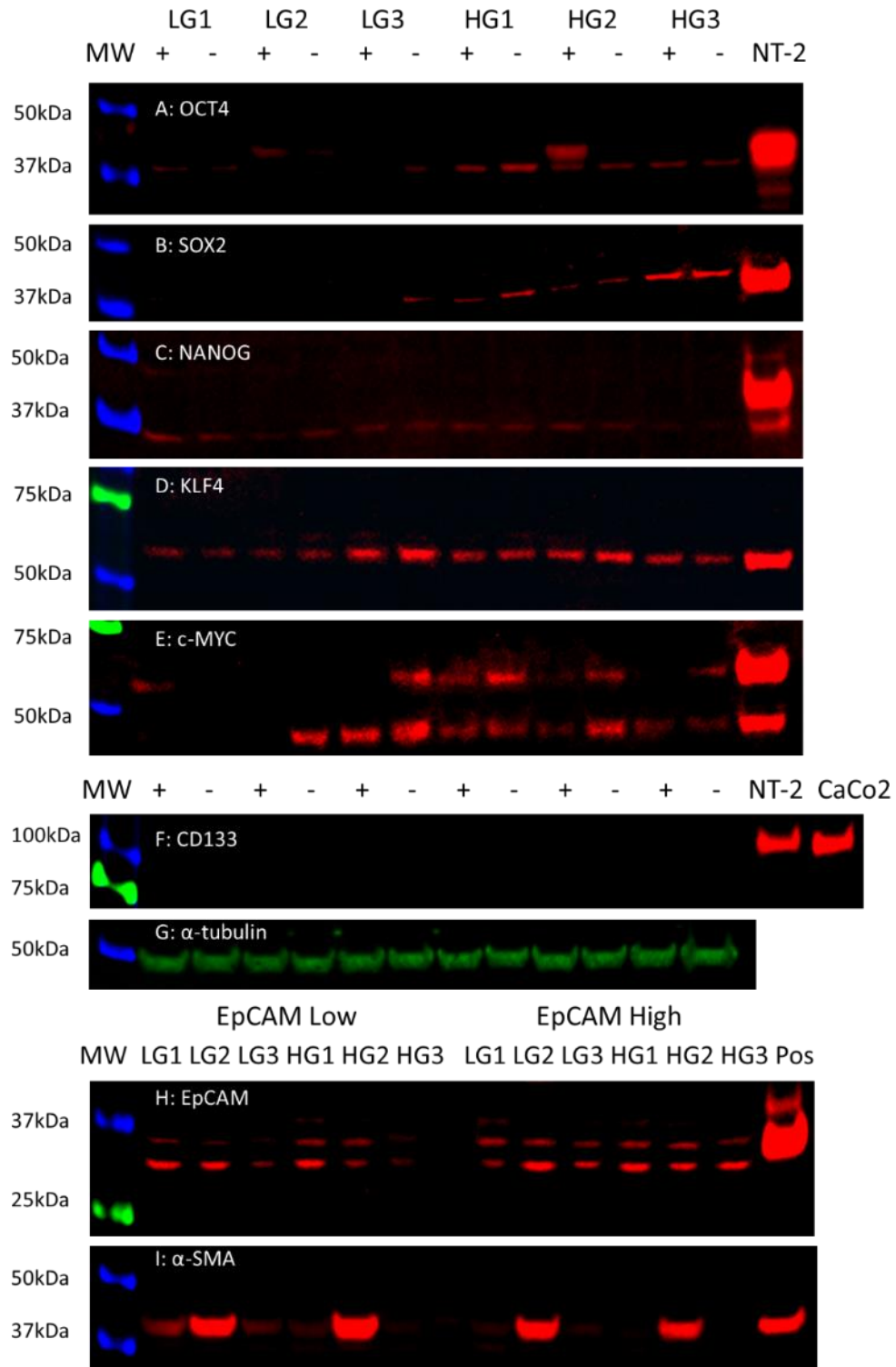
OCT4 (Fig 5.8A) was present in the EpCAM^{High} and EpCAM^{Low} cells derived from 2 of the LGCA samples, and only the EpCAM^{Low} cells derived from the third sample. Similarly, SOX2 (Fig 5.8B) was detected in the EpCAM^{Low} cells derived from 1 LGCA sample, but it was undetectable in the other 2 LGCA samples. Very faint bands corresponding to the NANOG protein (Fig 5.8C) were seen in EpCAM^{High} cells derived from 2 LGCA and 2 HGCA samples, and EpCAM^{Low} cells derived from 3 LGCA and 2 HGCA samples. KLF4 (Fig 5.8D) was the only marker to be detected in both the EpCAM^{High} and EpCAM^{Low} cells from all

LGCA and HGCA samples. c-MYC (Fig 5.8E) was present in the EpCAM^{High} and EpCAM^{Low} cells derived from 1 LGCA sample. Of the other 2 LGCA samples, c-MYC was only expressed in EpCAM^{High} cells derived from 1 sample and only in the EpCAM^{Low} cells derived from the other sample. Overall, there was a greater amount of c-MYC protein found in EpCAM^{Low} cells than EpCAM^{High} cells.

When comparing the levels of marker expression between LGCA and HGCA cell lines, only SOX2 showed a statistically significant increase (Fig 5.9C; $p=0.0037$). The difference in OCT4 expression between LGCA and HGCA cell lines was noticeable but was just below statistical significance (Fig 5.9A; $p=0.07$). There were no statistically significant differences in marker expression between EpCAM^{High} and EpCAM^{Low} cells (Fig 5.9B,D,F,H,J).

CD133 was not detected by WB in any of the cells, but a strong band was present in the NTERA-2 and CaCo2 positive controls (Fig 5.8F).

Figure 5.8: Western blotting for iPSC and stem cell markers in cells (page 95). Protein extractions from EpCAM^{High} (+) and EpCAM^{Low} (-) cells derived from 3 LGCA and 3 HGCA cases were probed for OCT4 (A; 40 kDa), SOX2 (B; 40-43 kDa), NANOG (C; 37-40 kDa), KLF4 (D; 54 kDa), c-MYC (E; 42 kDa and 57 kDa) and CD133 (F; 97 kDa). NTERA-2 cells were used as the positive control for all iPSC markers. NTERA-2 and CaCo2 cells were used as the positive controls for CD133. α -Tubulin was used as a loading control (G; 50 kDa). EpCAM^{High} and EpCAM^{Low} cells were also probed for their expression of EpCAM (H; bands from ~30-40kDa) and α -SMA (I; 42kDa). HepG2 and 3T3 cells were used as the positive controls for EpCAM and α -SMA, respectively.



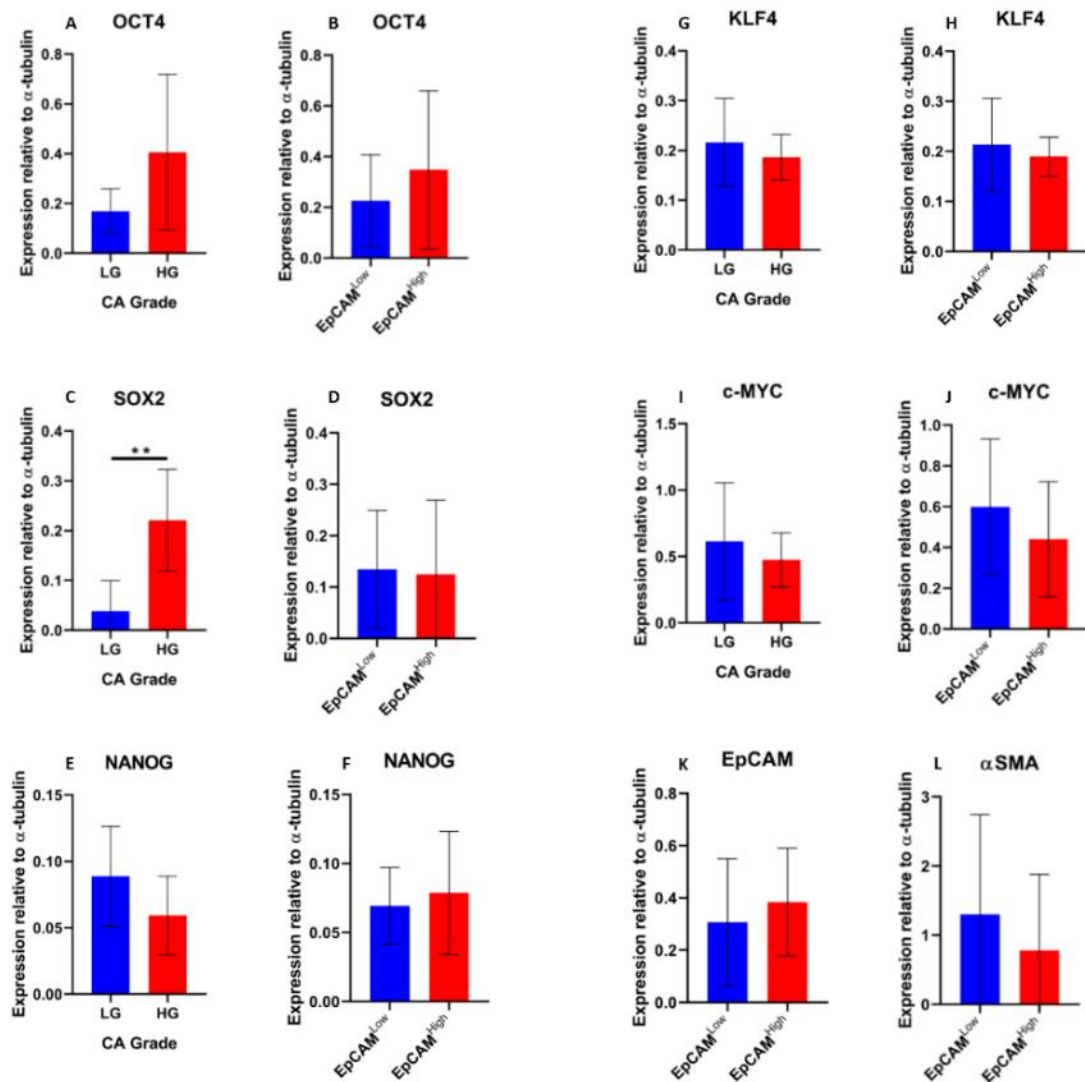


Figure 5.9: Densitometry performed on western blot. Densitometry provided semi-quantitative data for protein abundance. The intensity values of all LGCA cell lines (both EpCAM^{High} and EpCAM^{Low}) and all HGCA cell lines (both EpCAM^{High} and EpCAM^{Low}) were combined and the average intensity calculated, and these are shown for OCT4 (A), SOX2 (C), NANOG (E), KLF4 (G) and c-MYC (I). The intensity values of all EpCAM^{Low} cells (both LGCA and HGCA) and all EpCAM^{High} cells (both LGCA and HGCA) were combined and the average intensity calculated, and these are shown for OCT4 (B), SOX2 (D), NANOG (F), KLF4 (H), c-MYC (J), EpCAM (K) and α-SMA (L). Individual intensity values were normalised against the loading control α-tubulin before being combined and averaged. Error bars show standard deviation (SD).

To assess the efficiency of the EpCAM sort and the characteristics of the resultant subpopulations of cells, protein extracts from all 6 EpCAM^{Low} and 6 EpCAM^{High} cell lines were probed for EpCAM (Fig 5.8H) and α-SMA (Fig 5.8I). All the cells expressed EpCAM, however the corresponding bands for EpCAM were

stronger in the EpCAM^{High} cells than the EpCAM^{Low} cells when assessed using densitometry (Fig 5.9K). The observed bands seemingly correspond to un-glycosylated EpEX (the extracellular domain of EpCAM; 30kDa), glycosylated EpEX and un-glycosylated full-length EpCAM (~35kDa), and glycosylated full-length EpCAM (40kDa in HepG2). The same cells were then probed using an antibody against α -SMA, a stromal marker, to see whether there might be an inverse relationship between the expression levels of EpCAM and α -SMA. This appeared to be the case, with stronger staining for α -SMA in EpCAM^{Low} cells than in EpCAM^{High} cells (Fig 5.9L).

5.3 Discussion

Based on results demonstrating the expression of iPSC markers in primary LCGA and HGCA tissue samples (Chapter 4), this chapter assessed whether primary cell lines derived from CA tissue samples contain cells with stem cell characteristics, including tumoursphere formation and multilineage differentiation.

When induced to differentiate down the three embryonic lineages, these CA-derived cells were capable of undergoing endodermal, ectodermal and mesodermal differentiation, as evidenced by their expression of SOX17 (endoderm) and Otx2 (ectoderm), and formation of calcium deposits (mesoderm/osteogenesis) which are specifically bound by Alizarin Red dye at pH 4.2. This was validated by comparing with negative controls in which the cells were grown in their regular culture media, as well as by the omission of the primary antibody. However, somewhat unexpectedly, cells were found to be widely positive for Otx2 when grown in both the ectoderm differentiation media and regular media. Notably, there have been some concerns regarding the specificity of Otx2 as a marker for ectoderm (Kuang et al., 2019). The Human Protein Atlas suggests Otx2 is expressed weakly in the colon, specifically by the goblet cells. There is also some evidence suggesting Otx2 is expressed by cells within the renal tubules, bile ducts and seminiferous ducts of the testis, and neuronal and glial cells, as highlighted by the Human Protein Atlas.

In Chapter 4, the expression pattern of EpCAM in NC and CA tissues was assessed by IHC staining, which revealed that EpCAM expression is restricted to epithelial cells in NC and CA tissues. Therefore, the next step was to culture cells derived from these tissues and then isolate the epithelial and stromal subpopulations by sorting using an anti-EpCAM antibody bound to magnetic beads. After 2 passages post-sorting to allow cells to recover, total protein was extracted from EpCAM^{High} and EpCAM^{Low} cells from all CA-derived primary cell lines for WB. When probing for EpCAM, bands were detected at the expected molecular weight in EpCAM^{High} cells, but also in the EpCAM^{Low} cells. Furthermore, when WB was

performed for the stromal marker α -SMA, it was observed that both EpCAM^{Low} and EpCAM^{High} populations expressed α -SMA. However, EpCAM expression was higher in EpCAM^{High} cells than in EpCAM^{Low}, and α -SMA expression was higher in EpCAM^{Low} cells than in EpCAM^{High}. This suggests that there are some phenotypic differences between the two subpopulations resulting from the EpCAM sort.

When growing a primary cell line from a tissue sample, it has been observed that the expression of up to 10% of genes is altered within five passages (E. Neumann et al., 2010). In general, there is a selection pressure which favours an adherent and proliferative phenotype (Januszyk et al., 2015). The expression profile of these cells tends to drive them towards a more robust and stem-like phenotype (Januszyk et al., 2015). In fact, it has been widely observed that cells derived from different tissues all become more similar to each other when grown in culture (Mehrian Shai et al., 2005; Zaitseva et al., 2006). Furthermore, Sandberg and Ernberg (Sandberg et al., 2005) have shown that when gene expression is analysed, cancer tissue samples are more similar to their patient-matched normal tissue samples than to the primary cell strains derived from them. Amongst the genes most commonly upregulated in cell culture relative to their primary tissue source are the adhesion molecules. The CA-derived cells used in this study had been in culture for 7-10 passages at the time of sorting, so it is possible that any EpCAM-negative cells present when establishing the primary cell culture were outcompeted or acquired EpCAM expression before sorting. Alternatively, there could be an emergence of EpCAM expression in the EpCAM^{Low} cells after sorting (Zaitseva et al., 2006). This raises the possibility that most of the cells were already expressing EpCAM and the antibodies on the beads became saturated, leaving some EpCAM positive (or possibly EpCAM^{Low}) cells to be collected in the EpCAM “negative” fraction.

The method of establishing cell lines before banking contributes to selection for certain cell phenotypes. The GMRITB uses an explant method, in which a small piece of tumour tissue is embedded in Matrigel and cells migrate from the tissue into the matrix. Following this, the matrix and tissue piece are dissociated using dispase and pelleted by centrifugation. A range of cell types are present when the pellet is transferred to a culture flask, including red blood cells which are eliminated after the first passage, suggesting that multiple tumour cell types could be present. Indeed, a range of cell morphologies are observed at early passages. However, this method of procurement may favour cells with a migratory and possibly proliferative phenotype at the expense of other cell types, thus not fully reflecting the cellular heterogeneity of the original tumour tissue. This potential limitation was taken into account when analysing our results and is a subject of future work, as is the establishment of tissue type-specific media formulations. Cell culture remains a valuable experimental tool, although when designing an experiment,

the effects of cell culture on gene and protein expression should be considered. These concerns may be mitigated by using cells with a very low passage number for functional work whenever possible, or by using tissue samples when available for assays such as RT-qPCR and WB.

Further evidence of the influence of EpCAM expression was observed in the ICC PSC marker assays and tumoursphere formation assays carried out in this study. The ICC PSC kit was employed as a standardised way to assess the expression of validated pluripotency markers (Abujarour et al., 2013; Boulting et al., 2011; International Stem Cell et al., 2007; Quintanilla et al., 2014; Valamehr et al., 2012). Validated pluripotency markers SSEA4 and TRA-1-60 were expressed by a subset of cells in all cell lines assayed, which also co-expressed OCT4 in 2 of the 6 EpCAM^{Low} and 4 of the 6 EpCAM^{High} cell lines and SOX2 in 4 of the 6 EpCAM^{Low} and 5 of the 6 EpCAM^{High} cell lines. The co-expression of OCT4, SOX2, SSEA4 and TRA-1-60 suggests that a subpopulation of stem-like cells is present within these cultured cell strains. There seemed to be a noticeable difference between EpCAM^{High} and EpCAM^{Low} cells in terms of OCT4 expression, which was seen in 4 of the 6 EpCAM^{High} cells lines but only in 2 of the 6 EpCAM^{Low} cell lines. However, this differed from the results of tissue samples analysed in Chapter 4, in which OCT4 was expressed in the tumour stroma by EpCAM-negative cells. This may indicate selection for, or drive towards, a stem-like phenotype and similarities in gene expression in culture, whereby EpCAM-negative cells may begin to express EpCAM, cells positive for EpCAM begin to express OCT4, or both. Interestingly, OCT4 mRNA levels were higher in LGCA-derived EpCAM^{Low} cells than EpCAM^{High} cells, a result which is consistent with the experiments in tissues, but was lower in the EpCAM^{Low} cells from 2 of the 3 HGCA-derived cell lines than in the EpCAM^{High} cells.

Furthermore, in the tumoursphere formation assays, EpCAM^{High} cells consistently performed better than EpCAM^{Low} cells, suggesting a functional difference between these two subpopulations, possibly due to the expression of EpCAM and the correlation between EpCAM and iPSC marker expression levels. When the average diameters were calculated across the three biological replicates from each condition group, the size of tumourspheres produced by each condition increased in the order of LGCA EpCAM^{Low}, HGCA EpCAM^{Low}, HGCA EpCAM^{High} and LGCA EpCAM^{High}, though the average for LGCA2 EpCAM^{High} cells was much higher than those for LGCA1 and LGCA3, which caused the overall average to be the highest but with a large SD value (Table 5.1). The results presented in Table 5.1 and Fig 5.3 show that there is variation between each of the biological replicates. The size of tumourspheres produced by each cell line was relatively proportional to their vigour of growth as a monolayer in culture, and indeed reflects the inherent heterogeneity that exists between tumours from different individuals. Whilst each

sample is graded as HGCA or LGCA for convenience, in reality tumours are represented on a spectrum which takes into account a wide range of factors and all cases of a given grade will be different to each other in various parameters. However, the large variation between samples, as revealed by the SD values in Table 5.1, suggests that caution must be taken when interpreting the above results and that it would be worthwhile repeating these experiments in the future with a larger sample size (≥ 10).

The functional ability of HGCA-derived cells to form tumourspheres of a larger size than LGCA-derived cells mirrors the WB results, which showed that HGCA-derived cells expressed higher levels of OCT4, SOX2, NANOG and c-MYC proteins than LGCA-derived cells. The iPSC markers, especially OCT4, SOX2 and NANOG, are linked to stem-like characteristics such as tumoursphere formation and maintenance of pluripotency (Amini et al., 2014; Hadjimichael et al., 2015; Munro, Wickremesekera, Peng, et al., 2017), and so it was not surprising to find that the cells which produce larger tumourspheres have higher expression of iPSC markers. It has been suggested that KLF4 expression is inversely correlated with CA tumour grade, with HGCA expressing less KLF4 than LGCA tumours, and the highest expression seen in the NC adjacent to tumours (R. Hu et al., 2011). This was also demonstrated in the PCR results, where two HGCA samples had significantly lower levels of KLF4 mRNA than all 3 of the LGCA samples. WB showed that protein abundance was relatively similar in LGCA and HGCA cell lines. Another interesting example of this is the LGCA2 sample, which when compared to the other two LGCA samples had higher mRNA expression for *OCT4*, *SOX2* and *NANOG* and much lower levels of KLF4 mRNA, perhaps placing it somewhere near the boundary between low-grade and high-grade. There is a well-documented discrepancy between *SOX2* mRNA levels and protein level, whereby *SOX2* mRNA abundance is almost always much lower than the protein levels (Amini et al., 2014; Talebi et al., 2015), which was also observed in these CA-derived primary cell lines.

Due to the financial and time costs associated with animal studies, as well as the ethical implications of such work, alternatives to xenograft and teratoma experiments in animals is the focus of much review (Buta et al., 2013). While they remain valuable and perhaps essential for applications such as safety testing of stem cell therapies, there are a range of *in vitro* tests for assessing pluripotency that negate the need for animal testing. Teratoma assay protocols are often vague and inconsistent, and are not highly standardised and reproducible (Buta et al., 2013). To determine whether a cell population includes pluripotent cells, it is sufficient to employ directed or spontaneous differentiation, tumoursphere formation which can be sustained over multiple passages, and an analysis of pluripotency marker expression (Buta et al., 2013). It has become more acceptable to use markers including OCT4, SOX2,

NANOG, SSEA4 and TRA-1-60 to identify cells which are pluripotent (Kramer et al., 2013; Takahashi et al., 2007; Takahashi et al., 2006; J. Yu et al., 2007). Chapter 8 will explore further the serial passage of tumourspheres formed by CA-derived cell lines.

5.4 Conclusion

These experiments demonstrated that primary cell lines derived from LGCA and HGCA tissue samples are capable of forming tumourspheres, which can be recapitulated upon passaging, and can differentiate down the three embryonic lineages. This supports the presence of a CSC-like pluripotent subpopulation which was postulated in Chapter 4 in the CA tissue samples. The observation of iPSC markers having distinct expression profiles in HGCA-derived and LGCA-derived primary cell lines indicates that it may be possible to use iPSC markers in a prognostic context and to aid in grading of these tumours.

Chapter 6: Changes in the Proteomes of Colon Adenocarcinoma Tissues and Cell Lines Compared to the Normal Colon

6.1 Introduction

Proteomics is the study of all expressed proteins, including their function, abundance, modifications and interactions (Greening et al., 2017; Tyers et al., 2003). It is considered to be the most comprehensive method for the analysis of “almost everything post-genomic” (Tyers et al., 2003). Bottom-up proteomics, or discovery proteomics, is best suited to the high-throughput identification and quantification of proteins in an unbiased manner (Greening et al., 2017). The most effective tool for proteomics is mass spectrometry (MS), which enables the measurement of protein abundance, isoform expression ratios, turnover rate, subcellular localisation, PTMs and protein-protein interactions (Larance et al., 2015). Recent advancements allow for many parameters to be investigated simultaneously. MS is characterised by high sensitivity, throughput, speed and dynamic range (Larance et al., 2015).

For discovery proteomics, a complex protein sample is subjected to tryptic digestion and the resulting tryptic peptides are then separated through a column made of silicon or synthetic beads with grafted aliphatic chains of various lengths (e.g. C18) by high-performance liquid chromatography (HPLC). This allows the separation of the peptides based on their hydrophobicity (Yates et al., 2009). As the sample is eluted from the HPLC column, the peptides in an acidic solution are ionised into positively charged ions in the ion source, situated at the interface of the HPLC and the mass spectrometer, through a process called electrospray ionisation (ESI). These cationic precursor peptides are corralled into the mass spectrometer where they are resolved in the mass analyser (e.g. ion trap, quadrupole, orbitrap) according to their mass-to-charge (m/z) ratios and detected by the detector (e.g. multiplier). The intensities and the corresponding m/z values of the peptide ions detected are recorded with a computer and displayed as a spectrum; this is called an MS scan (precursor peptide ion scan). In the tandem mass spectrometry (MS/MS) mode, the precursor peptide ion of a particular m/z value from the MS scan is selected and fragmented into fragment ions via collision induced dissociation (CID), high-energy collision-induced dissociation (HCD), or electron-transfer dissociation (ETD), and the fragment ions are then resolved, detected and recorded, generating the MS/MS spectrum. Finally, unbiased annotation of MS/MS peaks via a database search is performed to identify and quantify the relative abundance of the proteins in the input sample (Yates et al., 2009). For this project, label-free quantitation LC-MS/MS was performed using

the state-of-the-art Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (ThermoFisher Scientific), which contains a continuous ESI source coupled to Quadrupole, Ion Trap and Orbitrap tribrid mass analysers in the mass spectrometer. The instrumental method was as described in Section 2.6 in the Method chapter. Briefly, the MS scan was detected in Orbitrap and the MS/MS scan in Ion Trap with top 20 most intense ions from the MS scan selected for MS/MS scans, which was the optimal method to give the maximum number of protein identifications.

The aim of this chapter was to investigate proteomic changes in CA relative to NC, and HGCA relative to LGCA, with the hypothesis that proteomic analysis would reveal changes to key processes in CA relative to the NC. MS was performed on the following biological replicates: tumour tissue samples from 4 patients with LGCA and 4 patients with HGCA and the corresponding 8 patient-matched NC samples, as well as primary cell lines derived from these 8 CA tissues plus 1 unmatched NC cell line. Identified proteins were filtered based on statistically significant changes in abundance. The effects of these changes were investigated by looking at the gene ontology terms assigned to the affected proteins.

6.2 Results

6.2.1 Protein identification

The spectral data output from the mass spectrometer was analysed using Proteome Discoverer 2.4 software (ThermoFisher Scientific), firstly using a workflow to identify which proteins were present. The output listed the protein groups identified in each of the 3 technical replicate LC-MS/MS runs for each biological replicate. Each protein group consists of proteins with very high sequence homology, such as different isoforms of a protein, meaning that they are essentially indistinguishable. Throughout this chapter, the protein groups identified by Proteome Discoverer will be referred to as “proteins”.

To assess the reproducibility of the MS runs, the 3 technical replicates from each cell line were compared using Venny. The number of proteins identified in each technical replicate for each of the cell lines are displayed in Table 6.1.

Across the 4 LGCA-derived cell lines, 66% of identified proteins were shared by the 3 technical replicates (Fig 6.1), and this was slightly lower for the 4 HGCA-derived cells at around 64% on average (Fig 6.2).

Table 6.1: Number of proteins identified in each technical replicate LC-MS/MS run for each CA-derived cell line

Biological replicates (cell lines)	Technical replicates			% Shared
LGCA1	3071	3043	3041	65.6%
LGCA2	3086	3111	3079	66.3%
LGCA3	2933	2986	2983	66.0%
LGCA4	2899	2881	2774	65.8%
HGCA1	2386	2465	2096	58.5%
HGCA2	1336	1429	1437	55.5%
HGCA3	3259	3208	-	75.5%
HGCA4	3227	3203	3232	67.9%

LC-MS/MS runs for each cell line biological replicate were repeated 3 times (3 technical replicates). The number of proteins identified in each cell line per technical replicate are displayed above, as well as the percentage of proteins shared between all 3 technical replicates.

Next, the combined lists for each biological replicate were consolidated to give lists of all proteins identified in each of the tissue samples and all proteins identified in each of the cell lines. The number of proteins identified in each biological replicate for cell lines and tissues are displayed in Table 6.2 and Table 6.3, respectively.

Table 6.2: Total number of proteins identified in each CA-derived cell line biological replicate across three LC-MS/MS technical replicate runs

	Biological replicates			
LGCA-derived cell lines	3764	3800	3662	3510
HGCA-derived cell lines	2976	1839	3685	3908

The total number of proteins identified across 3 LC-MS/MS runs (technical replicates) were calculated for each CA-derived cell line (biological replicates) and are displayed above.

Table 6.3: Total number of proteins identified in each CA tissue biological replicate across three LC-MS/MS technical replicate runs

	Biological replicates			
LGNC tissues	2732	2870	1491	2539
HGNC tissues	3324	3031	1447	2186
LGCA tissues	2274	2337	1793	2933
HGCA tissues	2422	3128	3270	3445

The total number of proteins identified across 3 LC-MS/MS runs (technical replicates) were calculated for each CA tissue (biological replicates) and are displayed above.

In this way, a total of 4767 proteins were identified across all tissue samples (4 HGCA, 4 LGCA, 16 NC), and a total of 4711 proteins were identified across cell lines (4 HGCA-derived, 4 LGCA-derived, 1 NC-derived). Complete lists of the proteins identified from tissue samples and from cell lines are available in Supplementary Files 1 and 2, respectively. The similarity of biological replicates for each condition were assessed for the tissues and cell lines. The 4 LGCA-derived cell lines shared 46.2% of identified proteins (Fig 6.3A), while the 4 HGCA-derived cell lines shared 30.6% (Fig 6.3B).

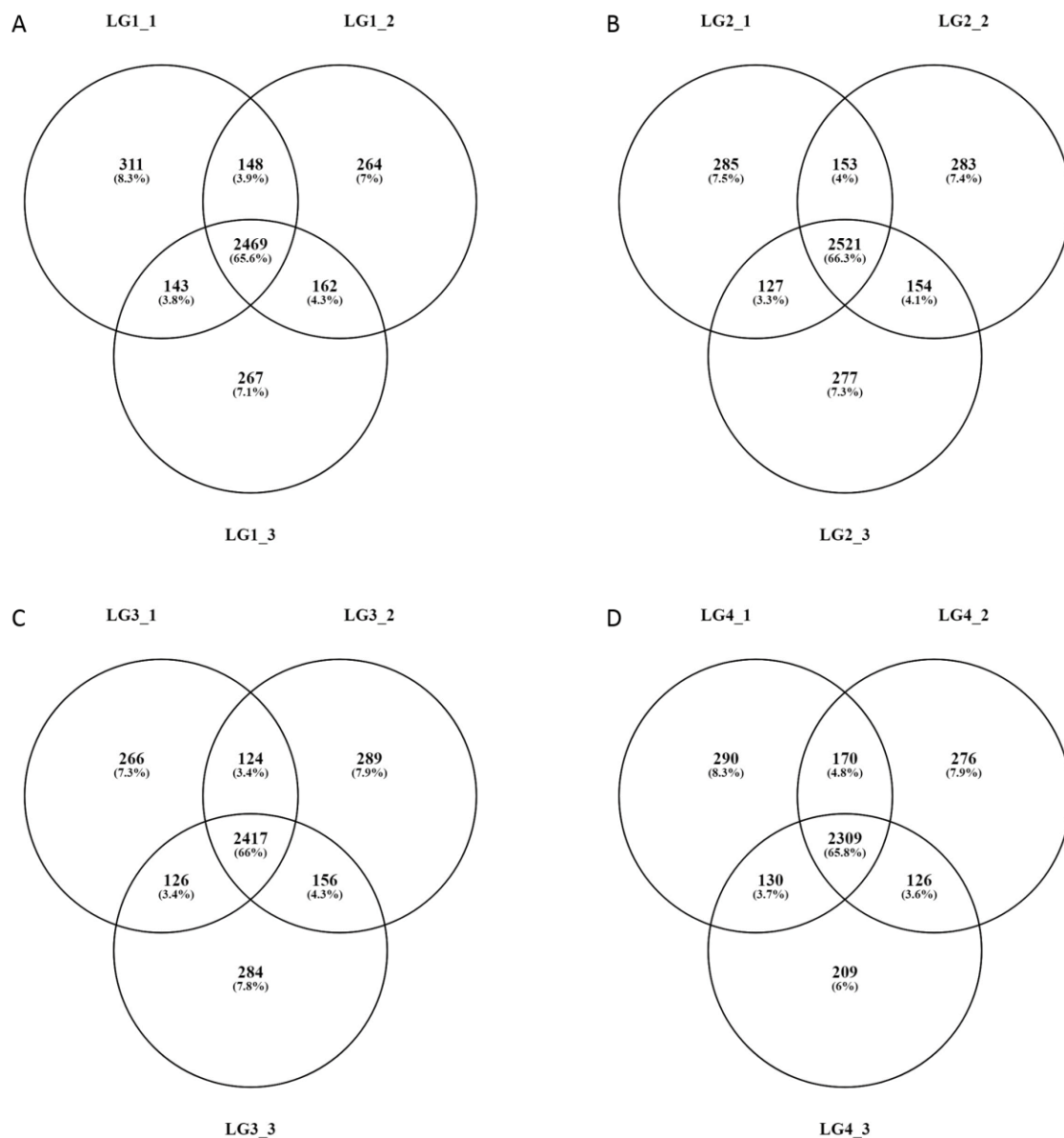


Figure 6.1: The distribution of proteins identified in each LC-MS/MS technical replicate run for each LGCA-derived cell line. Venn's diagrams showing the degree of overlap between the 3 technical replicates for each of the LGCA-derived cell line biological replicates, LGCA 1 (A), LGCA 2 (B), LGCA 3 (C) and LGCA 4 (D).

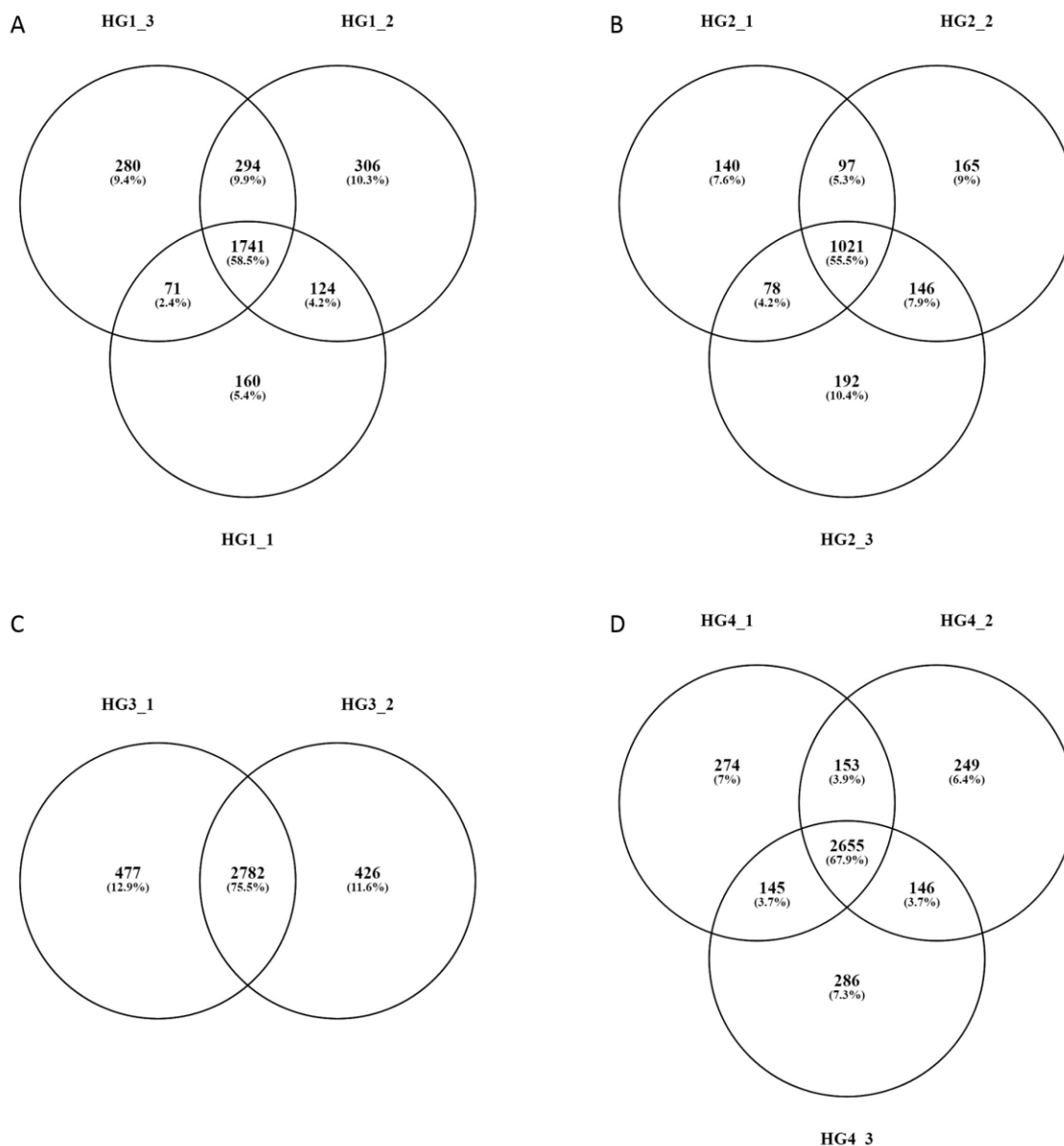
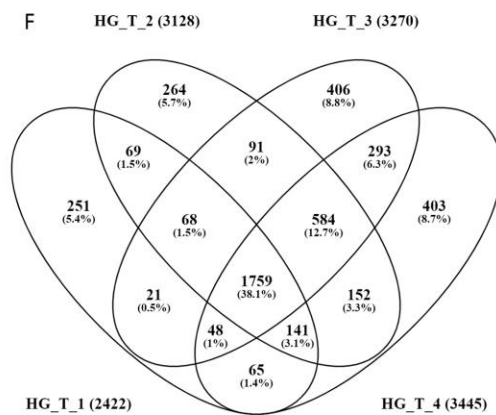
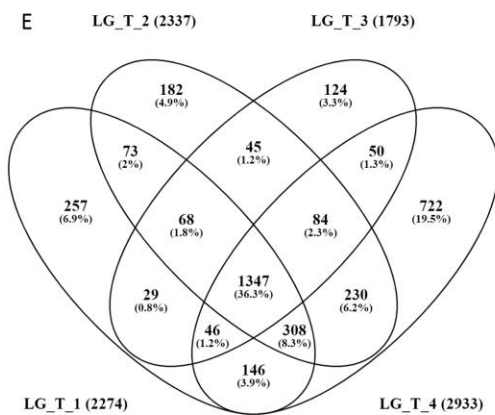
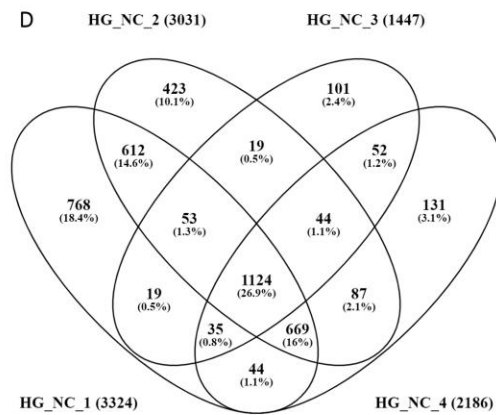
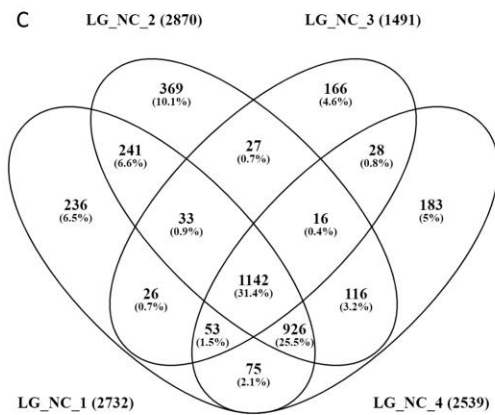
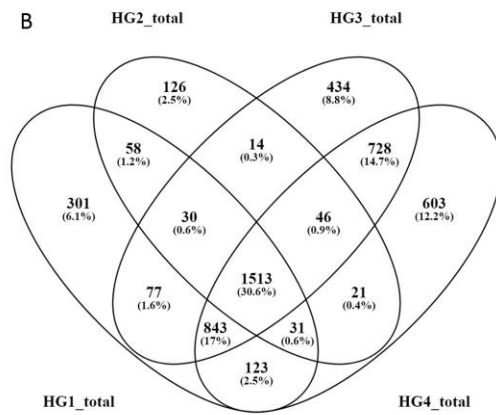
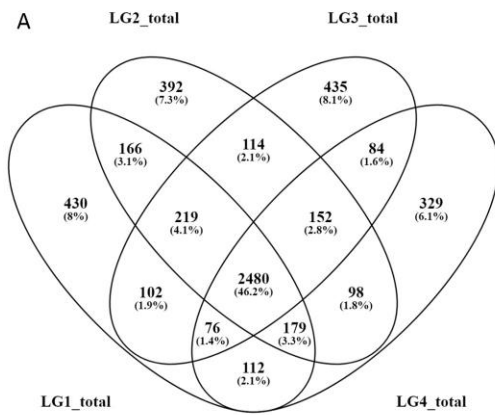


Figure 6.2: The distribution of proteins identified in each LC-MS/MS technical replicate run for each HGCA-derived cell line. Venn's diagrams showing the degree of overlap between the 3 technical replicates for each of the HGCA-derived cell line biological replicates, HGCA 1 (A), HGCA 2 (B), HGCA 3 (C; only 2 technical replicates) and HGCA 4 (D).

Figure 6.3: The distribution of proteins identified in each tissue biological replicate (page 108). Proteins identified by each technical replicate were combined to give a single list of identified proteins per biological replicate. The proteins shared between biological replicates for LGCA-derived cell lines (A) and HGCA-derived cell lines (B), LGNC tissues (C) and HGNC tissues (D), and LGCA tissues (E) and HGCA tissues (F), are shown.



The tissues were more variable than the cell lines. NC tissue sample biological replicates from LGCA and HGCA patients shared 31.4% and 26.9% of identified proteins, respectively (Fig 6.3C,D). The 4 LGCA tissues shared 36.3% of proteins and the 4 HGCA tissues had 38.1% in common (Fig 6.3E,F).

One of the strengths of this project is the proteomic analysis of primary tissue-derived cell lines, a point of difference from most proteogenomic studies which tend to analyse tissue samples and occasionally blood samples. The protein identification data revealed that the cell lines had a higher degree of similarity to each other than the tissue samples, representing a purer population compared to the heterogeneous nature of the tissue.

6.2.2 Label-free quantification

The raw protein identification data were subsequently run through a quantification workflow. Relative abundance was determined for the following comparisons: HGCA/LGCA, HGCA/NC, and LGCA/NC. The criteria for selecting identified proteins for analysis were a fold change of 2 or greater (\log_2 fold change ≥ 1), and a p-value of 0.05 or smaller ($-\log_{10}$ p-value ≥ 1.30103). This produced a list of proteins which were significantly upregulated or downregulated in one group compared to another. Each comparison shows the change in the first group versus the second, so that HGCA/LGCA shows the number of proteins higher and lower in HGCA than in LGCA, for example. A list of all the proteins with significant abundance changes can be found in Supplementary Files 1 (tissues) and 2 (cell lines), and summarised below in Table 6.4.

Table 6.4: Mass spectrometry summary data

Comparison		Number of upregulated proteins	Number of downregulated proteins
CA and NC tissues	HGCA v LGCA	124	137
	LGCA v NC	135	186
	HGCA v NC	148	243
Tissue-derived cell lines	HGCA v LGCA	192	152
	LGCA v NC	639	119
	HGCA v NC	642	142

Number of proteins with significantly different abundances in each group (HGCA, LGCA and NC) relative to each other group, for CA tissues and CA-derived cells.

The largest difference was seen between the LGCA and HGCA cell lines and the NC cell line, which had 639 and 642 significantly upregulated proteins, respectively.

6.2.3 Functional analysis of differentially expressed proteins

Gene Ontology (GO) is used to assign proteins to categories and terms which describe their functional role and distribution within the cell. The GO categories are “biological process” (BP), “molecular function” (MF) and “cellular component” (CC), and each of these categories comprise a list of descriptive terms, such as “RNA processing” (BP), “protein binding” (MF), or “nucleus” (CC). To determine which GO terms (BP, MF, CC) and KEGG pathways were enriched in each condition (NC, LGCA, HGCA), the string-db.org (STRING) online tool (Szkklarczyk et al., 2019) was utilised. This is a database of known protein interactions which produces an interaction map for the input list. From these interactions, the specific functions and processes whose alterations are enriched in each input group can be elucidated based on the GO terms and KEGG pathways assigned to the proteins and their interactions with each other.

Quantification data were visualised as heat maps to assess global changes in protein abundance between NC, LGCA and HGCA. Biological replicates were pooled to give 3 data sets – NC, LGCA and HGCA – for tissues and for cell lines. From these heat maps, various clusters of proteins with differential expression were examined. GO analysis was carried out on the lists of differentially expressed proteins identified from these clusters (Fig 6.4 and Fig 6.10) to gain insight into the collective contribution of these significantly altered proteins. Lists of proteins in each cluster from the heat maps are found in Supplementary Files 3 (tissues) and 4 (cell lines).

The branching above the heat maps (Fig 6.4 and Fig 6.10) revealed that the HGCA and LGCA tissues were more similar to each other than they were to the NC tissues. The branching on the y-axis represents the way in which proteins were clustered, and those which cluster together may have similar functions. Therefore, clusters within the heat maps that displayed differential expression between CA and NC tissues were selected, and GO analyses were performed using STRING to determine the functions enriched within each cluster.

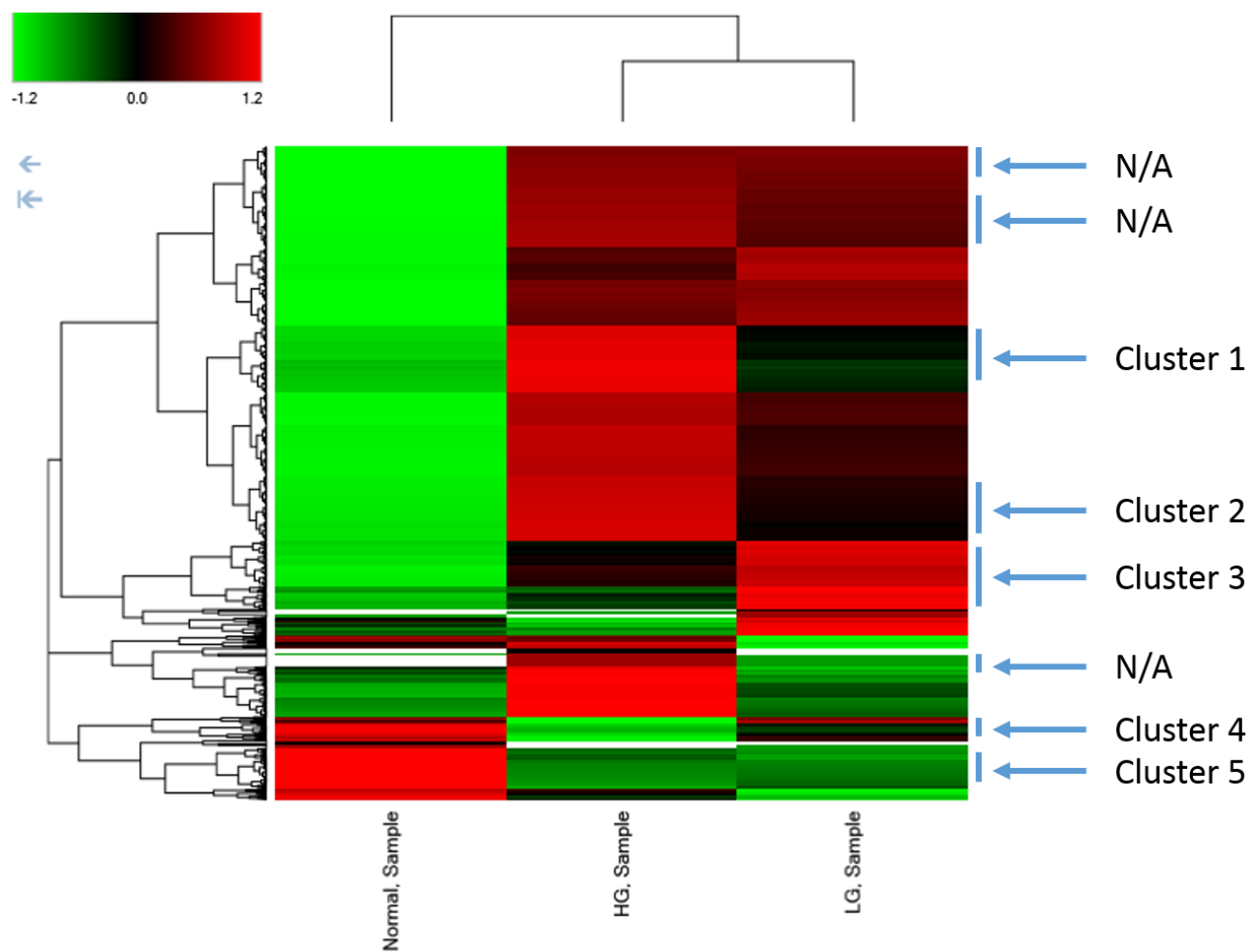


Figure 6.4: Heat map displaying the relative abundances of all quantified proteins in NC, HGCA and LGCA tissues. Heat map produced using Proteome Discoverer 2.4 showing the relative abundances of all proteins identified in tissue samples. Clusters of proteins with differential expression between conditions have been labelled, with 5 of these proceeding to further analysis. Biological and technical replicates were pooled to give 3 categories based on the conditions that were compared: NC, HG and LG. Relative abundances are shown as: green = low, black = moderate, red = high, white = not detected.

There were 8 clusters identified from the tissues heat map (Fig 6.4) which had differential expression across the three conditions. Upon analysis, 3 of these clusters did not show strong enrichment for any specific pathways or processes, and analysis was performed of the remaining 5 clusters which did exhibit significant enrichment for interesting and relevant functions.

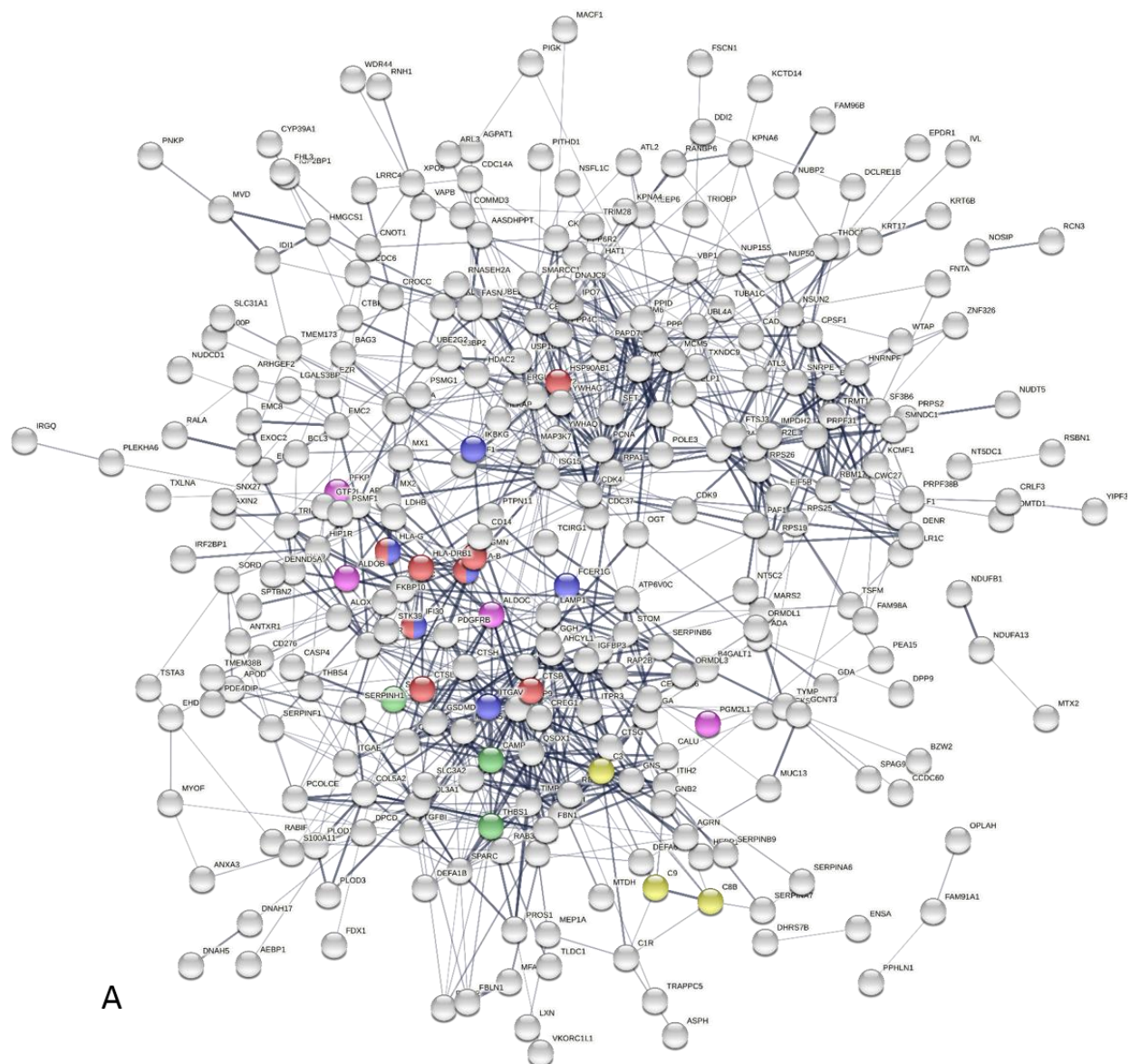


Figure 6.5: Cluster 1 from the tissue sample heat map. A) STRING analysis of proteins from cluster 1, with GO terms or KEGG pathways coloured as follows: red = antigen processing and presentation (KEGG), blue = antigen processing and presentation via MHC class I (BP), green = chronic inflammatory response (BP), yellow = complement activation (BP), purple = canonical glycolysis (BP). B) Table displaying the KEGG pathways that were upregulated in CA tissues relative to NC tissues. p-values subject to Benjamini-Hochberg correction.

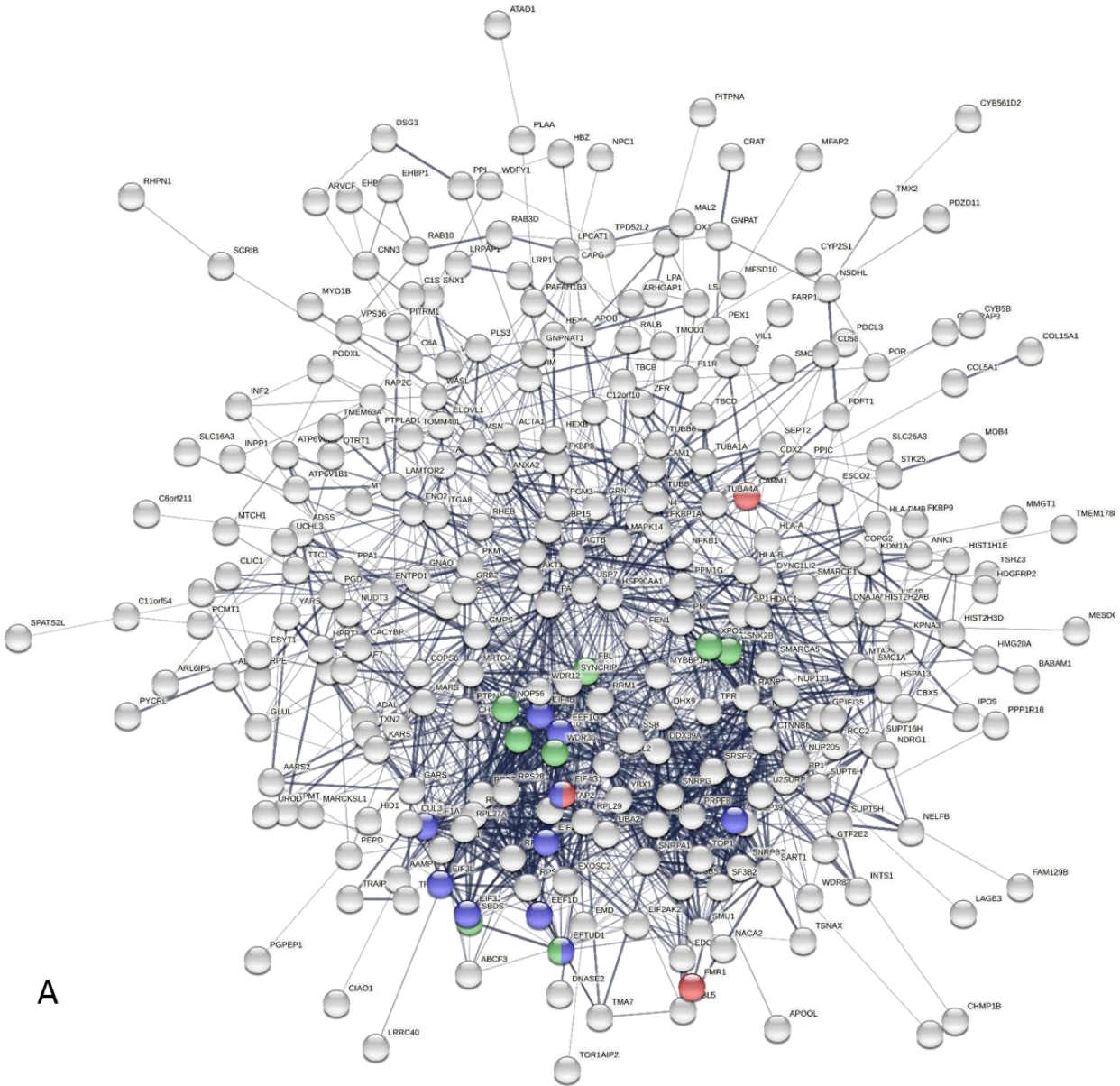


Figure 6.6: Cluster 2 from the tissue sample heat map. A) STRING analysis of proteins from cluster 2, with GO terms or KEGG pathways coloured as follows: red = regulation of RNA binding (BP), blue = translation factor activity, mRNA binding (MF), green = ribosome biogenesis in eukaryotes (KEGG). B) Table displaying the KEGG pathways that were upregulated in CA tissues relative to NC tissues. p-values subject to Benjamini-Hochberg correction.

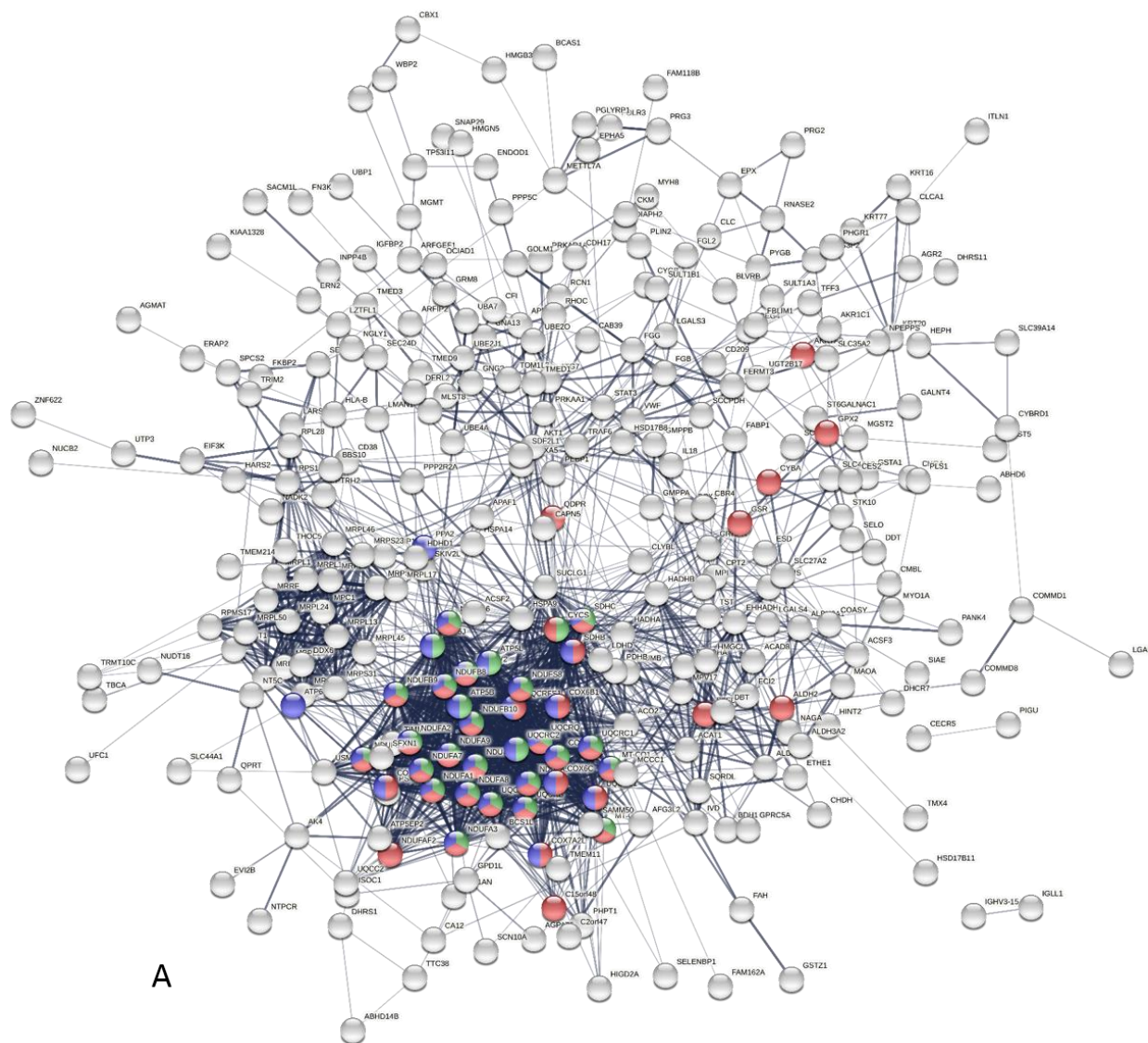
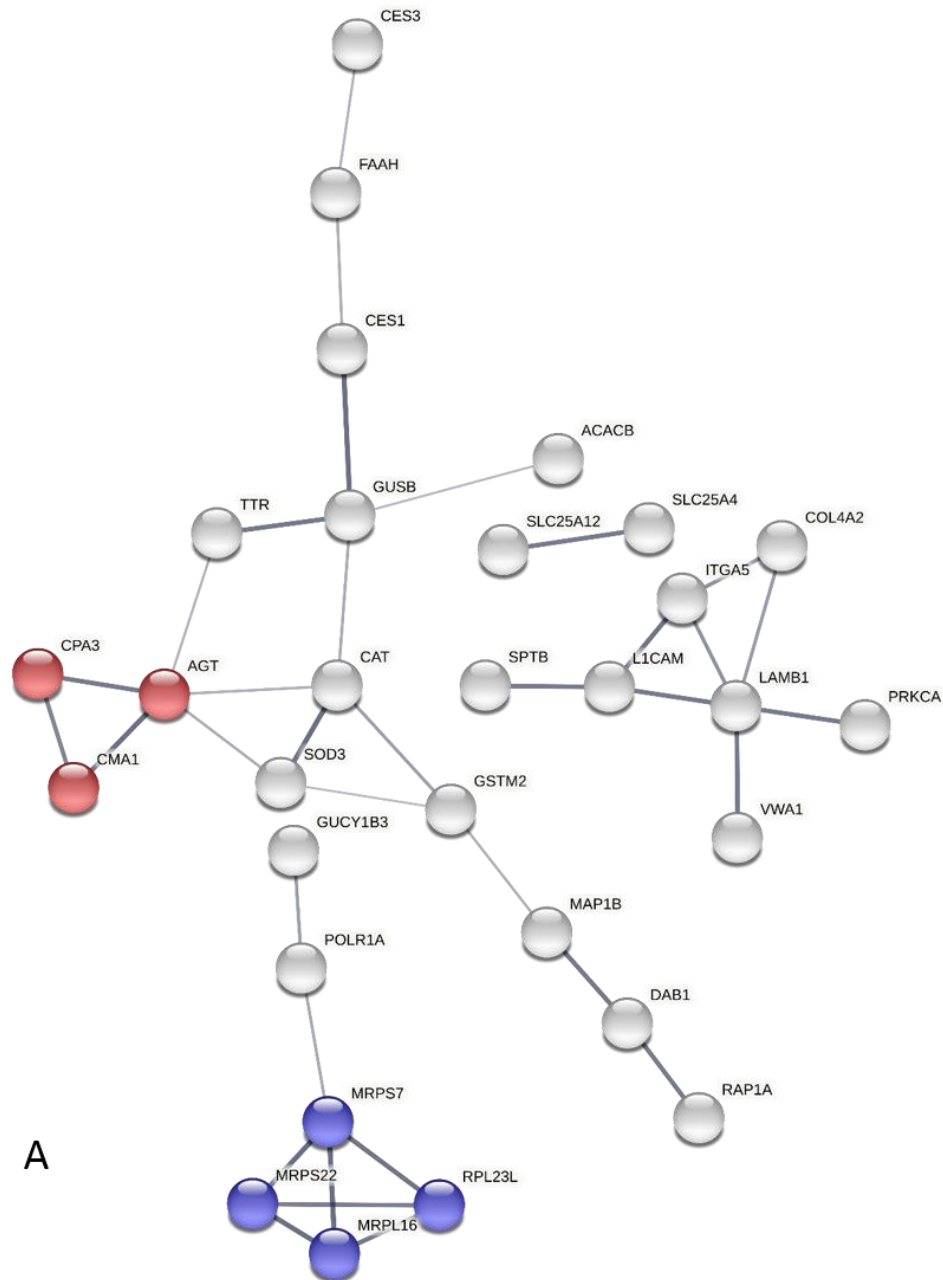


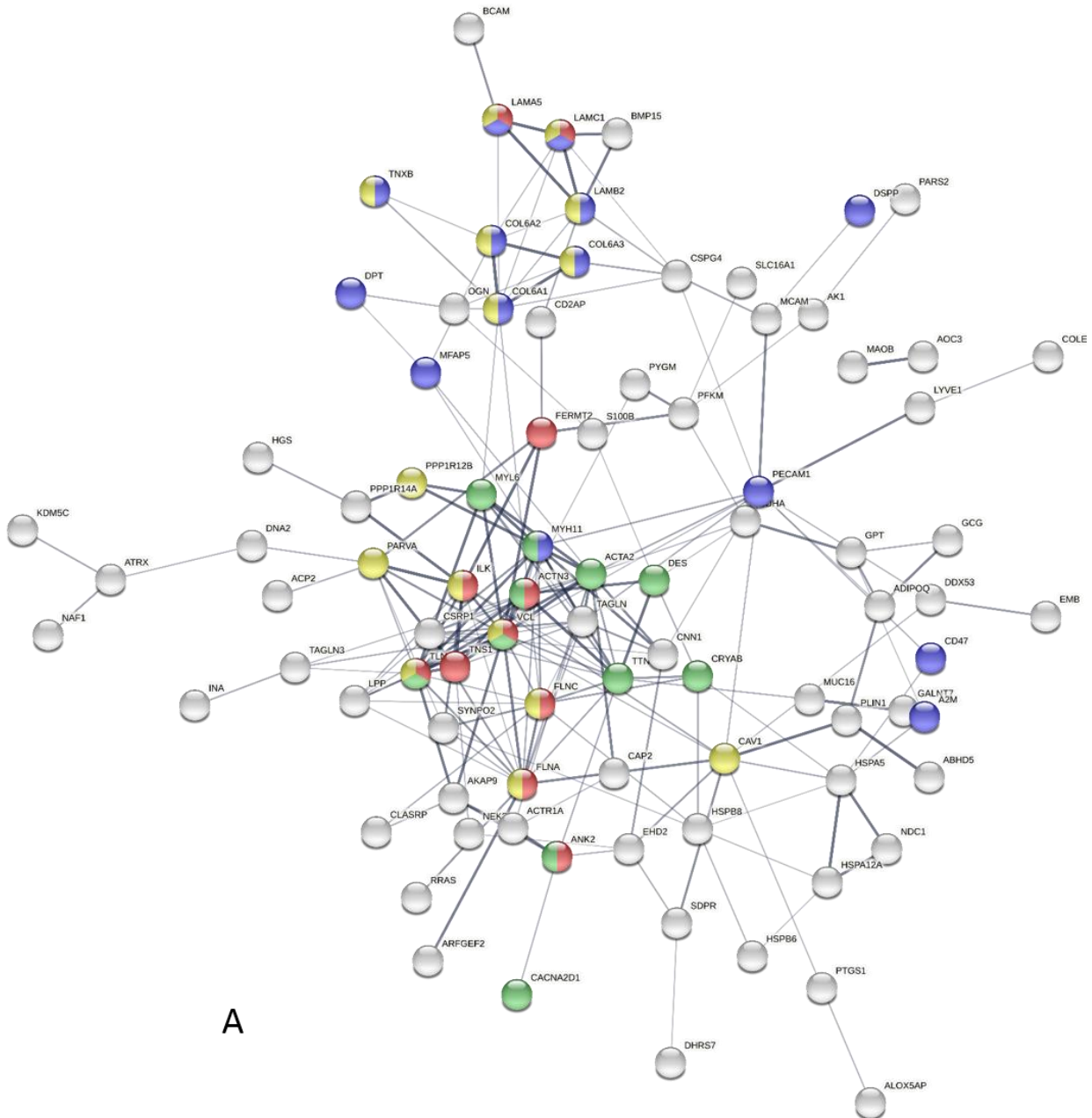
Figure 6.7 Cluster 3 from the tissue sample heat map. A) STRING analysis of proteins from cluster 3, with GO terms or KEGG pathways coloured as follows: green = oxidative phosphorylation (BP), blue = oxidative phosphorylation (KEGG), red = electron transport chain (KEGG). B) Table displaying the KEGG pathways that were upregulated in LGCA tissues relative to NC tissues, and downregulated in HGCA tissues relative to LGCA tissues. p-values subject to Benjamini-Hochberg correction.



A

B) KEGG pathways	Identifier	p-value
Renin-angiotensin system	hsa04614	0.0229

Figure 6.8: Cluster 4 from the tissue sample heat map. A) STRING analysis of proteins from cluster 4, with GO terms or KEGG pathways coloured as follows: red = renin-angiotensin system (KEGG), blue = mitochondrial ribosome (CC). B) Table displaying the KEGG pathways that were downregulated in CA tissues relative to NC tissues. p-values subject to Benjamini-Hochberg correction.



A

B) KEGG pathways	Identifier	p-value
Focal adhesion	hsa04510	5.53×10^{-10}
ECM-receptor interaction	hsa04512	5.08×10^{-6}
Proteoglycans in cancer	hsa05205	0.0433

Figure 6.9: Cluster 5 from the tissue sample heat map. A) STRING analysis of proteins from cluster 5, with GO terms or KEGG pathways coloured as follows: red = cell junction assembly (BP), blue = ECM organisation (BP), yellow = focal adhesion (KEGG), green = muscle contraction. B) Table displaying the KEGG pathways that were downregulated in CA tissues relative to NC tissues. p-values subject to Benjamini-Hochberg correction.

Cluster 1 showed that there were stepwise increases from NC to LGCA tissues and LGCA to HGCA tissues for the complement cascade, canonical glycolysis, and immune functions including antigen presentation and chronic inflammation (Fig 6.5). A similar pattern of sequential increase from NC to LGCA to HGCA was seen in cluster 2 for mRNA translation and the regulation of ribosomes binding to mRNA molecules (Fig 6.6). Cluster 3 contained proteins which were significantly upregulated in LGCA relative to NC tissues, but then decreased in HGCA compared to LGCA, and were primarily involved in mitochondrial respiration (Fig 6.7). Clusters 4 (Fig 6.8) and 5 (Fig 6.9) contained proteins which were less abundant in LGCA and HGCA compared to NC tissues. Cluster 4 was characterised by a reduction in mitochondrial ribosome components and RAS components AGT, chymase and mast cell carboxypeptidase A. Cluster 5 revealed significant downregulation of ECM organisation, cell junction assembly and focal adhesions.

The proteins significantly upregulated in CA tissues compared to NC showed enrichment for involvement in the functions and components of mitochondria. This was verified by focusing on heat map cluster 3 (Fig 6.7), which revealed a significant increase of mitochondrial respiration components in LGCA tissues relative to NC tissues, suggesting that the tumour cells are responding to an increased energy demand as the cancer becomes established. The shift from NC to LGCA also included a decrease in cytoskeletal binding and organisation, including cell junctions, focal adhesions, cell-matrix adhesion and ECM organisation (Fig 6.9). These trends hint at the increasing migratory capacity of cancer cells to enable invasion and metastasis. Similarly, dysregulation of mRNA translation highlighted by cluster 2 (Fig 6.6) is likely to be related to the changes in gene transcription inherent to cancer cells. Interestingly, cluster 1 revealed that HGCA tissues had the highest expression of proteins involved in immune activity, inflammation, and the complement cascade (Fig 6.6). These are characteristics of the CMS1 and CMS4 subtypes (Guinney et al., 2015), reflecting the DNA sequencing data in Chapter 6 which also aligned 3 of the 4 HGCA tissues with the CMS1 and CMS4 subtypes.

A heat map was also produced from cell line protein quantification data, from which 10 clusters were investigated. Of these clusters, 5 contained interesting protein abundance changes and were chosen for further analysis (Fig 6.10). As shown in Table 6.4, the majority of differentially expressed proteins in CA-derived cell lines relative to the NC cell line were upregulated, and similarly, the heat map clusters for downregulated proteins did not show as many relevant pathway enrichments as the clusters of upregulated proteins. Accordingly, the 5 clusters analysed below all contain proteins with a higher abundance in LGCA and HGCA-derived cell lines relative to the NC cell line.

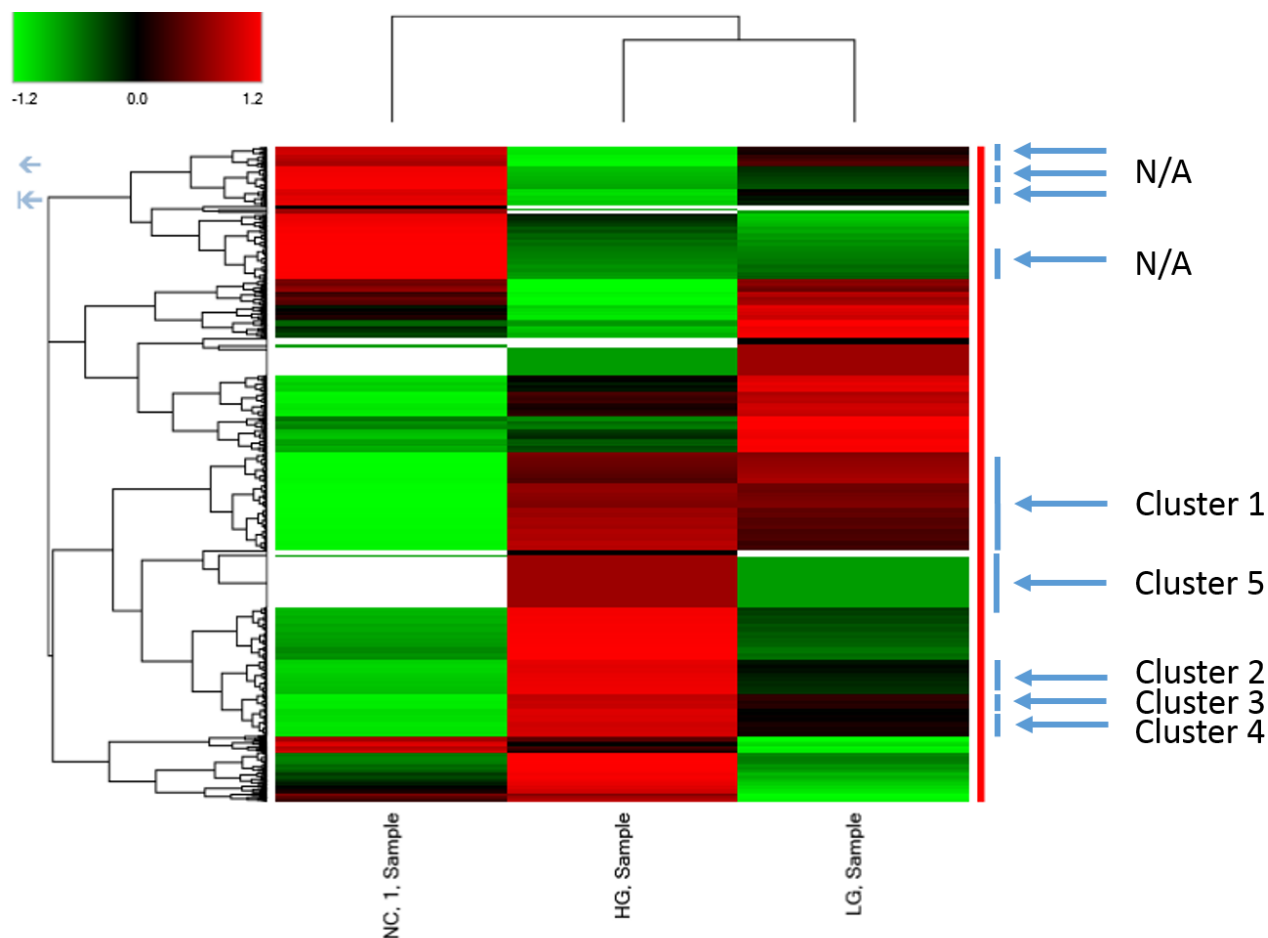


Figure 6.10: Heat map displaying the relative abundances of all quantified proteins in NC, HGCA and LGCA-derived cell lines. Heat map produced using Proteome Discoverer 2.4 showing the relative abundances of all proteins identified in tissue-derived cell lines. Clusters of proteins with differential expression between conditions have been labelled, with 5 of these proceeding to further analysis. Biological and technical replicates were pooled to give 3 categories based on the conditions that were compared: NC, HG and LG. Relative abundances are shown as: green = low, black = moderate, red = high, white = not detected.

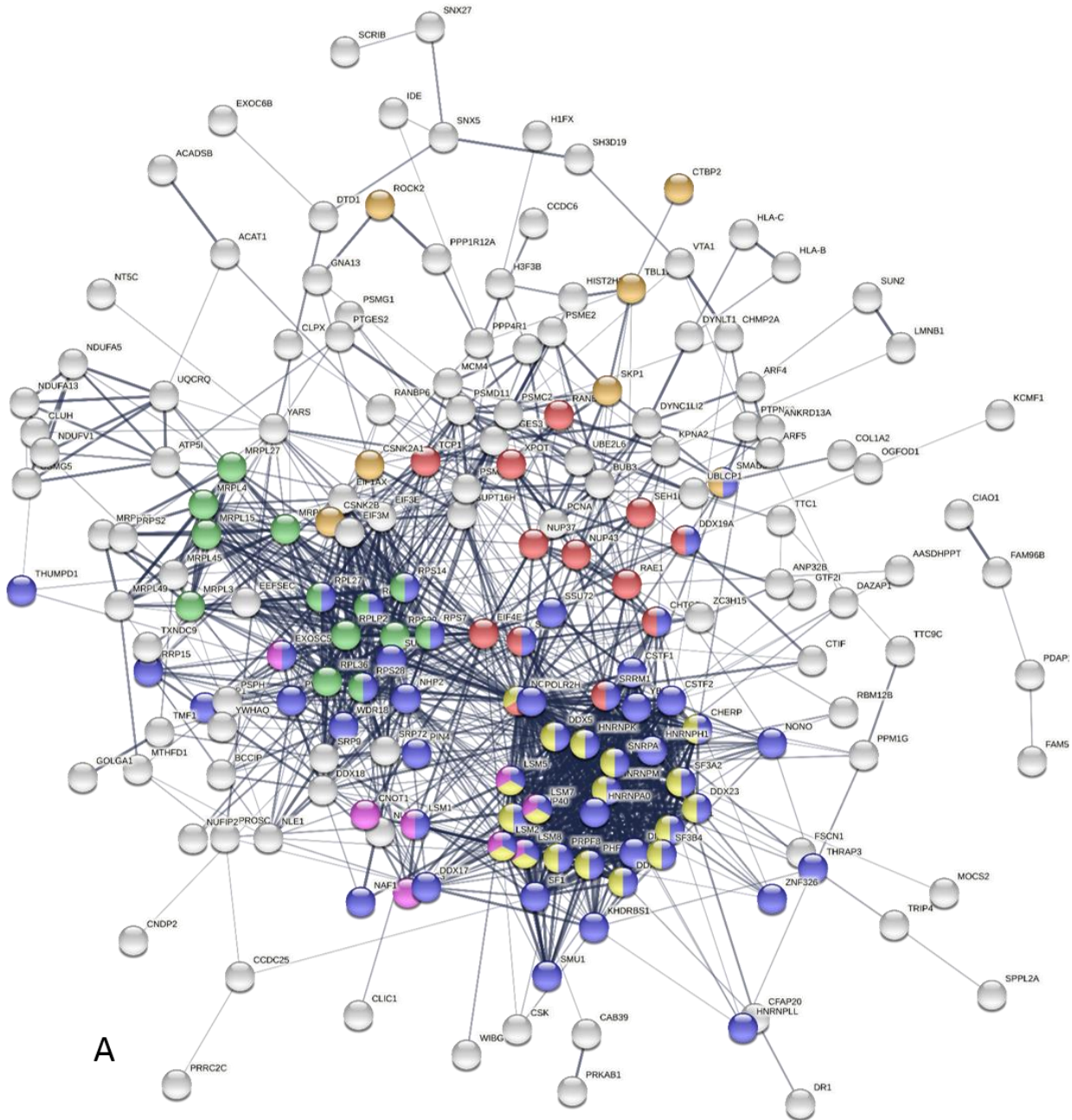


Figure 6.11: Cluster 1 from the cell lines heat map. A) STRING analysis of proteins from cluster 1, with GO terms or KEGG pathways coloured as follows: red = RNA localisation (BP), blue = RNA processing (BP), yellow = spliceosome (KEGG), green = ribosome (KEGG), purple = RNA degradation (KEGG), orange = Wnt signalling (KEGG). B) Table displaying the KEGG pathways that were upregulated in CA-derived cell lines relative to the NC cell line. p-values subject to Benjamini-Hochberg correction.

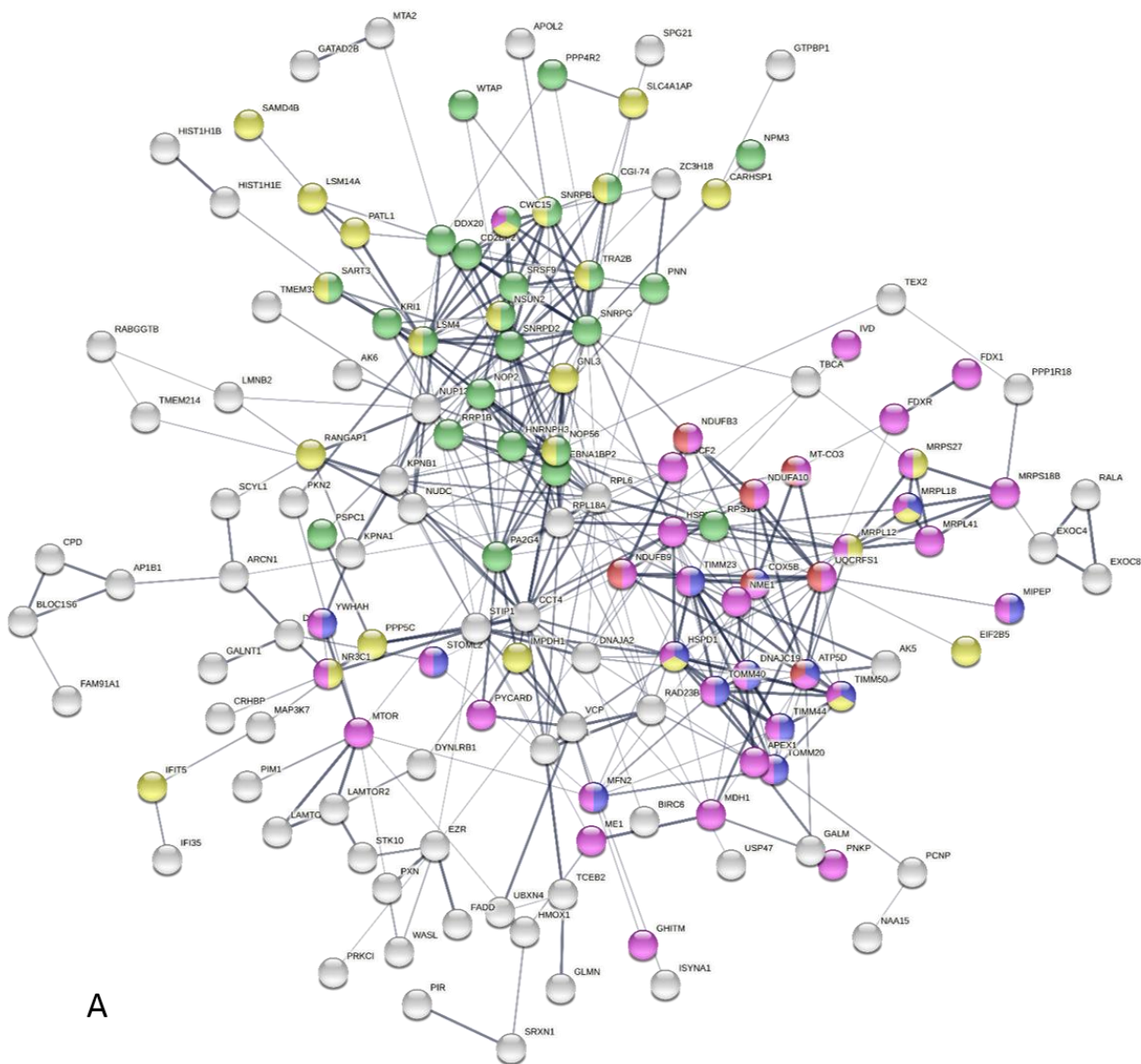


Figure 6.12: Cluster 2 from the cell lines heat map. A) STRING analysis of proteins from cluster 2, with GO terms or KEGG pathways coloured as follows: red = oxidative phosphorylation (KEGG), blue = mitochondrial transport (BP), yellow = RNA binding (BP), green = RNA processing (BP), purple = mitochondrion (CC). B) Table displaying the KEGG pathways that were upregulated in CA-derived cell lines relative to the NC cell line. p-values subject to Benjamini-Hochberg correction.

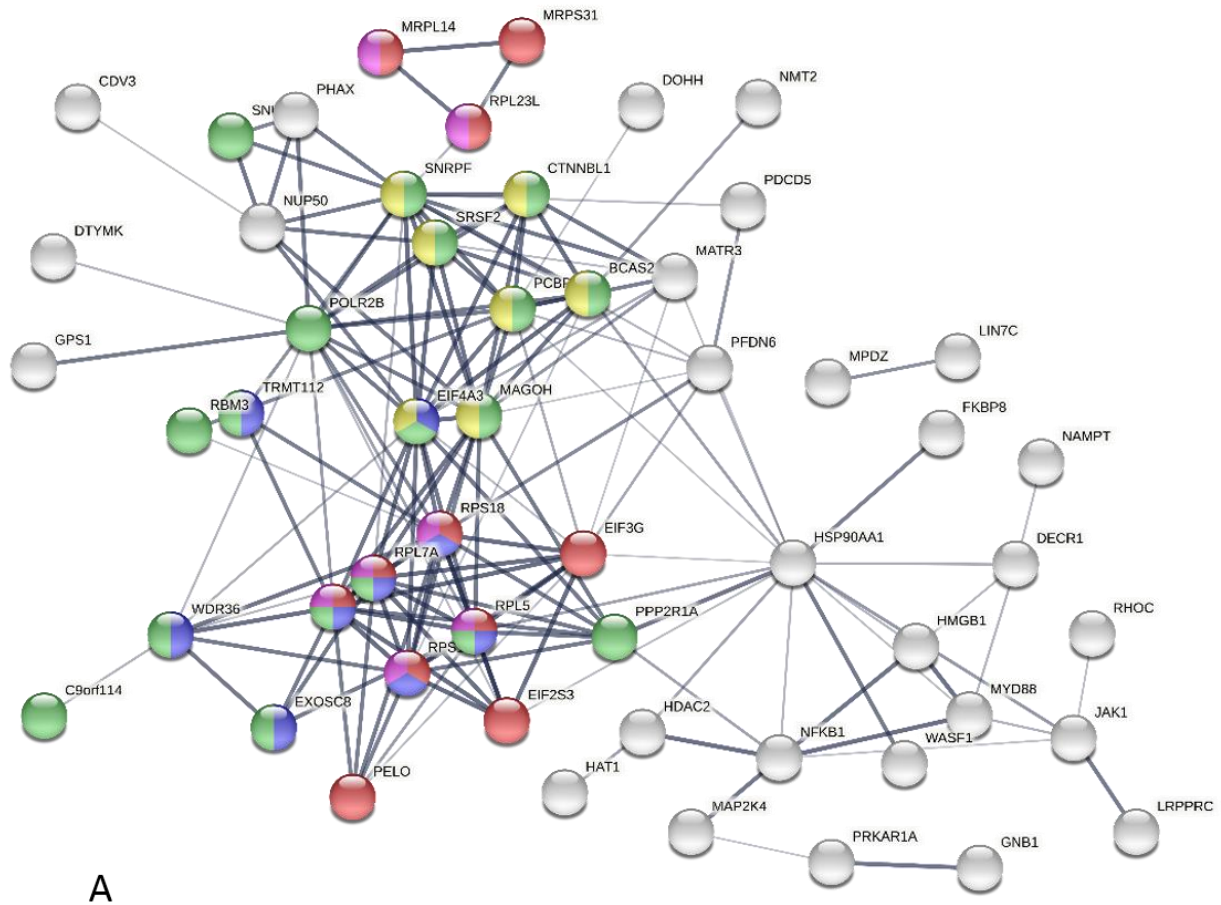


Figure 6.13: Cluster 3 from the cell lines heat map. A) STRING analysis of proteins from cluster 3, with GO terms or KEGG pathways coloured as follows: red = translation (BP), blue = ribosome biogenesis (BP), yellow = spliceosome (KEGG), purple = ribosome (KEGG), green = RNA processing (BP). B) Table displaying the KEGG pathways that were upregulated in CA-derived cell lines relative to the NC cell line. p-values subject to Benjamini-Hochberg correction.

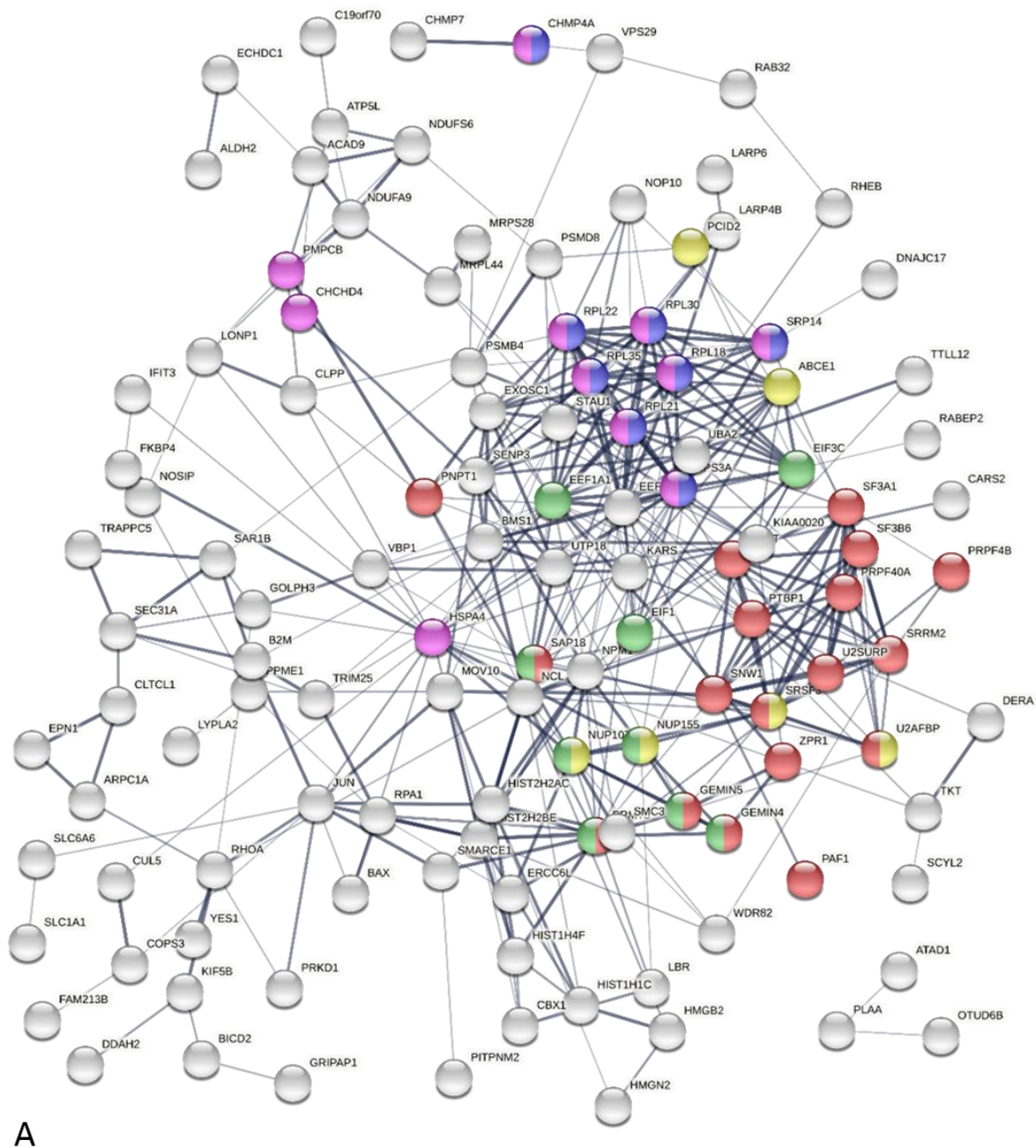
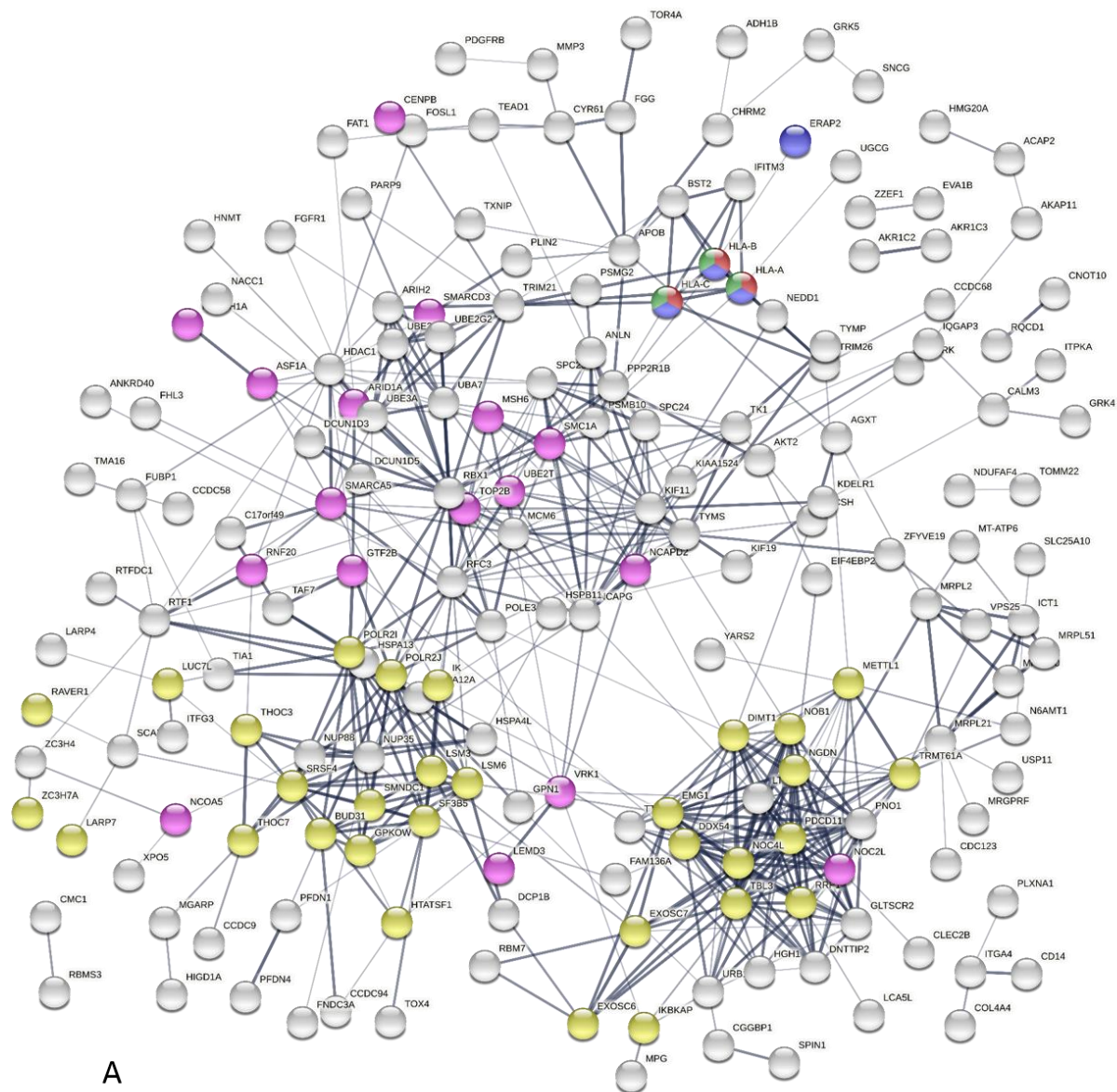


Figure 6.14: Cluster 4 from the cell lines heat map. A) STRING analysis of proteins from cluster 4, with GO terms or KEGG pathways coloured as follows: red = mRNA processing (BP), blue = protein targeting to ER (BP), yellow = protein export from the nucleus (BP), purple = protein targeting (BP), green = RNA transport (KEGG). B) Table displaying the KEGG pathways that were upregulated in CA-derived cell lines relative to the NC cell line. p-values subject to Benjamini-Hochberg correction.



B) KEGG pathways	Identifier	p-value
RNA degradation	hsa03018	0.0153

Figure 6.15: Cluster 5 from the cell lines heat map. A) STRING analysis of proteins from cluster 2, with GO terms or KEGG pathways coloured as follows: red, blue and green = unique pathways of antigen presentation (BP), yellow = RNA processing (BP), purple = chromatin binding (BP). B) Table displaying the KEGG pathways that were upregulated in CA-derived cell lines relative to the NC cell line. p-values subject to Benjamini-Hochberg correction.

Cluster 1 was characterised by proteins involved in RNA binding and RNA processing, including splicing, metabolic and catabolic processes, transport and localisation, as well as Wnt signalling (Fig 6.11). Similarly, clusters 2 (Fig 6.12), 3 (Fig 6.13) and 4 (Fig 6.14) revealed that there were stepwise increases from the NC cell line to the LGCA-derived cell lines and from the LGCA to HGCA-derived cell lines of proteins involved in RNA processing, transport and translation, protein transport, and mitochondrial components. Cluster 5 contained proteins which were absent from the NC cell line but detected at low abundance in the LGCA-derived cell lines and at high abundance in the HGCA-derived cell lines, and that were implicated in unique pathways by which endogenous and exogenous peptides are presented as antigens by MHC class I molecules (Fig 6.15).

In summary, when compared to the NC cell line, the LGCA and HGCA-derived cell lines predominantly exhibited increases in proteins involved in RNA-related processes, and were especially enriched for members of the spliceosome KEGG pathway. Multiple clusters of differentially expressed proteins as identified from the heat map were enriched for RNA binding and processing by ribosomes and spliceosomes, and the localisation, translation and degradation of RNA (Fig 6.11-14). This highlights that in the CA-derived cells, RNA processing is upregulated, which is likely to be a mechanism by which the cancer cells cope with higher levels of gene transcription.

To further investigate the proteomic changes between NC and LGCA and between LGCA and HGCA, all proteins with significantly differential expression (fold change ≥ 2 , p-value ≤ 0.05) between each condition were extracted from the lists of all proteins identified in cells lines and tissues, without performing clustering via a heat map. This was done to analyse whether specific processes or functions are implicated in tumour initiation (NC to LGCA) or progression (LGCA to HGCA). Venny was used to determine which proteins were unique to each list (e.g. only upregulated in HG compared to LG) and which were shared (e.g. upregulated in HG/LG and HG/NC).

Proteins that were significantly upregulated in HGCA and LGCA tissues relative to NC tissues were primarily enriched for involvement in RNA processing (GO:0006396; p=0.01) and RNA degradation (KEGG:hsa03018; p=0.033), and included the exosome complex components EXOSC1 and EXOSC6. Proteins significantly downregulated in HGCA and LGCA tissues compared to NC tissues were highly enriched for cytoskeletal and muscle systems (Table 6.5), in which tropomyosin 1 and 2 were strongly implicated. The loss of proteins involved in normal muscle function in both LGCA and HGCA tissues is a reflection of the considerable architectural changes which occur in the colon tissue as a result of CA, and the resulting impairment of normal function.

Table 6.5: Functions of proteins significantly downregulated in LGCA and HGCA tissues compared to NC tissues

GO category	Identifier	p-value
Cytoskeleton organisation	GO:0007010	0.00001
Cytoskeletal protein binding	GO:0008092	0.00001
Contractile fibre	GO:0043292	0.00008
Striated muscle thin filament	GO:0005868	0.0078
Muscle thin filament tropomyosin	GO:0005862	0.0088

GO categories that were enriched in the list of proteins that were significantly downregulated in both LGCA and HGCA compared to NC tissues. Significance is given by Benjamini-Hochberg corrected p-values.

Similarly, protein abundance changes that were unique to the comparison of LGCA tissues and HGCA tissues revealed processes that may be important in CA progression. Proteins that were significantly upregulated in HGCA relative to LGCA were enriched for protein degradation and antigen presentation (Table 6.6), whereas proteins which were downregulated in HGCA compared to LGCA were involved in drug metabolism (Table 6.7).

Table 6.6: Functions of proteins significantly upregulated in HGCA tissues compared to LGCA tissues

GO category / KEGG pathway	Identifier	p-value
Antigen processing and presentation	KEGG: hsa04612	0.0134
MHC protein complex	GO:0042611	0.0034
Lysosome	GO:0005764	0.001
Vacuole	GO:0005773	0.001

GO categories that were enriched in the list of proteins that were significantly upregulated in HGCA compared to LGCA tissues. Significance is given by Benjamini-Hochberg corrected p-values.

Table 6.7: Functions of proteins significantly downregulated in HGCA tissues compared to LGCA tissues

GO / reactome category	Identifier	p-value
Drug catabolic process	GO:0042737	0.0095
Antibiotic metabolic process	GO:0016999	0.0097
Cytosolic sulfonation of small molecules	Reactome: HSA-156584	0.00008

GO categories that were enriched in the list of proteins that were significantly downregulated in HGCA compared to LGCA tissues. Significance is given by Benjamini-Hochberg corrected p-values.

These changes would suggest that the progression from LGCA to HGCA in some way involves increased levels of protein turnover and a potential immune response to the tumour via the presentation

of antigens which may include cancer-specific neoantigens, and a reduction in drug metabolism that may indicate changes in cancer cell sensitivity to drugs.

In the cell lines, there was a significant enrichment for proteins involved in DNA packaging and replication in CA-derived cell lines relative to the NC cell line (Table 6.8). This was highlighted by the significant upregulation of histone subunits, histone chaperone ASF1A, and the interphase-to-mitosis condensing complex subunit 1 (NCAPD1). Conversely, the CA-derived cell lines had significantly reduced expression of peptidase inhibitors and other proteins which function in the lysosome (Table 6.9).

Table 6.8: Functions of proteins significantly upregulated in LGCA and HGCA-derived cell lines compared to the NC cell line

GO category / KEGG pathway	Identifier	p-value
DNA packaging	GO:0006323	5.13×10^{-7}
Chromatin binding	GO:0003682	0.00001
Nucleus	GO:0005634	8.12×10^{-16}
DNA replication	KEGG: hsa03030	0.0145

GO categories that were enriched in the list of proteins that were significantly upregulated in LGCA and HGCA-derived cell lines compared to the NC cell line. Significance is given by Benjamini-Hochberg corrected p-values.

Table 6.9: Functions of proteins significantly downregulated in LGCA and HGCA-derived cell lines compared to the NC cell line

GO category / KEGG pathway	Identifier	p-value
Negative regulation of endopeptidase activity	GO:0010951	1.27×10^{-12}
Complement and coagulation cascades	KEGG: hsa04610	1.73×10^{-6}
Lysosome	KEGG: hsa04142	0.00001

GO categories that were enriched in the list of proteins that were significantly downregulated in LGCA and HGCA-derived cell lines compared to the NC cell line. Significance is given by Benjamini-Hochberg corrected p-values.

The abundance of some proteins was only significantly different when comparing the HGCA-derived cells to the LGCA-derived cells. Upregulated proteins functioned in the endoplasmic reticulum, highlighted by CAV3 overexpression (Table 6.10), whereas downregulated proteins were involved in cell adhesion (Table 6.11), which included the membrane-cytoskeleton linker protein ANK3.

Overall, this shows that the CA-derived cell lines differ from the NC cell line in terms of increased chromatin remodelling to facilitate higher levels of gene transcription and DNA replication, and a reduction in peptidase inhibitors to allow greater activity within lysosomes. Furthermore, HGCA-derived

cells are in a more migratory state when compared to LGCA-derived cells as evidenced by their downregulation of adhesion proteins.

Table 6.10: Functions of proteins significantly upregulated in HGCA-derived cell lines compared to the LGCA-derived cell lines

GO category / KEGG pathway	Identifier	p-value
Endoplasmic reticulum	GO:0005783	0.00024
Endocytosis	KEGG: hsa04144	0.0435

GO categories that were enriched in the list of proteins that were significantly upregulated in HGCA-derived cell lines compared to the LGCA-derived cell lines. Significance is given by Benjamini-Hochberg corrected p-values.

Table 6.11: Functions of proteins significantly downregulated in HGCA-derived cell lines compared to the LGCA-derived cell lines

GO category / KEGG pathway	Identifier	p-value
Regulation of cell adhesion	GO:0030155	0.0334
Cell adhesion molecule binding	GO:0050839	0.0493
Cytoskeletal protein binding	GO:0008092	0.0445

GO categories that were enriched in the list of proteins that were significantly downregulated in HGCA-derived cell lines compared to the LGCA-derived cell lines. Significance is given by Benjamini-Hochberg corrected p-values.

6.2.4 Comparison of protein changes seen in CA tissues and CA-derived cells

Protein abundance changes in the CA tissues were compared to those in the CA-derived cells to give an insight into how similar the tumour-derived cell lines are to their parent tissue samples. Overall, the cells seemed to have less variability than the tissues, reflecting the fact that the tissue is a heterogeneous mix of cell types whereas the cell lines are comparatively homogenous. For example, there were extensive reductions in muscle structure and function, cell junctions and ECM organisation in the CA tissues which were not reflected in the CA-derived cells. Furthermore, the abundance of proteins involved in chronic inflammation and the complement cascade increased proportionally to CA grade in the tissues, but these enrichments were not seen in the tissue-derived cell lines, and in fact complement cascade components were reduced in CA-derived cell lines relative to the NC cell line.

However, the CA tissues and CA-derived cells both exhibited enrichment for RNA-related processes, as well as aspects of cellular metabolism including oxidative phosphorylation. The conservation of these changes in the CA-derived cells suggest that they are intrinsically important to the function of CA tumour cells.

Overall, this suggests that within the complexity of the tumour tissue there are many unique physiological and structural changes occurring that are important aspects in the context of normal

physiological function and the response of the body to the tumour, but that the specific changes which manifest in colon cancer cells relative to the normal colon cells predominantly drive intracellular processes such as increased mitochondrial respiration and alterations to gene transcription, mRNA splicing and translation.

6.2.5 Validation of mass spectrometry results by western blotting

In order to validate the MS results, WB was performed on selected proteins which had large differences in relative abundance with high statistical significance. Densitometry was carried out on blots, and this method of semi-quantification was compared with the MS quantification data for each individual sample. Ideally, protein samples used for WB and for MS would be derived from the same extraction. However, this was not possible, and a new extraction was performed from tissues for WB. Therefore, the relative abundances for MS and WB compared below were used mainly to provide a rough comparison between MS and WB.

CD44 is a cell surface receptor for hyaluronic acid, mediating adhesion to the ECM. It has been investigated as a CSC marker and is of particular interest in CRC as it seems to be expressed by the majority of colon tumour cells and its transcription is partially mediated through Wnt signalling (Jaggupilli et al., 2012; Wielenga et al., 1999). CD44 was selected to validate the LC-MS/MS result due to its significant upregulation in HGCA-derived cell lines relative to the LGCA cell lines, its relevance as a potential stem cell marker, and its abundance ratio of 2.3 for HGCA/LGCA. WB confirmed this trend, with a strong signal detected in the NC cell line and 4 HGCA cell lines, and weaker bands seen in the 4 LGCA cell lines (Fig 6.16A) at ~75 kDa. Densitometry revealed that the normalised intensity value ratio for HGCA/LGCA was 1.22 (Fig 6.16C), but this difference was not statistically significant (Fig 6.16B). However, the WB confirms the result seen in LC-MS/MS, in that CD44 was more abundant in the HGCA-derived cells than in the LGCA-derived cells.

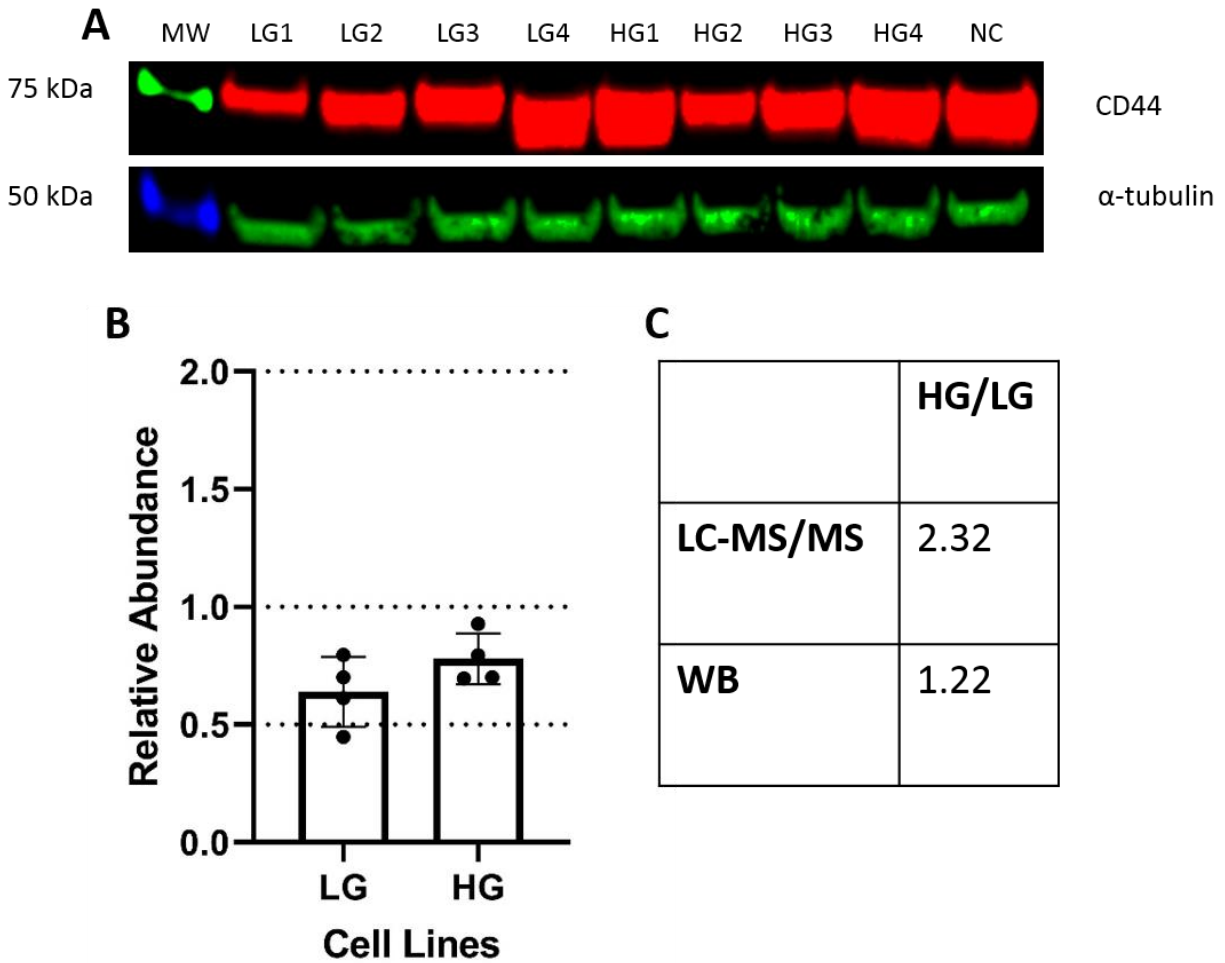


Figure 6.16: CD44 Densitometry for NC and CA-derived cell lines. A) Western blotting was performed for CD44 (A, red) on protein extracts from CA-derived cells, with α -tubulin as a loading control (A, green). B) Fluorescence of CD44 bands was normalised against α -tubulin, and then the abundance in CA-derived cell lines relative to the NC-derived cell line (y-axis) was calculated and displayed as an average \pm SD. C) The abundance ratios for HGCA/LGCA are displayed for LC-MS/MS (MS) and densitometry (WB) data.

Fibronectin (FN1) is one of many ECM proteins with aberrant expression in cancer (J. P. Wang et al., 2017). It is a large glycoprotein (260-280 kDa) which binds transmembrane proteins and influences cell growth, differentiation, migration, wound healing and blood coagulation (J. P. Wang et al., 2017; Ye et al., 2020; Yi et al., 2016). In cancers, FN1 is associated with angiogenesis, invasion via MMP activation, self-renewal, proliferation, and resistance to therapy (J. P. Wang et al., 2017; Yi et al., 2016). FN1 associates with various membrane proteins to activate PI3K/AKT signalling (via α/β 1-integrins), EMT gene expression (via α -integrins) and migration (via FAK/STAT3) (J. P. Wang et al., 2017). It also has a bi-directional

association with SOX2, whereby FN1/integrin binding upregulates SOX2 expression, and SOX2 overexpression leads to increased FN1 expression (J. P. Wang et al., 2017; Ye et al., 2020). FN1 was selected to validate the LC-MS/MS result in cell lines due to the reliability of the primary antibody and the significant downregulation of FN1 in HGCA cell lines relative to the LGCA cell lines, with an abundance ratio for HGCA/LGCA of 0.52. WB was performed for FN1 in the CA-derived cell lines (Fig 6.17A, red). When densitometry was performed (Fig 6.17B), it revealed a reduction in HGCA-derived cell lines relative to the LGCA-derived cell lines, with an intensity ratio for HGCA/LGCA of 0.54 (Fig 6.17C), which was very similar to the abundance ratio for LC-MS/MS (HGCA/LGCA = 0.52).

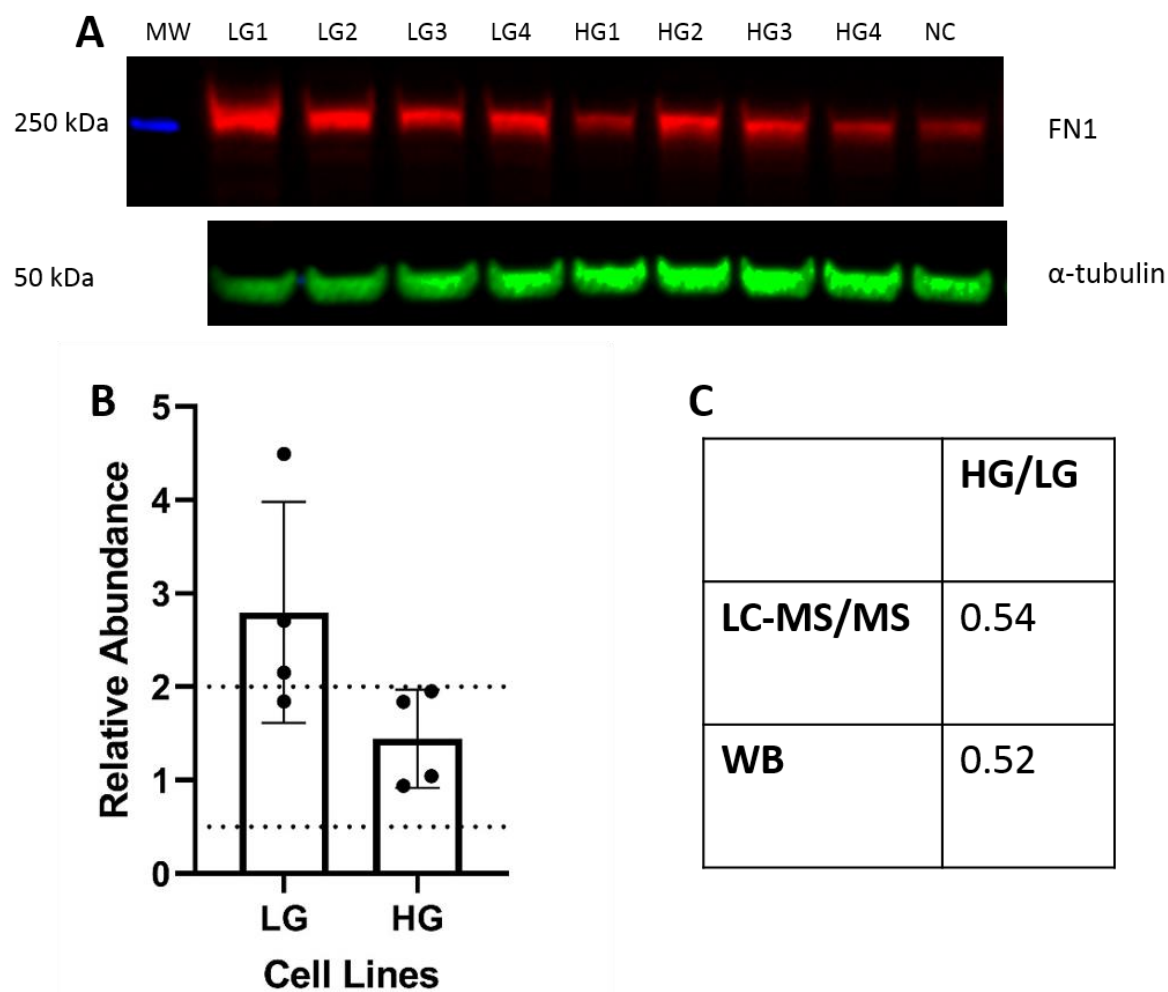


Figure 6.17: FN1 Densitometry for NC and CA-derived cell lines. A) Western blotting was performed for FN1 (A, red) on protein extracts from CA-derived cells, with α -tubulin as a loading control (A, green). B) Fluorescence of FN1 bands was normalised against α -tubulin, and then the abundance in CA-derived cell lines relative to the NC-derived cell line (y-axis) was calculated and displayed as an average \pm SD. C) The abundance ratios for HGCA/LGCA are displayed for LC-MS/MS (MS) and densitometry (WB) data.

S100A8 (MRP8) and S100A9 (MRP14) are commonly found as a heterodimer called calprotectin. Calprotectin binds Ca^{2+} and Zn^{2+} ions and plays an important role in inflammation caused by infection, autoimmunity or metabolic diseases (S. Wang et al., 2018). Together, S100A8 and S100A9 account for almost half of the total protein content of neutrophils and are released to induce chemotaxis of leukocytes, cytokine release and apoptosis (Nakatani et al., 2005; Ryckman et al., 2003; S. Wang et al., 2018). Expression of S100A8 and S100A9 is upregulated during inflammation, a key aspect of CRC (S. Wang et al., 2018). Accordingly, other studies have reported significant overexpression in CRC (Saleem et al., 2019; Vasaikar et al., 2019). S100A8 and S100A9 were selected to validate LC-MS/MS results in tissues because there was a sequential increase in abundance ratio for LGCA/NC (1.66 and 1.81), HGCA/LGCA (2.43 and 2.52) and HGCA/NC (4.42 and 4.86) in the tissue samples. This was verified by WB, which showed that HGCA tissues had stronger bands for S100A8 at ~16 kDa (Figure 6.18A) and S100A9 at ~13 kDa (Figure 6.18B) than the LGCA tissues, and bands were faint or absent in the NC tissues. This difference was quantified by densitometry, which revealed HGCA/LGCA abundance ratios of 1.41 for S100A8 and 2.81 for S100A9, which confirm the upregulation seen in LC-MS/MS with ratios of 2.80 (S100A8) and 2.40 (S100A9).

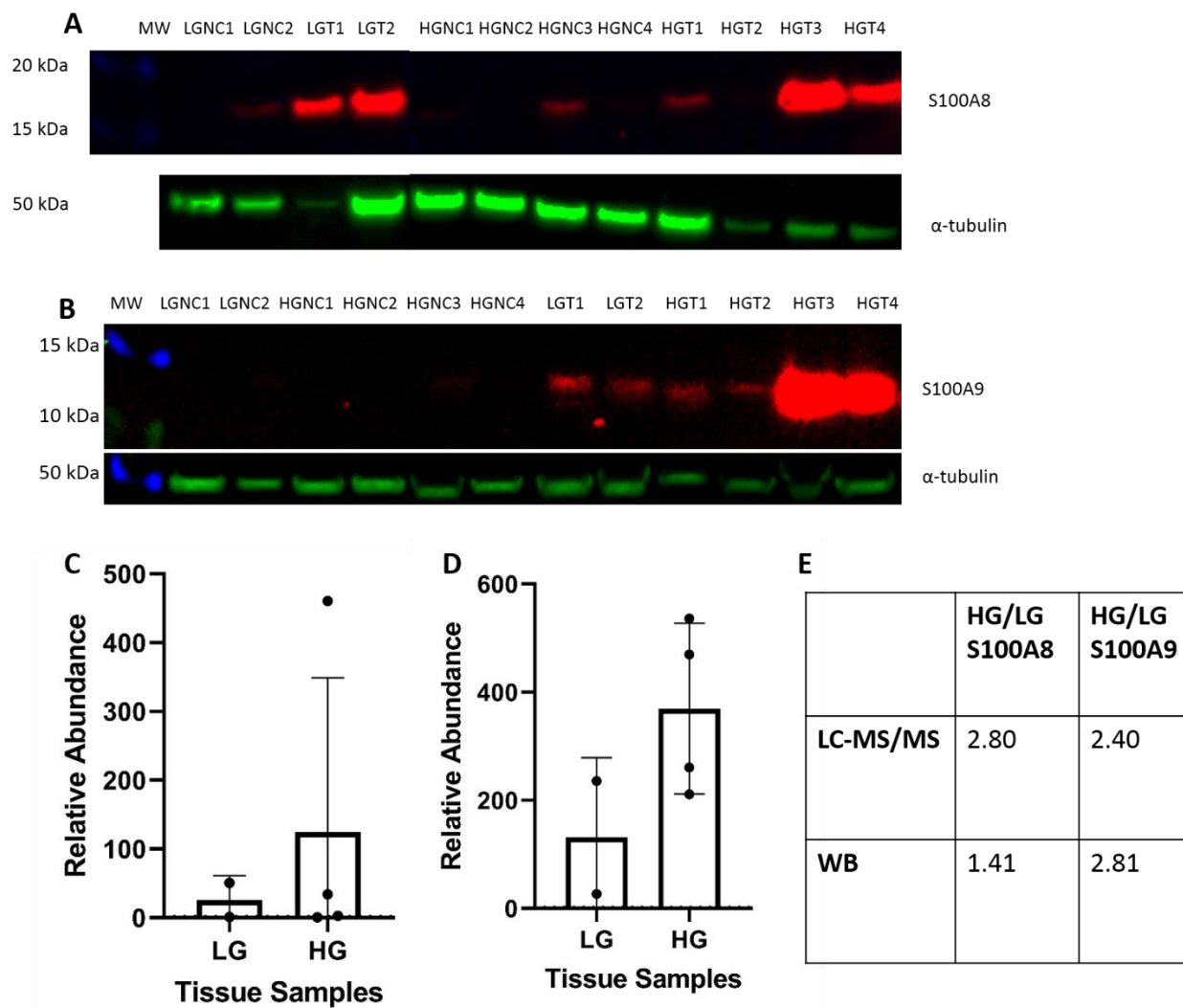


Figure 6.18: S100A8 and S100A9 Densitometry for tissues. A) Western blotting was performed for S100A8 (A, red) and S100A9 (B, red) on protein extracts from NC, LGCA and HGCA tissues, with α -tubulin as a loading control (A&B, green). Fluorescence of S100A8 (C) and S100A9 (D) bands were normalised against α -tubulin, and then the abundance in CA tissues relative to the NC tissues (y-axis) was calculated and displayed as an average \pm SD. E) The abundance ratios for HGCA/LGCA are displayed for LC-MS/MS (MS) and densitometry (WB) data.

6.3 Discussion

6.3.1 Comparing CA-derived cell lines to CA tissues

By performing a search against the UniProt *Homo sapiens* reference database, a total of 4767 proteins were identified across all tissue samples and a total of 4711 proteins were identified across cell lines. The biological replicates for LGCA and HGCA-derived cell lines shared 46.2% and 30.6% of identified proteins, and the LGCA and HGCA tissues shared 36.3% and 38.1% of proteins.

The tissue samples represent a complex collection of cell and tissue types, including muscle, blood vessels, immune cells and fat, leading to large changes in protein abundance due to the loss of normal architecture and function as a dense bulk of tumour cells forms. Accordingly, the CA tissues exhibited significant decreases in proteins involved in muscle structure and contraction (Table 6.5), which was not reflected in the CA-derived cells. Furthermore, increases in the CA tissues of proteins that regulate defence responses and immune system processes suggests that changes in the way the cancer cells interact with the immune system, in terms of suppression or avoidance of immune cells and tolerance of inflammation, is involved in tumour initiation and progression. These global changes are an important aspect of cancer biology; however, the cell lines, which are a pure population (without the muscle, blood vessels, fat, etc.) that has retained the mutational signatures of the original tumour tissue, may reveal changes which are more relevant to the signalling within and between tumour cells without these changes being overwhelmed by those seen in the tissues.

Looking beyond these general tissue changes, the CA tissues displayed increased levels of ribosomal components (MRPS2, MPRS6, MRPS18A), RNA processing (EXOSC1, EXOSC6), mitochondrial gene expression (MRPS7, MRPS22, MRPL16) and components of mitochondrial respiration (NADH dehydrogenases), and decreased cytoskeletal binding (ANK2) and focal adhesion. CA-derived cells also displayed upregulation of components of RNA processing and degradation (40s and 60s ribosomal proteins: RPL5, RPL7A, RPS2, RPS18), as well as chromatin binding (HIST1H1) and mitochondrial respiration (NADH dehydrogenases), and the HGCA-derived cells exhibited downregulation of cell adhesion proteins (ANK3). Collectively, these changes reflect that the tumour is in a state of heightened transcriptional and translational activity with increased energy demands associated with biogenesis. As the cancer cells possess a higher rate of proliferation, there is an increased demand for new and recycled cellular components, partially enabled by the downregulation of peptidase inhibitors (Table 6.9). In order to achieve this, the energy demands of the cell are greater, and this need appears to be met by increased oxidative phosphorylation and ATP production in the mitochondria. This is highlighted by significant

upregulation of various structural and functional components of the mitochondria in both the CA tissues and tissue-derived cell lines. Furthermore, this change in metabolic focus by the cancer cells appears to come at the expense of, or is enabled by changes to, normal physiological processes such as the regulation of peptidases by inhibitors (Table 6.12), RNA processing and splicing leading to alternative protein products, and proper protein folding and PTMs. This will affect the regulation of gene transcription in terms of which genes are transcribed and to what extent. In the context of the tissue, it is also apparent that the tumour cells are being primed for migration and invasion by downregulating adhesion components; however, the tissue-derived cell lines do not share these protein changes, most likely due to the differences in adhesion requirements between *in vivo* and *in vitro* conditions and for growth in a cell culture environment that lacks the structure present *in vivo*. Importantly, the similarity between tissues and cell lines with respect to the GO categories and KEGG pathways discussed above justifies the use of these CA-derived primary cells as an *in vitro* model.

Table 6.12: Peptidase inhibitors with differential expression in CA-derived cell lines compared to the NC cell line.

Gene Symbol	Accession Number	Abundance changes	GO categories
C3, SERPINE1 (PAI-1)	P01024 P05121	↓LG/NC; ↓HG/NC	Regulated exocytosis (BP) Negative regulation of endopeptidase activity (BP) Peptidase regulator activity (MF) Endopeptidase inhibitor activity (MF) Complement and coagulation cascade (KEGG) Secretion (BP)
A2M	P01023	↓LG/NC	Regulated exocytosis (BP) Negative regulation of endopeptidase activity (BP) Peptidase regulator activity (MF) Endopeptidase inhibitor activity (MF) Complement and coagulation cascade (KEGG)
TIMP1/3, AHSG, LTF, ITIH3	P01033 P35625 P02765 P02788 Q06033	↓LG/NC; ↓LG/NC	Regulated exocytosis (BP) Negative regulation of endopeptidase activity (BP) Peptidase regulator activity (MF) Endopeptidase inhibitor activity (MF) Secretion (BP)
SERPIND1, SERPINC1 (antithrombin III)	P05546 P01008	↓LG/NC	Negative regulation of endopeptidase activity (BP) Peptidase regulator activity (MF) Endopeptidase inhibitor activity (MF) Complement and coagulation cascade (KEGG)
SERPINA7, SERPINA10, AMBIP, ITIH1	P05543 Q9UK55 P02760 P19827	↓LG/NC	Negative regulation of endopeptidase activity (BP) Peptidase regulator activity (MF) Endopeptidase inhibitor activity (MF)

WB and densitometry were performed using a selection of proteins identified by the proteomic analysis in order to verify the MS results. However, there are two main limitations to the densitometry data. First, different protein extractions were used for the MS and for the WB. A trial blot was run prior to optimisation of the antibodies used for validation, which revealed sample degradation. Extractions were repeated and resulted in protein of higher quality, except for the LG3 and LG4 samples which were again degraded. It is suspected that these two samples may have been handled improperly between the time of surgery and fixation, which on occasion can exceed 18 hours. This was reflected in the DNA sequencing data in Chapter 3, for which no DNA of sufficient quality was obtained from LG4. Second, densitometry has inherent difficulties which cannot always be avoided. The images used for densitometry must not be “detector-saturated”, which occurs if the membrane is exposed for too long (Butler et al., 2019). Blots for S100A8 in tissue samples reached saturation within 1-2 sec, so they were repeated with 5 µg of protein per lane rather than 20 µg. Furthermore, the data itself can become “saturated”, after which point any further abundance increases are unable to be accurately measured, causing misleading comparisons (Butler et al., 2019). One important issue is that normalisation against α -tubulin is only valid if the intensity data fits a directly proportional model – intensity as measured by the detector does not have a linear relationship with the actual abundance of the protein at higher protein concentrations due to saturation (Butler et al., 2019), after which a hyperbolic model fits better than a linear model. One way to reduce the influence of saturation is to widen the area of band detection and reduce exposure time (Butler et al., 2019). Finally, band “ghosting” can occur at high loading concentrations, where the band appears washed out and the intensity of the band is underestimated. Blots of FN1 were repeated due to this “ghosting” effect by loading 4 µg per lane, which helped by reducing the intensity of bands. Overall, it is important to remember that densitometry is semi-quantitative and is not perfectly suited to determining the magnitude of differences in abundance.

6.3.2 Similarities with other publications

The proteomics findings from this chapter were compared with data from two other prominent CRC proteogenomics papers to identify similarities and differences.

The Cancer Genome Network (TCGN) carried out a comprehensive molecular characterisation of CRC in 2012 (Cancer Genome Atlas, 2012). This study identified frequently mutated genes and the pathways which they are involved in. They confirmed that common pathways implicated in CRC include Wnt, Ras-MAPK, PI3K, TGF- β , p53 and DNA MMR pathways. The Wnt pathway was altered in 93% of

tumours, usually due to APC inactivation and/or β -catenin activation. Tumours fell into two categories: hypermethylated, which were usually MSI-H and had CIMP; and non-hypermethylated, which were highly likely to be MSS and CIN tumours. More than half of the non-hypermethylated tumours had activating mutations to *KRAS*, *NRAS* and *BRAF*, which have a high degree of mutual exclusivity as previously discovered. These tumours also frequently had inactivating mutations in *APC*, *TP53*, *PIK3CA* and *SMAD4*.

A list of 67 genes from the TCGA study which had a high degree of correlation with either LG-CRC or HG-CRC was matched against the lists of differentially expressed proteins for the HGCA, LGCA and NC tissues and cell lines. Overall, 22 of 67 genes identified by TCGA were present in this data (Table 6.13), including DNA MMR genes *MSH2*, *MSH6* and *POLE*, CRC-associated *KRAS*, and PI3K signalling pathway members *PIK3AP1*, *PIK3C2A* and *PIK3R1*.

Table 6.13: Comparison with TCGA study.

Gene Symbol	Gene name/description	Abundance changes (cells)	Abundance changes (tissues)
POLE	DNA polymerase epsilon subunit	↑LG/NC; ↑HG/NC	
MSH2	DNA mismatch repair	↑LG/NC; ↑HG/NC	
MSH6	DNA mismatch repair	↑LG/NC; ↑HG/NC	↓LG/NC; ↓HG/NC
KRAS	GTPase KRAS	↑LG/NC; ↑HG/NC	
MYO1C	Unconventional myosin	↑LG/NC; ↑HG/NC	
PIK3R1	PI3K p85 regulatory subunit		↑LG/NC; ↑HG/NC
PIK3C2A	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit α	↑LG/NC; ↑HG/NC	
CLEC3B	Tetranectin: binds plasminogen	↓LG/NC; ↓HG/NC	
SLC38A7	Putative sodium-coupled neutral amino acid transporter 7		
GBP6	Guanylate-binding protein 6		↓LG/NC; ↓HG/NC
CDC27	Component of anaphase promoting complex/cyclosome: cell cycle control	↑LG/NC; ↓HG/LG	
GPC6	Glypican-6: cell surface glycoprotein co-receptor	↑LG/NC; ↓HG/LG	

SLC20A1	Sodium-dependent phosphate transporter 1		↓LG/NC; ↑HG/LG
SLC38A5	Sodium-coupled neutral amino acid transporter 5	↑HG/NC; ↑HG/LG	↓LG/NC; ↑HG/LG
ARL6IP4	ADP-ribosylation factor-like protein 6-interacting protein 4: modulates alternative pre-mRNA	↑LG/NC; ↓HG/LG	
TTN	Titin: key component of striated muscle	↑HG/NC; ↑HG/LG	
TSC22D	RNA polymerase II component	↑HG/NC; ↑HG/LG	
SLC39A8	Zinc transporter ZIP8	↑HG/NC; ↑HG/LG	
PIK3AP1	Phosphoinositide 3-kinase adapter protein		↑HG/NC; ↑HG/LG
CDIPT	CDP-diacylglycerol-inositol 3-phosphatidyltransferase	↓HG/NC; ↓HG/LG	
FBXO30	F-box only protein 30: substrate recognition component of an E3 ligase		↓HG/NC; ↓HG/LG
APOM	Apolipoprotein M: lipid transport		↓HG/NC; ↓HG/LG

Significantly mutated or CRC-correlated genes identified by TCGA whose protein products were significantly up- or down-regulated in either the CA tissues or CA-derived cells.

In 2019, Vasaikar *et al.* (Vasaikar et al., 2019) published the first comprehensive prospective proteogenomic study on CRC, which included label-free proteomics, TMT-labelled proteomics, phosphoproteomics, WXS, RNA-seq and miRNA-seq, and SNP arrays. They found that MS methods outperformed RNA-seq in terms of predicting gene function in CRC, though TMT labelling of the proteins was more effective than the label-free method. In this cohort, non-hypermethylated MSI-H tumours were enriched for DNA MMR, *POLE*, and *BRAF* activity. Chromosomal deletions were identified which were predicted to confer a repression of endocytosis; similarly, in the data from CA-derived cell lines there appeared to be an impairment of “regulated exocytosis” and “lysosome” functions in HGCA and LGCA cells relative to NC cells, but an increase in “endocytosis”-related proteins in HGCA cells relative to LGCA cells (Table 6.14). Encouragingly, these data confirmed the detection of CEACAM5, currently the most widely used CRC biomarker, as well as fellow family member and heterodimerisation partner CEACAM6.

From this data set, a list of 43 genes was matched against the tissue and cell line data, with 16 found to be shared between the two data sets (Table 6.14). Overall, the TCGA study had more overlap with proteins identified from the CA-derived cell lines than from the CA tissues. However, proteins identified by Vasaikar

et al. had greater overlap with the CA tissue data than data from the CA cell lines. Both studies analysed tissue samples rather than primary cell lines, and Vasaikar *et al.* also used blood samples.

Overall, the detection of many proteins that have previously been identified in CA proteogenomics studies validates the results presented in this chapter, and those which are unique to this study provide avenues for further investigation.

Table 6.14: Comparison with Vasaikar *et al.*

Gene Symbol	Gene name/description	Abundance (cells)	Abundance (tissues)
COL12A1	Type XII Collagen α	\uparrow LG/NC; \uparrow HG/NC	
S100P	Calcium sensor and contribution to calcium signalling		\uparrow LG/NC; \uparrow HG/NC
S100A8, S100A9	Calcium and zinc binding proteins		\uparrow HG/LG
CEACAM5, CEACAM6	Carcinoembryonic antigen-related cell adhesion molecules 5 and 6		\uparrow HG/NC; \uparrow HG/LG
SERPINB5	Tumour suppressor		\uparrow HG/NC; \uparrow HG/LG
SERPINH1	Binds collagen		\uparrow HG/LG
GPRC5C	Retinoic acid-inducible GPCR		\uparrow LG/NC; \uparrow HG/NC
FAP	Prolyl endopeptidase: ECM degradation, tumour growth	\downarrow LG/NC; \downarrow HG/NC	
PKM	Pyruvate kinase isoform M1	\downarrow HG/NC	
HMGA2	Transcriptional regulator		\uparrow LG/NC; \downarrow HG/LG
IGF2BP3	Insulin-like GF 2 mRNA-binding protein 3		\downarrow LG/NC; \uparrow HG/LG
MXRA5	Matrix-remodelling-associated protein 5		\uparrow HG/LG
THBS2	Thrombospondin-2: cell-cell and cell-matrix adhesion		\uparrow HG/LG
VSNL1	Visinin-like protein 1: calcium-dependent inhibition of rhodopsin phosphorylation		\uparrow HG/LG

Cancer-associated proteins identified by Vasaikar et al. (2019), and significantly up- or down-regulated in either the CA tissues or CA-derived cells.

6.3.3 Detection of RAS components by LC-MS/MS

A few RAS components were detected by MS, but none of the 5 iPSC markers were detected. In tissues samples (Table 6.15), AGT, ACE2, PRR, and cathepsins B (CTSB), D (CTSD) and G (CTSG) were detected. AGT abundance was higher in all LGCA tissues and lower in all HGCA tissues compared to their patient-matched NC tissues, but this was not statistically significance except for HGCA1 and HGCA3. ACE2 was detected in all HGCA tissues plus LGNC4, and abundances were significantly higher in all CA tissues relative to patient-matched NC tissues. Cathepsins B, D and G were detected at higher levels in all CA tissues compared to their patient-matched NC tissues, except for HGCA2 and LGCA4 for cathepsin G. However, there was no statistical significance for the differences in abundance of any cathepsins. The confidence levels for PRR detection in the tissues was low, and so no relative abundance data was produced. In the cell lines (Table 6.16), PRR was detected in LGCA1, HGCA3, HGCA4 and the NC cell line, with lower abundance in all CA-derived cells relative to the NC cells but with no statistical significance. Cathepsins B and D were detected in all CA-derived cells at lower levels than were detected in the NC cell line, except for cathepsin D in the LGCA2 cells. These abundance changes were only significant for cathepsin B in LGCA1, LGCA4 and HGCA3 cell lines.

Table 6.15: RAS components detected in tissues by LC-MS/MS

	Abundance ratios (p-values)							
	HGCA1 / HGNC1	HGCA2 / HGNC2	HGCA3 / HGNC3	HGCA4 / HGNC4	LGCA1 / LGNC1	LGCA2 / LGNC2	LGCA3 / LGNC3	LGCA4 / LGNC4
AGT	0.493 (0.041)	0.960 (0.258)	0.565 (0.042)	0.982 (0.110)	2.166 (0.916)	1.780 (0.735)	2.206 (0.982)	1.397 (0.992)
ACE2	100 (<0.0001)	100 (<0.0001)	100 (<0.0001)	100 (<0.0001)	100 (<0.0001)	100 (<0.0001)	100 (<0.0001)	6.091 (0.0008)
PRR	-	-	-	-	-	-	-	-
CTSB	2.773 (0.923)	8.216 (0.561)	14.758 (0.700)	9.967 (0.688)	1.718 (0.988)	7.048 (0.711)	2.002 (0.997)	0.527 (0.206)
CTSD	3.387 (0.850)	6.905 (0.673)	12.781 (0.762)	11.963 (0.591)	1.644 (0.999)	5.833 (0.817)	2.833 (0.827)	0.620 (0.438)
CTSG	1.129 (0.752)	0.577 (0.119)	4.324 (0.956)	2.837 (0.832)	2.535 (0.805)	1.473 (0.493)	1.598 (0.951)	0.551 (0.224)

Abundance ratios for RAS components in CA tissues relative to their patient-matched NC tissues, with p-values in brackets. A p-value<0.05 indicates a significant result.

Table 6.16: RAS components detected in cell lines by LC-MS/MS

	Abundance ratios (p-values)							
	HGCA1 / NC	HGCA2 / NC	HGCA3 / NC	HGCA4 / NC	LGCA1 / NC	LGCA2 / NC	LGCA3 / NC	LGCA4 / NC
PRR	0.01 (<0.0001)	0.01 (<0.0001)	0.449 (0.347)	0.542 (0.608)	0.379 (0.179)	0.01 (<0.0001)	0.01 (<0.0001)	0.01 (<0.0001)
CTSB	0.187 (0.503)	0.167 (0.428)	0.295 (0.022)	0.283 (0.093)	0.149 (<0.0001)	0.550 (0.537)	0.463 (0.635)	0.198 (0.009)
CTSD	0.106 (0.262)	0.147 (0.378)	0.467 (0.219)	0.334 (0.176)	0.381 (0.059)	1.915 (0.733)	0.625 (0.842)	0.398 (0.246)

Abundance ratios for RAS components in CA-derived cell lines relative to the NC-derived cell line, with p-values in brackets. A p-value<0.05 indicates a significant result.

6.3.4 Integration of the RAS with other signalling pathways

To assess the relevance to the scope of this thesis of proteins with significant abundance changes in the CA tissue and cell lines, the upstream and downstream signalling pathways that influence the RAS and pluripotency signalling were investigated.

The RAS is widely documented to sit both downstream and upstream of numerous signalling pathways. Most research has focussed on the AT₁R due to its association with adverse or cancer-related outcomes. AT₁R is activated when bound by its ligand, ATII. AT₁R is a G-protein coupled receptor (GPCR) and so it can directly activate G-proteins to initiate signalling cascades. ATII binds AT₁R to activate it, allowing AT₁R to phosphorylate G-proteins. However, AT₁R can also function independent of G-proteins by transactivating receptor tyrosine kinases (RTKs) such as EGFR, PDGFR and the insulin receptor (Balakumar et al., 2014; Mehta et al., 2007).

When ATII binds AT₁R it can cause receptor internalisation in either a β -arrestin-mediated or a caveolin-1 (CAV1)-mediated manner (Mehta et al., 2007), though receptor internalisation is not essential for signalling to occur (Hunyady et al., 2000). β -arrestin-mediated internalisation terminates G-protein signalling, but leads to activation of the MAP kinase ERK without requiring G-proteins (Mehta et al., 2007). ATII causes AT₁R to associate with CAV1 and be trafficked to caveolae, where AT₁R activates the G-protein Rac1 to initiate MAP kinase signalling, and non-receptor tyrosine kinase c-Src which then activates EGFR downstream signalling (Mehta et al., 2007).

Various caveola-related proteins were detected in the tissue and cell line samples, however there was a discrepancy between cells and tissues (Table 6.17). CAV1 and CAV2 were higher in LGCA cells than in the NC cell line, and CAV1 was higher again in HGCA cells than LGCA cells; however, CAV1 and CAV2

were lower in both LGCA and HGCA tissues relative to NC tissues. CAV3 was upregulated in HGCA cells relative to LGCA cells, and was unchanged in LGCA and HGCA tissue samples relative to NC tissues. The caveolin-associated proteins (CAVIN) CAVIN1 and CAVIN2 were both downregulated in LGCA and HGCA tissue relative to NC tissues; furthermore, CAVIN1 was lower in HGCA cells than LGCA cells. However, CAVIN2 was more abundant in LGCA cells than in the NC cell line. CAVIN3 was unchanged in the cell lines but downregulated in both the LGCA and HGCA tissues. CAVIN4 levels were lower in HGCA tissues and HGCA and LGCA cells relative to NC tissues and the NC cell line, respectively. Collectively, this suggests that in the tumour tissues the formation of caveolae is impaired, implying that β -arrestin-mediated internalisation may be favoured. However, either the cultured cells have regained their expression of caveola-related proteins, or growing the cells in culture conditions favoured those which retained their caveolae.

Table 6.17: Caveolae-related proteins detected in NC and CA tissues and tissue-derived cell lines

Gene name	Description	Abundance (cells)	Abundance (tissues)
CAV1	Forms a heterodimer with CAV2. Mediates recruitment of CAVINs and also β -catenin. Internalises and degrades TGFBR1.	\uparrow LG/NC; \uparrow HG/LG	\downarrow LG/NC; \downarrow HG/NC
CAV2	Required for insulin-stimulated nuclear translocation of MAPK1 and STAT3	\uparrow HG/NC; \uparrow HG/LG	\downarrow LG/NC; \downarrow HG/NC
CAV3	Mediated recruitment of CAVIN2 and 3	\uparrow HG/NC; \uparrow HG/LG	No change
CAVIN1	Essential for caveolae formation in all tissues and for CAV1 recruitment.	\downarrow HG/NC; \downarrow HG/LG	\downarrow LG/NC; \downarrow HG/NC
CAVIN2	Induces membrane curvature to form caveolae	\uparrow LG/NC	\downarrow LG/NC; \downarrow HG/NC
CAVIN3	Regulates caveolae trafficking and budding	No change	\downarrow LG/NC; \downarrow HG/NC
CAVIN4	Regulates localisation and stabilisation of CAV3. Recruits MAPK1/3.	\downarrow LG/NC; \downarrow HG/NC	\downarrow HG/NC; \downarrow HG/LG

AT₁R phosphorylates the regulatory p85 subunit of phosphatidylinositol (PtdIns) 3-kinases (PI3Ks) to initiate PI3K signalling (Fig 6.19) (Zhao et al., 2014). Once active, PI3Ks phosphorylate PtdIns4,5-

bisphosphate (PIP₂) to become PtdIns3,4,5-triphosphate (PIP₃), which binds AKT and brings it into contact with PDK1 and mTORC2, both of which phosphorylate AKT. PIP₂ plays its own role in signalling: phospholipase C (PLC) converts PIP₂ into diacylglycerol (DAG) and IP₃ (Mehta et al., 2007). AT₁R activates PLC, as well as PLA₂ and PLD, by phosphorylating them via coupling with G_α and G_{βγ} G-protein complexes (Balakumar et al., 2014; Mehta et al., 2007). DAG activates protein kinase C (PKC) which goes on to phosphorylate and activate many other signalling components, and IP₃ binds receptors on the smooth endoplasmic reticulum (ER) to release Ca²⁺ which is vital to AT₁R-mediated EGFR and JAK/STAT signalling (Balakumar et al., 2014).

The reverse reaction to that of PI3K is catalysed by PTEN, which converts PIP₃ back into PIP₂ (Papadatos-Pastos et al., 2015). This is an important regulatory mechanism to control the extent of PI3K signalling, and PTEN is a known tumour suppressor which is frequently mutated and silenced by promoter methylation in cancer (Papadatos-Pastos et al., 2015).

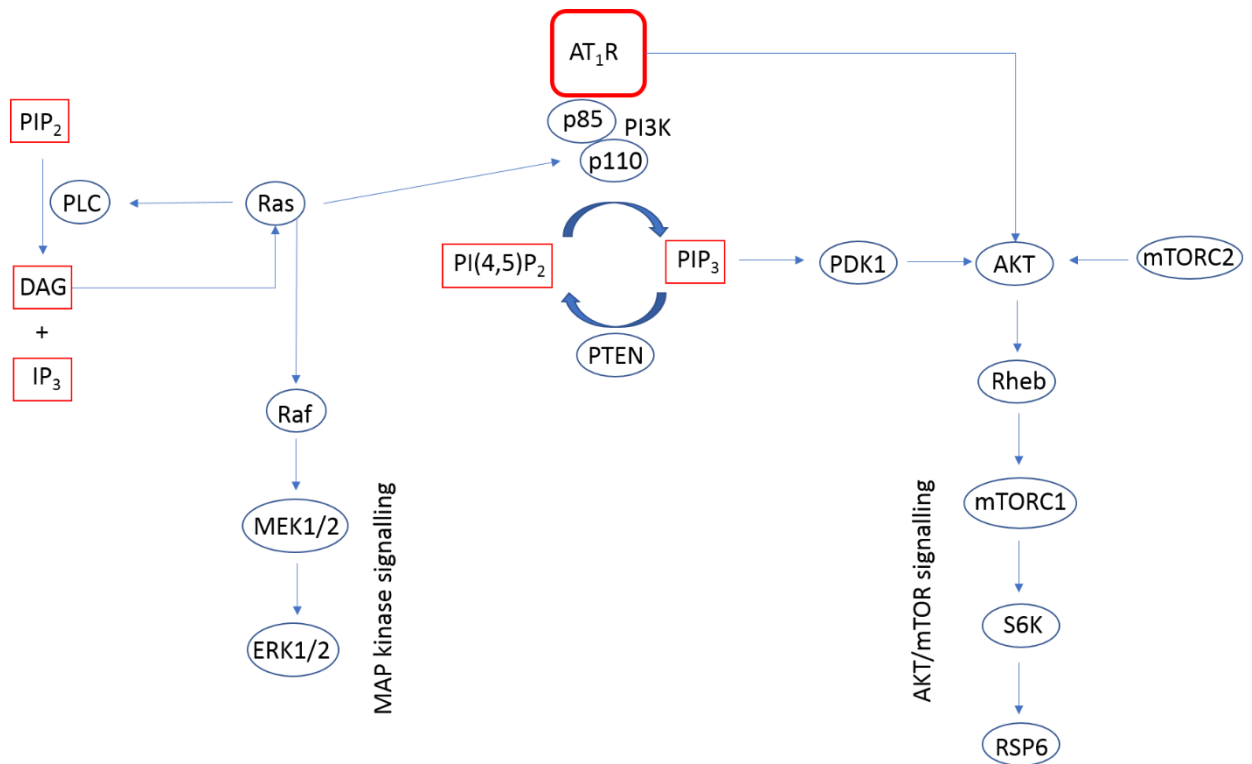


Figure 6.19: AT₁R-PI3K/AKT signalling. When activated by ATII, AT₁R can initiate AKT/mTOR signalling via PI3K.

As mentioned above, PI3K signalling activates AKT. AKT is at the top of the mTOR signalling cascade; it phosphorylates mTORC1 and mTORC2, leading to the transcription of antiapoptotic genes via NF- κ B and the degradation of the pro-apoptotic protein p53 (Papadatos-Pastos et al., 2015). There is a tendency for AKT mutations to occur in proximal CRC tumours which the TCGA study also associated with non-hypermethylation and alterations to *APC* and *KRAS*, than in distal tumours which TCGA found were more likely to be hypermutated, hypermethylated and MSI-H, and which tend to have *BRAF* mutations and a poorer prognosis (Cancer Genome Atlas, 2012; Johnson et al., 2010).

Four members of PI3K signalling were detected by MS (Table 6.18). The PI3K regulatory subunit p85 α was higher in LGCA tissues than NC tissues, and PIK3AP1, an adapter protein which links toll-like receptor and B-cell receptor signalling to PI3K/AKT pathways, was higher in HGCA tissues than LGCA and NC tissues. In LGCA and HGCA cell lines, there were higher levels of PI3K catalytic subunit p110 γ , and PIK3C2A, which releases insulin in response to glucose and generates PtdIns3-phosphate in response to insulin. Furthermore, CDP-diacylglycerol-inositol 3-phosphatidyltransferase (CDIPT), which helps to regulate the levels of free PtdIns and was identified in the TCGA study, was significantly reduced in LGCA and HGCA cells, highlighting a possible cause for increased PI3K signalling. AKT2 was detected at significantly higher levels in LGCA and HGCA cells relative to NC. Similarly, calcium sensing and binding proteins S100P, S100A8 and S100A9, also identified by Vasaiker *et al.* (Vasaikar et al., 2019), were significantly upregulated in LGCA and HGCA tissues relative to NC tissues.

Table 6.18: PI3K signalling-related proteins detected in NC and CA tissues and tissue-derived cell lines

Gene name	Description	Abundance (cells)	Abundance (tissues)
PIK3R1	PI3K regulatory p85 subunit α	No change	↑LG/NC
PIK3CG	PI3K catalytic p110 subunit γ	↑LG/NC	No change
PIK3C2A	Glucose-induced insulin release via AKT1. Insulin-induced PtdIns3P synthesis.	↑LG/NC; ↑HG/NC	No change
PIK3AP1	Adapter protein between TLR/BCR and PI3K/AKT	No change	↑HG/NC; ↑HG/LG
CDIPT	PtdIns biosynthesis and inositol exchange	↓LG/NC; ↓HG/NC	No change

AKT2	One of three closely related serine/threonine kinases (AKT1, AKT2, AKT3)	↑LG/NC; ↑HG/NC	No change
S100P	Contributes to cellular calcium signalling	No change	↑LG/NC; ↑HG/NC
S100A8/9	Calcium and zinc binding, regulation of inflammatory processes and immune response via ERK1/2, PI3K/AKT and NF-κB signalling. Usually found as a heterodimer called calprotectin.	No change	↑HG/LG

Ras is a small G-protein which can be activated by GPCRs and RTKs (Fig 6.20). The genes *KRAS* and *NRAS* code for ras proteins and commonly have activating mutations in CRC (Cancer Genome Atlas, 2012; Vanhaesebroeck et al., 2010). EGFR is one RTK that activates ras, allowing it to initiate the Raf-MEK1/2-ERK1/2 MAP kinase cascade and to activate PI3Ks (Papadatos-Pastos et al., 2015; Vanhaesebroeck et al., 2010). Similarly, PDGF is able to stimulate the production of PtdIns3-phosphate to create a pool of substrate for the formation of other PIP molecules. In addition, active ras causes PLC to become active and convert PIP₂ into DAG and IP₃. KRAS and EGFR were detected in CA-derived cells and in the CA tissues, respectively. KRAS was higher in HGCA and LGCA cells relative to the NC cell line, and EGFR was unchanged in cells but was lower in HGCA tissues than LGCA and NC tissues.

c-Src is a key mediator of ATII effects (Fig 6.21) (Mehta et al., 2007). It performs kinase roles in a non-RTK-dependent manner by associating with GPCRs such as AT₁R. Once recruited to active AT₁R, c-Src is phosphorylated and goes on to activate Pyk2 (Mehta et al., 2007). In this capacity, Pyk2 contributes to further c-Src activation, as well as activating PDK1 which is the link between PI3K and AKT (Mehta et al., 2007; Papadatos-Pastos et al., 2015). Furthermore, Pyk2 recruits c-Src to phosphorylate JAK2 at the top of the JAK/STAT signalling cascade (Balakumar et al., 2014). AT₁R can activate this pathway in two ways. The first is G-protein-independent, whereby AT₁R recruits the c-Src/SHP-2 complex to phosphorylate JAK2. The second involves AT₁R phosphorylating a G_{αq} protein complex to activate PLC and convert PIP₂ into DAG and IP₃, as discussed above. IP₃ causes Ca²⁺ release from the ER and DAG activates PKC-δ; then PKC-δ activates Pyk2 in a Ca²⁺-dependent manner (Balakumar et al., 2014). JAK2 phosphorylates STAT1/3/6 and Rho kinase to affect gene transcription and cause vasoconstriction, respectively (Balakumar et al., 2014). AT₁R can produce DAG via a second mechanism. AT₁R recruits G_{α12/13} and G_{βγ} complexes to phosphorylate c-Src and RhoA, which go on to activate PLD. PLD converts phosphatidylcholine into

phosphatidic acid (PA), which is subsequently converted by PA phosphohydrolase into DAG (Balakumar et al., 2014). As mentioned above, DAG-induced PKC activation and IP₃-induced Ca²⁺ release are required for AT₁R-mediated EGFR transactivation and JAK/STAT signalling (Balakumar et al., 2014). The JAK/STAT pathway is also downstream of MAP kinase signalling, as the STATs can be phosphorylated by ERK1/2, p38 and JNK. JAK1 was detected at significantly higher levels in LGCA and HGCA cells relative to NC; however, as opposed to JAK2, JAK1 is primarily involved in interferon signalling and not the JAK/STAT pathway.

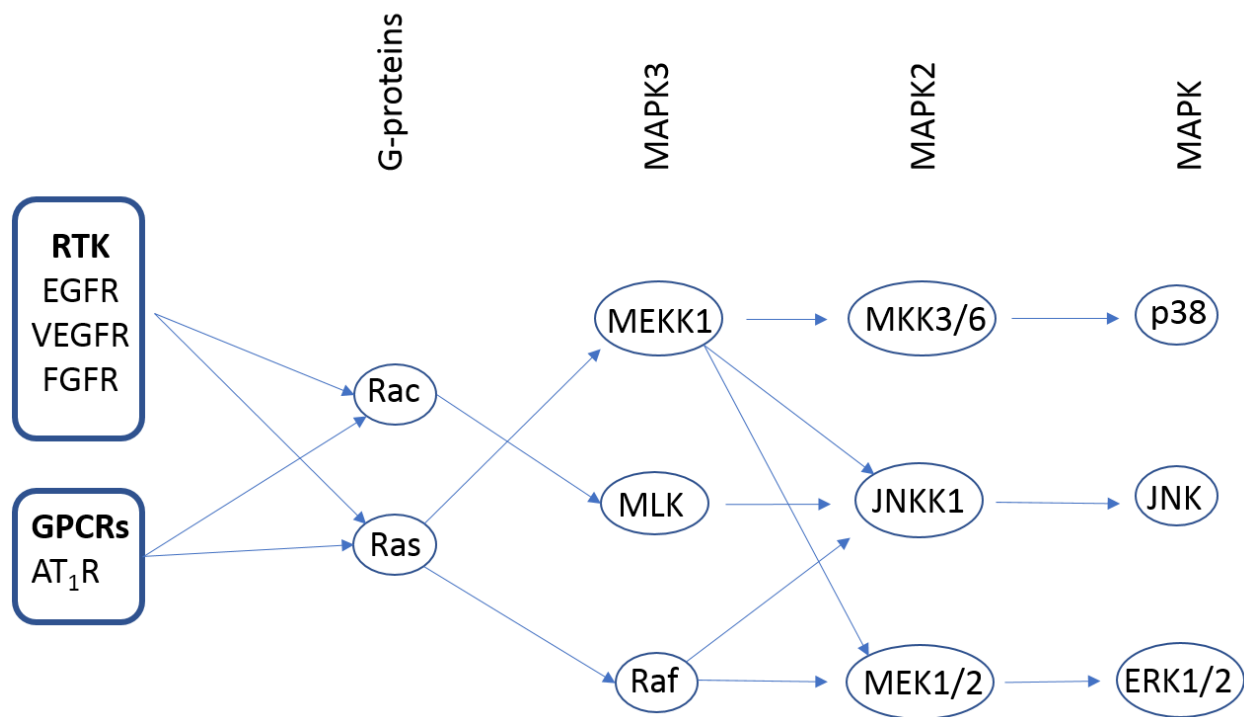


Figure 6.20: AT₁R-MAP kinase signalling. When activated by ATII, AT₁R can initiate MAP kinase signalling via ras and rac.

TGF- β signalling is involved in the synthesis of ECM components fibronectin and collagen and the production of tissue inhibitors of metalloproteinases (TIMPs), fibroblast proliferation, and EMT (Wolf, 2006). For these reasons, it is often implicated in the tumour microenvironment, including immunomodulation, fibrosis and tumour stroma production, and with migration and metastasis. TGF- β binds to the type 2 TGF- β receptor (TGF β R2), which then recruits the type 1 receptor (TGF β R1). TGF β R1 causes the phosphorylation of SMADs, which influence gene transcription.

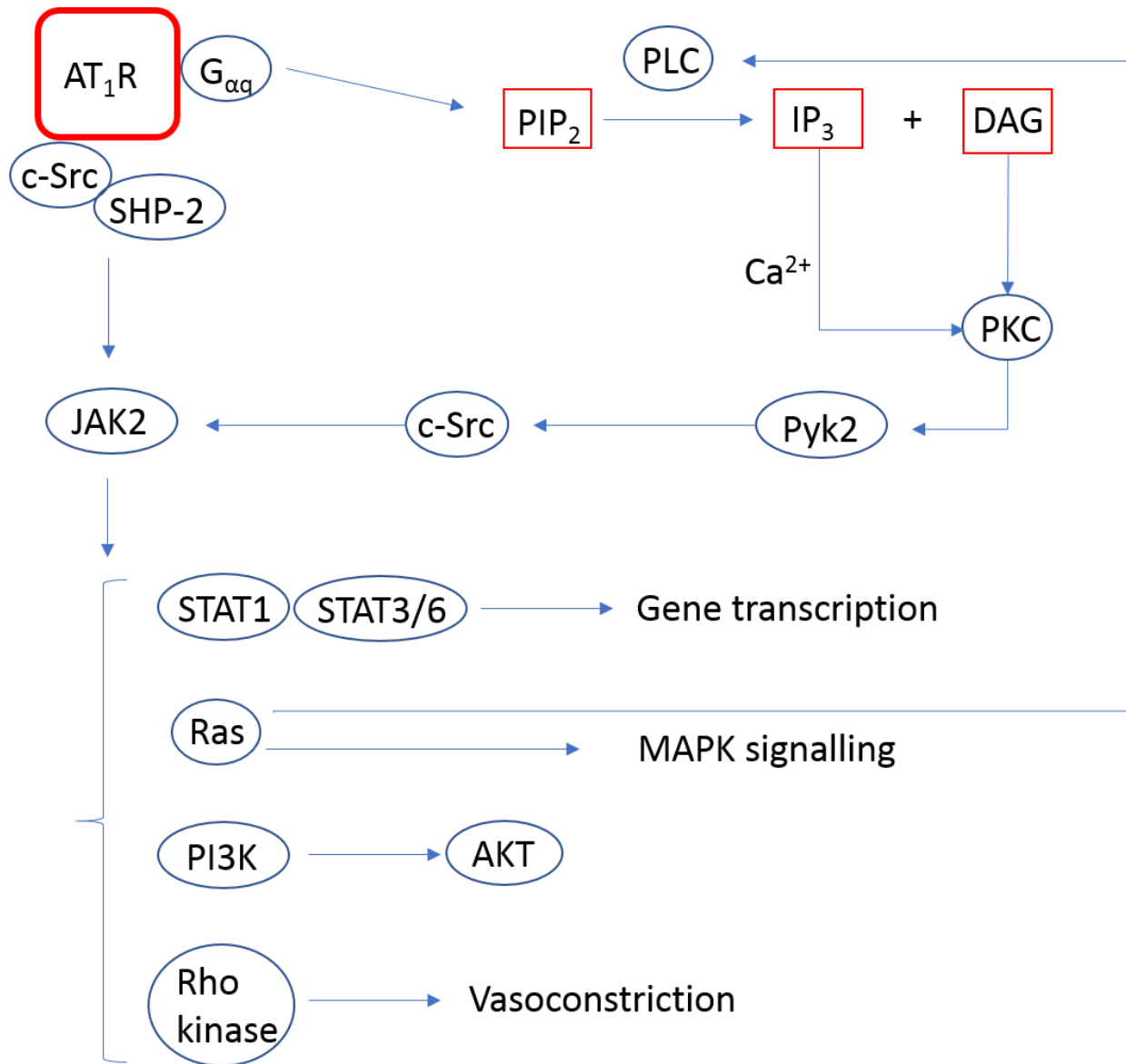


Figure 6.21: AT₁R-c-Src-JAK/STAT signalling. When activated by ATII, AT₁R can initiate JAK/STAT signalling via G-proteins or the non-receptor tyrosine kinase c-Src.

AT₁R sits upstream of EGFR-induced TGF- β signalling (Wolf, 2006). ATII stimulation of AT₁R causes the transactivation of EGFR to induce a cascade via Ras/Rac/MEKK1/JNKK to activate JNK, which then activates the c-Jun/c-Fos transcription factor complex (AP-1 early response transcription factor complex) to upregulate transcription of TGF- β . It has also been shown that AT₁R can directly phosphorylate SMADs in a TGF- β -independent manner, via ERK1/2 and the stress response p38 MAP kinase pathway (Wolf, 2006). Target genes of TGF- β signalling include those for p53, SMAD1 and c-Myc (Wolf, 2006).

Fibrosis, which can result from TGF- β signalling, is a physical barrier which prevents immune cell infiltration into a tumour, increases hypoxia and acidification, and exerts physical pressure on blood vessels (Pinter et al., 2017). ATII via AT₁R stimulates the release of IL-1, IL-6, IL-8 and VEGF from tumour-associated macrophages (TAMs) and cancer-associated fibroblasts (CAFs), stimulates TGF- β release and ECM deposition by CAFs, and promotes VEGF-mediated tumour angiogenesis (Pinter et al., 2017). The AT₁R blocker (ARB) losartan has been shown to partially alleviate these ATII-induced outcomes (Pinter et al., 2017).

PRR signalling also influences TGF- β -related pathways. When pro-renin or renin bind PRR, it activates p42/p44 MAP kinase signalling to upregulate profibrotic genes such as TGF- β , plasminogen activator inhibitor 1 (PAI-1/SERPINE1), FN1, collagen-1, and proinflammatory cytokines TNF- α and IL-1 (Balakumar et al., 2014). It also induces ERK1/2 and p38 MAP kinase signalling independent of the RAS (G. Nguyen, 2011); ERK1/2 activation by PRR leads to downstream profibrotic effects via TGF- β 1, plasminogen activator inhibitor 2 (PAI-2/SERPINE2), collagen, FN1 and COX2 (Ramkumar et al., 2019). Interestingly, angiotensin IV (ATIV) increases PAI-1/SERPINE1 via the type 4 angiotensin receptor (AT₄R) (Wolf, 2006); the SERPINS inhibit proteases including cathepsins and MMPs which are involved in RAS redundancy, cell migration and metastasis.

The relative abundance of PRR was not able to be determined in CA tissues due to the low confidence of identification, but it was found to be less abundant in CA cells relative to the NC cell line. PAI-1/SERPINE1 and PAI-2/SERPINE2 were significantly downregulated in HGCA cells relative to the NC cell line, and PAI-1/SERPINE1 was also significantly lowered in LGCA cells, but both were unchanged in tissue samples.

There were numerous RAS-related proteins detected in the cell lines and tissues. ACE2, which antagonises the adverse effects of AT₁R signalling by converting ATI into Ang1-9 and ATII into Ang1-7, was detected at elevated levels in LGCA and HGCA tissues relative to NC tissues. Aminopeptidase N (ANPEP) was detected with a greater abundance in LGCA cells than the NC cell line, which then fell again in HGCA cells, suggesting a role in tumour initiation. Cathepsins B and D are known to act as bypass mechanisms for the RAS, potentially allowing the RAS to continue functioning when classical targets are blocked by medications such as ARBs and ACEIs. Cathepsin B, which proteolytically activates renin, was significantly downregulated in CA-derived cells relative to the NC cell line. Similarly, there was a non-significant reduction of cathepsin D, which proteolytically activates renin and converts ATI to ATII, in HGCA and LGCA

cell lines. Taken together, these findings suggest that CRC tumours may have some capacity to bypass RAS blockade via ARBs or ACEIs.

6.3.5 Convergence of the RAS and Wnt signalling

PRR was found to be a vital component of the Wnt signalling receptor complex. The PRR transmembrane-ICD fragment and full length PRR provide a vital link between the Wnt receptor Frizzled (FZD) and low-density lipoprotein receptor-related protein (LRP) 5/6 (G. Nguyen, 2011; Nusse, 2005). Wnt proteins bind the FZD receptor, causing Dishevelled (Dvl) to be recruited and allowing LRP5/6 to polymerise with FZD (C. Gao et al., 2010; Khalek et al., 2010). PRR then provides a vital link between LRP5/6 and V-ATPase: to become active via phosphorylation, LRP5/6 must be in an acidic environment, and this is achieved through PRR co-localisation with proton pump V-ATPase, which acidifies the region in close proximity to LRP5/6 (G. Nguyen, 2011; Nusse, 2005; Ray, 2010). Once active, LRP5/6 binds Axin to prevent it from participating in the β -catenin destruction complex (Nusse, 2005). Importantly, PRR enables Wnt/ β -catenin signalling to occur via a mechanism which is independent of pro-renin or renin binding to PRR (L. Zhou et al., 2015). Interestingly, c-MYC, EpCAM, and RAS components ACE, AT₁R, AT₂R and PRR are among the Wnt-targeted genes which are transcribed by the β -catenin/Tcf/Lef complex (L. Zhou et al., 2015).

Two coiled-coil domain-containing (CCDC) proteins were found to influence the crosstalk between Wnt, RTK and GPCR signalling: CCDC88A/Girdin and CCDC88C/Daple.

Daple is a Dvl-binding protein which dissociates from Dvl when FZD is bound by the WNT5A ligand (Azhar et al., 2018). When this occurs, Daple binds FZD by displacing Dvl to stop canonical Wnt signalling; Daple then recruits G $_{\alpha i}$ protein complexes and releases G $_{\beta \gamma}$ complexes to inhibit cAMP signalling and initiate PI3K/AKT and MAP kinase signalling, respectively (Azhar et al., 2018). Dvl and G $_{\alpha i}$ bind to Daple in a mutually exclusive manner due to allosteric inhibition, so that when Dvl is overexpressed it is capable of saturating available Daple, preventing it from associating with G $_{\alpha i}$ and thereby favouring canonical Wnt signalling (Azhar et al., 2018). In this mechanism of action, Daple compartmentalises β -catenin with other Wnt-signalling components to enhance canonical Wnt (Azhar et al., 2017). However, multiple RTKs including EGFR, and non-RTKs including c-Src, are able to phosphorylate Daple in a manner which favours G $_{\alpha i}$ binding over Dvl. This increases downstream signalling via Rac1 and AKT, promotes EMT, and leads to β -catenin accumulation with E-cadherin at the cell membrane, therefore reducing canonical Wnt

signalling (Aznar et al., 2018; Aznar et al., 2017). Disruption of the Daple-AKT axis allows canonical Wnt signalling to proceed without restriction (Aznar et al., 2017).

Girdin is a guanine nucleotide exchange factor (GEF) which associates with RTKs and non-RTKs. Similar to Daple, Girdin binds $G_{\alpha i}$ complexes to inhibit cAMP and therefore reduce proliferation, and releases $G_{\beta \gamma}$ complexes which induce PI3K/AKT signalling (Gupta et al., 2016). Girdin is able to recruit phosphorylated p85 PI3K subunits to the EGFR to enable EMT, cell migration and metastasis via actin remodelling (Garcia-Marcos et al., 2011; Jun et al., 2013). It also induces PI3K to activate AKT1, which phosphorylates GSK3 β and reduces β -catenin degradation. Unlike Daple, which accelerates EGFR degradation, the Girdin/ $G_{\alpha i}$ complex prolongs EGFR signalling from the cell membrane, and when Girdin is not bound to $G_{\alpha i}$ complexes it prolongs EGFR signalling from endosomes via MAPK/ERK (Gupta et al., 2016).

Daple and Girdin were both detected by MS. Girdin, which enables metastasis and is a prognostic indicator in CRC (Garcia-Marcos et al., 2011; Jun et al., 2013), was significantly more abundant in HGCA cells than LGCA cells. Daple, a key regulator of Wnt signalling, was significantly upregulated in LGCA tissues relative to NC. WNT5A was found to be significantly upregulated in HGCA and LGCA cells and was the only Wnt ligand to be detected.

6.3.6 The RAS and pluripotency

As discussed above, the RAS via ATII/AT₁R stimulates signalling pathways including PI3K/AKT, JAK/STAT and TGF- β pathways (Fig 6.22). It has also been shown that these pathways are involved in complex networks which influence the transcription of pluripotency markers.

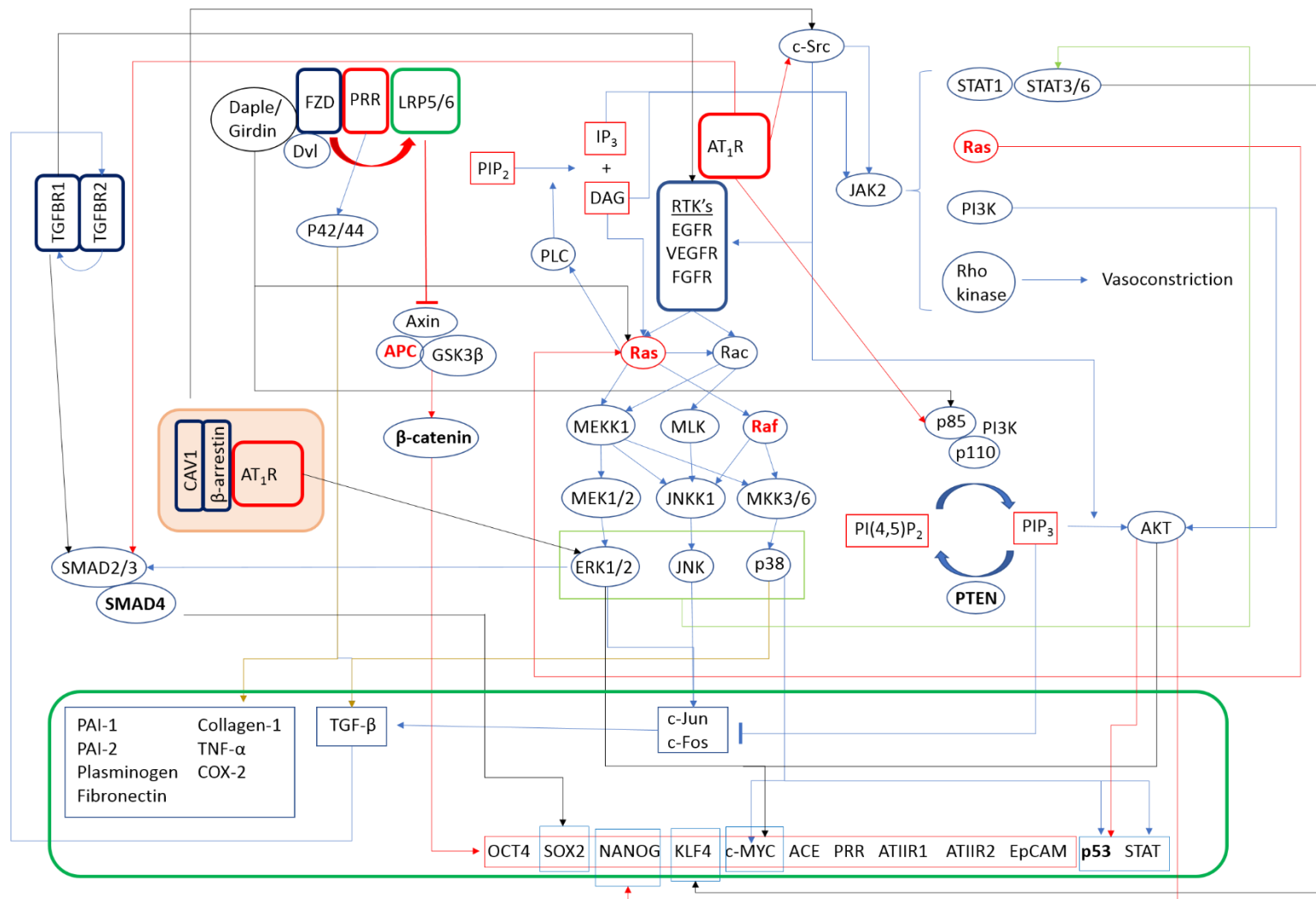


Figure 6.22: Synthesis of relevant pathways. Colours are simply to make it easier to see; however, bold type signifies commonly mutated proteins in CRC.

The core pluripotency factors are OCT4, SOX2 and NANOG (Takahashi et al., 2007; Takahashi et al., 2006). The OCT4/SOX2 complex is capable of transcribing NANOG, and the SOX2/NANOG complex is able to control OCT4 transcription (Masui et al., 2007; Niwa et al., 2009; G. Pan et al., 2007). In mouse ESCs, Niwa *et al.* (Niwa et al., 2009) demonstrated that the JAK/STAT pathway activates *KLF4* which then activates *SOX2*, and the PI3K-AKT pathway leads to *NANOG* activation. The OCT4/SOX2/NANOG complex transcribes *STAT3* to further potentiate this pathway, as well as transcribing components of TGF- β signalling which in turn increases *SOX2* expression via SMADs (Ikushima et al., 2009). Expression of *c-MYC* is also induced by STATs, as well as by AKT and ERK (Mehta et al., 2007). However, SMADs reduce *c-MYC* transcription and MAP kinase signalling lowers AKT-induced *NANOG* transcription (Niwa et al., 2009).

As outlined above, the activation of the RAS via ATII/ AT₁R initiates signalling via RTKs and GPCRs, including EGFR, TGF β R1/2 and AT₁R itself. This occurs through pathways such as PI3K/AKT, MAPK/ERK, JAK/STAT and TGF- β /SMAD. Each of these pathways, plus β -catenin/Wnt signalling, contribute in some way to the transcription of pluripotency genes OCT4, SOX2, NANOG, KLF4 and *c-MYC*, and RAS components PRR, ACE, AT₁R and AT₂R. Therefore, taking into account the downregulation of RAS bypass members (CTSB/D, ANPEP) and upregulation of the beneficial ACE2 in CA samples, modulation of the RAS via classic targets such as AT₁R and ACE may be able to target CSCs by reducing the expression of pluripotency drivers and thereby attenuate their function.

6.4 Conclusions

Proteomic analysis revealed that when compared to their NC controls, the CA tissues and CA-derived cell lines are enriched for proteins which are involved in RNA processing and mitochondrial respiration. MS identified abundance changes in members from a range of different signalling pathways which integrate with the RAS, often being activated via PRR or AT₁R. This includes the Wnt signalling pathway which is crucial to CA and initiates gene transcription for RAS components and iPSC markers by the β -catenin transcription complex. Various components which can bypass the RAS or cause it to branch, including ACE2 and cathepsins B and D, were detected in CA tissue and cell samples, highlighting that treatment with RAS modulators could be circumvented. The alignment of MS data from both the CA tissues and CA-derived cell lines with each other and with published data from other CRC proteogenomic studies suggests that the cell lines will be a useful *in vitro* model for CA tumours, though they are not a complete reflection of the complex CA tissues.

Chapter 7: Primary Colon Adenocarcinoma Stem Cells Express Components of the Renin-Angiotensin System

7.1 Introduction

The classical RAS maintains blood pressure and fluid levels by controlling sodium absorption, vascular tone and hormone release (Fountain et al., 2019), and its involvement in cancer is an emerging field of research (Munro, Wickremesekera, Davis, et al., 2017). The RAS is summarised in Section 1.5 of the general discussion (Fig 1.4). Many studies have shown that the RAS is involved in cancer-related processes such as angiogenesis, proliferation, tumourigenesis and metastasis (Munro, Wickremesekera, Davis, et al., 2017; Roth et al., 2019). Furthermore, downstream effects of AT₁R and AT₂R seem to play antagonistic roles – AT₁R has an association with adverse or cancer-related outcomes, especially when dysregulated, whereas AT₂R may promote activity which prevents or reduces malignancy (Fountain et al., 2019; Munro, Wickremesekera, Davis, et al., 2017). AT₁R and AT₂R are localised to the plasma membrane and AT₂R to the nuclei of CRC cells, promoting ATII-induced tumour growth, invasion and VEGF-A secretion (Zhou et al., 2014). Investigations into the role of the RAS in CRC have included retrospective studies showing that CRC patients taking RAS inhibitors display reduced incidence of polyp formation, CRC and metastasis (Childers, 2015). Specifically, AT₁R inhibition has been observed to have the greatest effect in reducing the development and metastasis of solid tumours (Childers, 2015). A large-scale meta-analysis found a positive trend in CRC patient survival associated with angiotensin receptor blockers (ARBs) and ACE inhibitors (ACEIs), with up to 20% increases in DFS, PFS and OS rates (Sun et al., 2017).

EMT is thought to be vital to metastasis (L. Nguyen et al., 2016). The downstream effects of ATII binding to AT₁R are consistent with those seen in EMT, including increased cellular migration (L. Nguyen et al., 2016). This suggests that the RAS may play a key role in CRC metastasis.

Cathepsins are known to be involved in cancer by activating angiogenic growth factors and proliferative cytokines and by degrading the ECM to allow tumour invasion and metastasis (S. Chen et al., 2017; Ruan et al., 2015; Turk et al., 2012). However, there is evidence of redundancy between the cathepsins, making them a difficult therapeutic target, especially given their crucial roles in normal physiology (Aggarwal et al., 2014; S. Chen et al., 2017). Cathepsins B and D are two of the most commonly studied cathepsins in a range of cancers (S. Chen et al., 2017). Despite there being elevated levels of cathepsins in CRC relative to normal tissue, there is no correlation between cathepsin levels and tumour grade in CRC (Kuester et al., 2008). In fact, despite a lack of focus on the activity of cathepsins in the stroma, there is evidence that the expression levels of cathepsin D by tumour-

associated macrophages has a greater influence on prognosis than its expression by cancer cells (Kuester et al., 2008). There is some evidence that cathepsin B influences local invasion more than distant metastasis (Kuester et al., 2008), including in CRC (Bian et al., 2016). This is partly through its degradation of E-cadherin to reduce cell-cell adhesion (Ruan et al., 2015), thereby leading to tumour budding, the process by which single or small clusters of tumour cells migrate away from the tumour nests (Kuester et al., 2008). Tumour budding is associated with poor prognosis and is strongly correlated with the expression levels of cathepsin B, but not that of cathepsin D, by the budding cells (Kuester et al., 2008). Interestingly, cathepsin G activity increases the formation of E-cadherin/ β -catenin complexes and therefore cell-cell adhesion, reducing cell motility and causing cells to clump together (Kudo et al., 2009). Research regarding the role of cathepsins in distant metastasis has found that they enhance the success of metastatic cells once they arrive at their secondary site by deriving growth factors, amino acids and nutrients from the ECM via nutrient recycling (Garcia et al., 1996).

Cathepsin B is upregulated by the p38, MAP kinase and K-ras signalling pathways (Kuester et al., 2008). Cathepsin B is expressed at low levels in the epithelium of normal colon where it is found near the luminal surface; however, it is upregulated in CRC and found deeper in the crypts, often seen diffusely in the cytoplasm of poorly-differentiated tumours (Gondi et al., 2013; Kuester et al., 2008). It can be secreted in exosomes by tumour cells to act extracellularly by degrading the ECM to enhance invasiveness and migration (Bian et al., 2016). Localisation of cathepsin B to the basal membrane and its secretion from the cell coincides with *KRAS* activation, a common mutation in CRC, and occurs around the time of malignant transformation from an adenoma to adenocarcinoma (Bian et al., 2016). However, an increase in cathepsin B mRNA levels occurs early in CRC, associated with mutation of the *APC* gene, which is considered to be one of the initiating mutations of CRC (Bian et al., 2016). Cathepsin B expression in the tumour cells of CRC has previously been investigated, collectively finding that more than 70% of CRC cases express cathepsin B but that its abundance does not increase with tumour grade (Bian et al., 2016; Chan et al., 2010; Khan et al., 1998). In mouse models of pancreatic and breast cancer, cathepsin B depletion reduces tumour burden (Turk et al., 2012). Furthermore, it has been proposed as a prognostic marker in human cancers, especially in breast cancer, where its serum levels correlate with the presence of metastasis (Garcia et al., 1996; Turk et al., 2012).

Pro-cathepsin D is secreted extracellularly by cancer cells and tumour-associated macrophages where it has a pro-tumoural effect by increasing proliferation, invasion and survival of fibroblasts, and activation of MAP kinase signalling (Benes et al., 2008; Liaudet-Coopman et al., 2006). These actions are independent of its proteolytic activity and are likely to be induced through receptor binding, because cathepsin D does not have enzymatic activity at a neutral pH (Liaudet-Coopman et al., 2006). Secreted inactive pro-cathepsin D binds to the extracellular domain of LRP1 on fibroblasts,

preventing the phosphorylation and activation of LRP1 and triggering fibroblast outgrowth (Derocq et al., 2012; Liaudet-Coopman et al., 2006).

Kanber et al. (Kanber et al., 2002) used IHC to analyse the expression of cathepsin D in CA and found that 100% of their samples expressed cathepsin D in the stroma, and 90.3% showed weak to moderate staining of the tumour epithelial cells.

Although cathepsins B and D are expressed in a range of cancers by tumour cells, infiltrating immune cells and sometimes microvessels, cathepsin G appears to be restricted to neutrophils and mast cells (Featherston et al., 2017; Koh et al., 2017; Mehrotra et al., 2018; R. M. A. Rahman et al., 2019), and so its ability to produce ATII *in vivo* depends on cathepsin G being released into the tumour microenvironment.

The aim of this chapter was to assess the expression and localisation of components of the RAS, including cathepsins, in CA tissue samples, tissue-derived primary cell lines, and patient-matched NC tissue. The first hypothesis was that RAS components, including cathepsins, would be expressed by CA tissues and cells, and that some of the RAS components would be expressed by the proposed CSCs that also express iPSC markers. Furthermore, it was also hypothesised that as the CA grade increased, the abundance of PRR, ACE and the cathepsins would increase and the abundance of ACE2 and AT₂R would decrease.

IHC staining was carried out to determine protein localisation within CA tissue. 10 LGCA and 8 HGCA tissue samples were investigated for the distribution and abundance of PRR, ACE and AT₂R. A sub-cohort of 6 LGCA and 6 HGCA tissues were probed for cathepsins B, D and G. A selection of 3 LGCA and 3 HGCA tissues were stained for ACE2. EpCAM-sorted CA-derived primary cell lines, as described in Chapter 5, were interrogated for their expression of RAS components and cathepsins B and D using WB. RT-qPCR confirmed gene transcription in tissues and cell lines. Finally, IF staining was carried out to co-localise components of the RAS with cells expressing iPSC markers in the tumour epithelium and stroma. Based on the protein expression of iPSC markers as outlined in Chapter 4, a subpopulation of CSCs in the CA epithelium has been proposed, expressing NANOG, SOX2, KLF4, c-MYC, CD133 and EpCAM. A subpopulation of cells in the stroma expressed OCT4, and may represent migratory cancer cells, cancer-associated fibroblasts, or cells which are undergoing induced de-differentiation (Munro et al., 2019). This chapter sought to clarify whether these subpopulations expressed components of the RAS.

7.2 Results

7.2.1 ACE and ACE2 are expressed variably by CA tissues and CA-derived cell lines

ACE was not detected by IHC staining in NC tissues (Fig 7.1B,D,F,H), except for the endothelial layer of microvessels. In CA tissues, ACE appeared to be weakly positive in the epithelial cytoplasm, and more pronounced on the luminal surface of epithelial cell membranes. Overall, the staining was weaker in HGCA tissues (Fig 7.1E,G) compared to LGCA tissues (Fig 7.1A,C). The antibody was also non-specifically seen in red blood cells. Only one NC sample displayed positive staining for ACE2 (Fig 7.2F), where it was seen on the luminal surface of crypt epithelial cells. In all 6 CA tissues, epithelial cells (Fig 7.2A,C,E,G) exhibited moderate staining in the cytoplasm, with stronger membranous staining at the luminal surface and occasionally at the stromal surface.

Neither ACE nor ACE2 were detected by WB in any cell lines, but were present in the positive controls at approximately 195 kDa and 110 kDa, respectively (Fig 7.3A,B).

ACE was detected by RT-qPCR in all LGCA tissues and patient-matched NC tissues (Fig 7.4A) with mRNA levels significantly higher in 3 of the tumours relative to NC, not statistically different in 2 cases, and significantly reduced in 1 case. ACE mRNA was detected in all HGCA tissues and patient-matched NC tissues, with 1 sample displaying significantly higher expression than its matched NC tissue and another having significantly decreased expression in the tumour. There was no difference in ACE mRNA expression between the CA tissues and their matched NC tissues from the other 4 patients.

In the cell lines, ACE mRNA was detected by RT-qPCR in all of the CA-derived EpCAM^{High} (+) and EpCAM^{Low} (-) cells, but at significantly lower levels than the pool of NC tissues used as a reference (Fig 7.5A). ACE2 mRNA was detected in EpCAM^{High} (+) and EpCAM^{Low} (-) cells from samples LGCA2 and HGCA2, and in the EpCAM^{Low} (-) cells from LGCA1 and the EpCAM^{High} (+) cells from LGCA3. However, the abundance was much lower than in the pooled NC tissue reference (Fig 7.5B).

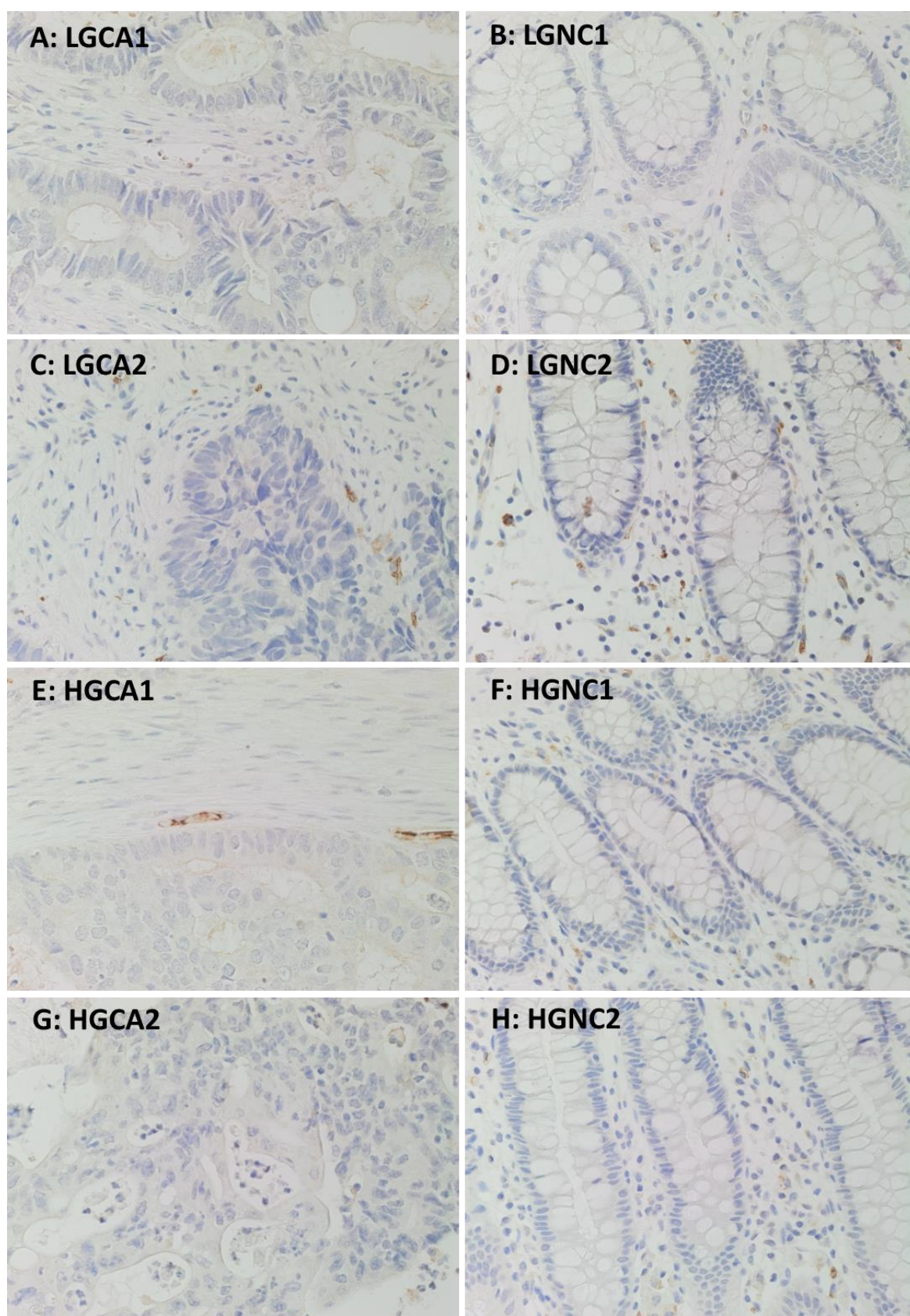


Figure 7.1: IHC staining for ACE. Representative images showing IHC staining for ACE, including 2 LGCA (A, C) and their patient-match NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

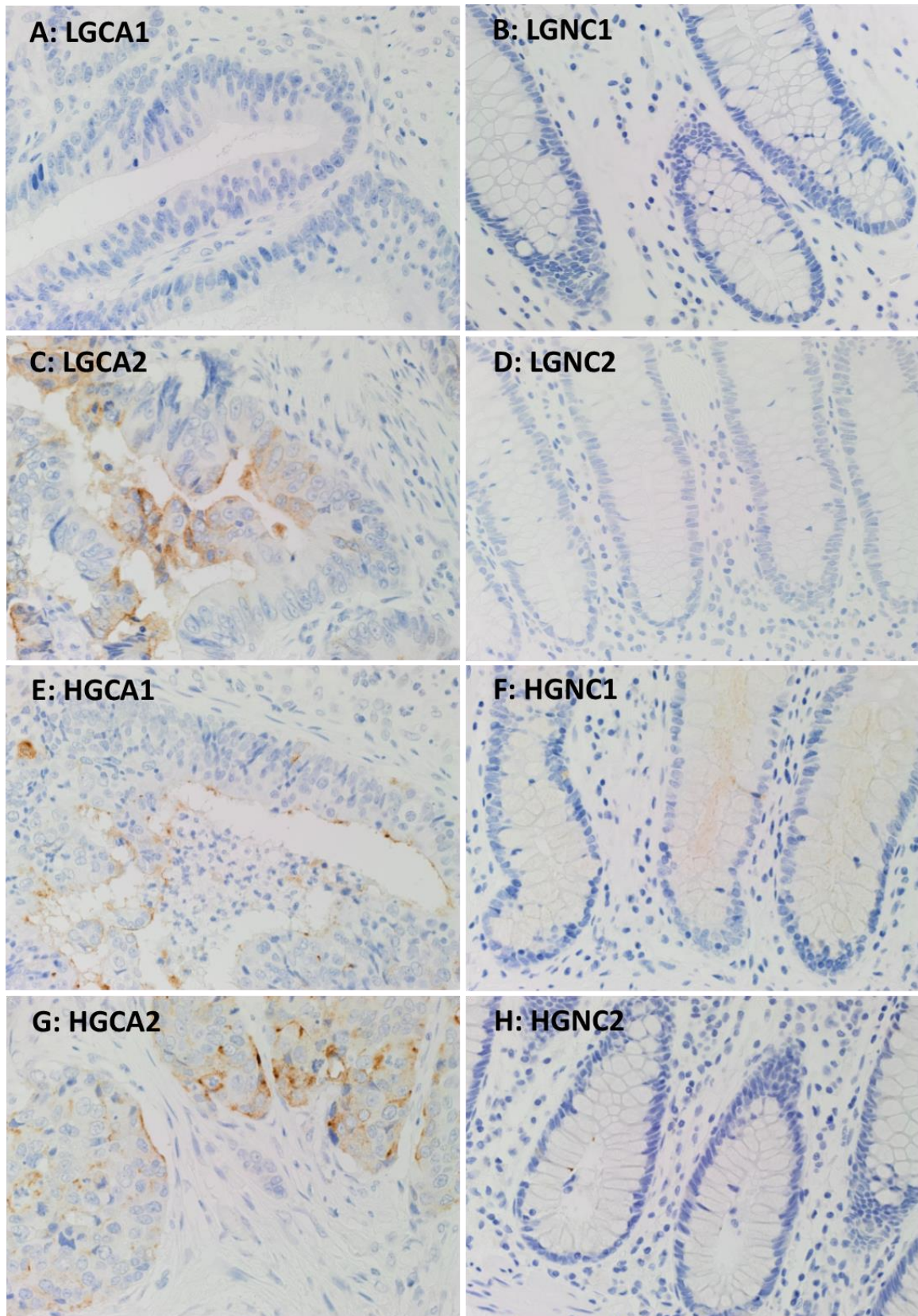


Figure 7.2: IHC staining for ACE2. Representative images showing IHC staining for ACE2, including 2 LGCA (A, C) and their patient-match NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

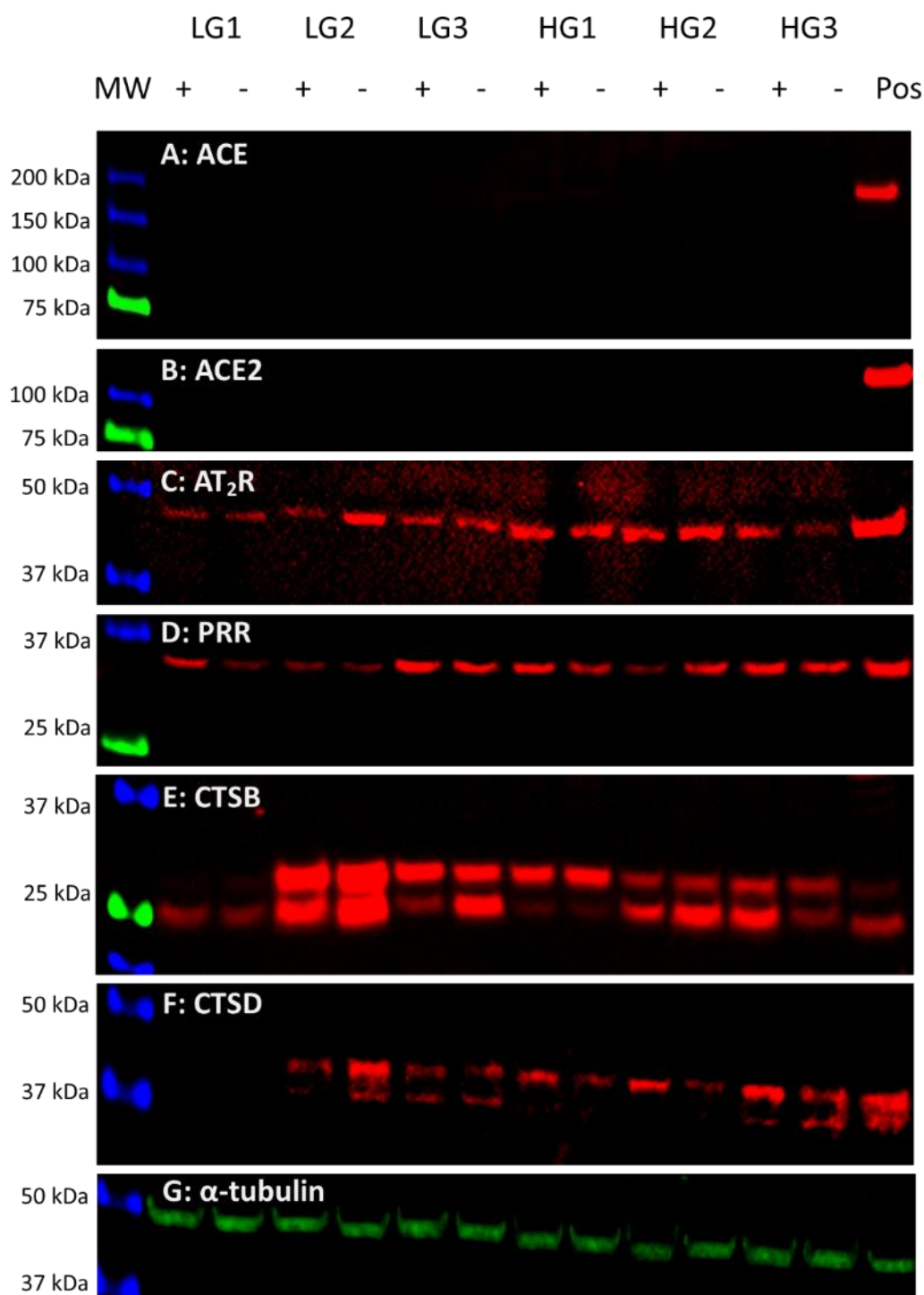


Figure 7.3: Western blot for RAS components. Protein extractions from LGCA (n=3) and HGCA (n=3) cell lines that had been sorted into EpCAM^{High} (+) and EpCAM^{Low} (-) subpopulations were probed for ACE (A; 195 kDa), ACE2 (B; 110 kDa), AT₂R (C; 45 kDa), PRR (D; 35 kDa), cathepsin B (E; 24 kDa and 27 kDa) and cathepsin D (F; 43 kDa and 46 kDa). Positive controls included tonsil (PRR), mouse lung (ACE), human kidney (ACE2) and HepG2 cells (AT₂R, Cath B, Cath D). α-Tubulin (G; 50 kDa) was used as a loading control.

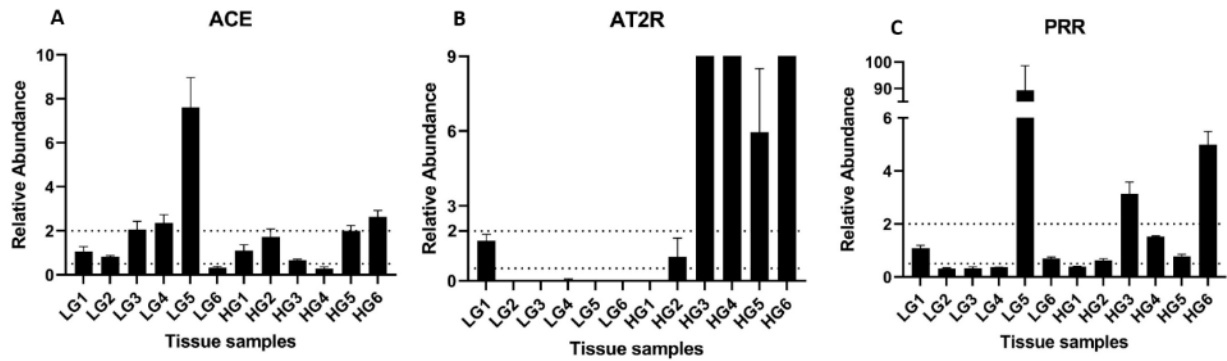


Figure 7.4: RT-qPCR for RAS components in CA tissues. RNA was extracted from LGCA (n=6) and HGCA (n=6) tissue samples and patient-matched NC tissue, and RT-qPCR was carried out to measure the mRNA levels of ACE (A), AT₂R (B) and PRR (C). The abundance of each marker in each tumour tissue was measured relative to their patient-matched NC tissue sample. The average relative abundance from triplicate values are displayed, with error bars representing standard deviation.

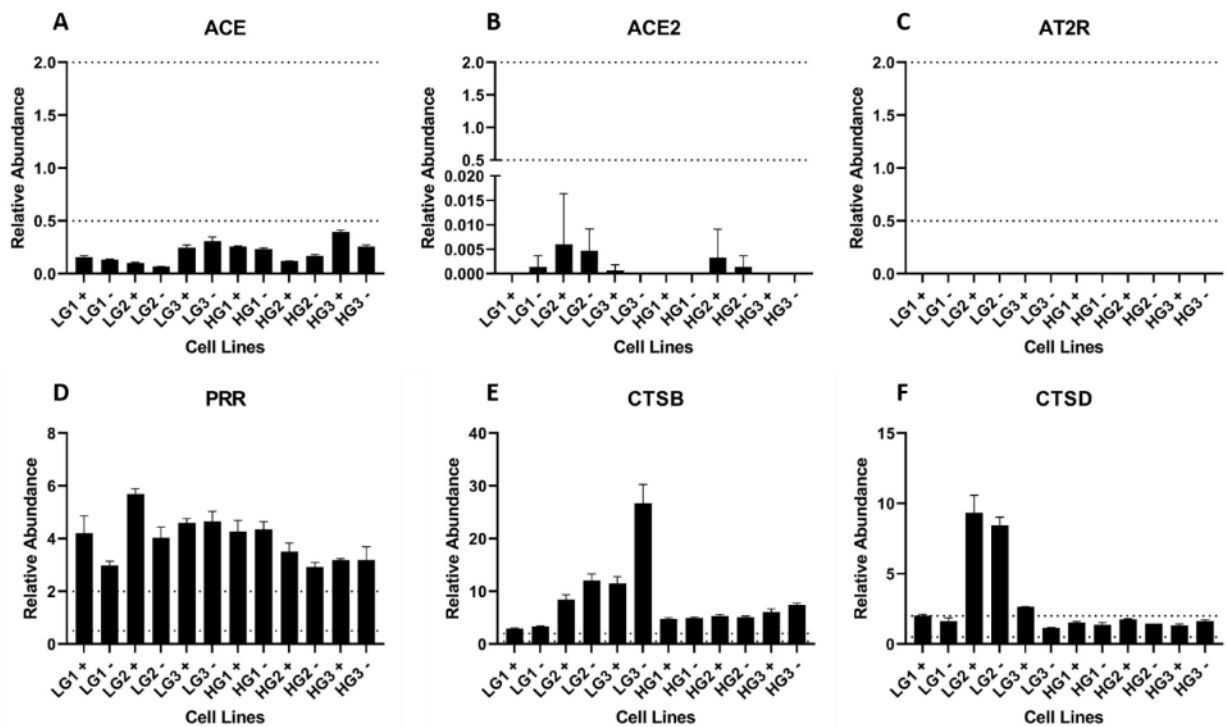


Figure 7.5: RT-qPCR for RAS components in CA-derived cells. RNA was extracted from EpCAM^{High} (+) and EpCAM^{Low} (-) cells derived from 3 LGCA and 3 HGCA cases, and PCR was performed to measure the mRNA levels of ACE (A), ACE2 (B), AT₂R (C), PRR (D), Cathepsin B (E) and Cathepsin D (F). The abundance of each marker was measured relative to the pooled data generated from 4 NC tissues, all run separately as triplicates in each PCR reaction. The average relative abundance from triplicates are displayed, with error bars representing standard deviation.

7.2.2 AT₂R expression in CA tissues is proportional to grade

In NC tissues, AT₂R (Fig 7.6B,D,F,H) IHC staining was negative-to-weak overall, and seemed to be cytoplasmic or concentrated to the luminal membrane of epithelial cells when present, with some possible nuclear staining. Staining for AT₂R across CA samples (Fig 7.6A,C,E,G) ranged from negative to strong; most positive cases displayed uniform moderate staining in the cytoplasm of epithelial cells, with some cases also displaying a more granular pattern within the cytoplasm. Some stromal cells stained positive, most of which could be attributed to immune cells, though some elongated fibroblast-like cells also had staining.

A band corresponding to AT₂R was detected by WB in all EpCAM^{High} and EpCAM^{Low} cells (Fig 7.3C) at approximately 45 kDa.

AT₂R mRNA was detected in 2 of the 6 LGCA tissues, with 1 showing similar abundance to its matched NC tissue, and 1 with significantly reduced levels (Fig 7.4B). Conversely, it was detected in 5 of the 6 HGCA tissues but only in 2 of the patient-matched NC samples, where AT₂R mRNA levels were lower in 1 NC and similar in the other NC relative to their matched CA tissues. AT₂R mRNA was below detectable levels in all CA-derived EpCAM^{High} (+) and EpCAM^{Low} (-) cell lines (Fig 7.5C).

7.2.3 PRR is upregulated in CA tissues and CA-derived cells

IHC staining for PRR in NC tissues (Fig 7.7B,D,F,H) was strong in the layer of smooth muscle (muscularis mucosae) below the crypts and stroma (Fig 7.7B), and negative or weak in the cytoplasm of crypt epithelial cells. In CA samples (Fig 7.7A,C,E,G), staining was predominantly moderate in the cytoplasm of epithelial cells, and often moderate-to-strong at the cell membrane.

PRR was detected by WB at the expected molecular weight of 35 kDa (Fig 7.3D) in all EpCAM^{High} and EpCAM^{Low} cell lines.

PRR mRNA was detected in NC, LGCA and HGCA tissues (Fig 7.4C). Expression levels were significantly upregulated in 1 LGCA tumour, significantly downregulated in 3 LGCA tumours, and at similar levels in 2 LGCA tumours relative to their matched NC tissues. Similarly, PRR mRNA was significantly upregulated in 2 HGCA tissues, downregulated in 1 HGCA tissue, and not significantly different in the remaining 3 HGCA tissues relative to their patient-matched NC tissues. All CA-derived EpCAM^{High} (+) and EpCAM^{Low} (-) cells expressed significantly more PRR mRNA than the pooled NC tissue reference (Fig 7.5D).

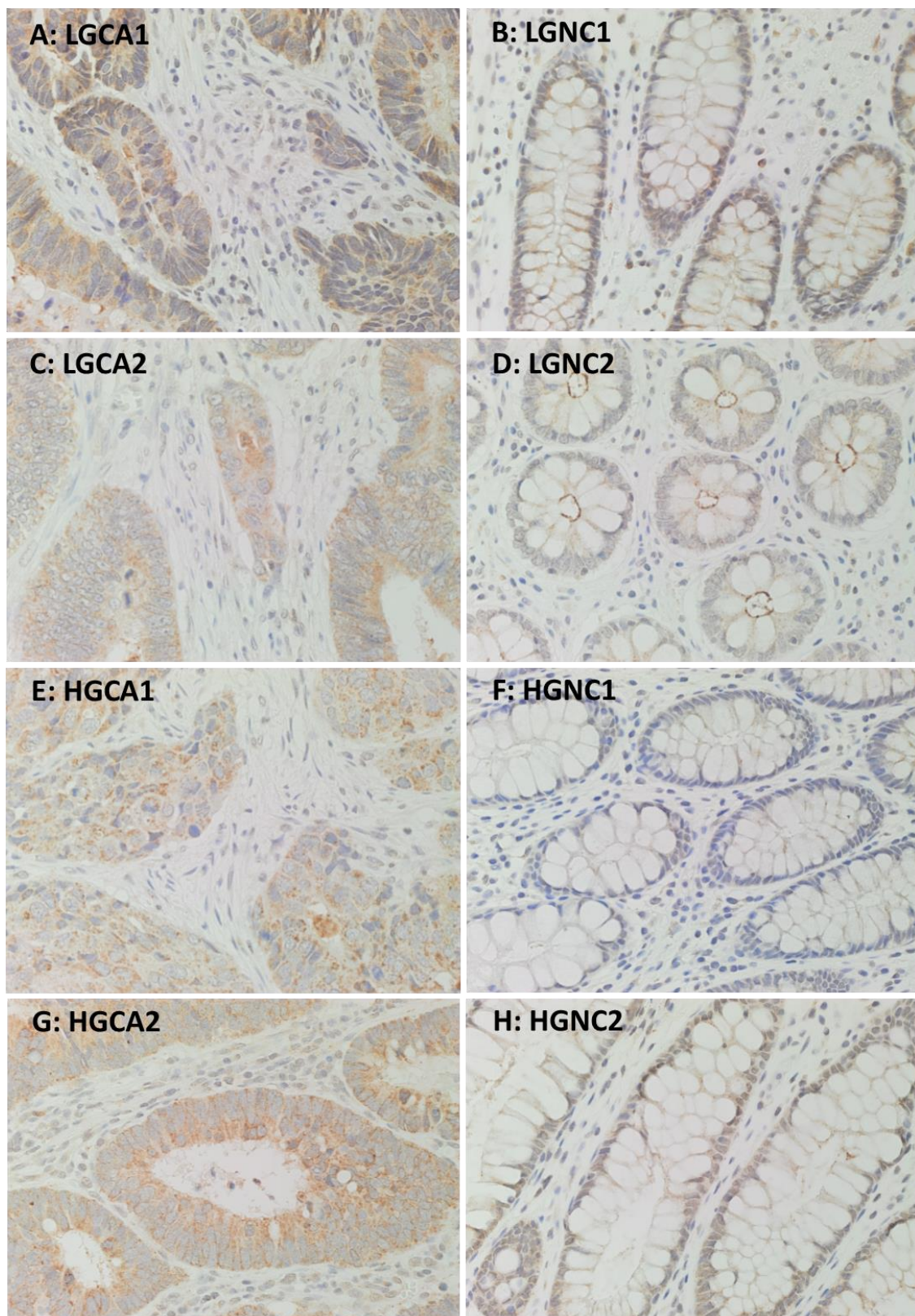


Figure 7.6: IHC staining for AT₂R. Representative images showing IHC staining for AT₂R, including 2 LGCA (A, C) and their patient-matched NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

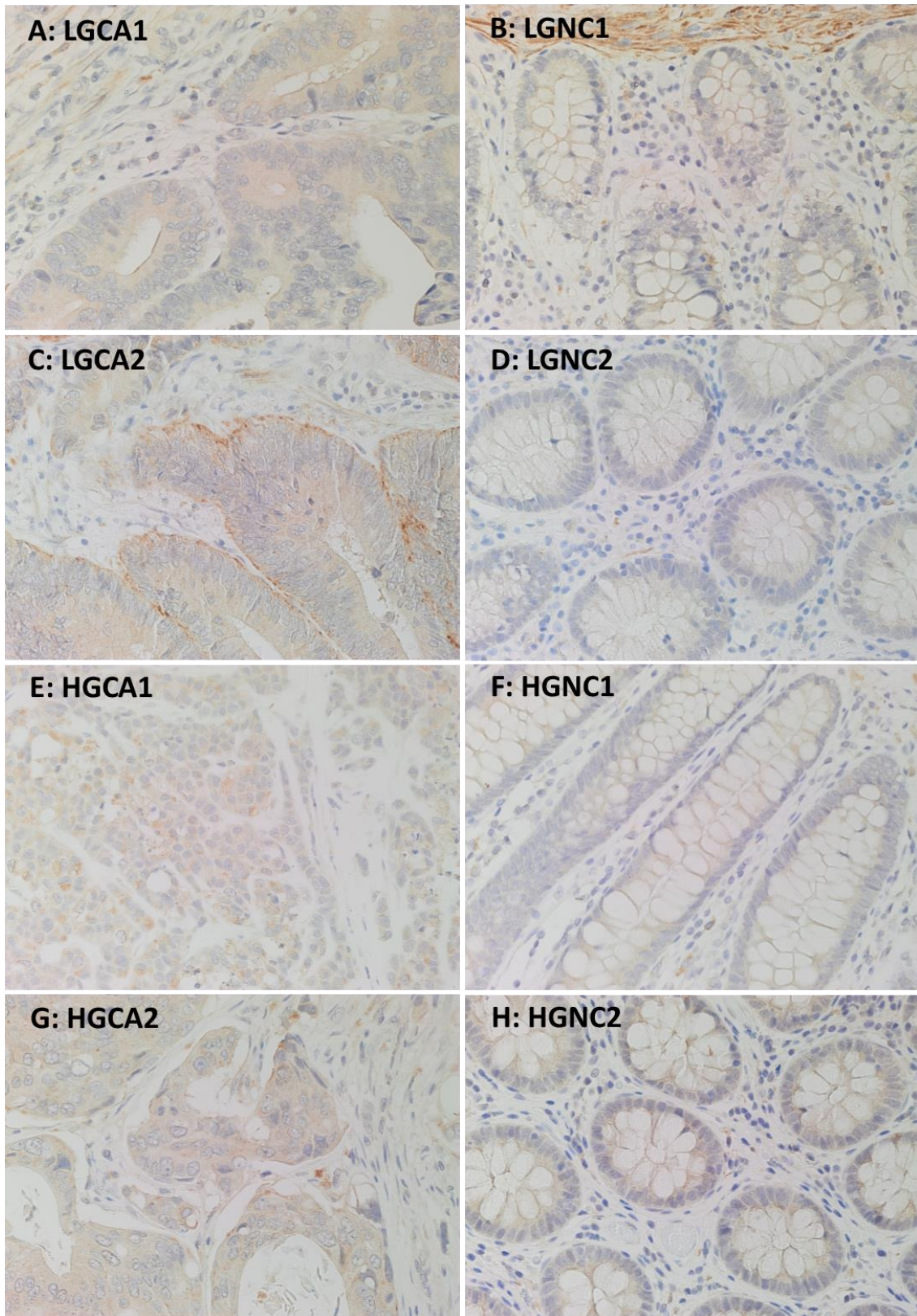


Figure 7.7: IHC staining for PRR. Representative images showing IHC staining for PRR, including 2 LGCA (A, C) and their patient-matched NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

7.2.4 Cathepsins are expressed in the CA epithelium and stroma

IHC staining found that cathepsin B (Fig 7.8) and cathepsin D (Fig 7.9) were highly expressed by immune cells in the stroma of NC and CA tissues, as well as by occasional crypt cells in NC tissues thought to be neuroendocrine cells. Epithelial cells in NC samples expressed low levels of cathepsin B, seen as a weak stain (Fig 7.8B,D,F,H). Cathepsin D was not expressed on the epithelium of most NC samples, and was weakly positive in a minority (Fig 7.9B,D,F,H). However, the epithelial cells of CA tissues stained weak-to-moderate for both cathepsins B (Fig 7.8A,C,E,G) and D (Fig 7.9A,C,E,G), with cathepsin D appearing more granular with spots of strong staining, possibly showing lysosomes or endosomes in which cathepsins generally function. Interestingly, staining for both cathepsins B and D was stronger in the epithelial cells of LGCA tissues than HGCA tissues. Cathepsin G (Fig 7.10) was expressed by scattered cells within the stroma, thought to be mast cells which typically express cathepsin G, as well as some tumour infiltrating immune cells. The stain also appeared non-specifically in necrotic areas (Fig 7.10A).

Bands for Cathepsin B (Fig 7.3E) were detected by WB in all CA-derived EpCAM^{High} and EpCAM^{Low} cell lines. Two prominent bands were observed, corresponding to the heavy chain of cathepsin B with and without glycosylation (27 kDa and 24 kDa, respectively). Cathepsin D (Fig 7.3F) was detected in all HGCA-derived cell lines, and in two of the three LGCA-derived cell lines (both EpCAM^{High} and EpCAM^{Low} cells), where it was predominantly more abundant in EpCAM^{High} cells. Bands of two different sizes were detected, which represented pre-pro-cathepsin D (43 kDa), and glycosylated pro-cathepsin D (46 kDa).

RT-qPCR detected mRNA for Cathepsins B and D in all CA-derived EpCAM^{High} (+) and EpCAM^{Low} (-) cells. All of the cell lines had significantly higher levels of cathepsin B mRNA relative to the pooled NC tissue reference (Fig 7.5E). Cathepsin D expression in cell lines was similar to that of the NC pool for 3 LGCA-derived and all 6 HGCA-derived cell lines, and significantly upregulated in the EpCAM^{High} (+) and EpCAM^{Low} (-) cells from LG2 and the EpCAM^{High} (+) cells from LG3 (Fig 7.5F).

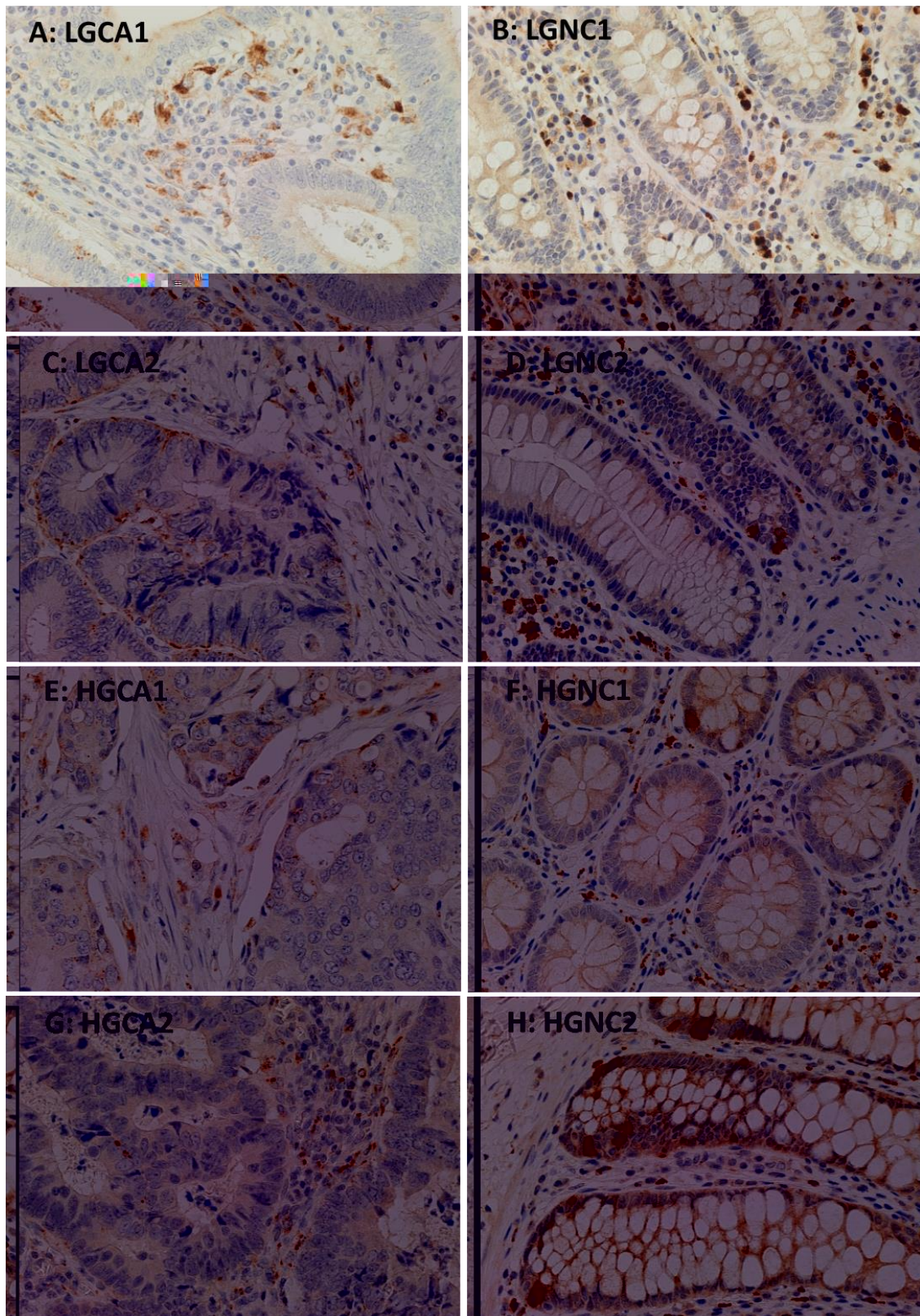


Figure 7.8: IHC staining for cathepsin B. Representative images showing IHC staining for cathepsin B, including 2 LGCA (A, C) and their patient-match NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

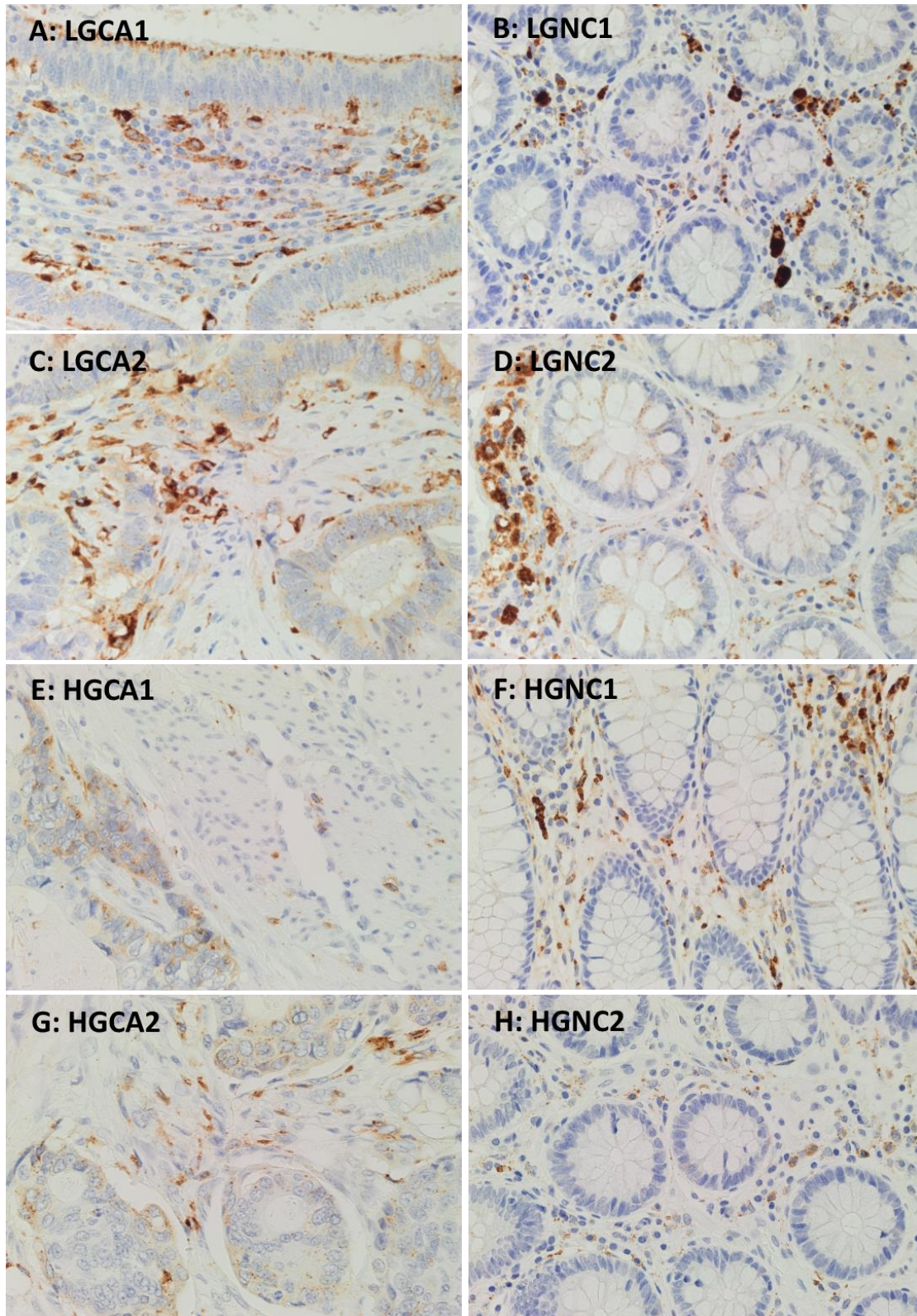


Figure 7.9: IHC staining for cathepsin D. Representative images showing IHC staining for cathepsin D, including 2 LGCA (A, C) and their patient-match NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

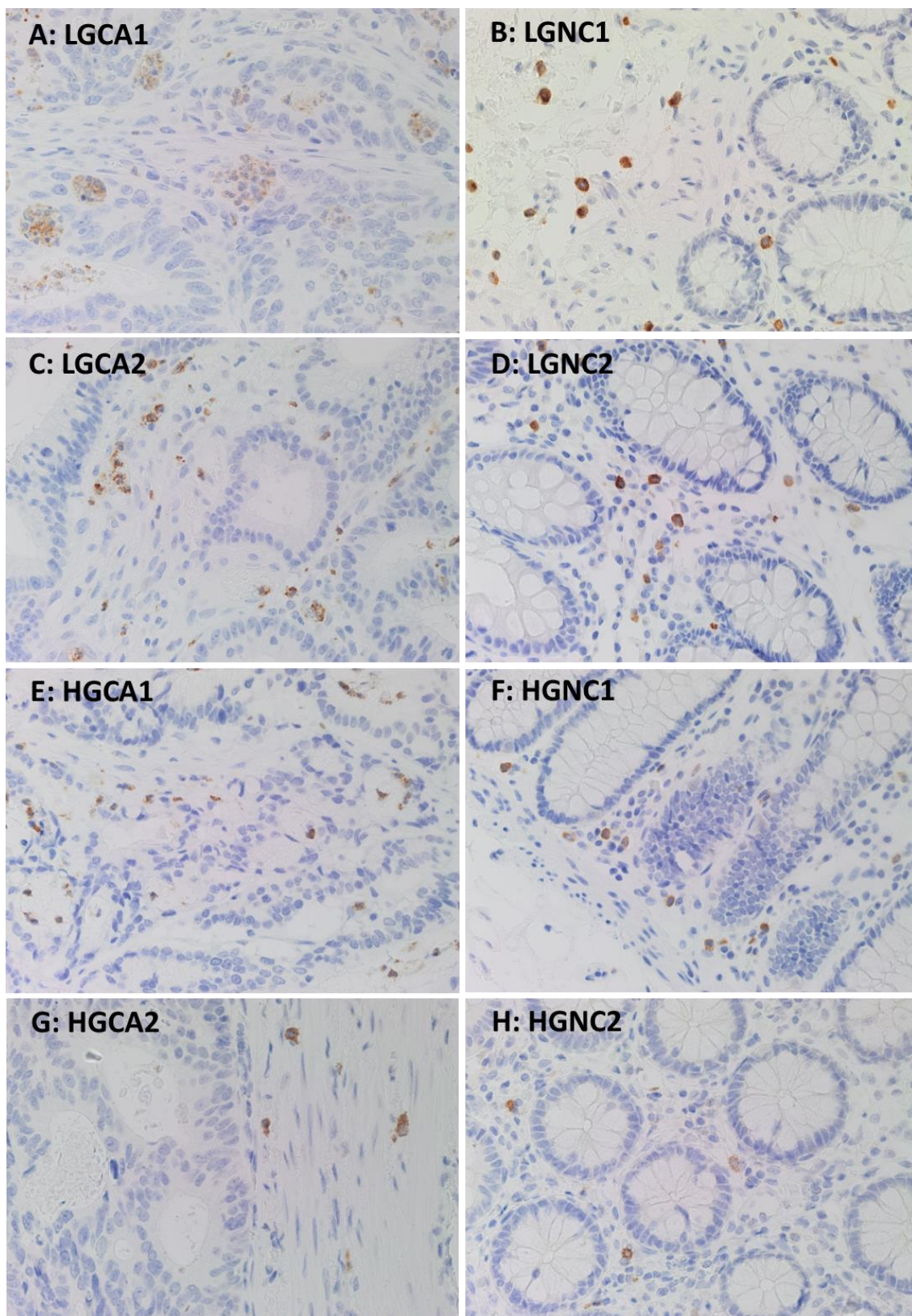


Figure 7.10: IHC staining for cathepsin G. Representative images showing IHC staining for cathepsin G, including 2 LGCA (A, C) and their patient-match NC tissue (B, D), and 2 HGCA (E, G) and their patient-matched NC tissue (F, H). Original magnification: 400x.

7.2.5 Cells expressing OCT4 and NANOG co-express RAS components

IF staining was carried out using combinations of various iPSC markers with RAS components to determine whether colon stem-like cells express components of the RAS.

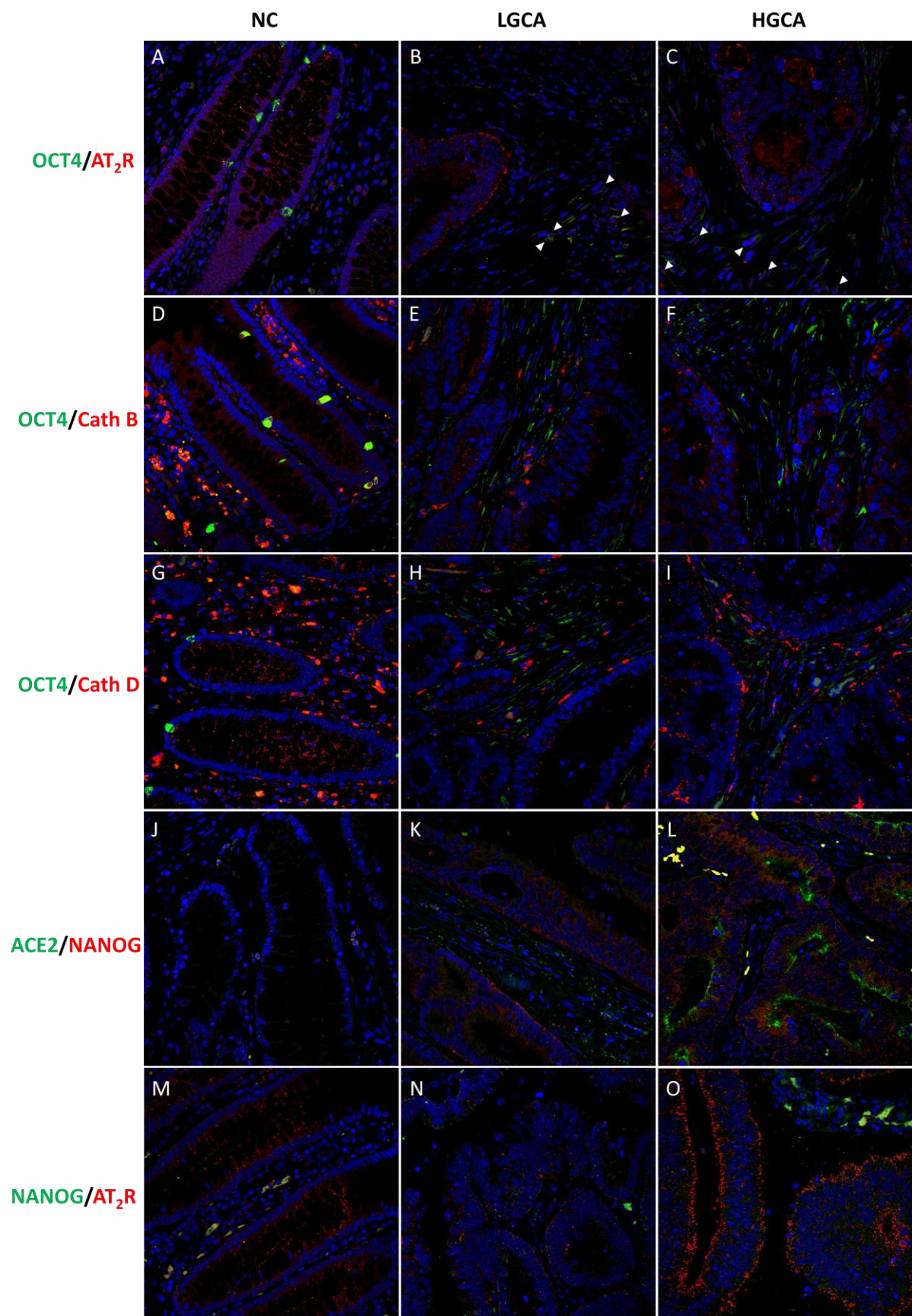
OCT4 (Fig 7.11A-I; green) was expressed by neuroendocrine cells within the crypts of NC tissues, and in the cytoplasm of elongated stromal cells in LGCA and HGCA tissues. Some of these OCT4⁺ cells stained positive for AT₂R in the nucleus (Fig 7.11B,C; red; arrowheads). AT₂R was also present in the cytoplasm and plasma membrane of epithelial cells in CA tissues.

Cathepsins B (Fig 7.11D-F; red) and D (Fig 7.11G-I; red) were expressed in the cytoplasm of epithelial cells in NC and CA tissues, and were abundant within immune cells in the stroma. There did not appear to be any co-expression of either cathepsin B or cathepsin D with OCT4 in the CA stromal cell subpopulation.

NANOG (Fig 7.11J-L; red) was not present in NC tissues but was seen in the cytoplasm and nuclei of epithelial cells in CA tissues. ACE2 (Fig 7.11J-L; green) was localised to the cytoplasm, and particularly to the luminal membrane, of CA crypt epithelial cells that also expressed NANOG. The mouse anti-NANOG primary antibody (Fig 7.11M-O; green) produced a weaker stain than the rabbit anti-NANOG (Fig 7.11J-L; red), but was detected in the cytoplasm of tumour epithelial cells, which also expressed AT₂R (Fig 7.11M-O; red) in the cytoplasm, often concentrated around the plasma membrane.

Based on IHC staining data, AT₂R, PRR and cathepsin D were overexpressed in CA tissues relative to NC tissues, whereas ACE and cathepsin B had a similar degree of staining in CA tissues compared to NC tissues, with cathepsin G predominantly negative. Overall, the IF staining suggests that the OCT4⁺ and NANOG⁺ subpopulations both co-express AT₂R, and that the NANOG⁺ subpopulation also expresses ACE2 and is likely to express cathepsins B and D. Taking into account the IHC and IF staining from Chapter 4, the proposed epithelial CSC subpopulation expresses NANOG/SOX2/KLF4/c-MYC/CD133/EpCAM/AT₂R/ACE2, and possibly LGR5, PRR and cathepsins B and D.

Figure 7.11: Co-localisation of iPSC markers with RAS components by IF staining (page 169). Representative immunofluorescence images showing protein expression of OCT4 (A-I, green) with AT₂R (A-C, red), cathepsin B (D-F, red) and cathepsin D (G-I, red). A rabbit anti-NANOG antibody (J-L, red) was co-stained with ACE2 (J-L, green). A mouse anti-NANOG antibody (M-O, green) was co-stained with AT₂R (A-C, red). NC (A,D,G,J,M), LGCA (B,E,H,K,N), HGCA (C,F,I,L,O). Cell nuclei were counterstained with DAPI (blue). Original magnification: 400x.



7.2.6 Cathepsin activity assays

The presence and distribution of cathepsins B and D were determined by IHC staining, IF staining and WB. However, the presence of cathepsins is only relevant if they are active within the CA tissues and CA-derived cells. They have a wide range of substrates and physiological roles, including involvement in the RAS. Cathepsin activity assays were performed by extracting total protein from 3 LGCA and 3 HGCA tissues, and EpCAM^{High} and EpCAM^{Low} cell lines derived from all 6 of these tissue samples. Briefly, protein extract was diluted in assay buffer and incubated with substrate at 37°C for 30 min. The average fluorescence was normalised against the amount of protein added, giving a relative fluorescence intensity unit (FIU) per µg of protein. Tonsil tissue was used as a positive control.

Activity of cathepsins B and D was detected in all tissue and cell samples (Figure 7.12A-D). In CA tissues, cathepsin B activity ranged from 2.84 to 6.95 FIU per µg of protein, with tonsil having 3.58 FIU per µg (Figure 7.12A). Cathepsin D activity was between 0.09 and 0.56 FIU per µg, with tonsil having 0.41 FIU per µg (Figure 7.12B). It was expected that the tissues would have higher cathepsin activity due to the presence of immune cells throughout the stroma as identified by IHC. However, the cells had higher cathepsin activity per µg of total protein than the tissues. The range of activity in cells was 2.27 to 33.23 FIU per µg for cathepsin B (Figure 7.12C) and 0.61 to 2.70 FIU per µg for cathepsin D (Figure 7.12D).

Cathepsin B activity was higher in the EpCAM^{Low} cells derived from 3 CA tissues and in the EpCAM^{High} cells from 2 CA tissues, with equal activity in the EpCAM^{High} and EpCAM^{Low} cells derived from the other tissue. Similarly, cathepsin D activity was higher in EpCAM^{Low} cells derived from 4 CA tissues and in the EpCAM^{High} cells from the other 2 tissues.

These results show that the cathepsins present in CA tissues and cell lines are active and theoretically capable of functioning in the RAS, and that their activity tends to be higher in EpCAM^{Low} cells than in EpCAM^{High} cells.

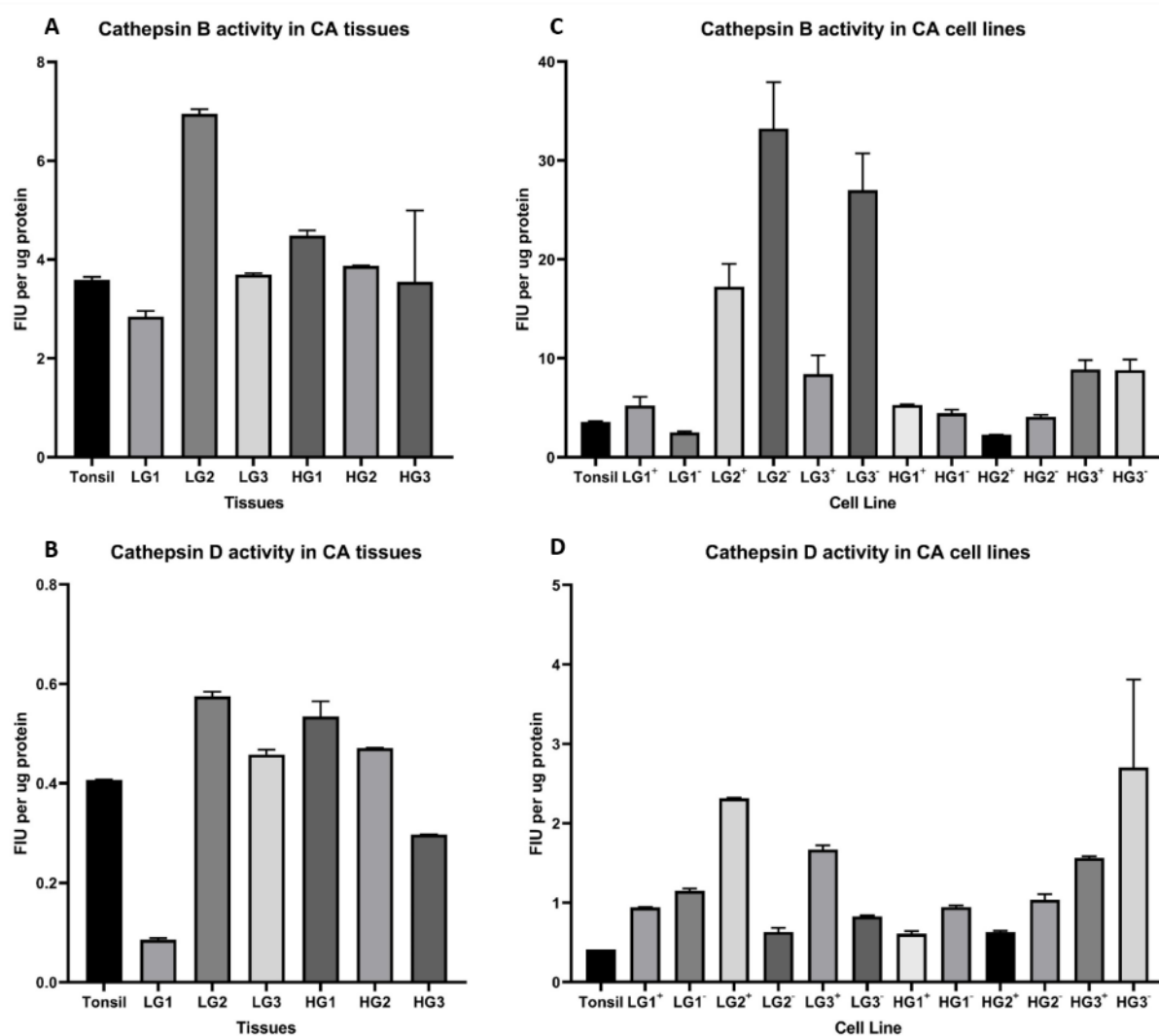


Figure 7.12: Cathepsin activity assays. Activity assays were performed for cathepsin B CA tissues (A) and cells (C), and for cathepsin D in CA tissues (B) and CA-derived cells (D). Fluorescence intensity units (FIU) were divided by the total amount of protein added to give a measure of activity per μg of protein. The average FIU per μg of protein across 2 technical replicates is displayed, with error bars representing standard deviation.

7.3 Discussion

This chapter investigated the expression and localisation of components of the RAS to determine whether there were any differences between CA tissues and their patient-matched NC tissues. Furthermore, based on the presence of a potential CSC subpopulation in the tumour epithelium and an OCT4⁺ subpopulation in the tumour stroma, combinations of iPSC markers with RAS components were visualised simultaneously via IF staining to determine whether RAS components are expressed by CSCs.

IHC staining revealed higher levels of PRR and cathepsins B and D in CA tissues than in their matched NC samples, and they were predominantly seen in the cytoplasm or luminal membrane of epithelial cells. Contrary to the hypothesis, ACE2 and AT₂R proteins were also more abundant in CA tissues. This may indicate that ACE2 and AT₂R act in a manner contrary to predictions, potentially due to differential outcomes dependent on the context. Interestingly, cathepsins B and D were more abundant in the LGCA epithelium than the HGCA epithelium, suggesting that they influence early tumourigenesis more than the progression from LGCA to HGCA. This aligns with previous work which determined that mRNA levels for cathepsins B and D increase concurrently with APC gene mutations that initiate the adenoma to adenocarcinoma sequence (Bian et al., 2016). This also appeared to be the case for ACE, which was rarely seen in the NC except for on the endothelium of blood vessels. However, it was occasionally seen on the luminal membranes of CA tumour epithelial cells and was weaker in the HGCA epithelium than in LGCA. However, when visualising ACE expression via IF staining, the tumours appeared negative, raising concerns that the weak blush seen by IHC staining may not have been true positive staining because IF staining is considered to be more sensitive than IHC staining.

The overexpression of PRR mRNA and protein in CA-derived cells reflects its importance in facilitating Wnt signalling. Furthermore, its ability to activate pro-renin to participate in the initial reactions of the RAS may hint at the relevance of the RAS in these cells.

Cathepsin G acts redundantly to ACE but is mainly expressed by mast cells and other immune cells. The lack of cathepsin G expression by CA tumour cells may relate to the finding that it increases cell-cell adhesion and therefore reduces metastasis (Kudo et al., 2009).

AT₁R is usually associated with adverse effects, unlike AT₂R which opposes these outcomes. Therefore, AT₁R is of interest when investigating the ways in which the RAS may influence cancer. However, there are currently no reliable or specific antibodies available to carry out IHC staining, IF staining or WB. One laboratory tested six commonly-cited anti-AT₁R antibodies using various knock-out models and cells known to be negative for AT₁R, and found that the antibodies all detected bands at around 43 kDa (the predicted size of AT₁R) in these knock-outs, and produced a variety of staining intensities and localisations when used for IHC staining which were evidently not specific for AT₁R (Benicky et al., 2012). AT₁R was not detected in CA tissue or cell line samples when analysed by mass spectrometry, as discussed in Chapter 6. Therefore, despite its promise as a target for RAS modulation in the context of cancer, measurements of AT₁R localisation and abundance were excluded from this project.

The species in which primary antibodies were raised limited the possible combinations for IF staining and meant that OCT4, one of the few primaries raised in mouse, was used in many

combinations. The most common species for raising primary antibodies is rabbit, and primary antibodies raised in mouse are less-frequently cited. NANOG and LGR5 antibodies raised in mouse were purchased to allow for more combinations. However, other unresolved issues with the IF workflow prevented much of this work from going ahead. Specifically, the 594 (red) channel on the confocal microscope was not performing as expected, at times showing no signal in positive control tissues and with discrepancies between IHC and IF staining, especially when using the c-MYC and AT₂R rabbit primary antibodies. At the time of writing, the GMRI was exploring options and the issue had not yet been resolved, so no further staining was performed. This may have included combinations of mouse anti-NANOG with rabbit-raised antibodies against SOX2, KLF4, c-MYC, PRR, CD133 and cathepsins B and D, to further verify the inferred co-expression phenotype of the epithelial subpopulation.

An LGR5 antibody raised in mouse was purchased to allow for co-localisation with iPSC markers in an effort to relate iPSC marker expression to colon CSC marker expression, as was done with CD133. However, the LGR5 antibody was unable to pass the in-house validation workflow due to issues with WB, and so because of this and the IF issues, LGR5 WB and IF staining were abandoned. When used for WB, a band for LGR5 was expected at a molecular weight of 84-100 kDa, however a band at 70 kDa was observed. After consulting the technical support personnel at two independent suppliers, it was recommended that the clean band at 70 kDa was evidence of specificity for the LGR5 protein and that some variation from the expected size is anticipated due to variable WB conditions. However, the band was not always present in the positive controls and so the antibody was abandoned for WB and IF staining. The LGR5 antibody was extensively validated for IHC staining by using multi-tissue blocks which included known positive tissues. The staining patterns were reproducible and compared favourably to those expected based on the literature and databases including the product data sheet references, European Bioinformatics Institute and Human Protein Atlas, and so the IHC data for LGR5 was considered reliable.

Overall, IF staining for iPSC markers and RAS components revealed that the cells which were characterised as being positive for NANOG/SOX2/KLF4/c-MYC/CD133 and likely to express EpCAM and LGR5 were also positive for AT₂R and ACE2, and possibly PRR and cathepsins B and D. Expression of these RAS components was typically higher in CA tissues than patient matched NC tissues. This result was encouraging as it provided some rationale for targeting potential CSCs with RAS modulators. Chapter 8 will explore the effects of RAS modulators on the metabolism of CA-derived primary cell lines and their expression of iPSC marker mRNA.

Research investigating cathepsins in CRC has focussed on their ability to facilitate cancer cell migration by degrading ECM components, leading to tumour budding, as well as growth factor

activation and nutrient recycling (S. Chen et al., 2017; Kuester et al., 2008; Ruan et al., 2015; Turk et al., 2012). However, angiotensins are also proven substrates for cathepsins, with cathepsin G recently found to have a similar capacity as ACE for converting ATI to ATII (Rykl et al., 2006). The finding that cathepsins B and D are active in CA tissues, that their activity is higher in CA-derived cell lines, and that they are expressed by the majority of tumour epithelial cells suggests that the cathepsins are likely to be participating in the RAS in cancer, and possibly in the CSCs.

There appeared to be differential cathepsin activity between the two sorted cell populations. EpCAM^{Low} cells derived from LGCA2 and LGCA3 both had much higher cathepsin B activity but lower cathepsin D activity than their patient-matched EpCAM^{High} cells. HGCA2-derived EpCAM^{Low} cells had higher activity of both cathepsins B and D than EpCAM^{High} cells. EpCAM^{Low} cells derived from the 3 remaining tissues had similar or lower cathepsin B activity levels, but significantly higher cathepsin D activity, than EpCAM^{High} cells. This observation aligns with the proposed function of cathepsins in the migration of cancer cells. The precise function of cathepsins in these tissues and cells is beyond the scope of this project, but the presence of active cathepsins lends support to their involvement in the RAS in CA. This implies RAS inhibitors could be bypassed by cathepsins. For example, despite administering ACEIs, cathepsin G can produce ATII from AGT and ATI. Similarly, aliskiren can be used to inhibit renin, but cathepsin D is still capable of converting AGT into ATI. Therefore, blockade at either end of the pathway via β -blockers (to prevent pro-renin production) and antagonists of AT₁R (ARBs) and AT₂R may be the best way to prevent RAS bypass.

Finally, the quality and quantity of RNA extracted from CA tissues was low, which limited the extent of markers analysed by RT-qPCR. The snap-frozen tissue pieces from the GMRITB had a high fat content resulting in low RNA yields and depletion of available tissue. Therefore, DNA sequencing and mass spectrometry were prioritised for the remaining tissue.

7.4 Conclusions

The aim of this chapter was to investigate the expression of RAS components, including cathepsins B, D and G, in CA tissue samples, CA-derived cells, and patient-matched NC tissue. AT₂R, PRR and cathepsin D were found to be upregulated in CA tissues, whereas the abundance of ACE and cathepsin B was similar in CA tissues compared to NC tissues. This suggests that RAS signalling is present within CA, possibly to a greater extent than in the healthy colon. Co-localisation via IF staining revealed that the NANOG⁺ subpopulation of CSCs expresses AT₂R and ACE2 and is likely to express PRR and cathepsins B and D, and that OCT4⁺ stromal cells express AT₂R which is localised to the nucleus. RAS modulating drugs may therefore be used to target CSCs in CA and hinder their function by reducing the influence of signalling pathways mediated by the RAS, such as those outlined in Chapter 6.

Furthermore, β -blockers and AT₂R antagonists may have the greatest effects because they are not bypassed by cathepsins.

Chapter 8: The Effects of Renin-Angiotensin System Modulation on Metabolism and Stem Cell Phenotype of Colon Adenocarcinoma derived Primary Cell Lines

8.1 Introduction

The RAS maintains blood pressure and fluid volumes, and various components of the RAS can be pharmacologically targeted to treat hypertension. Three drug classes are commonly used to lower blood pressure: β -blockers, which reduce renin production; ACE inhibitors (ACEIs); and angiotensin receptor blockers (ARBs), which antagonise AT₁R.

There has been recent interest in the role of the RAS in cancer and the influence of RAS modulators on cancer incidence and outcomes. This has led to retrospective studies investigating the incidence of CRC in patients who take RAS modulators compared to those who do not, and the outcome of CRC patients with or without RAS modulator administration. These studies have reported that RAS modulators lead to an overall reduction in polyp formation and lower incidence of primary or metastatic CRC (Childers, 2015). AT₁R antagonism via ARBs has been reported to have the greatest effect in reducing the development and metastasis of solid tumours (Childers, 2015). A large-scale meta-analysis found a positive trend in the survival of CRC patients administered ARBs and ACEIs, and increases in DFS, PFS and OS rates of up to 20% (Sun et al., 2017). However, it is still unclear whether the enhanced outcomes are due to the effect of RAS modulators on cancer cells directly, or by alleviating the comorbidities which they classically target, namely hypertension.

8.1.1 β -Blockers

β -blockers are antagonists of the adrenergic receptors which bind epinephrine and norepinephrine as a part of the sympathetic nervous system (Weber, 2005). It is a diverse class of drugs, each having specific pharmacologic properties, and therefore different β -blockers produce different clinical outcomes and side effects. The three main adrenergic receptors are β_1 , β_2 and α_1 . The β_1 receptors are most commonly expressed in the heart and kidneys (Wiysonge et al., 2007). Second-generation selective β -blockers have a higher affinity for β_1 receptors, which are the main target for hypertension treatment (Akbar et al., 2014). First-generation β -blockers have an equal affinity for the β_1 , β_2 and α_1 receptors and are known as non-selective β -blockers (Wiysonge et al., 2007). The β_2 adrenergic receptors are predominantly expressed in the lungs and gastrointestinal tract (Wiysonge et al., 2007). The α_1 receptors are expressed by endothelial cells and cause vasoconstriction when activated, so

selective antagonism of these receptors by third-generation β -blockers causes vasodilation to relieve hypertension (Weber, 2005; Wiysonge et al., 2007). As a drug class, β -blockers are relevant to the RAS because they block one of the three mechanisms by which juxtaglomerular cells produce renin, those being prolonged adrenergic stimulation, reduced blood supply (ischaemia), and low sodium (Sparks et al., 2014).

The first β -blocker available for clinical use was propranolol, a non-selective adrenergic receptor antagonist originally used to treat angina before its role in the sympathetic nervous system was elucidated (Akbar et al., 2014; Srivastava et al., 1964). The binding of epinephrine or norepinephrine to adrenergic receptors causes muscle contractions via G-protein/adenylyl cyclase signalling, which manifests as increased heart rate and vasoconstriction (Akbar et al., 2014). Therefore, blocking adrenergic receptors with β -blockers relaxes the muscles in the blood vessels and the heart. Propranolol is lipophilic and undergoes rapid uptake in the GI tract before first-pass hepatic clearance, leaving about 25-35% of the dose bioavailable (Pantziarka et al., 2016). Propranolol's lipophilicity allows it to cross the blood-brain barrier (Aronson, 2008; Mantyla et al., 1983). It has a plasma half-life of 3-6 h and a peak plasma concentration 1.5-3 h after administration (Pantziarka et al., 2016). Propranolol has two enantiomers: S(-)-Propranolol and R(+)-Propranolol, though it is administered as a racemic mixture due to the costs associated with purifying the two forms. S-propranolol has up to 100x greater binding affinity than R-propranolol for β -adrenergic receptors, and due to stereoselective hepatic clearance more of the R-propranolol is eliminated (Gilmore et al., 1992; Overman et al., 2019). Due to this difference in activity, S-propranolol is more effective for alleviating hypertension; however, R-propranolol is better suited to treating arrhythmia (Stark et al., 1989; Yang et al., 2015). It is unproven whether the different enantiomers of propranolol would affect cancer in different ways.

Propranolol is one of the most highly studied β -blockers in relation to cancer. A 12-year large cohort study with over 12,000 patients taking propranolol and over 12,000 who were not taking propranolol reported lower incidence of head and neck, oesophageal, gastric, colorectal and prostate cancers in those taking propranolol (Chang et al., 2015). Conversely, a more recent meta-analysis of post-operative administration of non-selective β -blockers, including propranolol, to cancer patients reported no effect on recurrence rates (Yap et al., 2018). Furthermore, DFS and OS outcomes varied between cancer types, with better outcomes for melanoma and ovarian cancer patients, but poorer OS in lung cancer and DFS in endometrial cancer (Yap et al., 2018).

Timolol is a non-selective β -blocker which is primarily used to treat glaucoma, but can also be prescribed for hypertension and anxiety (Dunn et al., 1981; Maenpaa et al., 2016; Mantyla et al., 1983). Timolol differs from propranolol in a number of ways. It has no membrane stabilising activity

which can dampen the cardiac membrane action potential (Dunn et al., 1981). Timolol is less lipid soluble than propranolol, which leads to comparatively little timolol being removed by the liver during first-pass clearance (Aronson, 2008). However, up to 90% is metabolised into 4 different metabolites (Mantyla et al., 1983). It has a half-life in plasma of around 3 h (Mantyla et al., 1983) and it can induce vasodilation of peripheral blood vessels, whereas it is unclear whether propranolol can do the same (Aronson, 2008). Timolol has occasionally been used for large studies investigating the effect of β -blockers on cancer outcomes (Yap et al., 2018), and although there has been little research into the different effects of the R- and S-enantiomers, they seem to produce comparable outcomes (Czechowicz-Janicka et al., 1998). Both the R- and S-enantiomers of each β -blocker were tested in this project to see whether they have different effects on the growth of CA-derived cells, which has not been tested before.

Due to the variability of results *in vivo*, attempts to elucidate the action of β -blockers on cancer cells have been continued *in vitro*. Propranolol has been shown to restore the sensitivity of resistant cells to first-line treatments such as doxorubicin, 5-FU and paclitaxel (Pantziarka et al., 2016). It also reduces SW480 cell migration, while β_1 -selective β -blockers were unable to do the same, suggesting a mechanism specific to β_2 adrenergic receptors (Masur et al., 2001). Proliferation and viability of colon cancer cell lines were both decreased by propranolol, with IC50 values of 119.5 μ M (SW620), 86.38 μ M (Colo205), 69.1 μ M and 65.4 μ M (both HT29) (Coelho et al., 2015; Pantziarka et al., 2016).

8.1.2 ACE Inhibitors

Captopril was the first orally active ACEI to be developed (Duchin et al., 1988). It is one of the only ACEI compounds that is not administered as a prodrug (Duchin et al., 1988). Captopril has a relatively short half-life of 2-4 h and a bioavailability of 65% when taken without food; oral administration at mealtimes decreases the bioavailability but also reduces clearance (Duchin et al., 1988). Captopril has a 30,000x greater affinity for ACE than for AT₁R, demonstrating high specificity for its intended target (Duchin et al., 1988).

Cilazapril is administered as a prodrug and is metabolised to cilazaprilat (Waterfall, 1989). When administered as cilazapril, the bioavailability of the resulting cilazaprilat is approximately 57% (Waterfall, 1989). It is more potent than captopril and has better absorption following oral administration (Waterfall, 1989). Cilazaprilat remains active in the plasma for up to 8 h, where it binds ACE with high specificity (Fasanella d'Amore et al., 1987).

8.1.3 Antagonists of AT₁R and AT₂R

ARBs are a class of drugs that antagonise AT₁R and are used for treating hypertension. They bind competitively over ATII and have a slow dissociation from their target (Aulakh et al., 2007).

Losartan was the first clinically approved oral ARB (Aulakh et al., 2007). Most of its action is due to the active metabolite EXP3174, which has a longer half-life (6-9 h) compared to Losartan (2 h) and is up to 40x as potent as unmetabolised Losartan (Aulakh et al., 2007). It binds AT₁R with 1000x more affinity than for AT₂R. Candesartan is administered as a prodrug to increase absorption, and in the GI tract it is completely metabolised to its active form which binds AT₁R with 10x greater affinity than EXP3174 (Aulakh et al., 2007). It has a relatively long half-life of 9 h (Aulakh et al., 2007).

There are currently no AT₂R antagonists used clinically. However, one candidate, EMA401, has undergone multiple phase I clinical trials to prove its tolerability and a phase II trial for treatment of neuropathic pain (Keppel Hesselink et al., 2017; Smith et al., 2015). EMA401, also known as olodanrigan, is the S-enantiomer of the EMA400 compound (PD126055) (Smith et al., 2013). Its binding affinity for AT₂R is 20-30x greater than that of the R-enantiomer, EMA402 (Smith et al., 2015), and its binding affinity for AT₂R is >10,000x higher than for AT₁R (Keppel Hesselink et al., 2017). Aside from having the greatest affinity for AT₂R, EMA401 also has a bioavailability of 30% after oral administration, higher than the 6-7% for related compounds EMA200 (PD123319) and EMA300 (PD121981) (Smith et al., 2015). EMA401 relieves neuropathic pain by preventing activation of ATII/AT₂R-mediated p38 and p42/p44 MAP kinase signalling pathways (Smith et al., 2015). This signalling can be driven by AT₂R expressed by either the peripheral nerves or macrophages, suggesting some interplay between the RAS and inflammation in neuropathic pain, and possibly in inflammatory cancers such as CRC (Shepherd et al., 2018; Smith et al., 2015).

The second AT₂R antagonist to be used in this project is the proprietary compound SMM02, developed by our collaborator Dr Sean Mackay (University of Otago).

Results from Chapter 7 suggest that AT₂R antagonists may have a greater effect than ACEIs on the CA-derived primary cell lines used in this study because WB and IF staining revealed that these cells expressed AT₂R but not ACE. It is expected that AT₂R antagonism would have a positive effect on cancer cell growth by allowing AT₁R signalling to proceed unopposed. Despite the presence of ACE expression being in doubt, ACEIs were administered to the cells due to reports that they reduce the incidence and mortality of CRC. As discussed, AT₁R protein abundance and localisation could not be assessed due to a lack of reliable antibodies. Despite this, ARBs were still administered during cell assays to determine whether they affect these CA-derived cells.

Renin can be directly antagonised using aliskiren, however administration of aliskiren is generally avoided in favour of drugs with fewer side effects and higher bioavailability (Parving et al.,

2012). Therefore, rather than directly blocking renin, cells were exposed to β -blockers to reduce the production of renin. It is important to keep in mind that because of their wide-ranging physiological significance, antagonism of β -adrenergic receptors may lead to changes in cell metabolism or growth unrelated to renin levels.

This chapter aimed to explore the effects of RAS modulators on the metabolism of CA-derived primary cell lines and their iPSC marker mRNA expression. Based on the literature regarding RAS modulators and cancer outcomes, it was hypothesised that β -blockers and ARBs would be able to reduce CA cell metabolism, but that due to ACE appearing negative in the cells ACEIs would have little effect. Furthermore, because AT_2R opposes AT_1R , administration of AT_2R antagonists was hypothesised to increase CA cell metabolism. Finally, it was hypothesised that if RAS modulators were capable for targeting CSCs, this would manifest as reductions in iPSC marker transcription and tumoursphere formation. Drugs from the four classes outlined above were selected, including two drugs from each class in order to determine whether any effect is drug-specific or a possible class effect. These were propranolol and timolol, captopril and cilazapril, losartan and candesartan, and EMA401 and SMM02. Cell metabolism was measured by using the RealTime-Glo™ Cell Viability Assay (Promega), and the mRNA levels of *OCT4*, *SOX2*, *NANOG* and *KLF4* were assessed by RT-qPCR.

8.2 Results

8.2.1 Cell metabolism assays

An initial assay was carried out on 1 HGCA-derived primary cell line to determine the appropriate range of doses for each drug, based on previous work in our lab. The assay involves a proprietary substrate which is reduced by metabolically active cells, allowing it to participate in a reaction catalysed by luciferase to produce a luminescent signal that is measured by a plate reader and used to assess the metabolic activity of the cells. The assay is non-lytic and lasts for 72 h, so the same plate was able to be read multiple times over the 72 h time course. Any increase or decrease in metabolism of less than 25% relative to the control cells (grown in the absence of any drug) was considered to be within the natural variation expected between cultures and not a significant result. If metabolism increased or decreased by more than 25%, the drug was considered to have had an effect on cell metabolism. In the initial assay, propranolol had the largest effect on both metabolism and confluency, and doses of 50 μ M, 10 μ M and 1 μ M were chosen. All other drugs were administered at 100 μ M, 50 μ M and 10 μ M. The final assays were performed on 4 HGCA-derived primary cell lines. Cells were dosed every 24 h for 3 d, beginning 24 h after seeding to allow for the cells to adhere.

Propranolol administered at 50 μ M consistently led to a 96-99% reduction in cell metabolism and visible cell death in the CA-derived primary cells (Figure 8.1A-L), an effect greater than that

expected given the IC₅₀ of approximately 65 μ M as reported by Coelho et al. (Coelho et al., 2015). In contrast, timolol, which has been reported to be 10x as potent as propranolol at antagonising β -adrenergic receptors (McMahon et al., 2012), appeared to have a lesser effect on cell metabolism, with only the 100 μ M doses causing a decrease of greater than 25% (Figure 8.2A-H). Interestingly, the two enantiomers of each β -blocker produced almost identical results. The R- and S-enantiomers of propranolol were administered separately and as a racemic mixture, and a 50 μ M dose of either enantiomer alone or of the racemic combination caused greater than 90% inhibition of metabolic activity by 48 h in all 4 cell lines. Doses of 10 μ M and 1 μ M did not make a significant difference at any time point. Similarly, 72 h after a 100 μ M dose, metabolism was inhibited by 30-50% by the R-enantiomer of timolol in 3 cell lines, and by the S-enantiomer in 2 of these same cell lines. The metabolism of the HGCA3 cell line was unaffected by any dose of either timolol enantiomer (Figure 8.2C,G).

Due to the uncertainty around ACE protein expression as determined by WB and IF staining, ACEIs were not expected to have an effect on the metabolism of the CA-derived primary cell lines. Accordingly, neither captopril nor cilazapril caused any significant changes to metabolism in any of the 4 cell lines at any concentration (Figure 8.3A-H).

As outlined in Chapter 7, there are currently no reliable antibodies for AT₁R (Benicky et al., 2012) and so its presence in these samples is undetermined. However, previous work in published literature suggests that ARBs are beneficial to CRC patients. Surprisingly, candesartan and losartan were unable to affect cell metabolism in any of the 4 cell lines by the 72 h time point (Figure 8.4A-H). The HGCA1 and HGCA2 cell lines appeared to have lower metabolism relative to control cells 48 h after beginning daily 100 μ M doses of either candesartan or losartan, but this had normalised by 72 h, suggesting that the cells were able to overcome any early effects of the drugs despite daily dosing (Figure 8.4A,B,E,F).

AT₂R is known to oppose the actions of AT₁R; therefore, it was expected that antagonism of AT₂R would allow AT₁R signalling to proceed unchecked and lead to increased metabolism and proliferation. However, EMA401 led to a clear reduction of metabolism, with the level of inhibition proportional to the dose administered (Figure 8.5A-D). In all 4 cell lines, doses of 100 μ M and 50 μ M consistently caused reductions in metabolism of around 80% and 50%, respectively, whereas the 10 μ M dose did not affect metabolism. SMM02 was trending in the same direction until the 72 h time point, where despite the inhibition remaining above 25% for the 100 μ M dose in all 4 cell lines, metabolism seemed to be recovering (Figure 8.5E-H).

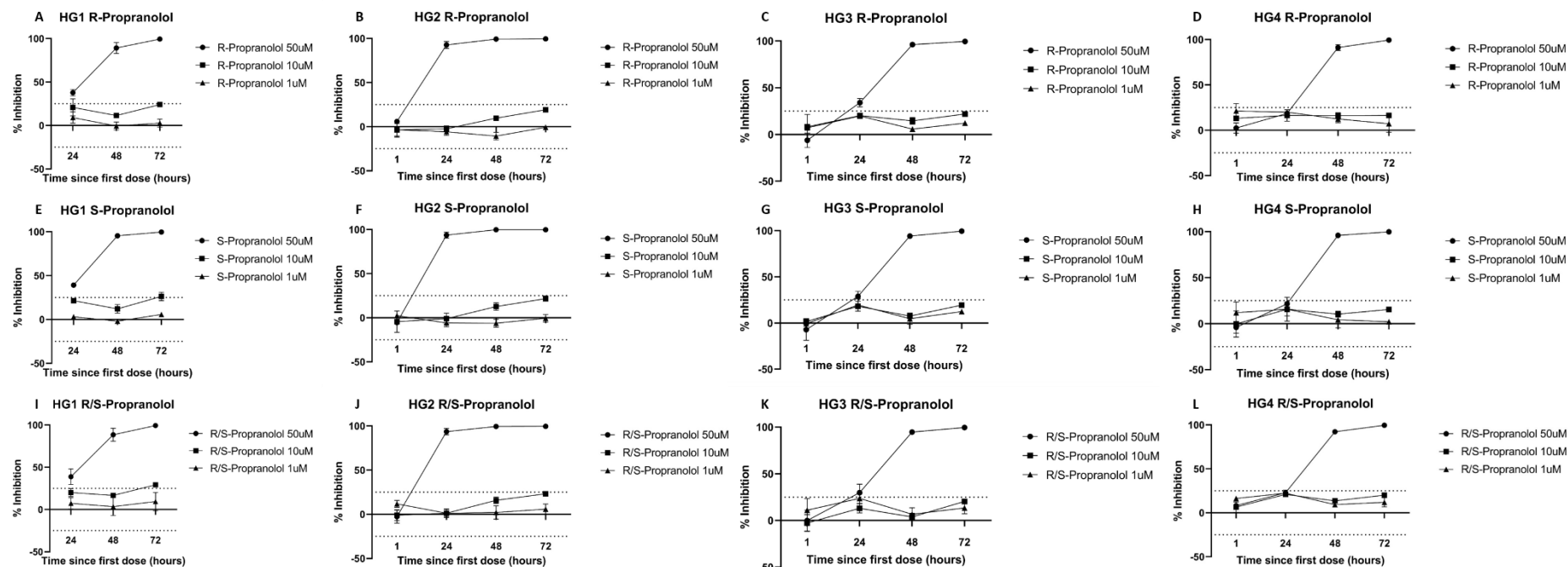


Figure 8.1: Effect of propranolol on the metabolism of HGCA-derived cell lines. Four HGCA-derived cells lines were exposed to R-propranolol (A-D), S-propranolol (E-H) and racemic propranolol (I-L) at doses of 50 μM (\bullet), 10 μM (\blacksquare) and 1 μM (\blacktriangle). Luminescence produced by cells exposed to propranolol was measured and used to measure the inhibitory effect of treatment on metabolic activity. The y-axis represents the extent of metabolic inhibition in the treated cells relative to the untreated control cells, and because the assay seeks to measure metabolic inhibition, the positive values indicate inhibition and the negative values indicate increased metabolic activity in the drug-treated cells. The x-axis shows the time after initial dose of each luminescence measurement. Each data point is the average of 3 technical replicates, with error bars showing standard deviation.

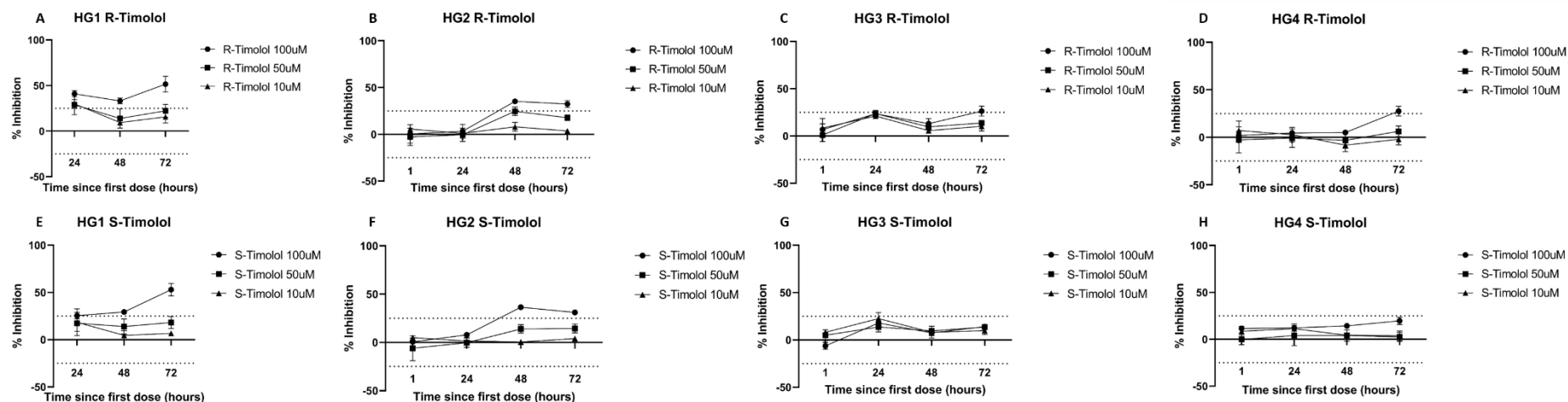


Figure 8.2: Effect of timolol on the metabolism of HGCA-derived cell lines. Four HGCA-derived cells lines were exposed to R-timolol (A-D) and S-timolol (E-H) at doses of 100 μM (●), 50 μM (■) and 10 μM (▲). Luminescence produced by cells exposed to propranolol was measured and used to measure the inhibitory effect of treatment on metabolic activity. The assay was developed to measure metabolic inhibition, and so the positive values on the y-axis represent a positive outcome in terms of inhibiting metabolism. Therefore, positive values show the extent of metabolic inhibition in the treated cells relative to the untreated control cells, and the negative values indicate increased metabolic activity in the drug-treated cells. The x-axis shows the time after initial dose of each luminescence measurement. Each data point is the average of 3 technical replicates, with error bars showing standard deviation.

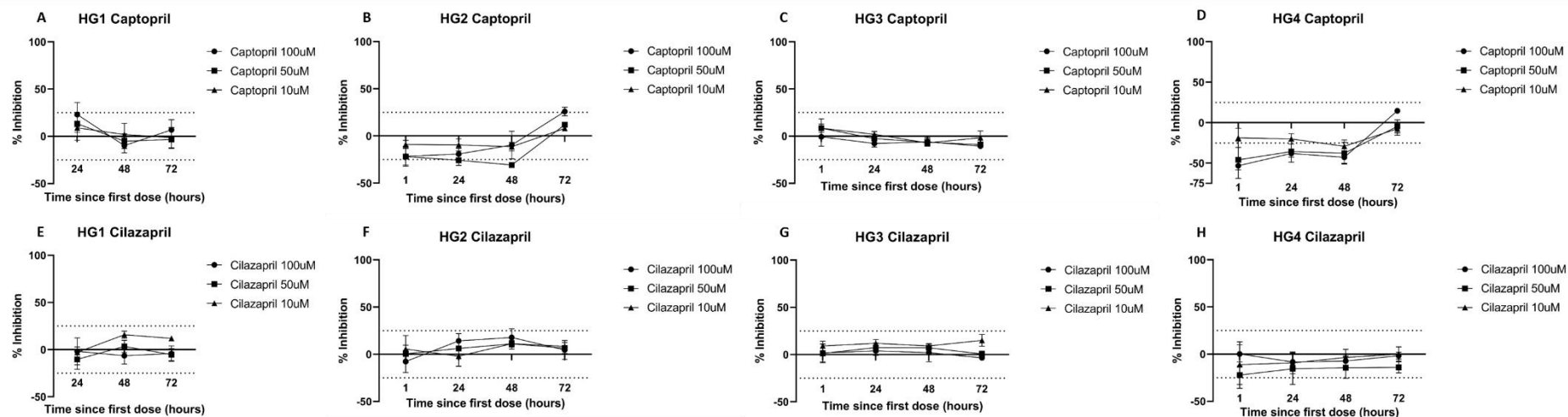


Figure 8.3: Effect of ACEIs on the metabolism of HGCA-derived cell lines. Four HGCA-derived cells lines were exposed to captopril (A-D) and cilazapril (E-H) at doses of 100 μM (●), 50 μM (■) and 10 μM (▲). Luminescence produced by cells exposed to ACEIs was measured and used to measure the inhibitory effect of treatment on metabolic activity. The assay was developed to measure metabolic inhibition, and so the positive values on the y-axis represent a positive outcome in terms of inhibiting metabolism. Therefore, positive values show the extent of metabolic inhibition in the treated cells relative to the untreated control cells, and the negative values indicate increased metabolic activity in the drug-treated cells. The x-axis shows the time after initial dose of each luminescence measurement. Each data point is the average of 3 technical replicates, with error bars showing standard deviation.

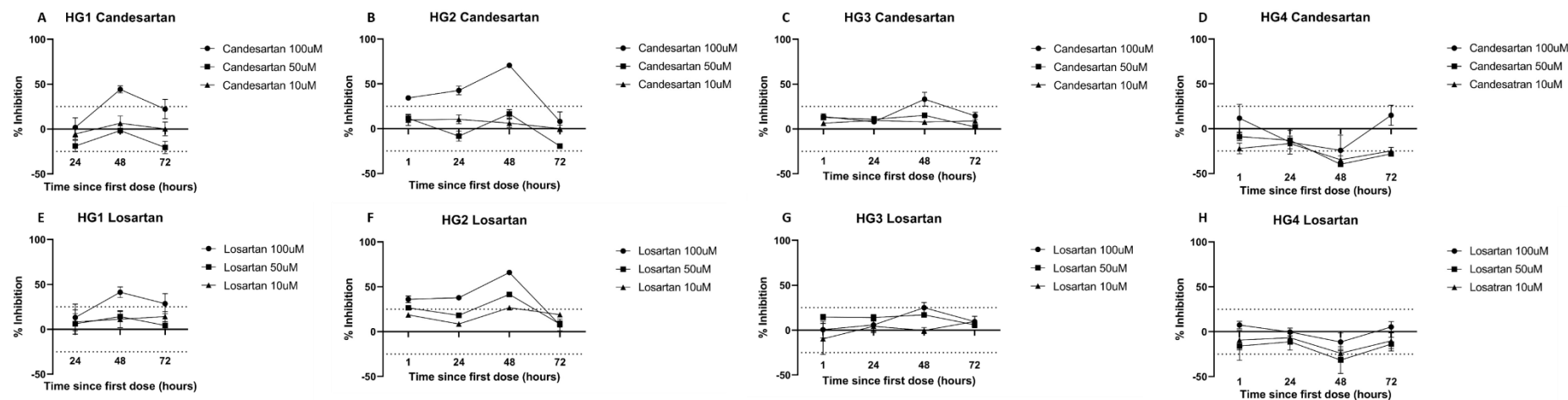


Figure 8.4: Effect of ARBs on the metabolism of HGCA-derived cell lines. Four HGCA-derived cells lines were exposed to candesartan (A-D) and losartan (E-H) at doses of 100 μM (●), 50 μM (■) and 10 μM (▲). Luminescence produced by cells exposed to ARBs was measured and used to measure the inhibitory effect of treatment on metabolic activity. The assay was developed to measure metabolic inhibition, and so the positive values on the y-axis represent a positive outcome in terms of inhibiting metabolism. Therefore, positive values show the extent of metabolic inhibition in the treated cells relative to the untreated control cells, and the negative values indicate increased metabolic activity in the drug-treated cells. The x-axis shows the time after initial dose of each luminescence measurement. Each data point is the average of 3 technical replicates, with error bars showing standard deviation.

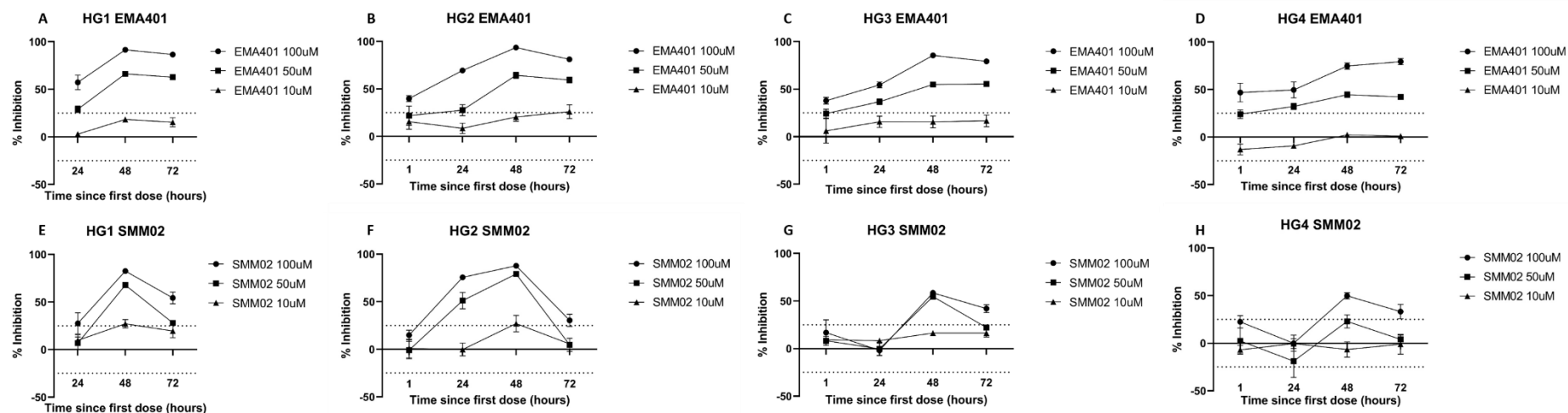


Figure 8.5: Effect of AT₂R antagonists on the metabolism of HGCA-derived cell lines. Four HGCA-derived cells lines were exposed to EMA401 (A-D) and SMM02 (E-H) at doses of 100 μM (●), 50 μM (■) and 10 μM (▲). Luminescence produced by cells exposed to AT₂R antagonists was measured and used to measure the inhibitory effect of treatment on metabolic activity. The assay was developed to measure metabolic inhibition, and so the positive values on the y-axis represent a positive outcome in terms of inhibiting metabolism. Therefore, positive values show the extent of metabolic inhibition in the treated cells relative to the untreated control cells, and the negative values indicate increased metabolic activity in the drug-treated cells. The x-axis shows the time after initial dose of each luminescence measurement. Each data point is the average of 3 technical replicates, with error bars showing standard deviation.

Overall, the R-, S- and racemic propranolol consistently caused reductions in cellular metabolism at doses of 50 μ M, as did EMA401 at doses of 100 μ M and 50 μ M. ACEIs did not affect metabolism, and ARBs had an early influence in 2 cell lines before metabolism returned to levels similar to control cells. SMM02 may also inhibit metabolism, but the cells appeared to be adjusting or recovering partially by 72 h.

8.2.2 Effect of RAS modulation on tumoursphere formation

Based on the initial metabolic assays, cells were dosed with a concentration of drug estimated to cause a significant change in metabolism without extensive cell death, to allow a sufficient number of cells to be harvested and seeded for tumoursphere assays. Cells were seeded in 24-well plates and exposed to R-propranolol (20 μ M), EMA401 (50 μ M) and losartan (100 μ M).

Due to questionable ACE expression and no effect seen on cell metabolism when using ACEIs, cells were not exposed to ACEIs for this assay. However, despite losartan having no effect on metabolism, it was chosen for this assay to see whether it might alter the level of iPSC marker mRNA expression without influencing metabolism. The R- and S-enantiomers of both β -blockers showed similar effects on metabolism, so R-propranolol was selected over S-propranolol for this assay because when administered to patients it would not cause the cardiac effects of the S-enantiomer. Cells grown without any RAS modulators were seeded concurrently as tumoursphere assay controls.

Table 8.1. Analysis of tumoursphere formation from treated cells.

Sample	Passage	Average maximum diameter \pm SD			
		Control	R-propranolol	EMA401	Losartan
HGCA1	1	55.39 (3) \pm 11.92	29.06 (4) \pm 4.76	50.71 (4) \pm 15.95	74.24 (4) \pm 26.07
	2	32.30 (2) \pm 7.44	N/A	34.22 (2) \pm 4.43	N/A
	3	N/A	N/A	N/A	N/A
HGCA2	1	60.38 (7) \pm 12.86	50.58 (3) \pm 6.80	45.15 (3) \pm 9.02	51.17 (3) \pm 11.78
	2	44.32 (3) \pm 9.77	N/A	N/A	N/A
	3	N/A	N/A	N/A	N/A
HGCA3	1	72.20 (4) \pm 18.87	42.75 (3) \pm 5.10	43.87 (4) \pm 10.70	46.06 (3) \pm 9.72
	2	53.92 (4) \pm 13.70	N/A	N/A	N/A
	3	32.25 (3) \pm 7.56	N/A	N/A	N/A
HGCA4	1	59.52 (4) \pm 14.21	52.90 (3) \pm 11.15	41.31 (3) \pm 11.12	56.82 (3) \pm 12.32
	2	38.12 (6) \pm 7.29	31.75 (3) \pm 3.70	N/A	N/A
	3	N/A	N/A	N/A	N/A

Primary cell lines derived from 4 HGCA tissue samples were exposed to 3 RAS modulators before being seeded for tumoursphere assays. Tumoursphere diameter was measured in μ m. The number of days taken to reach the maximum diameter are shown in brackets. Diameter values represent the average diameter of all measured tumourspheres across multiple technical replicates for each biological replicate, with SD values shown.

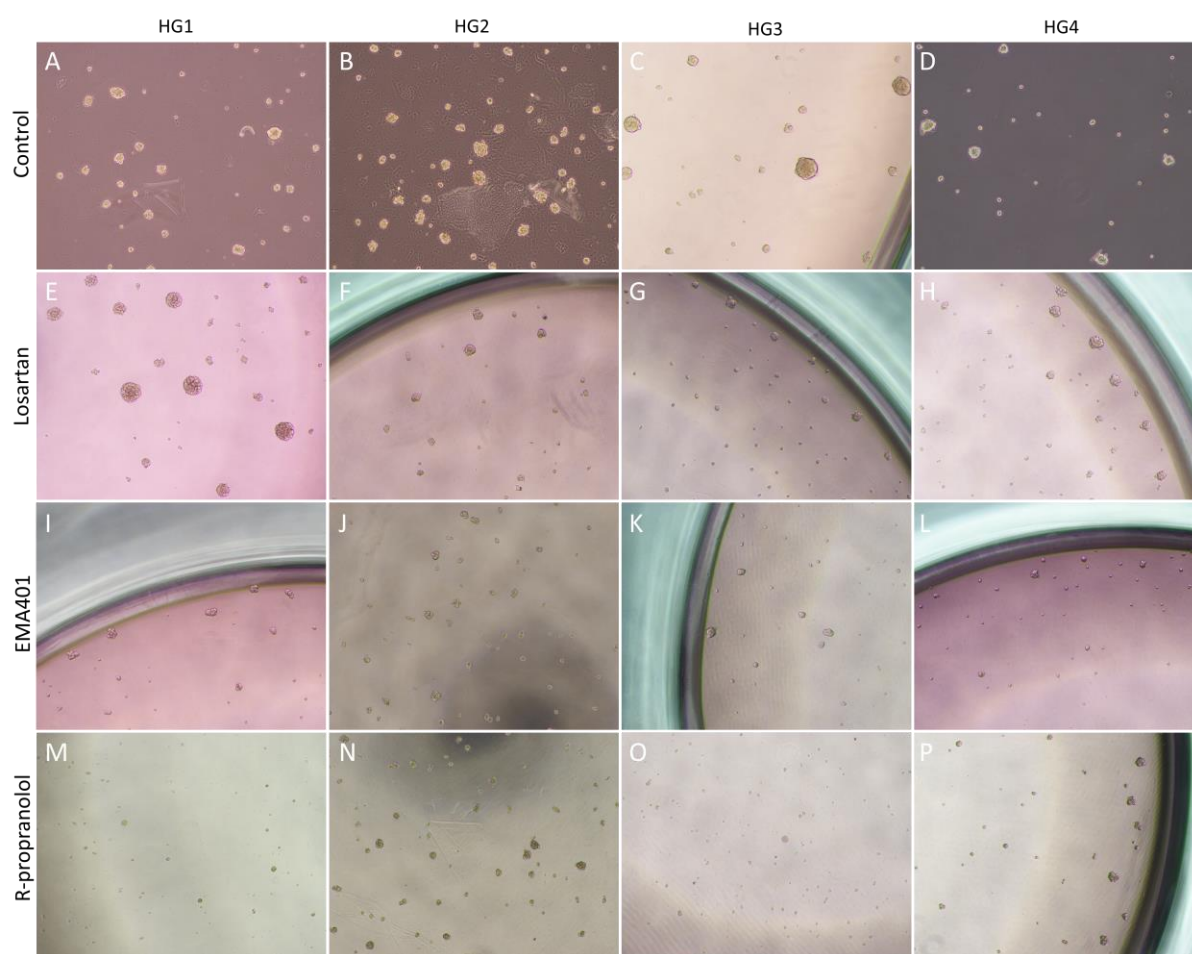


Figure 8.6: Effect of RAS modulators on tumoursphere formation. Four HGCA-derived cells lines were exposed to losartan (E-H), EMA401 (I-L) and R-propranolol (M-P). Untreated cells were used as a control (A-D). Images show first-passage tumourspheres. Original magnification: 100x.

An average diameter of 50 μm across measured tumourspheres was chosen as the threshold for successful tumoursphere formation by a cell line. Control cells from all 4 cell lines were able to reach the threshold in the first passage (Figure 8.6A-D), and while all 4 cell lines formed sphere-like structures in the second passage only HGCA3 reached the 50 μm threshold (Table 8.1). This result suggested that there were few sphere-forming cells present in each cell line. When exposed to losartan, each cell line still formed tumourspheres which had an average diameter near or greater than the threshold (Figure 8.6E-H), but could not be recapitulated in a second passage (Table 8.1). HGCA1 cells dosed with losartan produced larger tumourspheres than the control cells in passage 1 but failed in passage 2. Cells exposed to EMA401 reached the threshold in HGCA1 but fell just short in the other 3 cell lines (Figure 8.6I-L; Table 8.1). Similarly, cells dosed with R-propranolol reached the threshold in 2 cell lines (Figure 8.6M-P; Table 8.1). Overall, the RAS modulators had a variable effect on tumoursphere formation in the first passage, but appeared to limit the ability of cells to

recapitulate tumourspheres in the second passage, with EMA401 reducing tumoursphere size the most consistently.

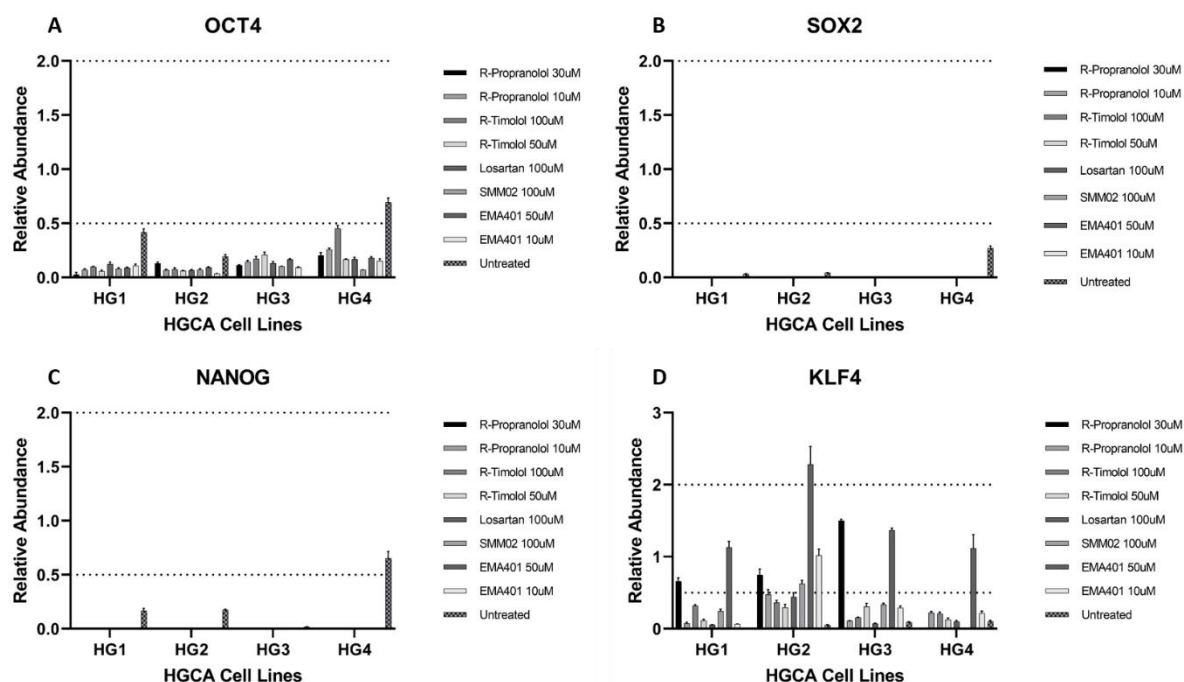


Figure 8.7: mRNA levels of iPSC markers in HGCA-derived cells treated with RAS modulators. RNA was extracted from 4 HGCA-derived cell lines that had been exposed to RAS modulators, as well as untreated controls, and the relative abundance of *OCT4* (A), *SOX2* (B), *NANOG* (C) and *KLF4* (D) mRNA from these cell lines relative to the NC tissues was calculated (y-axis). Error bars show standard deviation from the average of technical replicates after removing outliers (CT value ± 0.5 from average).

8.2.3 RT-qPCR on treated cells

Based on the initial metabolic assays, cells were dosed with a concentration of drug estimated to cause a significant change in metabolism without extensive cell death, to allow enough RNA to be collected for RT-qPCR. Cells were grown in 24-well plates and exposed to R-propranolol (30 μ M and 10 μ M), R-timolol (100 μ M and 50 μ M), EMA401 (50 μ M and 10 μ M), SMM02 (100 μ M) and losartan (100 μ M), before lysis and RNA collection.

OCT4 mRNA expression (Figure 8.7A) was lower in the untreated control cells than the pooled NC tissues they were derived from, and was unable to be detected in HGCA3 untreated cells. However, cells exposed to all 5 RAS modulators expressed significantly less *OCT4* than untreated cells in 3 of the cell lines. Conversely, *OCT4* mRNA was detected in treated HGCA3 cells despite being below the detection threshold in untreated cells.

SOX2 mRNA (Figure 8.7B) was expressed in 3 of the untreated cell lines, but was below detectable levels in HGCA3. Exposure to any of the RAS modulators caused a reduction in *SOX2* mRNA levels, which could not be detected in any treated cells.

Similarly, *NANOG* (Figure 8.7C) was present in all 4 untreated cell lines, though at very low levels in HGCA3, and was below the detection threshold in all cells treated with any RAS modulator.

The highest *KLF4* expression levels are reportedly observed in the NC, with decreased levels in LGCA and the lowest expression in HGCA. Cells treated with 50 μ M doses of EMA401 or 100 μ M doses of SMM02 exhibited upregulation of *KLF4* mRNA in 4 and 3 cell lines, respectively. Similarly, 30 μ M of R-propranolol or 100 μ M R-timolol caused an increase in *KLF4* mRNA levels in 3 and 4 cell lines, respectively. All 5 RAS modulators at all concentrations led to increased *KLF4* mRNA levels in the HGCA1 and HGCA2 cell lines.

Overall, the RAS modulators caused a reduction in the mRNA levels of pluripotency markers *OCT4*, *SOX2* and *NANOG*, and increased the expression of the goblet lineage differentiation-associated marker *KLF4*. Despite the lack of effect of the ARBs on metabolism, losartan was able to slightly hinder tumoursphere formation in 3 cell lines and to change the expression patterns of iPSC markers. R-propranolol, EMA401 and SMM02 inhibited metabolism and altered iPSC gene expression, while R-propranolol and EMA401 also reduced the size of tumourspheres formed by treated cells relative to untreated controls.

8.3 Discussion

The aim of this chapter was to assess the effects of RAS modulators on the metabolism, tumoursphere forming capacity and iPSC marker transcription in CA-derived primary cell lines. As hypothesised, β -blockers but not ACEIs were able to reduce cell metabolism. However, unexpectedly, ARBs had no effect on metabolism, but AT2R antagonists did cause a reduction in metabolic activity. Furthermore, although the ARB losartan did not affect metabolism, it was able to reduce iPSC gene transcription, as were R-propranolol, R-timolol, EMA401 and SMM02.

The tumoursphere assay is a method to enrich a cell line for pluripotent cells, which grow as colonies in suspension in low-attachment plates, as opposed to differentiated cells which undergo anoikis. Tumourspheres should be able to be maintained for many passages, and typically adapt to the conditions of suspension culture within 3 passages. Therefore, the aim for these tumoursphere assays was to complete 3 serial passages in which the average tumoursphere diameter reached 50 μ m. However, only 1 untreated cell line managed to reach the third passage, where it failed to reach

the threshold. This suggests that by the time the cell lines were used for the tumoursphere assays, there were few, if any, pluripotent cells remaining, or that the tumoursphere culture conditions were not ideal. The method by which a cell line is established is vital to ensuring that the desired cell types are present in the cell line. The cell lines in the GMRITB were derived by explant in Matrigel, as described in Chapter 5, and then expanded in DMEM media with 10% FCS supplement. These conditions tend to induce cellular differentiation and favour the growth of fibroblasts, and so it is possible that any CSCs in the parent tissue had differentiated or died in culture before the cell line was banked. However, the sequencing results from Chapter 3 showed that the cell lines do contain the same mutational signatures as the parent tissue, and so they are a suitable *in vitro* model of the original tumour. Furthermore, they were capable of being induced to differentiate down specific lineages, as presented in Chapter 5.

In this context it is interesting to consider pluripotency as a phenotype of a population rather than of individual cells, as touched on in Chapter 4 (MacArthur et al., 2013). The various differentiation media used in Chapter 5 are supplemented with factors that guide cells towards a particular lineage. For example, the endodermal differentiation kit includes FGF, Activin A and Wnt3a supplements. If the cancer cells are growing in an environment which favours the development and growth of endodermal cells, then some members of the population which are poised to revert to a pluripotent state may now have the impetus to do so, in order to differentiate down the endodermal lineage and exploit the changed conditions. This is analogous to cancer cells responding to a change in their microenvironment, such as the removal of tumour bulk by chemotherapy or radiotherapy, which tip the balance in favour of reversion to a more pluripotent state in order to replenish the lost tumour cells. In line with this, if the cancer cells were not grown in the correct media from the time they were isolated, then it may be possible to prime them into a more stem-like state by growing them in a defined media before performing the tumoursphere assays. Therefore, because the CA-derived cells retain the mutational signature of the original tumour, some of them could be induced into a sphere-forming capacity due to the StemXVivo™ media and suspension culture conditions, though at a lower capacity than cells maintained in the optimal conditions from the time of isolation.

With the exception of HGCA3, untreated cells expressed detectable levels of *OCT4*, *SOX2* and *NANOG*, which were all downregulated by administration of RAS modulators (Figure 8.7A-C). Furthermore, while *KLF4* was detected in untreated cells, its expression was significantly lower than in the NC tissue, and cells treated with R-propranolol or EMA401 recovered *KLF4* expression to levels close to that of the NC tissue (Figure 8.7D). This change in the expression of pluripotency genes was seen at doses which caused minor changes to metabolism. These results highlighted that RAS modulators may be capable of targeting CSCs by attenuating iPSC gene expression and guiding them

away from a pluripotent phenotype. If this is the case, it may enhance the effect of standard therapies on CA tumours by making CSCs more susceptible and therefore dampening their ability to cause recurrence and metastasis. In the future, it would be interesting to investigate whether CA-derived cells have lower resistance to drugs such as 5-FU, OX and IR if simultaneously exposed to RAS modulators, and the effect this might have on iPSC marker expression and stem cell functionality.

ACEIs have been shown to be beneficial in cancer patients. The ACE protein was not detected in CA-derived cells by WB or in CA tissues by IF staining, with weak staining detected on the luminal membrane of CA epithelial cells via IHC staining. This weak staining was considered to be a non-specific blush due to the lack of staining by the more sensitive IF method. In Chapter 7, RT-qPCR detected ACE mRNA, and generally at higher abundance in CA tissues than the matched NC samples, but mRNA was below detectable levels in the CA-derived cells. The expression of ACE mRNA in the tissues could suggest that the IHC staining was true and that despite a lack of effect from ACEIs on cells in the drug assay, there may be benefits *in vivo* where ACE is transcribed and possibly translated in CA tumours. Despite the lack of effect on cell metabolism, the ACEI losartan caused reductions in *OCT4*, *SOX2* and *NANOG* mRNA levels, but only increased *KLF4* relative to untreated cells for 1 HGCA cell line. This shows that an effect on metabolism does not necessarily accompany a change in gene expression, and so in the future it would be interesting to assess the effect of ARBs on iPSC gene expression. Without the detection of ACE in the cells, however, it might be that the cell lines used were derived from a cell type negative for ACE or have lost their expression of ACE, and are not a suitable model for studying ACEI efficacy on CA cells.

To determine whether any observations were specific to single drugs or were a class effect, 2 drugs from each class were investigated. Propranolol and timolol did not elicit comparable results, suggesting the effects of β -blockers may vary depending on their individual characteristics, such as lipophilicity, membrane stabilising effect or selectivity for adrenergic receptor types. This result was interesting because propranolol and timolol are both non-selective β -blockers and renin release is triggered via both the β_1 and β_2 adrenergic receptors (S. M. Kim et al., 2012). This indicates that the difference in efficacy is either due to the other differences in their properties, that the effects on metabolism are not related to renin release, or that the relative potency of propranolol and timolol is the reverse in cell culture compared to *in vivo*. Masur et al. (Masur et al., 2001) found that SW480 cell migration seems to be mediated by β_2 receptors because β -blockers selective for the β_1 receptor could not prevent migration whereas propranolol could, and propranolol also reduces colon cancer cell proliferation and viability (Coelho et al., 2015; Pantziarka et al., 2016). This indicates that the β_2 receptor is mediating the effects of β -blockers on colon cancer cells, which makes sense in that the type 2 receptors predominate in the GI tract (Wiysonge et al., 2007). Therefore, future experiments

using β_1 or β_2 receptor knock-out cells would reveal whether any effects are mediated by β -adrenergic receptors and, if so, clarify the significance of each receptor type in the response of cancer cells to β -blockers. It would also be interesting to measure renin levels before and after β -blocker treatment to see whether the effects are due to reduced renin and therefore reduced RAS signalling, or due to other β -adrenergic receptor-mediated outcomes.

Up until the 72 h time point, both AT₂R antagonists affected the cell lines very similarly, suggesting that it is a class effect of AT₂R antagonists. AT₁R is internalised and then recycled to the cell membrane when it binds ATII, whereas AT₂R is not internalised (Hein et al., 1997), therefore the difference in efficacy at 72 h should not be due to differing effects of the two drugs on the internalisation and recycling speed of the receptor. However, it could be due to the amount of time which the antagonist stays bound to AT₂R. One limitation of this experiment was the 72 h time frame; if repeated in the future, it would be worthwhile to modify the assay to enable readings beyond 72 hours. This may be achieved by setting up multiple plates and harvesting 1 per day, perhaps over the course of a week or more. A longer assay would provide a better idea of whether the drugs continue to affect the cells in the same way over time and resolve the observation of a possible plateau in efficacy or a loss of sensitivity, visualised in the drug assays for ARBs (Fig 8.4) and AT₂R antagonists (Fig 8.5). Furthermore, it would provide the drugs with more time to take effect, particularly on gene and protein expression.

Interestingly, the efficacy of β -blockers and AT₂R antagonists aligns with the hypothesis stated in Chapter 7, which proposed that cathepsins acting as bypass loops of the RAS would be capable of circumventing ACE inhibition, but that blockade at either end of the RAS would not allow cathepsins to act in this way. Accordingly, β -blockers and AT₂R antagonists seemed to have the greatest effects on metabolism, tumoursphere formation and iPSC mRNA expression by HGCA-derived cells.

8.4 Conclusions

The aim of this chapter was to investigate the effects of RAS modulators on the metabolism and CSC phenotype of primary CA-derived cell lines. Though there were smaller impacts on metabolism than expected based on published *in vivo* studies, the changes in gene expression and tumoursphere-forming capability suggest that RAS modulators may attenuate the functionality of CSCs. Furthermore, the role of AT₂R in CA may be greater than first thought, with EMA401 able to inhibit cancer cell metabolism in a dose-dependent manner, reduce iPSC gene expression and hinder tumoursphere formation. It is yet to be determined whether RAS modulators may work synergistically with standard chemotherapeutic agents by influencing treatment-resistant CSCs.

Chapter 9: General Discussion

9.1 Key findings

The CA-derived primary cell lines were compared to their parent CA tissues using DNA sequencing. Proteomics was performed to investigate the differences between proteomes of CA tissues and CA-derived cells, and each of these CA sample types relative to their matched NC counterparts. Expression and relative abundance of iPSC markers in the CA tissues and CA-derived cells was measured by WB, RT-qPCR and ISH. The localisation of iPSC markers and conventional CA CSC markers in CA tissues and patient-matched NC tissues was determined using IHC staining on tissue sections, and co-expression profiles were determined via IF staining. Furthermore, the RAS components were investigated using IHC staining, WB and RT-qPCR, and the co-localisation of iPSC markers with RAS components was visualised by IF staining. To assess whether the proposed CSCs identified using iPSC marker expression had stem cell function, tumoursphere formation assays and multilineage differentiation assays were performed. Tumoursphere formation, metabolic activity and iPSC marker gene transcription were then measured in CA-derived cell lines which had been treated with RAS modulators to determine whether RAS modulators could target CSCs. To the best of our knowledge, this is the first study to interrogate the expression of RAS components by colon CSCs, to test whether RAS modulators affect the metabolism or function of CSCs and to assess whether the R- and S- enantiomers of β -blockers behave differently to each other in this respect, and to examine the effects of RAS modulation on the expression of iPSC genes. Some key findings from the experiments outlined above are discussed below.

DNA sequencing was performed to validate the primary cell lines derived from the CA tissue samples. For the functional cell culture assays, it was important to know how representative the cell lines were of the patient-derived tumour tissue samples. Sequencing revealed that the majority of mutations detected within each patient were shared by both the FFPE CA tissues and the CA-derived cell lines, and were often shared between patients. The retention of identical mutations to key CA-related genes, including *APC* and *TP53*, confirms that the primary cell lines are a suitable model for *in vitro* experiments, but the lower TMB scores in cells suggest that they do not reflect the full heterogeneity of the tumour tissues.

The HGCA1 tissue sample and cell line clustered with the LGCA samples for multiple parameters. Initial analysis of the DNA sequencing data indicated similarity between HGCA1 and the 3 LGCA patients in terms of their alignment with the CMS3 category. Tissue samples from these 4 patients displayed a roughly equal proportion of epithelial and stromal cells with relatively low immune cell infiltration (Fig 9.1A-D), and shared the same c.483 C>T *KRAS* mutation which was absent

from the other 3 HGCA tissues. However, although the CMS2 and CMS3 categories are relatively similar to each other morphologically, CMS2 is predisposed to the distal colon whereas CMS3 is more common in the proximal colon (Loree et al., 2018; G. Martini et al., 2020). LGCA1, LGCA3 and HGCA1 were removed from the distal colon indicating they may be CMS2, whereas LGCA2 was from the proximal colon and is therefore more likely to be CMS3. CRC tumours are known to be heterogeneous and usually do not fit perfectly into a single CMS category but contain aspects of multiple categories, maybe reflecting transition between categories during tumour progression (Guinney et al., 2015; Sveen et al., 2019). Compared to the other 3 HGCA cell lines, untreated HGCA1 cells formed the smallest tumourspheres during passage 1 and 2, and its TMB score (7.75) was more similar to the 3 LGCA tissues (4.63, 5.92, 6.01) than to the other 3 HGCA tissues (10.49, 17.67, 38.65). As well as having a high mutation rate, these 3 HGCA tissues exhibited a much higher degree of immune cell infiltration and were located in the proximal colon, all of which seem to indicate they would be categorised as CMS1 (Guinney et al., 2015; Loree et al., 2018). HGCA3 and HGCA4 also bore similarity to CMS4 by containing dense nests of tumour epithelial cells surrounded by large amounts of stroma (Fig 9.1F,G) (Loree et al., 2018; G. Martini et al., 2020). The HGCA4 tissue and cell samples were from the only patient tested for MSI at the hospital during diagnosis where it appeared normal for MMR proteins by IHC staining, but the CMS categories for all samples would need to be verified by further MSI testing, DNA methylation analysis, or by analysing RNA-seq gene expression profiles using the “Single Sample Predictor” tool available as part of the CMS Subtype Classifier package in R software (Guinney et al., 2015). LGCA4 was unable to be used for DNA sequencing, and did not undergo IHC staining but was included for cell culture assays and the proteomics study.

IHC and IF staining were employed to compare the expression of iPSC markers and RAS components by CA tissues and their patient-matched NC tissues. OCT4 was found in the stroma of CA but never in the stroma of the NC, whereas NANOG was present in CA epithelial cells but absent in the NC tissues. The NC epithelial cells contained AT₂R and weakly expressed ACE and cathepsins B and D, while the muscle below the crypts stained strongly for PRR. Relative to patient-matched NC samples, CA tissues overexpressed AT₂R, PRR and cathepsin D, expressed ACE2 at normal levels, and ACE and cathepsin B at low levels. IF staining revealed the presence of a subpopulation of cells in the CA epithelium which co-expressed iPSC markers NANOG, SOX2, KLF4 and c-MYC, RAS components AT₂R and ACE2, and likely also express PRR, cathepsins B and D, and colon CSC markers LGR5, CD133 and EpCAM. Cells derived from CA tissues had variable expression profiles, especially for mRNA levels as measured by RT-qPCR. WB detected strong bands for KLF4, c-MYC, AT₂R, PRR and cathepsin B, and weak bands for NANOG and cathepsin D in cell lines derived from 3 LGCA and 3 HGCA tissues, while

OCT4 and SOX2 were detected in all HGCA-derived cell lines. However, ACE and ACE2 were below detectable levels in all cell lines.

There was alignment across the data from multiple techniques which validated these results. IHC staining of LGCA2 tissue sections showed moderate levels of cathepsins B and D in the epithelial cells. LGCA2 cells had the largest bands corresponding to cathepsin B and relatively strong bands for cathepsin D by WB, and moderate to high levels of mRNA for both cathepsins relative to the other cell lines as measured by PCR. Accordingly, LGCA2 had the highest activity levels of both cathepsins in both the cells and the tissues. Conversely, LGCA1 cells expressed the lowest levels of cathepsin B mRNA and similar levels of cathepsin D mRNA compared to the other cell lines, and weak epithelial staining for both cathepsins by IHC staining. Cathepsin activity corresponded to the mRNA and protein expression levels, with LGCA1 exhibiting the lowest activity of both cathepsins in tissues and the lowest cathepsin B activity in cells.

While quantitative assessment of iPSC marker expression in tissues by IHC staining was not undertaken, it was observed that the abundance of OCT4, SOX2 and c-MYC was generally higher in HGCA tissues than LGCA tissues. Furthermore, WB densitometry results confirmed that OCT4 and SOX2 were more abundant in HGCA cells than LGCA cells. Initially, manual cell counting was performed on the IHC stained CA tissues in this study (data not shown), and revealed that all cases could be accurately graded (canonical correlation $r=0.981$, Wilke's lambda variance $\Lambda=0.037$) based on 4 elements of data: KLF4 in the stroma of NC ($p=0.020$) and CA ($p=0.034$), and OCT4 ($p=0.001$) and NANOG ($p=0.026$) in CA epithelial cells. However, quantification from IHC stained sections was not completed due to re-staining with new SOX2, NANOG and c-MYC antibodies, and because of the biases inherent to manual quantitation. It was hoped that iPSC markers might be able to be used in a prognostic capacity, but this would need to be validated with further work. Ideally, this would involve scanning full images of IHC stained slides and uploading them to an unbiased whole-slide image analysis tool such as QuPath (Bankhead et al., 2017), in order to generate quantitative data regarding the abundances of iPSC proteins in tissue.

Large-scale label-free quantitative proteomic analysis was carried out on the CA tissues and CA-derived cell lines. The proteomics data revealed that CA tissues were enriched for components of mitochondrial respiration and oxidative phosphorylation, reflecting the increased energy demand of the cancer cells compared to the NC, and for inflammatory markers. Furthermore, the increased potential for migration, invasion and metastasis was highlighted by the significant reduction in adhesion molecules, and the effect on normal gut function was revealed by the loss of proteins involved in muscle structure and function. The CA-derived cell lines were particularly enriched for

proteins involved in RNA processing and translation, chromatin binding, DNA replication, and especially in spliceosome activity, which emphasise the higher levels of gene transcription, alternative mRNA splicing and DNA synthesis within cancer cells in response to dysregulated gene expression, uncontrolled cell proliferation, and tumour growth. Proteomic study found elevated levels of components of signalling pathways that are important to CA, including Wnt and PI3K signalling. PI3K subunits PIK3R1 and PIK3GC and Wnt-related proteins Rho-associated kinase 2 (ROCK2), casein kinase and SMADs were found to be upregulated, and these pathways are known to sit both upstream and downstream of the RAS. The influence of these pathways may be a factor in the results of RAS modulating experiments, where β -blockers and AT₂R antagonists caused reductions in iPSC marker transcription, as discussed below.

CA-derived cell lines were exposed to RAS modulators to assess whether they could alter their metabolism and stem cell characteristics. ACEIs and ARBs had no effect on cell metabolism, but propranolol and EMA401 doses greater than 10 μ M consistently reduced the metabolic activity of CA-derived cells. The effect of each RAS modulator on tumoursphere formation was cell line-specific. Relative to untreated cells, HGCA1 cells formed larger tumourspheres when treated with losartan, and smaller tumourspheres when treated with R-propranolol. R-propranolol, EMA401 and losartan all caused HGCA3 cells to form smaller tumourspheres. HGCA2 and HGCA4 cells were able to form tumourspheres of a similar or slightly reduced size compared to untreated cells when treated with any of the drugs. This highlights the heterogeneity of CA tumours and the fact that different patients respond differently to treatments. Interestingly, RAS modulators affected iPSC marker gene transcription even with no discernible change to metabolism. After treatment with R-propranolol, R-timolol, EMA401, SMM02 or losartan, expression of *SOX2* and *NANOG* mRNA could no longer be detected in any cell line, and expression of *OCT4* was reduced in 3 out of 4 cell lines. Furthermore, *KLF4* mRNA expression was increased by RAS modulator treatment. Across all the cell lines, HGCA4 had the highest TMB score and the highest *OCT4*, *SOX2* and *NANOG* mRNA expression by untreated cells. HGCA3 FFPE tissue had the highest TMB score but its matched cell line had a low TMB score, and the untreated cells expressed the lowest levels of *OCT4*, *SOX2* and *NANOG* mRNA. Interestingly, these results were mirrored in the tumoursphere assays following treatment with RAS modulators. HGCA4 appeared to be the least affected by RAS modulators in terms of tumoursphere forming capacity, and saw the smallest increases in *KLF4* mRNA expression, though still exhibited reductions in *OCT4*, *SOX2* and *NANOG* mRNA. Meanwhile, untreated HGCA3 cells reached the 50 μ m threshold in the first 2 passages but formed significantly smaller tumourspheres after treatment with each of the 3 RAS modulators tested. This could indicate that higher basal levels of iPSC marker expression are related to a higher tumour grade and confer resistance to RAS modulation, and that lower iPSC marker

expression leaves the cells more susceptible; yet it also reveals that RAS modulators have the potential to reduce iPSC marker gene expression. Furthermore, reductions to iPSC genes without affecting overall metabolism may suggest that the RAS modulators affect CSCs to a greater extent than they affect more differentiated tumour cells, or that they encourage tumour cells to shift towards a more differentiated phenotype.

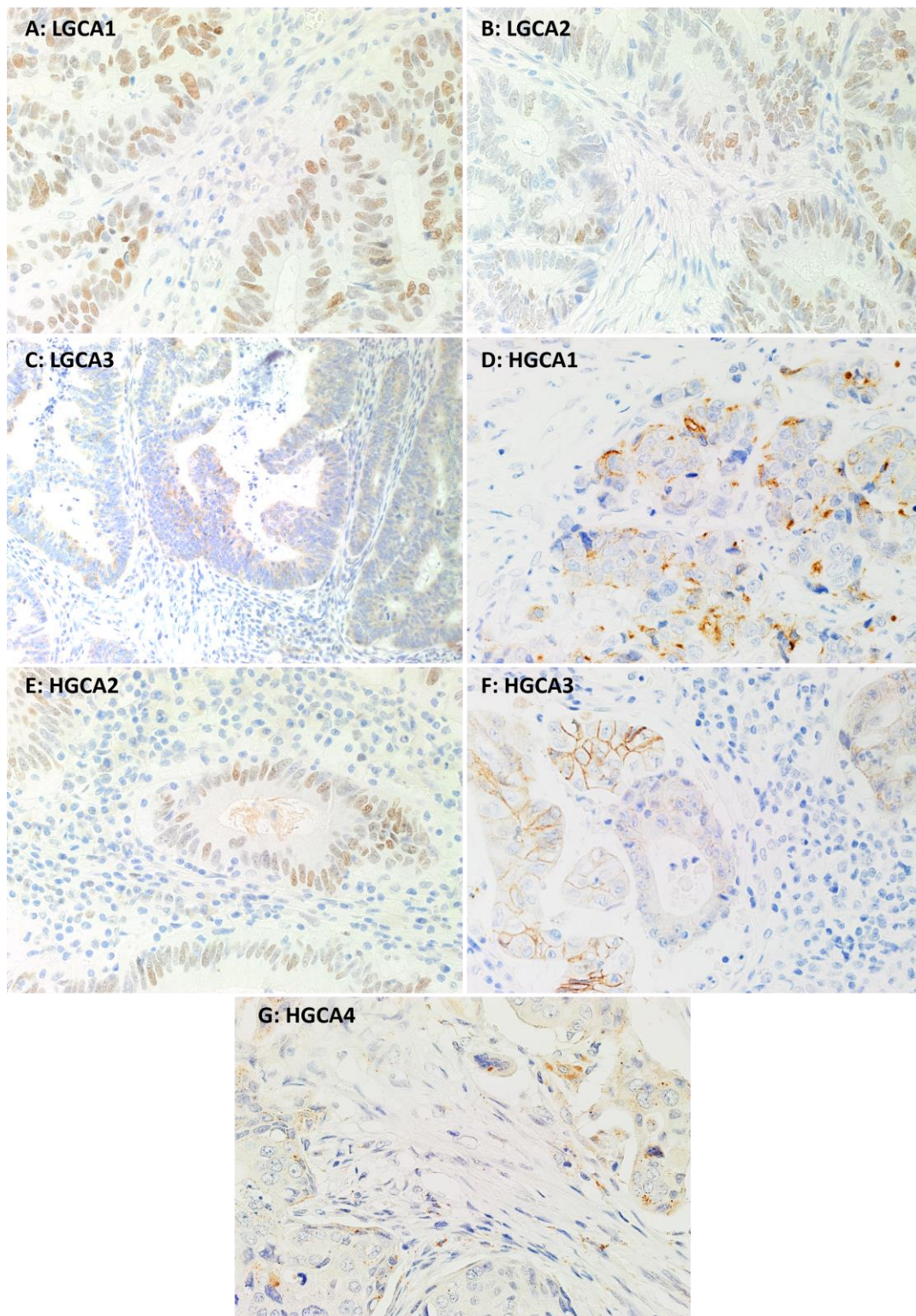


Figure 9.1: Representative images of CA cases used for DNA sequencing, mass spectrometry and cell culture (page 199). Representative images captured from sections of FFPE tissue, showing the general architecture of the LGCA1-3 (A-C) and HGCA1-4 (D-G) tumours. LGCA4 was unable to be used for DNA sequencing, and did not undergo any IHC staining but was included for cell culture assays and MS. Original magnification: 400x, except C (200x).

9.2 Clinical Significance

Over the last decade, there has been increasing focus on the role of the RAS in cancer. While many retrospective studies and meta-analyses have been published, there are increasing numbers of prospective clinical trials, studies in animal models and *in vitro* mechanistic experiments.

This project explored the influence of the RAS on just a few aspects of cancer cell biology, namely metabolic activity, iPSC marker expression and tumoursphere formation, whereas there are many published studies that offer a glimpse into other interactions. For example, breast cancer cells treated with ATII exhibit reduced expression of integrin proteins and cellular adhesion molecules, and this was reversed by the ARB losartan but not by the AT₂R antagonist PD123319 (Puddefoot et al., 2006). ARBs and ACEIs have been shown to act on cancer in two ways: by lowering cancer cell proliferation and by reducing neovascularisation (S. T. Kim et al., 2012). This is an important reminder that not all of the anti-cancer activity displayed by RAS modulators is due to direct action on tumour cells, but also on aspects of the microenvironment and the patient's normal physiology. In this respect, it is not surprising that only β -blockers and AT₂R antagonists showed consistent effects on CA-derived primary cell lines, especially given the relatively limited range of experiments undertaken. In Chapter 8, ACEIs and ARBs did not have significant influence over cell metabolism. However, most of the beneficial effects which they exert in a cancer context are not involved directly in cancer cell intracellular activity. For example, candesartan enhances the effects of radiation therapy by reducing intra-tumoural angiogenesis (Ohnuma et al., 2009). Nevertheless, the experiments reported in Chapter 8 have led to the conclusion that any beneficial effects which cancer patients receive from ACEI or ARB administration are not due to their direct influence on cancer cell metabolism, however, the β -blockers and AT₂R antagonists may act in such a way. Published *in vivo* data have shown that macrophage recruitment is stimulated by ATII/AT₁R signalling and reduced by ATII/AT₂R signalling, indicating that AT₂R antagonists may also aid the anti-tumour immune response by increasing immune infiltration (Ager et al., 2008). Collectively, this shows that the RAS has the potential to influence CSCs directly, such as by regulating gene transcription, or indirectly, by modulating their interactions with the tumour microenvironment and the immune system.

Attention has been given to the prospect that RAS modulators in combination with standard therapies can improve patient outcomes. Losartan and candesartan have been observed to increase the anti-tumour activity of gemcitabine in a mouse model of pancreatic cancer (Noguchi et al., 2009), and improve the OS and PFS of patients receiving gemcitabine for advanced pancreatic cancer (Nakai et al., 2010). In a mouse xenograft study, captopril alone reduced the size of gastric adenocarcinoma tumours by ~40%, which was enhanced to 70% when administered with cisplatin (Williams et al., 2005). Similarly, synergy between apoptosis-inducing platinum-based chemotherapeutics and anti-angiogenic ACEIs and ARBs improved control of bladder cancer in a mouse xenograft model (Kosugi et al., 2009), improved OS in patients with recurrent or metastatic gastric cancer (S. T. Kim et al., 2012), and increased median survival of non-small cell lung cancer (NSCLC) patients from 8.6 months to 11.7 months (Wilop et al., 2009). The same synergistic potential has been observed with β -blockers, most notably in combination with vinblastine, paclitaxel and/or radiation in inoperable cases of advanced angiosarcoma, a tumour of vascular origin which was hypothesised to be susceptible to the anti-angiogenic effects of RAS modulators (Pasquier et al., 2016). 5-FU, one of the front-line treatments for CA, had an additive effect with propranolol on reducing the proliferation of breast cancer cells (Pasquier et al., 2011). While there have been few prospective studies investigating 5-FU, or any other chemotherapeutics for CA, in combination with RAS modulators, these results collectively encourage further work in this area. To the best of our knowledge, there are no studies which investigate the effects of RAS modulators on the expression of the iPSC markers in CA, and the results of this thesis indicate that further investigation is merited.

Improved outcome in cancer patients administered RAS modulators is in part due to their ability to reduce the side effects of conventional treatments. Specifically, ATII has been associated with cardiotoxicity, especially left ventricular dysfunction, arrhythmia and structural changes to the heart (Pinter et al., 2018). Radiotherapy has been shown to increase ATII production, which induces fibrosis, inflammation and ROS generation (Pinter et al., 2018). RAS modulators could therefore lead to reduced doses of radiotherapy or chemotherapy without a loss of therapeutic effect, and prevent late-onset cardiac damage (Pinter et al., 2018). However, given that most of the cathepsin expression within the tumour is due to immune cells within the stroma, it is unclear whether RAS bypass loops will be more apparent *in vivo* and therefore attenuate the effects of ACEIs and ARBs to some degree.

9.3 Limitations and Future Directions

The cell lines in the GMRITB were derived from CA tumour tissues and had been banked between 2015 and 2018. One limitation of the project was that these cells had not been characterised before

being banked, and the explant method of cell line derivation meant that the nature of the cells was unclear. If this project were to be repeated, it would be better to isolate the cells via a dissociation and cell sorting method, followed by expansion in a culture media formulated specifically for the cell type of interest. In this way, there could be more confidence in the relevancy of results from cell-based experiments. However, this alternative was not suited to this PhD project due to uncontrollable factors such as the frequency of operations and tissue donation rates, which would have prevented experimental replicates from being performed in parallel and may have resulted in lower numbers of samples, and so it was beneficial to have a repository of pre-banked cell lines to draw from.

To fully explore the effects of RAS modulators on cancer cell activity, and specifically on CSC phenotype, further functional assays can be explored, including angiogenesis, migration and invasion experiments. This will allow for the roles of each class of RAS modulator to be delineated. The use of two members of each RAS modulator class in this project was insufficient to determine whether there were any “class effects”. While the two ARBs and the two ACEIs produced comparable results in the metabolic activity assay, the two β -blockers elicited differing responses, which may be due to potency, but does not confirm whether there is a class effect. Extending the experiment beyond 72 h may help to resolve these ambiguities. If pursued further in the future, these experiments would be enhanced by including a larger cohort to reduce the effect of variation between samples and strive for statistical significance, and a larger range of RAS modulators. ACEIs and ARBs may perform better *in vivo* because they seem to have fewer direct effects on tumour cells (S. T. Kim et al., 2012), and so animal studies may be required to simulate the complex *in vivo* environment in which these drugs may have a more apparent influence. This project tested the R- and S-enantiomers of propranolol and timolol to determine whether there was any difference in effect between the two isoforms. Previous studies have shown that only S-propranolol exerts anti-adrenergic effects, whereas R-propranolol is anti-angiogenic (Overman et al., 2019; Sasaki et al., 2019). R-propranolol acts as a direct inhibitor of mutant varieties of the SOX18 transcription factor, preventing the expression of hemangioma markers and inhibiting the differentiation of hemangioma stem cells into endothelial cells (Overman et al., 2019). Interestingly, the wild-type SOX18 protein is equally inhibited by the R- and S-enantiomers, and so the stereoselectivity is thought to be due to the change in conformation of mutant SOX18, leading to preferential binding of R-propranolol (Overman et al., 2019). It is possible that R- and S-propranolol act as small molecule inhibitors in CA to affect the transcription of iPSC genes. It appears that the R-enantiomer is responsible for reduced expression of VEGF and angiopoietin-like 4 to allow vascular regression, mediated by angiopoietin-2, in hemangioma patients (Sasaki et al., 2019). Racemic propranolol is more toxic than either of the enantiomers alone (Overman et al., 2019), so administration of R-propranolol alone would greatly reduce the side-effects caused by β -adrenergic

receptor antagonism and possibly allow for a higher dose. Propranolol was found to induce cell death independently of β -adrenergic receptor expression in hemangiomas (Overman et al., 2019), so it would be useful to confirm the expression of β -adrenergic receptors in CA tissues and CA-derived cells using IHC staining and RT-qPCR, and to then knock out the β 1 and β 2 adrenergic receptors to delineate the mechanism by which β -blockers affect CA-derived cells. Overall, these results suggest that RAS modulators, specifically β -blockers, may function through mechanisms which are currently unknown to target stem cells by modulating gene expression, independently of β -adrenergic receptors, and in a manner that might vary between drug isomers.

There is a need for alternative or combinatorial approaches to treat CA due to its high rates of mortality and metastasis, especially to the liver. CSCs are thought to be responsible for recurrence as they resist conventional cancer treatments. The RAS is now acknowledged to play a role in cancer, with many of its downstream effects promoting cancer functions, including proliferation, inflammation, EMT and migration. The expression of RAS components by CSCs potentially offers a novel therapeutic target using commonly available low-cost medications. One of the latest developments in RAS biology is the discovery of an intracellular RAS, alongside the classical endocrine RAS and the tissue-specific paracrine RAS. Evidence of an intracellular RAS emerged around 2004 with the observation that intracellular ATII produced from a plasmid caused cardiac hypertrophy in mice within 48 h, without causing increased blood pressure or ATII levels in circulation (Baker et al., 2004). Numerous studies have found that AT₁R is more abundant in the nuclear homogenate from brain cells than the whole cell homogenate, and that the nucleus and cytoplasm both contain AGT, AT₁R, AT₂R and PRR (Garrido-Gil et al., 2013; Villar-Cheda et al., 2017). IF staining in Chapter 7 revealed the AT₂R was occasionally localised to the nucleus of CA stromal cells. The intracellular RAS predominantly involves AT₁R localised to the nuclear membrane, which binds intracellular ATII leading to internalisation of the receptor to the nucleus (Villar-Cheda et al., 2017). This leads to increased gene transcription, specifically of TGF- β and IGF-1 (Villar-Cheda et al., 2017). It has been shown that the intracellular RAS counteracts the oxidative stress induced by paracrine ATII acting on AT₁R at the plasma membrane, partly through AT₁R acting in the nucleus and partly due to AT₂R in the mitochondria (Villar-Cheda et al., 2017). It is possible that this mechanism could control the expression of other genes and establish the early reprogramming events that lead to disease, and that ARBs might inhibit this process (Villar-Cheda et al., 2017). This may explain how losartan reduced the expression of iPSC markers *OCT4*, *SOX2* and *NANOG* and increased *KLF4* transcription in CA-derived cell lines without affecting overall metabolism.

The observations of reduced iPSC marker gene transcription in response to RAS modulators presented in Chapter 8 encourages further work to determine whether a combination of conventional

therapeutics and RAS modulators could be employed to target the tumour bulk and make CSCs more susceptible to conventional therapies, while reducing angiogenesis, enhancing anti-tumour immune infiltration and reducing cardiotoxic side effects to improve patient outcomes. In terms of the view of pluripotency as a population phenotype rather than that of single cells (MacArthur et al., 2013), the reduction in iPSC marker expression caused by RAS modulators may reflect a shift in the balance from a resistant, stem-like state to a state which is less adaptable and more susceptible to treatment.

There are two further avenues of research which warrant mention here. Standard 2D cell culture systems cannot reflect the complexity of human organ systems and have therefore led to a bottleneck between *in vitro* models and their clinical application (Kim et al., 2020). However, animal models do not always faithfully represent human physiology and have inherent ethical issues (Kim et al., 2020). Organoid cultures are spontaneously self-assembling 3D *in vitro* models of human organs, which can be produced from iPSCs or adult stem cells (ASCs) (Drost et al., 2018; Kim et al., 2020). Tissue biopsy-derived ASCs produce organoids that recapitulate key aspects of the organs they are derived from, including biological processes that cannot be modelled in animals or 2D cultures (Drost et al., 2018; Kim et al., 2020). ASC-derived organoids create simpler models than iPSC-derived organoids because iPSCs can form a wider range of cell types. iPSCs can be created by reprogramming patient-derived cells, allowing for a more representative drug screening platform for developing personalised treatments (Drost et al., 2018; Kim et al., 2020). Furthermore, organoids can be produced from normal cells and manipulated via CRISPR to address specific questions, such as the sequence in which different mutations arise during carcinogenesis (Ringel et al., 2020). This system would go some way towards addressing the discrepancies between the 2D *in vitro* experiments presented in this thesis and *in vivo* trials in terms of modelling the documented effects of RAS modulators in a more complex system. Secondly, it has been shown that certain bacteria are enriched in the microbiomes of CRC patients, and this dysbiosis is thought to influence gene expression and contribute to carcinogenesis (Purcell et al., 2017). In CRC, the specific bacterial species present varies depending on the CMS category of the tumour, and it is hypothesised that the microbiota which are present in the patient's gut contribute to the characteristics of the tumour (Purcell et al., 2017). In this context, it is possible that specific microorganisms may cause dysregulation of iPSC genes, encouraging a CSC phenotype and leading to carcinogenesis. Organoid culture could be employed to replicate the tumour environment and allow a CA patient-derived microbiome to survive *ex vivo* in order to explore the effects of the microbiome on drug efficacy.

9.4 Concluding Remarks

CRC is one of the most common and life-threatening cancers worldwide and its poor prognosis is linked to recurrence and metastasis. Therefore, treating the causes of these phenomena is a focus of much research. CSCs, one proposed origin of cancer, can resist therapy to initiate recurrence and metastasis.

This thesis explored two main hypotheses. The first was that CSCs in CA could be identified by their expression of the genes used to create iPSCs and that control pluripotency, a characteristic of ESCs and also CSCs. To address this, the thesis aimed to identify and characterise the CSC subpopulations in FFPE tissues via IHC staining, in-situ hybridisation and RT-pPCR, and to assess the stem cell properties of cell lines derived from the same tissues by performing tumoursphere and differentiation assays. It was found that iPSC marker abundance increased with CA grade, as did the tumoursphere formation capability of CA-derived cell lines. The second hypothesis was that these CSCs would express RAS components, and that the RAS may play a role in CSC regulation. IHC and IF staining were used to assess the co-expression of iPSC markers with RAS components, and suggested that CA CSCs express RAS components. When treated with RAS modulators, these cells exhibited reduced expression of iPSC markers. AT₂R antagonists showed unexpected promise as a class of compounds for reducing cancer cell metabolism and CSC phenotype, and warrants further exploration.

These results encourage future work to elucidate whether low-cost RAS modulating medications, which are already in clinical use, may be able to work in tandem with conventional therapies by dampening the cancer stem cell phenotype and thereby rendering them more sensitive to treatment.

References

- Abujarour, R., Valamehr, B., Robinson, M., Rezner, B., Vranceanu, F., & Flynn, P. (2013). Optimized surface markers for the prospective isolation of high-quality hiPSCs using flow cytometry selection. *Sci Rep*, 3, 1179. doi:10.1038/srep01179
- Adam, R., Delvart, V., Pascal, G., Vaneanu, A., Castaing, D., Azoulay, D., . . . Bismuth, H. (2004). Rescue surgery for unresectable colorectal liver metastases downstaged by chemotherapy: a model to predict long-term survival. *Ann Surg*, 240(4), 644-657; discussion 657-648. doi:10.1097/01.sla.0000141198.92114.f6
- Ager, E. I., Neo, J., & Christophi, C. (2008). The renin-angiotensin system and malignancy. *Carcinogenesis*, 29(9), 1675-1684. doi:10.1093/carcin/bgn171
- Aggarwal, N., & Sloane, B. F. (2014). Cathepsin B: multiple roles in cancer. *Proteomics Clin Appl*, 8(5-6), 427-437. doi:10.1002/prca.201300105
- Akbar, S., & Alorainy, M. S. (2014). The current status of beta blockers' use in the management of hypertension. *Saudi Med J*, 35(11), 1307-1317. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25399206>
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A*, 100(7), 3983-3988. doi:10.1073/pnas.0530291100
- Alexander, R. E., Cheng, L., Grignon, D. J., & Idrees, M. T. (2014). Cytoplasmic OCT4 staining is a sensitive marker of neuroendocrine differentiation. *Hum Pathol*, 45(1), 27-32. doi:10.1016/j.humpath.2013.08.006
- Amini, S., Fathi, F., Mobalegi, J., Sofimajidpour, H., & Ghadimi, T. (2014). The expressions of stem cell markers: Oct4, Nanog, Sox2, nucleostemin, Bmi, Zfx, Tcl1, Tbx3, Dppa4, and Esrrb in bladder, colon, and prostate cancer, and certain cancer cell lines. *Anat Cell Biol*, 47(1), 1-11. doi:10.5115/acb.2014.47.1.1
- Aronson, J. K. (2008). Changing beta-blockers in heart failure: when is a class not a class? *Br J Gen Pract*, 58(551), 387-389. doi:10.3399/bjgp08X299317
- Assou, S., Le Carrou, T., Tondeur, S., Strom, S., Gabelle, A., Marty, S., . . . De Vos, J. (2007). A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells*, 25(4), 961-973. doi:10.1634/stemcells.2006-0352
- Atlasi, Y., Mowla, S. J., Ziaee, S. A., Gokhale, P. J., & Andrews, P. W. (2008). OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells*, 26(12), 3068-3074. doi:10.1634/stemcells.2008-0530
- Aulakh, G. K., Sodhi, R. K., & Singh, M. (2007). An update on non-peptide angiotensin receptor antagonists and related RAAS modulators. *Life Sci*, 81(8), 615-639. doi:10.1016/j.lfs.2007.06.007
- Avery, S., Inniss, K., & Moore, H. (2006). The regulation of self-renewal in human embryonic stem cells. *Stem Cells Dev*, 15(5), 729-740. doi:10.1089/scd.2006.15.729
- Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev*, 17(1), 126-140. doi:10.1101/gad.224503
- Aznar, N., Ear, J., Dunkel, Y., Sun, N., Satterfield, K., He, F., . . . Ghosh, P. (2018). Convergence of Wnt, growth factor, and heterotrimeric G protein signals on the guanine nucleotide exchange factor Daple. *Sci Signal*, 11(519). doi:10.1126/scisignal.aao4220
- Aznar, N., Sun, N., Dunkel, Y., Ear, J., Buschman, M. D., & Ghosh, P. (2017). A Daple-Akt feed-forward loop enhances noncanonical Wnt signals by compartmentalizing beta-catenin. *Mol Biol Cell*, 28(25), 3709-3723. doi:10.1091/mbc.E17-06-0405

- Baillie, R., Itinteang, T., Yu, H. H., Brasch, H. D., Davis, P. F., & Tan, S. T. (2016). Cancer stem cells in moderately differentiated oral tongue squamous cell carcinoma. *J Clin Pathol*, 69(8), 742-744. doi:10.1136/jclinpath-2015-203599
- Baker, K. M., Chernin, M. I., Schreiber, T., Sanghi, S., Haiderzaidi, S., Booz, G. W., . . . Kumar, R. (2004). Evidence of a novel intracrine mechanism in angiotensin II-induced cardiac hypertrophy. *Regul Pept*, 120(1-3), 5-13. doi:10.1016/j.regpep.2004.04.004
- Balakumar, P., & Jagadeesh, G. (2014). A century old renin-angiotensin system still grows with endless possibilities: AT1 receptor signaling cascades in cardiovascular physiopathology. *Cell Signal*, 26(10), 2147-2160. doi:10.1016/j.cellsig.2014.06.011
- Bankhead, P., Loughrey, M. B., Fernandez, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., . . . Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Sci Rep*, 7(1), 16878. doi:10.1038/s41598-017-17204-5
- Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., . . . Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*, 449(7165), 1003-1007. doi:10.1038/nature06196
- Basakran, N. S. (2015). CD44 as a potential diagnostic tumor marker. *Saudi Med J*, 36(3), 273-279. doi:10.15537/smj.2015.3.9622
- Beggs, A. D., Jones, A., El-Bahrawy, M., Abulafi, M., Hodgson, S. V., & Tomlinson, I. P. (2013). Whole-genome methylation analysis of benign and malignant colorectal tumours. *J Pathol*, 229(5), 697-704. doi:10.1002/path.4132
- Beitia, M., Solano-Iturri, J. D., Errarte, P., Sanz, B., Perez, I., Etxezarraga, M. C., . . . Larrinaga, G. (2019). Altered expression of renin-angiotensin system receptors throughout colorectal adenoma-adenocarcinoma sequence. *Int J Med Sci*, 16(6), 813-821. doi:10.7150/ijms.32599
- Ben-Porath, I., Thomson, M. W., Carey, V. J., Ge, R., Bell, G. W., Regev, A., & Weinberg, R. A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet*, 40(5), 499-507. doi:10.1038/ng.127
- Benes, P., Vetvicka, V., & Fusek, M. (2008). Cathepsin D--many functions of one aspartic protease. *Crit Rev Oncol Hematol*, 68(1), 12-28. doi:10.1016/j.critrevonc.2008.02.008
- Benicky, J., Hafko, R., Sanchez-Lemus, E., Aguilera, G., & Saavedra, J. M. (2012). Six commercially available angiotensin II AT1 receptor antibodies are non-specific. *Cell Mol Neurobiol*, 32(8), 1353-1365. doi:10.1007/s10571-012-9862-y
- Bian, B., Mongrain, S., Cagnol, S., Langlois, M. J., Boulanger, J., Bernatchez, G., . . . Rivard, N. (2016). Cathepsin B promotes colorectal tumorigenesis, cell invasion, and metastasis. *Mol Carcinog*, 55(5), 671-687. doi:10.1002/mc.22312
- Boland, C. R., & Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology*, 138(6), 2073-2087.e2073. doi:10.1053/j.gastro.2009.12.064
- Bolocan, A., Ion, D., Stoian, R. V., & Serban, M. B. (2011). Map syndrome (MYH Associated Polyposis) colorectal cancer, etiopathological connections. *J Med Life*, 4(1), 109-111. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/21505584>
- Boman, B. M., & Huang, E. (2008). Human colon cancer stem cells: a new paradigm in gastrointestinal oncology. *J Clin Oncol*, 26(17), 2828-2838. doi:10.1200/JCO.2008.17.6941
- Booth, H. A., & Holland, P. W. (2004). Eleven daughters of NANOG. *Genomics*, 84(2), 229-238. doi:10.1016/j.ygeno.2004.02.014
- Boulting, G. L., Kiskinis, E., Croft, G. F., Amoroso, M. W., Oakley, D. H., Wainger, B. J., . . . Eggan, K. (2011). A functionally characterized test set of human induced pluripotent stem cells. *Nat Biotechnol*, 29(3), 279-286. doi:10.1038/nbt.1783
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., . . . Young, R. A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6), 947-956. doi:10.1016/j.cell.2005.08.020

- Bradshaw, A., Wickremsekera, A., Tan, S. T., Peng, L., Davis, P. F., & Itinteang, T. (2016). Cancer Stem Cell Hierarchy in Glioblastoma Multiforme. *Frontiers in Surgery*, 3(21). doi:10.3389/fsurg.2016.00021
- Bradshaw, A. R., Wickremsekera, A. C., Brasch, H. D., Chibnall, A. M., Davis, P. F., Tan, S. T., & Itinteang, T. (2016). Glioblastoma Multiforme Cancer Stem Cells Express Components of the Renin-Angiotensin System. *Front Surg*, 3, 51. doi:10.3389/fsurg.2016.00051
- Braun, M. S., & Seymour, M. T. (2011). Balancing the efficacy and toxicity of chemotherapy in colorectal cancer. *Ther Adv Med Oncol*, 3(1), 43-52. doi:10.1177/1758834010388342
- Burster, T., Macmillan, H., Hou, T., Boehm, B. O., & Mellins, E. D. (2010). Cathepsin G: roles in antigen presentation and beyond. *Mol Immunol*, 47(4), 658-665. doi:10.1016/j.molimm.2009.10.003
- Buta, C., David, R., Dressel, R., Emgard, M., Fuchs, C., Gross, U., . . . Weitzer, G. (2013). Reconsidering pluripotency tests: do we still need teratoma assays? *Stem Cell Res*, 11(1), 552-562. doi:10.1016/j.scr.2013.03.001
- Butler, T. A. J., Paul, J. W., Chan, E. C., Smith, R., & Tolosa, J. M. (2019). Misleading Westerns: Common Quantification Mistakes in Western Blot Densitometry and Proposed Corrective Measures. *Biomed Res Int*, 2019, 5214821. doi:10.1155/2019/5214821
- Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330-337. doi:10.1038/nature11252
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., & Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5), 643-655. doi:10.1016/s0092-8674(03)00392-1
- Chan, A. T., Baba, Y., Shima, K., Noshio, K., Chung, D. C., Hung, K. E., . . . Ogino, S. (2010). Cathepsin B expression and survival in colon cancer: implications for molecular detection of neoplasia. *Cancer Epidemiol Biomarkers Prev*, 19(11), 2777-2785. doi:10.1158/1055-9965.EPI-10-0529
- Chang, P. Y., Huang, W. Y., Lin, C. L., Huang, T. C., Wu, Y. Y., Chen, J. H., & Kao, C. H. (2015). Propranolol Reduces Cancer Risk: A Population-Based Cohort Study. *Medicine (Baltimore)*, 94(27), e1097. doi:10.1097/MD.0000000000001097
- Chen, S., Dong, H., Yang, S., & Guo, H. (2017). Cathepsins in digestive cancers. *Oncotarget*, 8(25), 41690-41700. doi:10.18632/oncotarget.16677
- Chen, Z. Y., Shie, J., & Tseng, C. (2000). Up-regulation of gut-enriched kruppel-like factor by interferon-gamma in human colon carcinoma cells. *FEBS Lett*, 477(1-2), 67-72. doi:10.1016/s0014-5793(00)01764-6
- Childers, W. K. (2015). Interactions of the renin-angiotensin system in colorectal cancer and metastasis. *Int J Colorectal Dis*, 30(6), 749-752. doi:10.1007/s00384-014-2118-1
- Clevers, H., & Batlle, E. (2006). EphB/EphrinB receptors and Wnt signaling in colorectal cancer. *Cancer Res*, 66(1), 2-5. doi:10.1158/0008-5472.CAN-05-3849
- Coelho, M., Moz, M., Correia, G., Teixeira, A., Medeiros, R., & Ribeiro, L. (2015). Antiproliferative effects of beta-blockers on human colorectal cancer cells. *Oncol Rep*, 33(5), 2513-2520. doi:10.3892/or.2015.3874
- Corvinus, F. M., Orth, C., Moriggl, R., Tsareva, S. A., Wagner, S., Pfitzner, E. B., . . . Friedrich, K. (2005). Persistent STAT3 activation in colon cancer is associated with enhanced cell proliferation and tumor growth. *Neoplasia*, 7(6), 545-555. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/16036105>
- Czechowicz-Janicka, K., Kaminska, E., Staszkievicz, J., Terelak-Borys, B., Oledzka, K., & Danyasz, A. (1998). [The influence of RS-Timolol in comparison with R-Timolol and S-Timolol on the intraocular pressure in water-loaded rabbits]. *Klin Oczna*, 100(5), 259-261. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9884519>
- Dai, X., Ge, J., Wang, X., Qian, X., Zhang, C., & Li, X. (2013). OCT4 regulates epithelial-mesenchymal transition and its knockdown inhibits colorectal cancer cell migration and invasion. *Oncol Rep*, 29(1), 155-160. doi:10.3892/or.2012.2086

- Dalerba, P., Dylla, S. J., Park, I. K., Liu, R., Wang, X., Cho, R. W., . . . Clarke, M. F. (2007). Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A*, 104(24), 10158-10163. doi:10.1073/pnas.0703478104
- Deng, S., Yang, X., Lassus, H., Liang, S., Kaur, S., Ye, Q., . . . Zhang, L. (2010). Distinct expression levels and patterns of stem cell marker, aldehyde dehydrogenase isoform 1 (ALDH1), in human epithelial cancers. *PLoS One*, 5(4), e10277. doi:10.1371/journal.pone.0010277
- Deng, Y., Zhou, J., Fang, L., Cai, Y., Ke, J., Xie, X., . . . Wang, J. (2014). ALDH1 is an independent prognostic factor for patients with stages II-III rectal cancer after receiving radiochemotherapy. *Br J Cancer*, 110(2), 430-434. doi:10.1038/bjc.2013.767
- Derocq, D., Prebois, C., Beaujouin, M., Laurent-Matha, V., Pattingre, S., Smith, G. K., & Liaudet-Coopman, E. (2012). Cathepsin D is partly endocytosed by the LRP1 receptor and inhibits LRP1-regulated intramembrane proteolysis. *Oncogene*, 31(26), 3202-3212. doi:10.1038/onc.2011.501
- Deshayes, F., & Nahmias, C. (2005). Angiotensin receptors: a new role in cancer? *Trends Endocrinol Metab*, 16(7), 293-299. doi:10.1016/j.tem.2005.07.009
- Do, H., & Dobrovic, A. (2015). Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem*, 61(1), 64-71. doi:10.1373/clinchem.2014.223040
- Doi, C., Egashira, N., Kawabata, A., Maurya, D. K., Ohta, N., Uppalapati, D., . . . Tamura, M. (2010). Angiotensin II type 2 receptor signaling significantly attenuates growth of murine pancreatic carcinoma grafts in syngeneic mice. *BMC Cancer*, 10, 67. doi:10.1186/1471-2407-10-67
- Dos Santos, W., Sobanski, T., de Carvalho, A. C., Evangelista, A. F., Matsushita, M., Berardinelli, G. N., . . . Guimaraes, D. P. (2019). Mutation profiling of cancer drivers in Brazilian colorectal cancer. *Sci Rep*, 9(1), 13687. doi:10.1038/s41598-019-49611-1
- Drost, J., & Clevers, H. (2018). Organoids in cancer research. *Nat Rev Cancer*, 18(7), 407-418. doi:10.1038/s41568-018-0007-6
- Du, L., Wang, H., He, L., Zhang, J., Ni, B., Wang, X., . . . Chen, Q. (2008). CD44 is of Functional Importance for Colorectal Cancer Stem Cells. *Clinical Cancer Research*, 14(21), 6751-6760. doi:10.1158/1078-0432.ccr-08-1034
- Du, Z., Jia, D., Liu, S., Wang, F., Li, G., Zhang, Y., . . . Hao, A. (2009). Oct4 is expressed in human gliomas and promotes colony formation in glioma cells. *Glia*, 57(7), 724-733. doi:10.1002/glia.20800
- Duan, Y., Beli, E., Li Calzi, S., Quigley, J. L., Miller, R. C., Moldovan, L., . . . Grant, M. B. (2018). Loss of Angiotensin-Converting Enzyme 2 Exacerbates Diabetic Retinopathy by Promoting Bone Marrow Dysfunction. *Stem Cells*, 36(9), 1430-1440. doi:10.1002/stem.2848
- Duchin, K. L., McKinstry, D. N., Cohen, A. I., & Migdalof, B. H. (1988). Pharmacokinetics of captopril in healthy subjects and in patients with cardiovascular diseases. *Clin Pharmacokinet*, 14(4), 241-259. doi:10.2165/00003088-198814040-00002
- Dunn, F. G., & Frohlich, E. D. (1981). Pharmacokinetics, mechanisms of action, indications, and adverse effects of timolol maleate, a nonselective beta-adrenoreceptor blocking agent. *Pharmacotherapy*, 1(3), 188-200. doi:10.1002/j.1875-9114.1981.tb02540.x
- Dylla, S. J., Beviglia, L., Park, I. K., Chartier, C., Raval, J., Ngan, L., . . . Gurney, A. L. (2008). Colorectal cancer stem cells are enriched in xenogeneic tumors following chemotherapy. *PLoS One*, 3(6), e2428. doi:10.1371/journal.pone.0002428
- Einaga, N., Yoshida, A., Noda, H., Suemitsu, M., Nakayama, Y., Sakurada, A., . . . Esumi, M. (2017). Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. *PLoS One*, 12(5), e0176280. doi:10.1371/journal.pone.0176280
- Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nat Rev Genet*, 5(6), 435-445. doi:10.1038/nrg1348
- Fasanella d'Amore, T., Bussien, J. P., Nussberger, J., Waeber, B., Turini, G. A., Brunner, H. R., . . . Francis, R. J. (1987). Effects of single doses of the converting enzyme inhibitor cilazapril in

- normal volunteers. *J Cardiovasc Pharmacol*, 9(1), 26-31. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/2434790>
- Featherston, T., Marsh, R. W., van Schaijik, B., Brasch, H. D., Tan, S. T., & Itinteang, T. (2017). Expression and Localization of Cathepsins B, D, and G in Two Cancer Stem Cell Subpopulations in Moderately Differentiated Oral Tongue Squamous Cell Carcinoma. *Front Med (Lausanne)*, 4, 100. doi:10.3389/fmed.2017.00100
- Fernandez, P. C., Frank, S. R., Wang, L., Schroeder, M., Liu, S., Greene, J., . . . Amati, B. (2003). Genomic targets of the human c-Myc protein. *Genes Dev*, 17(9), 1115-1129. doi:10.1101/gad.1067003
- Ferrario, C. M., & Schiavone, M. T. (1989). The renin angiotensin system: importance in physiology and pathology. *Cleve Clin J Med*, 56(4), 439-446. doi:10.3949/ccjm.56.4.439
- Fidler, M. M., Bray, F., Vaccarella, S., & Soerjomataram, I. (2017). Assessing global transitions in human development and colorectal cancer incidence. *Int J Cancer*. doi:10.1002/ijc.30686
- Fleming, M., Ravula, S., Tatishchev, S. F., & Wang, H. L. (2012). Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol*, 3(3), 153-173. doi:10.3978/j.issn.2078-6891.2012.030
- Fountain, J. H., & Lappin, S. L. (2019). Physiology, Renin Angiotensin System. In *StatPearls*. Treasure Island (FL).
- Fyhrquist, F., & Saijonmaa, O. (2008). Renin-angiotensin system revisited. *J Intern Med*, 264(3), 224-236. doi:10.1111/j.1365-2796.2008.01981.x
- Gage, F. H. (2000). Mammalian neural stem cells. *Science*, 287(5457), 1433-1438. doi:10.1126/science.287.5457.1433
- Gao, C., & Chen, Y. G. (2010). Dishevelled: The hub of Wnt signaling. *Cell Signal*, 22(5), 717-727. doi:10.1016/j.cellsig.2009.11.021
- Gao, S., Zhu, H., Zuo, X., & Luo, H. (2018). Cathepsin G and Its Role in Inflammation and Autoimmune Diseases. *Arch Rheumatol*, 33(4), 498-504. doi:10.5606/ArchRheumatol.2018.6595
- Garcia-Marcos, M., Jung, B. H., Ear, J., Cabrera, B., Carethers, J. M., & Ghosh, P. (2011). Expression of GIV/Girdin, a metastasis-related protein, predicts patient survival in colon cancer. *FASEB J*, 25(2), 590-599. doi:10.1096/fj.10-167304
- Garcia, M., Platet, N., Liaudet, E., Laurent, V., Derocq, D., Brouillet, J. P., & Rochefort, H. (1996). Biological and clinical significance of cathepsin D in breast cancer metastasis. *Stem Cells*, 14(6), 642-650. doi:10.1002/stem.140642
- Garrido-Gil, P., Valenzuela, R., Villar-Cheda, B., Lanciego, J. L., & Labandeira-Garcia, J. L. (2013). Expression of angiotensinogen and receptors for angiotensin and prorenin in the monkey and human substantia nigra: an intracellular renin-angiotensin system in the nigra. *Brain Struct Funct*, 218(2), 373-388. doi:10.1007/s00429-012-0402-9
- Gawlik-Rzemieniewska, N., & Bednarek, I. (2016). The role of NANOG transcriptional factor in the development of malignant phenotype of cancer cells. *Cancer Biol Ther*, 17(1), 1-10. doi:10.1080/15384047.2015.1121348
- Gazouli, M., Roubelakis, M. G., Theodoropoulos, G. E., Papailiou, J., Vaiopoulou, A., Pappa, K. I., . . . Anagnou, N. P. (2012). OCT4 spliced variant OCT4B1 is expressed in human colorectal cancer. *Mol Carcinog*, 51(2), 165-173. doi:10.1002/mc.20773
- George, A. J., Thomas, W. G., & Hannan, R. D. (2010). The renin-angiotensin system and cancer: old dog, new tricks. *Nat Rev Cancer*, 10(11), 745-759. doi:10.1038/nrc2945
- Ghaleb, A. M., Elkarim, E. A., Bialkowska, A. B., & Yang, V. W. (2016). KLF4 Suppresses Tumor Formation in Genetic and Pharmacological Mouse Models of Colonic Tumorigenesis. *Mol Cancer Res*, 14(4), 385-396. doi:10.1158/1541-7786.MCR-15-0410
- Ghoncheh, M., Mohammadian, M., Mohammadian-Hafshejani, A., & Salehiniya, H. The Incidence and Mortality of Colorectal Cancer and Its Relationship With the Human Development Index in Asia. *Annals of Global Health*, 82(5), 726-737. doi:10.1016/j.aogh.2016.10.004

- Gilmore, D. A., Gal, J., Gerber, J. G., & Nies, A. S. (1992). Age and gender influence the stereoselective pharmacokinetics of propranolol. *J Pharmacol Exp Ther*, 261(3), 1181-1186. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/1602383>
- Goldberg, R. M., Sargent, D. J., Morton, R. F., Fuchs, C. S., Ramanathan, R. K., Williamson, S. K., . . . Alberts, S. R. (2004). A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J Clin Oncol*, 22(1), 23-30. doi:10.1200/JCO.2004.09.046
- Gondi, C. S., & Rao, J. S. (2013). Cathepsin B as a cancer target. *Expert Opin Ther Targets*, 17(3), 281-291. doi:10.1517/14728222.2013.740461
- Greening, D. W., Xu, R., Gopal, S. K., Rai, A., & Simpson, R. J. (2017). Proteomic insights into extracellular vesicle biology - defining exosomes and shed microvesicles. *Expert Rev Proteomics*, 14(1), 69-95. doi:10.1080/14789450.2017.1260450
- Grivennikov, S., Karin, E., Terzic, J., Mucida, D., Yu, G. Y., Vallabhapurapu, S., . . . Karin, M. (2009). IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. *Cancer Cell*, 15(2), 103-113. doi:10.1016/j.ccr.2009.01.001
- Guinney, J., Dienstmann, R., Wang, X., de Reynies, A., Schlicker, A., Soneson, C., . . . Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nat Med*, 21(11), 1350-1356. doi:10.1038/nm.3967
- Guo, Y., Liu, S., Wang, P., Zhao, S., Wang, F., Bing, L., . . . Hao, A. (2011). Expression profile of embryonic stem cell-associated genes Oct4, Sox2 and Nanog in human gliomas. *Histopathology*, 59(4), 763-775. doi:10.1111/j.1365-2559.2011.03993.x
- Gupta, V., Bhandari, D., Leyme, A., Aznar, N., Midde, K. K., Lo, I. C., . . . Ghosh, P. (2016). GIV/Girdin activates Galphai and inhibits Galphas via the same motif. *Proc Natl Acad Sci U S A*, 113(39), E5721-5730. doi:10.1073/pnas.1609502113
- Hadjimichael, C., Chanoumidou, K., Papadopoulou, N., Arampatzi, P., Papamatheakis, J., & Kretsovali, A. (2015). Common stemness regulators of embryonic and cancer stem cells. *World J Stem Cells*, 7(9), 1150-1184. doi:10.4252/wjsc.v7.i9.1150
- Haggar, F. A., & Boushey, R. P. (2009). Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors. *Clinics in Colon and Rectal Surgery*, 22(4), 191-197. doi:10.1055/s-0029-1242458
- Halim, S., Markert, E. K., & Vazquez, A. (2018). Analysis of cell proliferation and tissue remodelling uncovers a KLF4 activity score associated with poor prognosis in colorectal cancer. *Br J Cancer*, 119(7), 855-863. doi:10.1038/s41416-018-0253-0
- Hampel, H., & de la Chapelle, A. (2011). The search for unaffected individuals with Lynch syndrome: do the ends justify the means? *Cancer Prev Res (Phila)*, 4(1), 1-5. doi:10.1158/1940-6207.CAPR-10-0345
- Han, C., Sun, B., Zhao, X., Zhang, Y., Gu, Q., Liu, F., . . . Wu, L. (2017). Phosphorylation of STAT3 Promotes Vasculogenic Mimicry by Inducing Epithelial-to-Mesenchymal Transition in Colorectal Cancer. *Technol Cancer Res Treat*, 16(6), 1209-1219. doi:10.1177/1533034617742312
- Haraldsdottir, S., Rafnar, T., Frankel, W. L., Einarsdottir, S., Sigurdsson, A., Hampel, H., . . . Stefansson, K. (2017). Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat Commun*, 8, 14755. doi:10.1038/ncomms14755
- Hashimoto, I., Nagata, T., Sekine, S., Moriyama, M., Shibuya, K., Hojo, S., . . . Tsukada, K. (2017). Prognostic significance of KLF4 expression in gastric cancer. *Oncol Lett*, 13(2), 819-826. doi:10.3892/ol.2016.5499
- Haydon, A. M., & Jass, J. R. (2002). Emerging pathways in colorectal-cancer development. *Lancet Oncol*, 3(2), 83-88. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11905459>

- Hein, L., Meinel, L., Pratt, R. E., Dzau, V. J., & Kobilka, B. K. (1997). Intracellular trafficking of angiotensin II and its AT1 and AT2 receptors: evidence for selective sorting of receptor and ligand. *Mol Endocrinol*, 11(9), 1266-1277. doi:10.1210/mend.11.9.9975
- Hemminki, A., Peltomaki, P., Mecklin, J. P., Jarvinen, H., Salovaara, R., Nystrom-Lahti, M., . . . Aaltonen, L. A. (1994). Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nat Genet*, 8(4), 405-410. doi:10.1038/ng1294-405
- Hou, P. C., Li, Y. H., Lin, S. C., Lin, S. C., Lee, J. C., Lin, B. W., . . . Tsai, S. J. (2017). Hypoxia-Induced Downregulation of DUSP-2 Phosphatase Drives Colon Cancer Stemness. *Cancer Res*, 77(16), 4305-4316. doi:10.1158/0008-5472.CAN-16-2990
- Hsu, H. T., Wu, P. R., Chen, C. J., Hsu, L. S., Yeh, C. M., Hsing, M. T., . . . Yeh, K. T. (2014). High cytoplasmic expression of Kruppel-like factor 4 is an independent prognostic factor of better survival in hepatocellular carcinoma. *Int J Mol Sci*, 15(6), 9894-9906. doi:10.3390/ijms15069894
- Hu, J., Li, J., Yue, X., Wang, J., Liu, J., Sun, L., & Kong, D. (2017). Expression of the cancer stem cell markers ABCG2 and OCT-4 in right-sided colon cancer predicts recurrence and poor outcomes. *Oncotarget*, 8(17), 28463-28470. doi:10.18632/oncotarget.15307
- Hu, R., Zuo, Y., Zuo, L., Liu, C., Zhang, S., Wu, Q., . . . Wang, Y. (2011). KLF4 Expression Correlates with the Degree of Differentiation in Colorectal Cancer. *Gut Liver*, 5(2), 154-159. doi:10.5009/gnl.2011.5.2.154
- Huang, E. H., Hynes, M. J., Zhang, T., Ginestier, C., Dontu, G., Appelman, H., . . . Boman, B. M. (2009). Aldehyde dehydrogenase 1 is a marker for normal and malignant human colonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. *Cancer Res*, 69(8), 3382-3389. doi:10.1158/0008-5472.CAN-08-4418
- Huang, L., Yang, Y., Yang, F., Liu, S., Zhu, Z., Lei, Z., & Guo, J. (2018). Functions of EpCAM in physiological processes and diseases (Review). *Int J Mol Med*, 42(4), 1771-1785. doi:10.3892/ijmm.2018.3764
- Humphries, A., & Wright, N. A. (2008). Colonic crypt organization and tumorigenesis. *Nat Rev Cancer*, 8(6), 415-424. doi:10.1038/nrc2392
- Hunyady, L., Catt, K. J., Clark, A. J., & Gaborik, Z. (2000). Mechanisms and functions of AT(1) angiotensin receptor internalization. *Regul Pept*, 91(1-3), 29-44. doi:10.1016/s0167-0115(00)00137-3
- Huth, C., Kloor, M., Voigt, A. Y., Bozukova, G., Evers, C., Gaspar, H., . . . Blaker, H. (2012). The molecular basis of EPCAM expression loss in Lynch syndrome-associated tumors. *Mod Pathol*, 25(6), 911-916. doi:10.1038/modpathol.2012.30
- Ibrahim, E. E., Babaei-Jadidi, R., Saadeddin, A., Spencer-Dene, B., Hossaini, S., Abuzinadah, M., . . . Nateri, A. S. (2012). Embryonic NANOG activity defines colorectal cancer stem cells and modulates through AP1- and TCF-dependent mechanisms. *Stem Cells*, 30(10), 2076-2087. doi:10.1002/stem.1182
- Ikushima, H., Todo, T., Ino, Y., Takahashi, M., Miyazawa, K., & Miyazono, K. (2009). Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell Stem Cell*, 5(5), 504-514. doi:10.1016/j.stem.2009.08.018
- Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., . . . Bird, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet*, 6(9), e1001134. doi:10.1371/journal.pgen.1001134
- International Stem Cell, I., Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P. W., . . . Zhang, W. (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat Biotechnol*, 25(7), 803-816. doi:10.1038/nbt1318
- Jaggupilli, A., & Elkord, E. (2012). Significance of CD44 and CD24 as cancer stem cell markers: an enduring ambiguity. *Clin Dev Immunol*, 2012, 708036. doi:10.1155/2012/708036
- Jang, E., & Chung, D. C. (2010). Hereditary colon cancer: lynch syndrome. *Gut Liver*, 4(2), 151-160. doi:10.5009/gnl.2010.4.2.151

- Januszyk, M., Rennert, R. C., Sorkin, M., Maan, Z. N., Wong, L. K., Whittam, A. J., . . . Gurtner, G. C. (2015). Evaluating the Effect of Cell Culture on Gene Expression in Primary Tissue Samples Using Microfluidic-Based Single Cell Transcriptional Analysis. *Microarrays (Basel)*, 4(4), 540-550. doi:10.3390/microarrays4040540
- Johnson, S. M., Gulhati, P., Rampy, B. A., Han, Y., Rychahou, P. G., Doan, H. Q., . . . Evers, B. M. (2010). Novel expression patterns of PI3K/Akt/mTOR signaling pathway components in colorectal cancer. *J Am Coll Surg*, 210(5), 767-776, 776-768. doi:10.1016/j.jamcollsurg.2009.12.008
- Jun, B. Y., Kim, S. W., Jung, C. K., Cho, Y. K., Lee, I. S., Choi, M. G., . . . Oh, S. T. (2013). Expression of girdin in human colorectal cancer and its association with tumor progression. *Dis Colon Rectum*, 56(1), 51-57. doi:10.1097/DCR.0b013e31826b9b7e
- Kanber, Y., Demirbag, N. R., Sam, A. D., & Aydin, N. (2002). Cathepsin D expression in colorectal adenocarcinomas and adenomas. *Int J Biol Markers*, 17(3), 165-168. doi:10.1177/172460080201700304
- Katsuno, Y., Lamouille, S., & Derynck, R. (2013). TGF-beta signaling and epithelial-mesenchymal transition in cancer progression. *Curr Opin Oncol*, 25(1), 76-84. doi:10.1097/CCO.0b013e32835b6371
- Kemper, K., Prasetyanti, P. R., De Lau, W., Rodermond, H., Clevers, H., & Medema, J. P. (2012). Monoclonal antibodies against Lgr5 identify human colorectal cancer stem cells. *Stem Cells*, 30(11), 2378-2386. doi:10.1002/stem.1233
- Keppel Hesselink, J. M., & Schatman, M. E. (2017). EMA401: an old antagonist of the AT2R for a new indication in neuropathic pain. *J Pain Res*, 10, 439-443. doi:10.2147/JPR.S128520
- Khalek, F. J. A., Gallicano, G. I., & Mishra, L. (2010). Colon Cancer Stem Cells. *Gastrointestinal Cancer Research : GCR*(Suppl 1), S16-S23. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3047031/>
- Khan, A., Krishna, M., Baker, S. P., & Banner, B. F. (1998). Cathepsin B and tumor-associated laminin expression in the progression of colorectal adenoma to carcinoma. *Mod Pathol*, 11(8), 704-708. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9720496>
- Kilmister, E. J., Patel, J., van Schaijik, B., Bockett, N., Brasch, H. D., Paterson, E., . . . Tan, S. T. (2020). Cancer Stem Cell Subpopulations Are Present Within Metastatic Head and Neck Cutaneous Squamous Cell Carcinoma. *Front Oncol*, 10, 1091. doi:10.3389/fonc.2020.01091
- Kim, J., Koo, B. K., & Knoblich, J. A. (2020). Human organoids: model systems for human biology and medicine. *Nat Rev Mol Cell Biol*, 21(10), 571-584. doi:10.1038/s41580-020-0259-3
- Kim, S. M., Briggs, J. P., & Schnermann, J. (2012). Convergence of major physiological stimuli for renin release on the Gs-alpha/cyclic adenosine monophosphate signaling pathway. *Clin Exp Nephrol*, 16(1), 17-24. doi:10.1007/s10157-011-0494-1
- Kim, S. T., Park, K. H., Oh, S. C., Seo, J. H., Kim, J. S., Shin, S. W., & Kim, Y. H. (2012). How does inhibition of the renin-angiotensin system affect the prognosis of advanced gastric cancer patients receiving platinum-based chemotherapy? *Oncology*, 83(6), 354-360. doi:10.1159/000337979
- Koh, S. P., Wickremesekera, A. C., Brasch, H. D., Marsh, R., Tan, S. T., & Itinteang, T. (2017). Expression of Cathepsins B, D, and G in Isocitrate Dehydrogenase-Wildtype Glioblastoma. *Front Surg*, 4, 28. doi:10.3389/fsurg.2017.00028
- Kosugi, M., Miyajima, A., Kikuchi, E., Kosaka, T., Horiguchi, Y., Murai, M., & Oya, M. (2009). Angiotensin II type 1 receptor antagonist enhances cis-dichlorodiammineplatinum-induced cytotoxicity in mouse xenograft model of bladder cancer. *Urology*, 73(3), 655-660. doi:10.1016/j.urology.2008.10.031
- Kramer, A. S., Harvey, A. R., Plant, G. W., & Hodgetts, S. I. (2013). Systematic review of induced pluripotent stem cell technology as a potential clinical therapy for spinal cord injury. *Cell Transplant*, 22(4), 571-617. doi:10.3727/096368912X655208

- Kreso, A., & Dick, J. E. (2014). Evolution of the cancer stem cell model. *Cell Stem Cell*, 14(3), 275-291. doi:10.1016/j.stem.2014.02.006
- Kuan, H., Liang, K. H., Wang, Y. P., Kuo, T. W., Meir, Y. J., Wu, S. C., . . . Wu, H. C. (2017). EpEX/EpCAM and Oct4 or Klf4 alone are sufficient to generate induced pluripotent stem cells through STAT3 and HIF2alpha. *Sci Rep*, 7, 41852. doi:10.1038/srep41852
- Kuang, Y. L., Munoz, A., Nalula, G., Santostefano, K. E., Sanghez, V., Sanchez, G., . . . Medina, M. W. (2019). Evaluation of commonly used ectoderm markers in iPSC trilineage differentiation. *Stem Cell Res*, 37, 101434. doi:10.1016/j.scr.2019.101434
- Kudo, T., Kigoshi, H., Hagiwara, T., Takino, T., Yamazaki, M., & Yui, S. (2009). Cathepsin G, a neutrophil protease, induces compact cell-cell adhesion in MCF-7 human breast cancer cells. *Mediators Inflamm*, 2009, 850940. doi:10.1155/2009/850940
- Kuester, D., Lippert, H., Roessner, A., & Krueger, S. (2008). The cathepsin family and their role in colorectal cancer. *Pathol Res Pract*, 204(7), 491-500. doi:10.1016/j.prp.2008.04.010
- Larance, M., & Lamond, A. I. (2015). Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol*, 16(5), 269-280. doi:10.1038/nrm3970
- Lassmann, S., Schuster, I., Walch, A., Gobel, H., Jutting, U., Makowiec, F., . . . Werner, M. (2007). STAT3 mRNA and protein expression in colorectal cancer: effects on STAT3-inducible targets linked to cell survival and proliferation. *J Clin Pathol*, 60(2), 173-179. doi:10.1136/jcp.2005.035113
- Le Magnen, C., Bubendorf, L., Ruiz, C., Zlobec, I., Bachmann, A., Heberer, M., . . . Mengus, C. (2013). Klf4 transcription factor is expressed in the cytoplasm of prostate cancer cells. *Eur J Cancer*, 49(4), 955-963. doi:10.1016/j.ejca.2012.09.023
- Le Rolle, A. F., Chiu, T. K., Zeng, Z., Shia, J., Weiser, M. R., Paty, P. B., & Chiu, V. K. (2016). Oncogenic KRAS activates an embryonic stem cell-like program in human colon cancer initiation. *Oncotarget*, 7(3), 2159-2174. doi:10.18632/oncotarget.6818
- Lee, K. S., Kwak, Y., Nam, K. H., Kim, D. W., Kang, S. B., Choe, G., . . . Lee, H. S. (2015). c-MYC Copy-Number Gain Is an Independent Prognostic Factor in Patients with Colorectal Cancer. *PLoS One*, 10(10), e0139727. doi:10.1371/journal.pone.0139727
- Leng, Z., Tao, K., Xia, Q., Tan, J., Yue, Z., Chen, J., . . . Zheng, H. (2013). Kruppel-like factor 4 acts as an oncogene in colon cancer stem cell-enriched spheroid cells. *PLoS One*, 8(2), e56082. doi:10.1371/journal.pone.0056082
- Li, H., Qi, Y., Li, C., Braseth, L. N., Gao, Y., Shabashvili, A. E., . . . Sumners, C. (2009). Angiotensin type 2 receptor-mediated apoptosis of human prostate cancer cells. *Mol Cancer Ther*, 8(12), 3255-3265. doi:10.1158/1535-7163.MCT-09-0237
- Li, I. C., Chan, C. T., Lu, Y. F., Wu, Y. T., Chen, Y. C., Li, G. B., . . . Hwang, S. P. (2011). Zebrafish kruppel-like factor 4a represses intestinal cell proliferation and promotes differentiation of intestinal cell lineages. *PLoS One*, 6(6), e20974. doi:10.1371/journal.pone.0020974
- Li, Z. (2013). CD133: a stem cell biomarker and beyond. *Exp Hematol Oncol*, 2(1), 17. doi:10.1186/2162-3619-2-17
- Liang, K. H., Tso, H. C., Hung, S. H., Kuan, H., Lai, J. K., Ke, F. Y., . . . Wu, H. C. (2018). Extracellular domain of EpCAM enhances tumor progression through EGFR signaling in colon cancer cells. *Cancer Lett*, 433, 165-175. doi:10.1016/j.canlet.2018.06.040
- Liaudet-Coopman, E., Beaujouin, M., Derocq, D., Garcia, M., Glondou-Lassis, M., Laurent-Matha, V., . . . Vignon, F. (2006). Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis. *Cancer Lett*, 237(2), 167-179. doi:10.1016/j.canlet.2005.06.007
- Lin, L., Liu, A., Peng, Z., Lin, H. J., Li, P. K., Li, C., & Lin, J. (2011). STAT3 is necessary for proliferation and survival in colon cancer-initiating cells. *Cancer Res*, 71(23), 7226-7237. doi:10.1158/0008-5472.CAN-10-4660
- Lin, Q., Lai, R., Chirieac, L. R., Li, C., Thomazy, V. A., Grammatikakis, I., . . . Amin, H. M. (2005). Constitutive activation of JAK3/STAT3 in colon carcinoma tumors and cell lines: inhibition of

- JAK3/STAT3 signaling induces apoptosis and cell cycle arrest of colon carcinoma cells. *Am J Pathol*, 167(4), 969-980. doi:10.1016/S0002-9440(10)61187-X
- Liu, C. W., Li, C. H., Peng, Y. J., Cheng, Y. W., Chen, H. W., Liao, P. L., . . . Yeng, M. H. (2014). Snail regulates Nanog status during the epithelial-mesenchymal transition via the Smad1/Akt/GSK3beta signaling pathway in non-small-cell lung cancer. *Oncotarget*, 5(11), 3880-3894. doi:10.18632/oncotarget.2006
- Liu, W., Ahmad, S. A., Jung, Y. D., Reinmuth, N., Fan, F., Bucana, C. D., & Ellis, L. M. (2002). Coexpression of ephrin-Bs and their receptors in colon carcinoma. *Cancer*, 94(4), 934-939. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11920461>
- Liu, Y., Zheng, B., Zhang, X. H., Nie, C. J., Li, Y. H., & Wen, J. K. (2013). Localization and function of KLF4 in cytoplasm of vascular smooth muscle cell. *Biochem Biophys Res Commun*, 436(2), 162-168. doi:10.1016/j.bbrc.2013.05.067
- Liu, Z., Yang, H., Luo, W., Jiang, Q., Mai, C., Chen, Y., . . . Fang, W. (2013). Loss of cytoplasmic KLF4 expression is correlated with the progression and poor prognosis of nasopharyngeal carcinoma. *Histopathology*, 63(3), 362-370. doi:10.1111/his.12176
- Loboda, A., Nebozhyn, M. V., Watters, J. W., Buser, C. A., Shaw, P. M., Huang, P. S., . . . Yeatman, T. J. (2011). EMT is the dominant program in human colon cancer. *BMC Med Genomics*, 4, 9. doi:10.1186/1755-8794-4-9
- Loree, J. M., Pereira, A. A. L., Lam, M., Willauer, A. N., Raghav, K., Dasari, A., . . . Kopetz, S. (2018). Classifying Colorectal Cancer by Tumor Location Rather than Sidedness Highlights a Continuum in Mutation Profiles and Consensus Molecular Subtypes. *Clin Cancer Res*, 24(5), 1062-1072. doi:10.1158/1078-0432.CCR-17-2484
- Lu, T. Y., Lu, R. M., Liao, M. Y., Yu, J., Chung, C. H., Kao, C. F., & Wu, H. C. (2010). Epithelial cell adhesion molecule regulation is associated with the maintenance of the undifferentiated phenotype of human embryonic stem cells. *J Biol Chem*, 285(12), 8719-8732. doi:10.1074/jbc.M109.077081
- Luo, W., Li, S., Peng, B., Ye, Y., Deng, X., & Yao, K. (2013). Embryonic stem cells markers SOX2, OCT4 and Nanog expression and their correlations with epithelial-mesenchymal transition in nasopharyngeal carcinoma. *PLoS One*, 8(2), e56324. doi:10.1371/journal.pone.0056324
- Luo, Y., Ohmori, H., Shimomoto, T., Fujii, K., Sasahira, T., Chihara, Y., & Kuniyasu, H. (2011). Anti-angiotensin and hypoglycemic treatments suppress liver metastasis of colon cancer cells. *Pathobiology*, 78(5), 285-290. doi:10.1159/000330169
- Ma, I., & Allan, A. L. (2011). The role of human aldehyde dehydrogenase in normal and cancer stem cells. *Stem Cell Rev*, 7(2), 292-306. doi:10.1007/s12015-010-9208-4
- Ma, S., Chan, K. W., Lee, T. K., Tang, K. H., Wo, J. Y., Zheng, B. J., & Guan, X. Y. (2008). Aldehyde dehydrogenase discriminates the CD133 liver cancer stem cell populations. *Mol Cancer Res*, 6(7), 1146-1153. doi:10.1158/1541-7786.MCR-08-0035
- Ma, Y., Wu, L., Liu, X., Xu, Y., Shi, W., Liang, Y., . . . Zhang, J. (2017). KLF4 inhibits colorectal cancer cell proliferation dependent on NDRG2 signaling. *Oncol Rep*, 38(2), 975-984. doi:10.3892/or.2017.5736
- MacArthur, B. D., & Lemischka, I. R. (2013). Statistical mechanics of pluripotency. *Cell*, 154(3), 484-489. doi:10.1016/j.cell.2013.07.024
- Maenpaa, J., & Pelkonen, O. (2016). Cardiac safety of ophthalmic timolol. *Expert Opin Drug Saf*, 15(11), 1549-1561. doi:10.1080/14740338.2016.1225718
- Maetzel, D., Denzel, S., Mack, B., Canis, M., Went, P., Benk, M., . . . Gires, O. (2009). Nuclear signalling by tumour-associated antigen EpCAM. *Nat Cell Biol*, 11(2), 162-171. doi:10.1038/ncb1824
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., . . . Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133(4), 704-715. doi:10.1016/j.cell.2008.03.027

- Mantyla, R., Mannisto, P., Nykanen, S., Koponen, A., & Lamminsivu, U. (1983). Pharmacokinetic interactions of timolol with vasodilating drugs, food and phenobarbitone in healthy human volunteers. *Eur J Clin Pharmacol*, 24(2), 227-230. doi:10.1007/BF00613822
- Markowitz, S. D., & Bertagnolli, M. M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med*, 361(25), 2449-2460. doi:10.1056/NEJMra0804588
- Martini, G., Dienstmann, R., Ros, J., Baraibar, I., Cuadra-Urteaga, J. L., Salva, F., . . . Elez, E. (2020). Molecular subtypes and the evolution of treatment management in metastatic colorectal cancer. *Ther Adv Med Oncol*, 12, 1758835920936089. doi:10.1177/1758835920936089
- Martini, M., Basso, M., Cocomazzi, A., Cenci, T., Strippoli, A., Fiorentino, V., . . . Larocca, L. M. (2016). c-Myc expression as a key-marker in the colorectal cancer resistance to EGFR inhibitors. *Journal of Clinical Oncology*, 34(15_suppl), e15034-e15034. doi:10.1200/JCO.2016.34.15_suppl.e15034
- Martowicz, A., Seeber, A., & Untergasser, G. (2016). The role of EpCAM in physiology and pathology of the epithelium. *Histol Histopathol*, 31(4), 349-355. doi:10.14670/HH-11-678
- Mason, S. D., & Joyce, J. A. (2011). Proteolytic networks in cancer. *Trends Cell Biol*, 21(4), 228-237. doi:10.1016/j.tcb.2010.12.002
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., . . . Niwa, H. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol*, 9(6), 625-635. doi:10.1038/ncb1589
- Masur, K., Niggemann, B., Zanker, K. S., & Entschladen, F. (2001). Norepinephrine-induced migration of SW 480 colon carcinoma cells is inhibited by beta-blockers. *Cancer Res*, 61(7), 2866-2869. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11306460>
- May, C. L., & Kaestner, K. H. (2010). Gut endocrine cell development. *Mol Cell Endocrinol*, 323(1), 70-75. doi:10.1016/j.mce.2009.12.009
- McMahon, P., Oza, V., & Frieden, I. J. (2012). Topical timolol for infantile hemangiomas: putting a note of caution in "cautiously optimistic". *Pediatr Dermatol*, 29(1), 127-130. doi:10.1111/j.1525-1470.2011.01685.x
- Mehrian Shai, R., Reichardt, J. K., Ya-Hsuan, H., Kremen, T. J., Liau, L. M., Cloughesy, T. F., . . . Nelson, S. F. (2005). Robustness of gene expression profiling in glioma specimen samplings and derived cell lines. *Brain Res Mol Brain Res*, 136(1-2), 99-103. doi:10.1016/j.molbrainres.2005.01.017
- Mehrotra, S., Wickremesekera, S. K., Brasch, H. D., Van Schaijik, B., Marsh, R. W., Tan, S. T., & Itinteang, T. (2018). Expression and Localization of Cathepsins B, D and G in Cancer Stem Cells in Liver Metastasis From Colon Adenocarcinoma. *Front Surg*, 5, 40. doi:10.3389/fsurg.2018.00040
- Mehta, P. K., & Griendling, K. K. (2007). Angiotensin II cell signaling: physiological and pathological effects in the cardiovascular system. *Am J Physiol Cell Physiol*, 292(1), C82-97. doi:10.1152/ajpcell.00287.2006
- Merlos-Suarez, A., Barriga, F. M., Jung, P., Iglesias, M., Cespedes, M. V., Rossell, D., . . . Batlle, E. (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell*, 8(5), 511-524. doi:10.1016/j.stem.2011.02.020
- Miraglia, S., Godfrey, W., Yin, A. H., Atkins, K., Warnke, R., Holden, J. T., . . . Buck, D. W. (1997). A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning. *Blood*, 90(12), 5013-5021. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9389721>
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., . . . Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113(5), 631-642. doi:10.1016/s0092-8674(03)00393-3
- Miyazono, K. (2009). Transforming growth factor-beta signaling in epithelial-mesenchymal transition and progression of cancer. *Proc Jpn Acad Ser B Phys Biol Sci*, 85(8), 314-323. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/19838011>

- Morath, I., Hartmann, T. N., & Orian-Rousseau, V. (2016). CD44: More than a mere stem cell marker. *Int J Biochem Cell Biol*, 81(Pt A), 166-173. doi:10.1016/j.biocel.2016.09.009
- Morris, B. J. (1978). Activation of human inactive ("pro-") renin by cathepsin D and pepsin. *J Clin Endocrinol Metab*, 46(1), 153-157. doi:10.1210/jcem-46-1-153
- Muller, M., Hermann, P. C., Liebau, S., Weidgang, C., Seufferlein, T., Kleger, A., & Perkhofer, L. (2016). The role of pluripotency factors to drive stemness in gastrointestinal cancer. *Stem Cell Res*, 16(2), 349-357. doi:10.1016/j.scr.2016.02.005
- Munro, M. J., Wickremesekera, A. C., Davis, P. F., Marsh, R., Tan, S. T., & Itinteang, T. (2017). Renin-angiotensin system and cancer: a review. *Integr Cancer Sci Ther*, 4(2), 1-6.
- Munro, M. J., Wickremesekera, S. K., Peng, L., Marsh, R. W., Itinteang, T., & Tan, S. T. (2019). Cancer stem cell subpopulations in primary colon adenocarcinoma. *PLoS One*, 14(9), e0221963. doi:10.1371/journal.pone.0221963
- Munro, M. J., Wickremesekera, S. K., Peng, L., Tan, S. T., & Itinteang, T. (2017). Cancer stem cells in colorectal cancer: a review. *J Clin Pathol*. doi:10.1136/jclinpath-2017-204739
- Munz, M., Baeuerle, P. A., & Gires, O. (2009). The emerging role of EpCAM in cancer and stem cell signaling. *Cancer Res*, 69(14), 5627-5629. doi:10.1158/0008-5472.CAN-09-0654
- Nakai, Y., Isayama, H., Ijichi, H., Sasaki, T., Sasahira, N., Hirano, K., . . . Koike, K. (2010). Inhibition of renin-angiotensin system affects prognosis of advanced pancreatic cancer receiving gemcitabine. *Br J Cancer*, 103(11), 1644-1648. doi:10.1038/sj.bjc.6605955
- Nakatani, Y., Yamazaki, M., Chazin, W. J., & Yui, S. (2005). Regulation of S100A8/A9 (calprotectin) binding to tumor cells by zinc ion and its implication for apoptosis-inducing activity. *Mediators Inflamm*, 2005(5), 280-292. doi:10.1155/MI.2005.280
- Neo, J. H., Ager, E. I., Angus, P. W., Zhu, J., Herath, C. B., & Christophi, C. (2010). Changes in the renin angiotensin system during the development of colorectal cancer liver metastases. *BMC Cancer*, 10, 134. doi:10.1186/1471-2407-10-134
- Neumann, E., Riepl, B., Knedla, A., Lefevre, S., Tarner, I. H., Grifka, J., . . . Muller-Ladner, U. (2010). Cell culture and passaging alters gene expression pattern and proliferation rate in rheumatoid arthritis synovial fibroblasts. *Arthritis Res Ther*, 12(3), R83. doi:10.1186/ar3010
- Neumann, J., Bahr, F., Horst, D., Kriegl, L., Engel, J., Luque, R. M., . . . Jung, A. (2011). SOX2 expression correlates with lymph-node metastases and distant spread in right-sided colon cancer. *BMC Cancer*, 11, 518. doi:10.1186/1471-2407-11-518
- Neves, F. A., Duncan, K. G., & Baxter, J. D. (1996). Cathepsin B is a prorenin processing enzyme. *Hypertension*, 27(3 Pt 2), 514-517. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/8613195>
- Nguyen, G. (2011). Renin, (pro)renin and receptor: an update. *Clin Sci (Lond)*, 120(5), 169-178. doi:10.1042/CS20100432
- Nguyen, G., & Muller, D. N. (2010). The biology of the (pro)renin receptor. *J Am Soc Nephrol*, 21(1), 18-23. doi:10.1681/ASN.2009030300
- Nguyen, L., Ager, E. I., Neo, J., & Christophi, C. (2016). Regulation of colorectal cancer cell epithelial to mesenchymal transition by the renin angiotensin system. *J Gastroenterol Hepatol*, 31(10), 1773-1782. doi:10.1111/jgh.13307
- Nichols, J., Zevnik, B., Anastasiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., . . . Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95(3), 379-391. doi:10.1016/s0092-8674(00)81769-9
- Niwa, H., Ogawa, K., Shimosato, D., & Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature*, 460(7251), 118-122. doi:10.1038/nature08113
- Noguchi, R., Yoshiji, H., Ikenaka, Y., Namisaki, T., Kitade, M., Kaji, K., . . . Fukui, H. (2009). Synergistic inhibitory effect of gemcitabine and angiotensin type-1 receptor blocker, losartan, on murine pancreatic tumor growth via anti-angiogenic activities. *Oncol Rep*, 22(2), 355-360. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/19578777>

- Nusse, R. (2005). Wnt signaling in disease and in development. *Cell Res*, 15(1), 28-32. doi:10.1038/sj.cr.7290260
- O'Brien, C. A., Pollett, A., Gallinger, S., & Dick, J. E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature*, 445(7123), 106-110. doi:10.1038/nature05372
- Ohnuma, Y., Toda, M., Fujita, M., Hosono, K., Suzuki, T., Ogawa, Y., . . . Majima, M. (2009). Blockade of an angiotensin type I receptor enhances effects of radiation on tumor growth and tumor-associated angiogenesis by reducing vascular endothelial growth factor expression. *Biomed Pharmacother*, 63(2), 136-145. doi:10.1016/j.biopha.2007.11.005
- Overman, J., Fontaine, F., Wylie-Sears, J., Moustaqil, M., Huang, L., Meurer, M., . . . Francois, M. (2019). R-propranolol is a small molecule inhibitor of the SOX18 transcription factor in a rare vascular syndrome and hemangioma. *Elife*, 8. doi:10.7554/eLife.43026
- Pan, G., & Thomson, J. A. (2007). Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res*, 17(1), 42-49. doi:10.1038/sj.cr.7310125
- Pan, M., Schinke, H., Luxemburger, E., Kranz, G., Shakhmourad, J., Libl, D., . . . Gires, O. (2018). EpCAM ectodomain EpEX is a ligand of EGFR that counteracts EGF-mediated epithelial-mesenchymal transition through modulation of phospho-ERK1/2 in head and neck cancers. *PLoS Biol*, 16(9), e2006624. doi:10.1371/journal.pbio.2006624
- Pantziarka, P., Bouche, G., Sukhatme, V., Meheus, L., Rooman, I., & Sukhatme, V. P. (2016). Repurposing Drugs in Oncology (ReDO)-Propranolol as an anti-cancer agent. *Ecancermedicalscience*, 10, 680. doi:10.3332/ecancer.2016.680
- Papadatos-Pastos, D., Rabbie, R., Ross, P., & Sarker, D. (2015). The role of the PI3K pathway in colorectal cancer. *Crit Rev Oncol Hematol*, 94(1), 18-30. doi:10.1016/j.critrevonc.2014.12.006
- Parving, H. H., Brenner, B. M., McMurray, J. J., de Zeeuw, D., Haffner, S. M., Solomon, S. D., . . . Investigators, A. (2012). Cardiorenal end points in a trial of aliskiren for type 2 diabetes. *N Engl J Med*, 367(23), 2204-2213. doi:10.1056/NEJMoa1208799
- Pasquier, E., Andre, N., Street, J., Chougule, A., Rekhi, B., Ghosh, J., . . . Banavali, S. D. (2016). Effective Management of Advanced Angiosarcoma by the Synergistic Combination of Propranolol and Vinblastine-based Metronomic Chemotherapy: A Bench to Bedside Study. *EBioMedicine*, 6, 87-95. doi:10.1016/j.ebiom.2016.02.026
- Pasquier, E., Ciccolini, J., Carre, M., Giacometti, S., Fanciullino, R., Pouchy, C., . . . Andre, N. (2011). Propranolol potentiates the anti-angiogenic effects and anti-tumor efficacy of chemotherapy agents: implication in breast cancer treatment. *Oncotarget*, 2(10), 797-809. doi:10.18632/oncotarget.343
- Patriarca, C., Macchi, R. M., Marschner, A. K., & Mellstedt, H. (2012). Epithelial cell adhesion molecule expression (CD326) in cancer: a short review. *Cancer Treat Rev*, 38(1), 68-75. doi:10.1016/j.ctrv.2011.04.002
- Pinter, M., & Jain, R. K. (2017). Targeting the renin-angiotensin system to improve cancer treatment: Implications for immunotherapy. *Sci Transl Med*, 9(410). doi:10.1126/scitranslmed.aan5616
- Pinter, M., Kwanten, W. J., & Jain, R. K. (2018). Renin-Angiotensin System Inhibitors to Mitigate Cancer Treatment-Related Adverse Events. *Clin Cancer Res*, 24(16), 3803-3812. doi:10.1158/1078-0432.CCR-18-0236
- Pizon, M., Schott, D., Pachmann, U., & Pachmann, K. (2016). The number of tumorspheres cultured from peripheral blood is a predictor for presence of metastasis in patients with breast cancer. *Oncotarget*, 7(30), 48143-48154. doi:10.18632/oncotarget.10174
- Popat, S., Hubner, R., & Houlston, R. S. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol*, 23(3), 609-618. doi:10.1200/JCO.2005.01.086
- Prentice, L. M., Miller, R. R., Knaggs, J., Mazloomian, A., Aguirre Hernandez, R., Franchini, P., . . . Sheffield, B. S. (2018). Formalin fixation increases deamination mutation signature but

- should not lead to false positive mutations in clinical practice. *PLoS One*, 13(4), e0196434. doi:10.1371/journal.pone.0196434
- Puddefoot, J. R., Udeozo, U. K., Barker, S., & Vinson, G. P. (2006). The role of angiotensin II in the regulation of breast cancer cell adhesion and invasion. *Endocr Relat Cancer*, 13(3), 895-903. doi:10.1677/erc.1.01136
- Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S., & Frizelle, F. A. (2017). Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci Rep*, 7(1), 11590. doi:10.1038/s41598-017-11237-6
- Quintanilla, R. H., Jr., Asprer, J. S., Vaz, C., Tanavde, V., & Lakshmipathy, U. (2014). CD44 is a negative cell surface marker for pluripotent stem cell identification during human fibroblast reprogramming. *PLoS One*, 9(1), e85419. doi:10.1371/journal.pone.0085419
- Rahman, A. (2014). FOLFOXIRI and bevacizumab in metastatic colorectal cancer. *Lancet Oncol*, 15(13), e590. doi:10.1016/S1470-2045(14)71112-X
- Rahman, R. M. A., van Schaijik, B., Brasch, H. D., Marsh, R. W., Wickremesekera, A. C., Johnson, R., . . . Itinteang, T. (2019). Expression of Cathepsins B, D, and G in WHO Grade I Meningioma. *Front Surg*, 6, 6. doi:10.3389/fsurg.2019.00006
- Ram, R., Brasch, H. D., Dunne, J. C., Davis, P. F., Tan, S. T., & Itinteang, T. (2017). The Identification of Three Cancer Stem Cell Subpopulations within Moderately Differentiated Lip Squamous Cell Carcinoma. *Front Surg*, 4, 12. doi:10.3389/fsurg.2017.00012
- Ram, R. S., Brasch, H. D., Dunne, J. C., Davis, P. F., Tan, S. T., & Itinteang, T. (2017). Cancer Stem Cells in Moderately Differentiated Lip Squamous Cell Carcinoma Express Components of the Renin-Angiotensin System. *Front Surg*, 4, 30. doi:10.3389/fsurg.2017.00030
- Ramkumar, N., & Kohan, D. E. (2019). The (pro)renin receptor: an emerging player in hypertension and metabolic syndrome. *Kidney Int*, 95(5), 1041-1052. doi:10.1016/j.kint.2018.10.042
- Ray, L. B. (2010). Of Wnt, Prorenin Receptor, and V-ATPase. *Science Signaling*, 3(106), ec29-ec29. doi:10.1126/scisignal.3106ec29
- Ricci-Vitiani, L., Lombardi, D. G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C., & De Maria, R. (2007). Identification and expansion of human colon-cancer-initiating cells. *Nature*, 445(7123), 111-115. doi:10.1038/nature05384
- Ringel, T., Frey, N., Ringnalda, F., Janjuha, S., Cherkaoui, S., Butz, S., . . . Schwank, G. (2020). Genome-Scale CRISPR Screening in Human Intestinal Organoids Identifies Drivers of TGF-beta Resistance. *Cell Stem Cell*, 26(3), 431-440 e438. doi:10.1016/j.stem.2020.02.007
- Roth, I. M., Wickremesekera, A. C., Wickremesekera, S. K., Davis, P. F., & Tan, S. T. (2019). Therapeutic Targeting of Cancer Stem Cells via Modulation of the Renin-Angiotensin System. *Front Oncol*, 9, 745. doi:10.3389/fonc.2019.00745
- Ruan, H., Hao, S., Young, P., & Zhang, H. (2015). Targeting Cathepsin B for Cancer Therapies. *Horiz Cancer Res*, 56, 23-40. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/26623174>
- Ryckman, C., Vandal, K., Rouleau, P., Talbot, M., & Tessier, P. A. (2003). Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *J Immunol*, 170(6), 3233-3242. doi:10.4049/jimmunol.170.6.3233
- Rykl, J., Thiemann, J., Kurzawski, S., Pohl, T., Gobom, J., Zidek, W., & Schluter, H. (2006). Renal cathepsin G and angiotensin II generation. *J Hypertens*, 24(9), 1797-1807. doi:10.1097/01.hjh.0000242404.91332.be
- Sakoda, M., Ichihara, A., Kaneshiro, Y., Takemitsu, T., Nakazato, Y., Nabi, A. H., . . . Itoh, H. (2007). (Pro)renin receptor-mediated activation of mitogen-activated protein kinases in human vascular smooth muscle cells. *Hypertens Res*, 30(11), 1139-1146. doi:10.1291/hypres.30.1139
- Saleem, S., Tariq, S., Aleem, I., Sadr-Ul, S., Tahseen, M., Atiq, A., . . . Sutton, C. (2019). Proteomics analysis of colon cancer progression. *Clin Proteomics*, 16, 44. doi:10.1186/s12014-019-9264-y

- Salem, M. E., Bodor, J. N., Puccini, A., Xiu, J., Goldberg, R. M., Grothey, A., . . . Hall, M. J. (2020). Relationship between MLH1, PMS2, MSH2 and MSH6 gene-specific alterations and tumor mutational burden in 1057 microsatellite instability-high solid tumors. *Int J Cancer*, 147(10), 2948-2956. doi:10.1002/ijc.33115
- Salem, M. E., Puccini, A., Xiu, J., Raghavan, D., Lenz, H. J., Korn, W. M., . . . Goldberg, R. M. (2018). Comparative Molecular Analyses of Esophageal Squamous Cell Carcinoma, Esophageal Adenocarcinoma, and Gastric Adenocarcinoma. *Oncologist*, 23(11), 1319-1327. doi:10.1634/theoncologist.2018-0143
- Sanchez, J. A., Vogel, J. D., Kalady, M. F., Bronner, M. P., Skacel, M., & Church, J. M. (2008). Identifying Lynch syndrome: we are all responsible. *Dis Colon Rectum*, 51(12), 1750-1756. doi:10.1007/s10350-008-9414-1
- Sandberg, R., & Ernberg, I. (2005). Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). *Proc Natl Acad Sci U S A*, 102(6), 2052-2057. doi:10.1073/pnas.0408105102
- Sasaki, M., North, P. E., Elsey, J., Bubley, J., Rao, S., Jung, Y., . . . Arbiser, J. L. (2019). Propranolol exhibits activity against hemangiomas independent of beta blockade. *NPJ Precis Oncol*, 3, 27. doi:10.1038/s41698-019-0099-9
- Sato, T., van Es, J. H., Snippert, H. J., Stange, D. E., Vries, R. G., van den Born, M., . . . Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature*, 469(7330), 415-418. doi:10.1038/nature09637
- Sayeski, P. P., & Bernstein, K. E. (2001). Signal transduction mechanisms of the angiotensin II type AT(1)-receptor: looking beyond the heterotrimeric G protein paradigm. *J Renin Angiotensin Aldosterone Syst*, 2(1), 4-10. doi:10.3317/jraas.2001.007
- Schepers, A. G., Snippert, H. J., Stange, D. E., van den Born, M., van Es, J. H., van de Wetering, M., & Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science*, 337(6095), 730-735. doi:10.1126/science.1224676
- Schmohl, J. U., & Vallera, D. A. (2016). CD133, Selectively Targeting the Root of Cancer. *Toxins (Basel)*, 8(6). doi:10.3390/toxins8060165
- Schnell, U., Cirulli, V., & Giepmans, B. N. (2013). EpCAM: structure and function in health and disease. *Biochim Biophys Acta*, 1828(8), 1989-2001. doi:10.1016/j.bbamem.2013.04.018
- Schnell, U., Kuipers, J., & Giepmans, B. N. (2013). EpCAM proteolysis: new fragments with distinct functions? *Biosci Rep*, 33(2), e00030. doi:10.1042/BSR20120128
- Seaberg, R. M., & van der Kooy, D. (2003). Stem and progenitor cells: the premature desertion of rigorous definitions. *Trends Neurosci*, 26(3), 125-131. doi:10.1016/S0166-2236(03)00031-6
- Shepherd, A. J., Mickle, A. D., Golden, J. P., Mack, M. R., Halabi, C. M., de Kloet, A. D., . . . Mohapatra, D. P. (2018). Macrophage angiotensin II type 2 receptor triggers neuropathic pain. *Proc Natl Acad Sci U S A*, 115(34), E8057-E8066. doi:10.1073/pnas.1721815115
- Shi, G., & Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res Ther*, 1(5), 39. doi:10.1186/scrt39
- Shimokawa, M., Ohta, Y., Nishikori, S., Matano, M., Takano, A., Fujii, M., . . . Sato, T. (2017). Visualization and targeting of LGR5+ human colon cancer stem cells. *Nature*, 545(7653), 187-192. doi:10.1038/nature22081
- Shimomoto, T., Ohmori, H., Luo, Y., Chihara, Y., Denda, A., Sasahira, T., . . . Kuniyasu, H. (2012). Diabetes-associated angiotensin activation enhances liver metastasis of colon cancer. *Clin Exp Metastasis*, 29(8), 915-925. doi:10.1007/s10585-012-9480-6
- Shmelkov, S. V., Butler, J. M., Hooper, A. T., Hormigo, A., Kushner, J., Milde, T., . . . Rafii, S. (2008). CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. *J Clin Invest*, 118(6), 2111-2120. doi:10.1172/JCI34401
- Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G. S., Barzi, A., & Jemal, A. (2017). Colorectal cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, n/a-n/a. doi:10.3322/caac.21395

- Smith, M. T., & Muralidharan, A. (2015). Targeting angiotensin II type 2 receptor pathways to treat neuropathic pain and inflammatory pain. *Expert Opin Ther Targets*, 19(1), 25-35. doi:10.1517/14728222.2014.957673
- Smith, M. T., Wyse, B. D., & Edwards, S. R. (2013). Small molecule angiotensin II type 2 receptor (AT₂R) antagonists as novel analgesics for neuropathic pain: comparative pharmacokinetics, radioligand binding, and efficacy in rats. *Pain Med*, 14(5), 692-705. doi:10.1111/pme.12063
- Som, A., Bloch, S., Ippolito, J. E., & Achilefu, S. (2016). Acidic extracellular pH of tumors induces octamer-binding transcription factor 4 expression in murine fibroblasts in vitro and in vivo. *Sci Rep*, 6, 27803. doi:10.1038/srep27803
- Soond, S. M., Kozhevnikova, M. V., Townsend, P. A., & Zamyatnin, A. A., Jr. (2019). Cysteine Cathepsin Protease Inhibition: An update on its Diagnostic, Prognostic and Therapeutic Potential in Cancer. *Pharmaceuticals (Basel)*, 12(2). doi:10.3390/ph12020087
- Sparks, M. A., Crowley, S. D., Gurley, S. B., Mirotsoy, M., & Coffman, T. M. (2014). Classical Renin-Angiotensin system in kidney physiology. *Compr Physiol*, 4(3), 1201-1228. doi:10.1002/cphy.c130040
- Spitzner, M., Roesler, B., Bielfeld, C., Emons, G., Gaedcke, J., Wolff, H. A., . . . Grade, M. (2014). STAT3 inhibition sensitizes colorectal cancer to chemoradiotherapy in vitro and in vivo. *Int J Cancer*, 134(4), 997-1007. doi:10.1002/ijc.28429
- Srivastava, S. C., Dewar, H. A., & Newell, D. J. (1964). Double-Blind Trial of Propranolol (Inderal) in Angina of Effort. *Br Med J*, 2(5411), 724-725. doi:10.1136/bmj.2.5411.724
- Stark, G., Stark, U., Lueger, A., Bertuch, H., Pilger, E., Pietsch, B., . . . Lindner, W. (1989). The effects of the propranolol enantiomers on the intracardiac electrophysiological activities of Langendorff perfused hearts. *Basic Res Cardiol*, 84(5), 461-468. doi:10.1007/BF01908198
- Strate, L. L., & Syngal, S. (2005). Hereditary colorectal cancer syndromes. *Cancer Causes Control*, 16(3), 201-213. doi:10.1007/s10552-004-3488-4
- Sun, H., Li, T., Zhuang, R., Cai, W., & Zheng, Y. (2017). Do renin-angiotensin system inhibitors influence the recurrence, metastasis, and survival in cancer patients?: Evidence from a meta-analysis including 55 studies. *Medicine (Baltimore)*, 96(13), e6394. doi:10.1097/MD.0000000000006394
- Suo, G., Han, J., Wang, X., Zhang, J., Zhao, Y., Zhao, Y., & Dai, J. (2005). Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun*, 337(4), 1047-1051. doi:10.1016/j.bbrc.2005.09.157
- Sveen, A., Cremolini, C., & Dienstmann, R. (2019). Predictive modeling in colorectal cancer: time to move beyond consensus molecular subtypes. *Ann Oncol*, 30(11), 1682-1685. doi:10.1093/annonc/mdz412
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., . . . Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47(D1), D607-D613. doi:10.1093/nar/gky1131
- Tack, V., Ligtenberg, M. J. L., Siebers, A. G., Rombout, P. D. M., Dabir, P. D., Weren, R. D. A., . . . Dequeker, E. M. C. (2018). RAS testing for colorectal cancer patients is reliable in European laboratories that pass external quality assessment. *Virchows Arch*. doi:10.1007/s00428-017-2291-z
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5), 861-872. doi:10.1016/j.cell.2007.11.019
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4), 663-676. doi:10.1016/j.cell.2006.07.024

- Takashima, S., Gold, D., & Hartenstein, V. (2013). Stem cells and lineages of the intestine: a developmental and evolutionary perspective. *Dev Genes Evol*, 223(1-2), 85-102. doi:10.1007/s00427-012-0422-8
- Talebi, A., Kianersi, K., & Beiraghdar, M. (2015). Comparison of gene expression of SOX2 and OCT4 in normal tissue, polyps, and colon adenocarcinoma using immunohistochemical staining. *Adv Biomed Res*, 4, 234. doi:10.4103/2277-9175.167958
- Tang, D. G. (2012). Understanding cancer stem cell heterogeneity and plasticity. *Cell Res*, 22(3), 457-472. doi:10.1038/cr.2012.13
- Todaro, M., Alea, M. P., Di Stefano, A. B., Cammareri, P., Vermeulen, L., Iovino, F., . . . Stassi, G. (2007). Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. *Cell Stem Cell*, 1(4), 389-402. doi:10.1016/j.stem.2007.08.001
- Tomita, H., Tanaka, K., Tanaka, T., & Hara, A. (2016). Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget*, 7(10), 11018-11032. doi:10.18632/oncotarget.6920
- Tonnesen, M. G., Klempner, M. S., Austen, K. F., & Wintroub, B. U. (1982). Identification of a human neutrophil angiotensin II-generating protease as cathepsin G. *J Clin Invest*, 69(1), 25-30. doi:10.1172/jci110437
- Trzpis, M., McLaughlin, P. M., de Leij, L. M., & Harmsen, M. C. (2007). Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *Am J Pathol*, 171(2), 386-395. doi:10.2353/ajpath.2007.070152
- Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., & Turk, D. (2012). Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta*, 1824(1), 68-88. doi:10.1016/j.bbapap.2011.10.002
- Tyers, M., & Mann, M. (2003). From genomics to proteomics. *Nature*, 422(6928), 193-197. doi:10.1038/nature01510
- Uchiyama, Y. (2001). Autophagic cell death and its execution by lysosomal cathepsins. *Arch Histol Cytol*, 64(3), 233-246. doi:10.1679/aohc.64.233
- Valamehr, B., Abujarour, R., Robinson, M., Le, T., Robbins, D., Shoemaker, D., & Flynn, P. (2012). A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs. *Sci Rep*, 2, 213. doi:10.1038/srep00213
- Valenta, T., Hausmann, G., & Basler, K. (2012). The many faces and functions of beta-catenin. *EMBO J*, 31(12), 2714-2736. doi:10.1038/emboj.2012.150
- van der Gun, B. T., Melchers, L. J., Ruiters, M. H., de Leij, L. F., McLaughlin, P. M., & Rots, M. G. (2010). EpCAM in carcinogenesis: the good, the bad or the ugly. *Carcinogenesis*, 31(11), 1913-1921. doi:10.1093/carcin/bgq187
- van der Wath, R. C., Gardiner, B. S., Burgess, A. W., & Smith, D. W. (2013). Cell organisation in the colonic crypt: a theoretical comparison of the pedigree and niche concepts. *PLoS One*, 8(9), e73204. doi:10.1371/journal.pone.0073204
- Vanhaesebroeck, B., Guillermet-Guibert, J., Graupera, M., & Bilanges, B. (2010). The emerging mechanisms of isoform-specific PI3K signalling. *Nat Rev Mol Cell Biol*, 11(5), 329-341. doi:10.1038/nrm2882
- Vasaikar, S., Huang, C., Wang, X., Petyuk, V. A., Savage, S. R., Wen, B., . . . Clinical Proteomic Tumor Analysis, C. (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell*, 177(4), 1035-1049 e1019. doi:10.1016/j.cell.2019.03.030
- Villar-Cheda, B., Costa-Besada, M. A., Valenzuela, R., Perez-Costas, E., Melendez-Ferro, M., & Labandeira-Garcia, J. L. (2017). The intracellular angiotensin system buffers deleterious effects of the extracellular paracrine system. *Cell Death Dis*, 8(9), e3044. doi:10.1038/cddis.2017.439
- Vogelstein, B., Fearon, E. R., Kern, S. E., Hamilton, S. R., Preisinger, A. C., Nakamura, Y., & White, R. (1989). Allelotype of colorectal carcinomas. *Science*, 244(4901), 207-211. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/2565047>

- Wang, B., Chen, Q., Cao, Y., Ma, X., Yin, C., Jia, Y., . . . Fan, W. (2016). LGR5 Is a Gastric Cancer Stem Cell Marker Associated with Stemness and the EMT Signature Genes NANOG, NANOGP8, PRRX1, TWIST1, and BMI1. *PLoS One*, 11(12), e0168904. doi:10.1371/journal.pone.0168904
- Wang, J. P., & Hielscher, A. (2017). Fibronectin: How Its Aberrant Expression in Tumors May Improve Therapeutic Targeting. *J Cancer*, 8(4), 674-682. doi:10.7150/jca.16901
- Wang, P. H., Do, Y. S., Macaulay, L., Shinagawa, T., Anderson, P. W., Baxter, J. D., & Hsueh, W. A. (1991). Identification of renal cathepsin B as a human prorenin-processing enzyme. *J Biol Chem*, 266(19), 12633-12638. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/2061332>
- Wang, S., Song, R., Wang, Z., Jing, Z., Wang, S., & Ma, J. (2018). S100A8/A9 in Inflammation. *Front Immunol*, 9, 1298. doi:10.3389/fimmu.2018.01298
- Waterfall, J. F. (1989). A review of the preclinical cardiovascular pharmacology of cilazapril, a new angiotensin converting enzyme inhibitor. *Br J Clin Pharmacol*, 27 Suppl 2, 139S-150S. doi:10.1111/j.1365-2125.1989.tb03475.x
- Weber, M. A. (2005). The role of the new beta-blockers in treating cardiovascular disease. *Am J Hypertens*, 18(12 Pt 2), 169S-176S. doi:10.1016/j.amjhyper.2005.09.009
- Wegman-Ostrosky, T., Soto-Reyes, E., Vidal-Millan, S., & Sanchez-Corona, J. (2015). The renin-angiotensin system meets the hallmarks of cancer. *J Renin Angiotensin Aldosterone Syst*, 16(2), 227-233. doi:10.1177/1470320313496858
- Wei, J., Li, G., Dang, S., Zhou, Y., Zeng, K., & Liu, M. (2016). Discovery and Validation of Hypermethylated Markers for Colorectal Cancer. *Dis Markers*, 2016, 2192853. doi:10.1155/2016/2192853
- Wen, K., Fu, Z., Wu, X., Feng, J., Chen, W., & Qian, J. (2013). Oct-4 is required for an antiapoptotic behavior of chemoresistant colorectal cancer cells enriched for cancer stem cells: effects associated with STAT3/Survivin. *Cancer Lett*, 333(1), 56-65. doi:10.1016/j.canlet.2013.01.009
- Werner, J., & Heinemann, V. (2016). Standards and Challenges of Care for Colorectal Cancer Today. *Visc Med*, 32(3), 156-157. doi:10.1159/000447070
- Wielenga, V. J., Smits, R., Korinek, V., Smit, L., Kielman, M., Fodde, R., . . . Pals, S. T. (1999). Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. *Am J Pathol*, 154(2), 515-523. doi:10.1016/S0002-9440(10)65297-2
- Williams, R. N., Parsons, S. L., Morris, T. M., Rowlands, B. J., & Watson, S. A. (2005). Inhibition of matrix metalloproteinase activity and growth of gastric adenocarcinoma cells by an angiotensin converting enzyme inhibitor in in vitro and murine models. *Eur J Surg Oncol*, 31(9), 1042-1050. doi:10.1016/j.ejso.2005.04.003
- Wilop, S., von Hobe, S., Crysandt, M., Esser, A., Osieka, R., & Jost, E. (2009). Impact of angiotensin I converting enzyme inhibitors and angiotensin II type 1 receptor blockers on survival in patients with advanced non-small-cell lung cancer undergoing first-line platinum-based chemotherapy. *J Cancer Res Clin Oncol*, 135(10), 1429-1435. doi:10.1007/s00432-009-0587-3
- Wilson, T. J., Nannuru, K. C., Futakuchi, M., & Singh, R. K. (2010). Cathepsin G-mediated enhanced TGF-beta signaling promotes angiogenesis via upregulation of VEGF and MCP-1. *Cancer Lett*, 288(2), 162-169. doi:10.1016/j.canlet.2009.06.035
- Wiysonge, C. S., Volmink, J., & Opie, L. H. (2007). Beta-blockers and the treatment of hypertension: it is time to move on. *Cardiovasc J Afr*, 18(6), 351-352. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18092107>
- Wolf, G. (2006). Renal injury due to renin-angiotensin-aldosterone system activation of the transforming growth factor-beta pathway. *Kidney Int*, 70(11), 1914-1919. doi:10.1038/sj.ki.5001846
- Wu, Y., & Wu, P. Y. (2009). CD133 as a marker for cancer stem cells: progresses and concerns. *Stem Cells Dev*, 18(8), 1127-1134. doi:10.1089/scd.2008.0338
- Xu, J., Lamouille, S., & Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. *Cell Res*, 19(2), 156-172. doi:10.1038/cr.2009.5

- Xu, X., Chai, S., Wang, P., Zhang, C., Yang, Y., Yang, Y., & Wang, K. (2015). Aldehyde dehydrogenases and cancer stem cells. *Cancer Lett*, 369(1), 50-57. doi:10.1016/j.canlet.2015.08.018
- Yamashita, T., Budhu, A., Forgues, M., & Wang, X. W. (2007). Activation of hepatic stem cell marker EpCAM by Wnt-beta-catenin signaling in hepatocellular carcinoma. *Cancer Res*, 67(22), 10831-10839. doi:10.1158/0008-5472.CAN-07-0908
- Yang, J., Wang, E., Zhou, S., & Yang, Q. (2015). Effects of (R)- and (S)-propranolol hydrochloride enantiomers on the resonance Rayleigh scattering spectra with erythrosine B as probe and their analytical applications. *Talanta*, 134, 754-760. doi:10.1016/j.talanta.2014.12.030
- Yap, A., Lopez-Olivo, M. A., Dubowitz, J., Pratt, G., Hiller, J., Gottumukkala, V., . . . Schier, R. (2018). Effect of beta-blockers on cancer recurrence and survival: a meta-analysis of epidemiological and perioperative studies. *Br J Anaesth*, 121(1), 45-57. doi:10.1016/j.bja.2018.03.024
- Yates, J. R., Ruse, C. I., & Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*, 11, 49-79. doi:10.1146/annurev-bioeng-061008-124934
- Ye, Y., Zhang, R., & Feng, H. (2020). Fibronectin promotes tumor cells growth and drugs resistance through a CDC42-YAP-dependent signaling pathway in colorectal cancer. *Cell Biol Int*, 44(9), 1840-1849. doi:10.1002/cbin.11390
- Yeung, T. M., Gandhi, S. C., Wilding, J. L., Muschel, R., & Bodmer, W. F. (2010). Cancer stem cells from colorectal cancer-derived cell lines. *Proc Natl Acad Sci U S A*, 107(8), 3722-3727. doi:10.1073/pnas.0915135107
- Yi, W., Xiao, E., Ding, R., Luo, P., & Yang, Y. (2016). High expression of fibronectin is associated with poor prognosis, cell proliferation and malignancy via the NF-kappaB/p53-apoptosis signaling pathway in colorectal cancer. *Oncol Rep*, 36(6), 3145-3153. doi:10.3892/or.2016.5177
- Yin, A. H., Miraglia, S., Zanjani, E. D., Almeida-Porada, G., Ogawa, M., Leary, A. G., . . . Buck, D. W. (1997). AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood*, 90(12), 5002-5012. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9389720>
- Yu, H., Pardoll, D., & Jove, R. (2009). STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer*, 9(11), 798-809. doi:10.1038/nrc2734
- Yu, H. H., Featherston, T., Tan, S. T., Chibnall, A. M., Brasch, H. D., Davis, P. F., & Itinteang, T. (2016). Characterization of Cancer Stem Cells in Moderately Differentiated Buccal Mucosal Squamous Cell Carcinoma. *Front Surg*, 3, 46. doi:10.3389/fsurg.2016.00046
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., . . . Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science*, 318(5858), 1917-1920. doi:10.1126/science.1151526
- Yu, T., Ma, Y., & Wang, H. (2017). EpCAM Intracellular Domain Promotes Porcine Cell Reprogramming by Upregulation of Pluripotent Gene Expression via Beta-catenin Signaling. *Sci Rep*, 7, 46315. doi:10.1038/srep46315
- Yu, X., Lin, Y., Yan, X., Tian, Q., Li, L., & Lin, E. H. (2011). CD133, Stem Cells, and Cancer Stem Cells: Myth or Reality? *Curr Colorectal Cancer Rep*, 7(4), 253-259. doi:10.1007/s11888-011-0106-1
- Zaitseva, M., Vollenhoven, B. J., & Rogers, P. A. (2006). In vitro culture significantly alters gene expression profiles and reduces differences between myometrial and fibroid smooth muscle cells. *Mol Hum Reprod*, 12(3), 187-207. doi:10.1093/molehr/gal018
- Zhang, J., Espinoza, L. A., Kinders, R. J., Lawrence, S. M., Pfister, T. D., Zhou, M., . . . Jessup, J. M. (2013). NANOG Modulates Stemness in Human Colorectal Cancer. *Oncogene*, 32(37), 4397-4405. doi:10.1038/onc.2012.461
- Zhang, S., & Cui, W. (2014). Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J Stem Cells*, 6(3), 305-311. doi:10.4252/wjsc.v6.i3.305
- Zhang, W., Sui, Y., Ni, J., & Yang, T. (2016). Insights into the Nanog gene: A propeller for stemness in primitive stem cells. *Int J Biol Sci*, 12(11), 1372-1381. doi:10.7150/ijbs.16349

- Zhao, Y., Wang, H., Li, X., Cao, M., Lu, H., Meng, Q., . . . Cai, L. (2014). Ang II-AT1R increases cell migration through PI3K/AKT and NF-kappaB pathways in breast cancer. *J Cell Physiol*, 229(11), 1855-1862. doi:10.1002/jcp.24639
- Zheng, S., Xin, L., Liang, A., & Fu, Y. (2013). Cancer stem cell hypothesis: a brief summary and two proposals. *Cytotechnology*, 65(4), 505-512. doi:10.1007/s10616-012-9517-3
- Zhou, L., Li, Y., Hao, S., Zhou, D., Tan, R. J., Nie, J., . . . Liu, Y. (2015). Multiple genes of the renin-angiotensin system are novel targets of Wnt/beta-catenin signaling. *J Am Soc Nephrol*, 26(1), 107-120. doi:10.1681/ASN.2014010085
- Zhou, L., Luo, Y., Sato, S., Tanabe, E., Kitayoshi, M., Fujiwara, R., . . . Kuniyasu, H. (2014). Role of two types of angiotensin II receptors in colorectal carcinoma progression. *Pathobiology*, 81(4), 169-175. doi:10.1159/000362092
- Zhou, X., Wang, G., & Sun, Y. (2015). A reliable parameter to standardize the scoring of stem cell spheres. *PLoS One*, 10(5), e0127348. doi:10.1371/journal.pone.0127348

Appendices

Appendix A: Supplementary figures and tables

Case	Sample	DoD	Mets/ recur?	Age	Ethnicity	Gender	Stage/Grade	Lab report	Molecular testing	Comorbidity	Medication
15-199	LG1	N/A	No	53	NZ Euro	M	pT3 N1a Mx / LGCA	Irregular cribriform glands; pleomorphic columnar cells	Normal for MLH1, MSH2, MSH6, PMS2	Yes	Cilazapril
15-242	-	N/A	No	79	NZ Euro	M	pT3 N1a Mx / LGCA	Ulceration, aberrant crypt foci, lymphoid aggregates. High grade dysplasia, intramucosal	Unknown	Yes	Atenolol, Atorvastatin
15-258	-	28/08/17	Yes – mets to chest, pelvis, abdomen	85	NZ Euro	M	pT3 pN1b Mx / LGCA	Mucinous foci. Irregular and cribriform glands; Invades within desmoplastic stroma	Diagnostic criteria for testing not met (no IHC or seq)	Yes	Cilazapril
15-259	-	N/A	Yes – hyperplastic polyp removed 08/18	54	NZ Euro	F	pT4a pN2b pM0 / LGCA	Ulcerated, mucinous. Invades within desmoplastic stroma	KRAS mutation, not BRAF. Abnormal loss on MSI enzyme testing	Yes - Lynch syndrome	Cilazapril
15-361	-	24/01/18	No	91	NZ Euro	F	pT3 pN0 M0 / HGCA	Ulcerated, embedded in a desmoplastic stroma; trabecular, mucinous, transmural lymphoid aggregates	S-100 positive; neg for CD34, SMA, desmin, EMA, GLUT1 = benign mucosal Schwann cell proliferation	Yes	?
15-398	-	N/A	No	75	NZ Euro	M	pT3 pN1a M0 / LGCA	Cribriform glands lined by moderately pleomorphic columnar epithelium within a desmoplastic stroma	Unknown	Yes	?
16-059	-	N/A	No	69	NZ Euro	M	T3 N1 Mx / LGCA	Mucinous features. Invades within desmoplastic stroma.	Unknown	Yes - Prostate carcinoma + others	Cilazapril, Simvastatin
16-127	HG1	26/05/18	Yes – liver, lung	50	NZ Euro	M	pT3 N0 Mx / HGCA	No records - presumed private surgery	Unknown	?	?

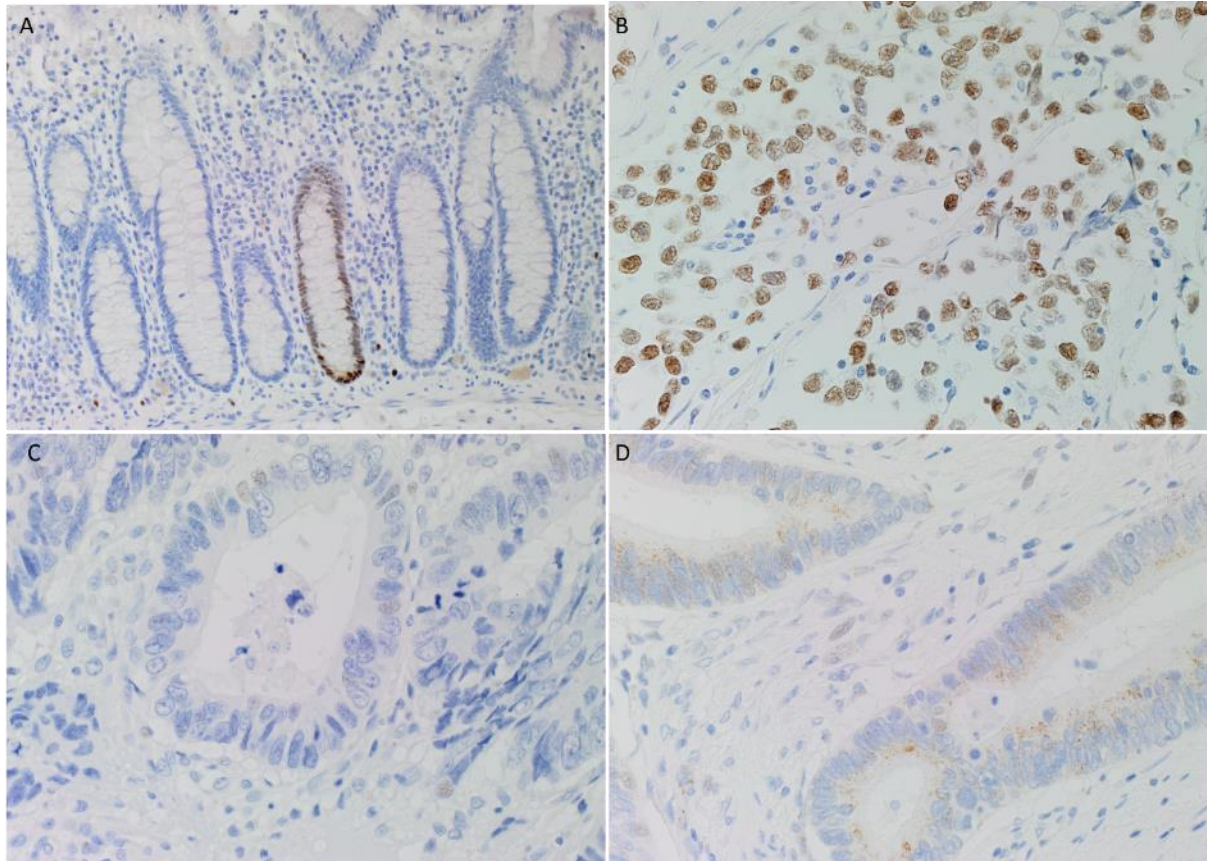
16-146	HG2	8/10/17	Yes	81	NZ Euro	F	? / HGCA	Mucinous component	Unknown	?	?
16-243	-	N/A	Yes – treated 2019, clear 01/20	78	NZ Euro	M	pT3 pN2a Mx / LGCA	Mod diff adenoca., no lymph nodes (0/7)	Unknown	?	?
16-285	-	29/10/17	Yes – liver, lung	71	Chinese	M	pT2 pN0 Mx / LGCA	Fungating; desmoplastic stroma.	Unknown	Yes	?
16-337	LG2	N/A	No	67	NZ Euro	M	pT3 pN1a+ Mx / LGCA	Several small foci of high grade noted.	Unknown	Yes	?
17-020	HG3	N/A	No	74	Indian	F	T3 N1 Mx / HGCA	Tumour invades through muscularis propria into the serosa	Unknown	?	?
17-146	-	N/A	Yes – liver: treated 2018, clear 12/19	78	NZ Euro	F	pT4 N2b Mx / LGCA	Mucin, strips of neoplastic mildly pleomorphic columnar epithelium	MSH6 and MSH2 positive staining; MLH1 and PSM2 show abnormal loss; BRAF not yet tested	Yes - Meningioma grade 2	Metoprolol, Quinapril
17-228	-	4/03/18	Yes – liver (not treatable)	87	NZ Euro	F	T3 N0 Mx / HGCA	Sheets and cords of moderate pleomorphic polygonal cells with some signet ring forms	SATB2 positive, CDX2 negative	?	?
17-235	-	N/A	?	84	NZ Euro	F	T4b N1b Mx / HGCA	Invades through muscularis propria	MLH1, PMS2, MSH6, MSH2 normal (positive)	?	Amolodopine
17-279	LG3	N/A	No	60	NZ Euro	F	T3 N2a Mx / LGCA	Spread to pericolic fat but not invasive though muscularis	MLH1, PMS2, MSH6, MSH2 normal (positive)	Yes – COPD, hyperthyroid, benign papillary mesothelioma	Carbimazole, Salbutamol, Striverdi respimat
17-455	-	N/A	?	37	NZ Euro	F	pT3 N0 Mx / HGCA	Predominantly low grade but foci of high grade. Private operation - few details	MLH1, PMS2, MSH6, MSH2 normal (positive)	?	?

18-069	HG4	N/A	?	70	NZ Euro	M	pT4b N2a Mx / HGCA	Invades through muscularis propria. Treated privately	MLH1, PMS2, MSH6, MSH2 normal (positive)	?	?
18-282	LG4	N/A	?	73	NZ Euro	?	T3 N1b Mx / LGCA	1/32 lymph nodes, extramural tumour deposits; perineural invasion. Treated privately	MLH1, PMS2, MSH6, MSH2 normal (positive)	?	?

Appendix Table A1.1: Patient data

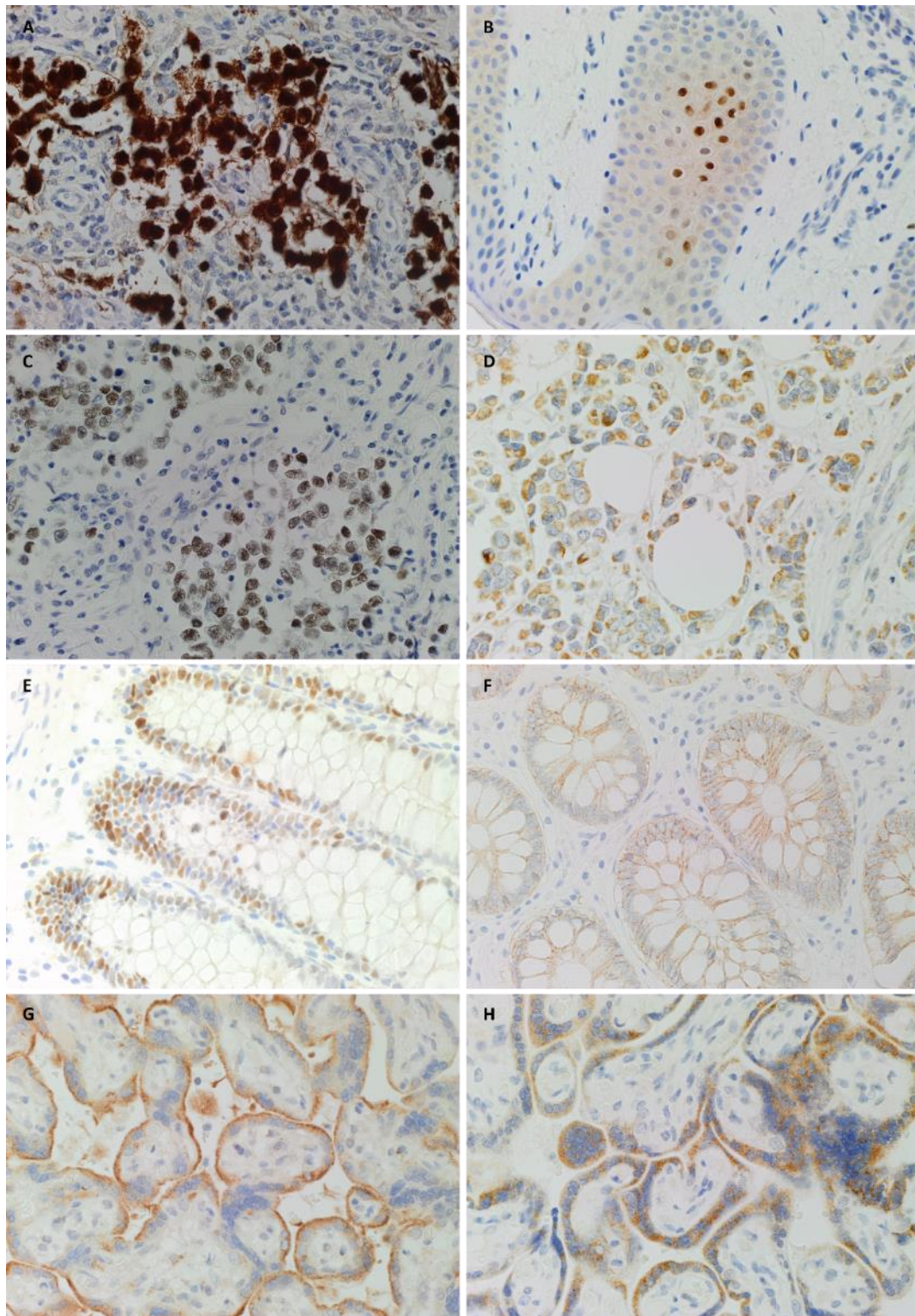
Case = unique anonymised identification number assigned at GMRI. Sample = name which the sample is referred to within the thesis. DoD = date of death. Mets/recr? = has the tumour metastasised, or recurred since surgery? Stage/grade = TNM staging and tumour grade (LG or HG).

Appendix Figure A4.1: Supplementary iPSC marker IHC-stained images.



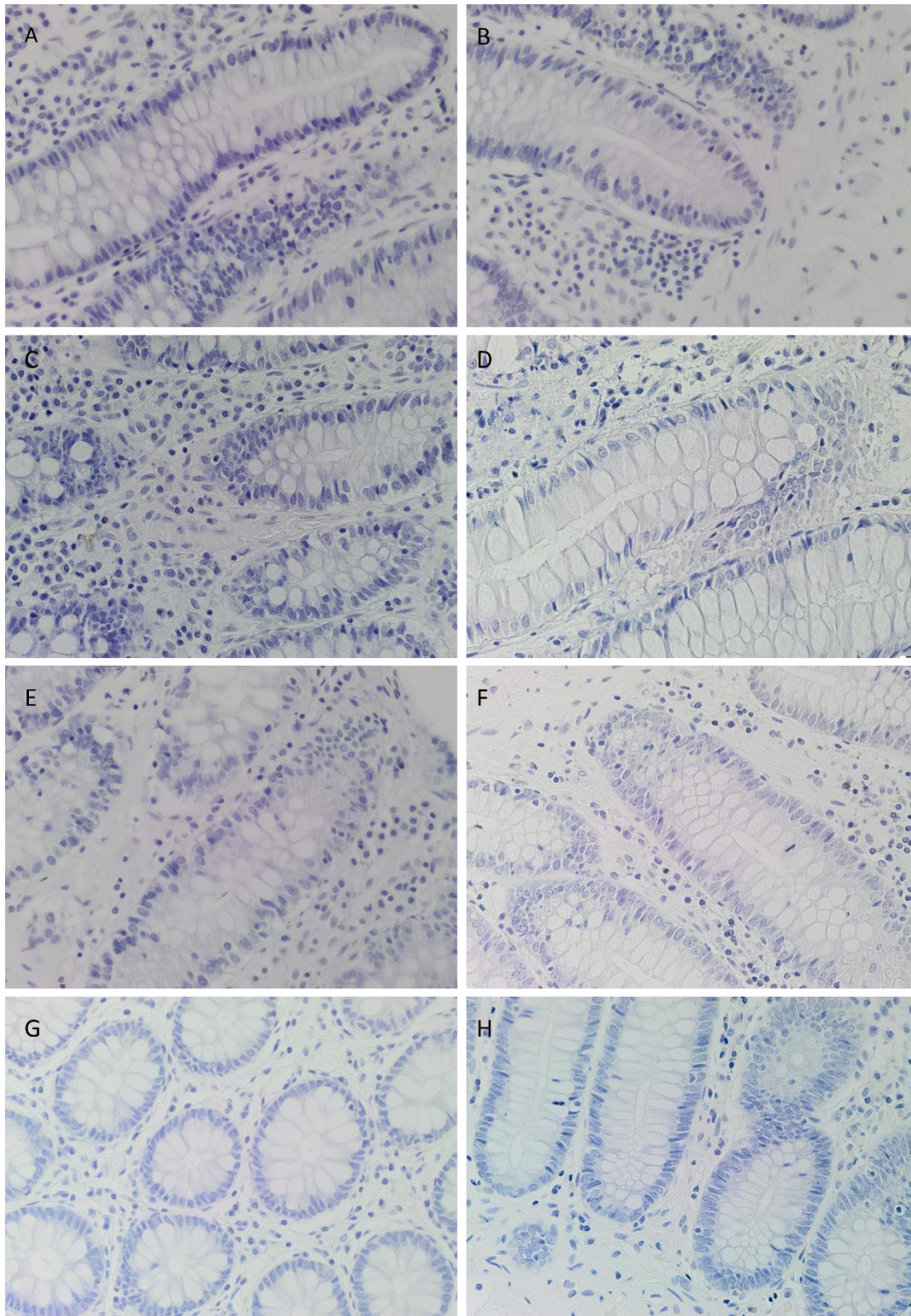
Representative IHC staining images showing nuclear SOX2 expression (brown) in one crypt of a normal colon sample (A); seminoma positive control stained with the second NANOG antibody (ab62734) (B); and representative IHC stained images of LGCA (C) and HGCA (D) stained with the second NANOG antibody (ab62734). Nuclei were counter-stained with hematoxylin (blue). Original magnification: 400x.

Appendix Figure A4.2: iPSC marker IHC staining positive controls.



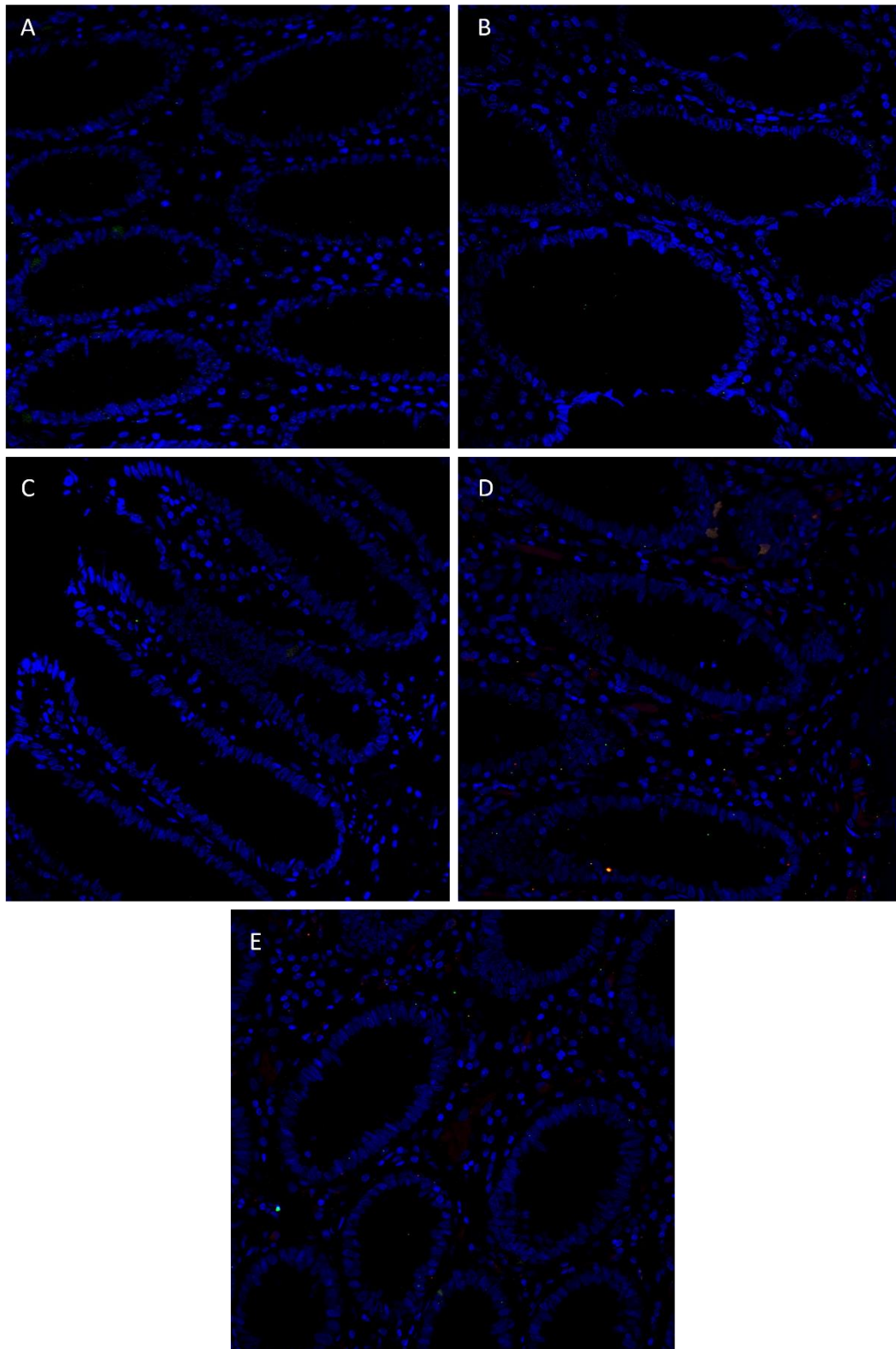
Representative IHC stained images of human positive control tissues demonstrating the expected staining patterns on seminoma for OCT4 (A) and NANOG (C), skin for SOX2 (B), normal breast tissue for KLF4 (D), normal colon for c-MYC (E) and EpCAM (F), and placenta for CD133 (G) and LGR5 (H). Nuclei were counterstained with hematoxylin (blue). Original magnification: 400X.

Appendix Figure A4.3: iPSC marker IHC staining negative controls.



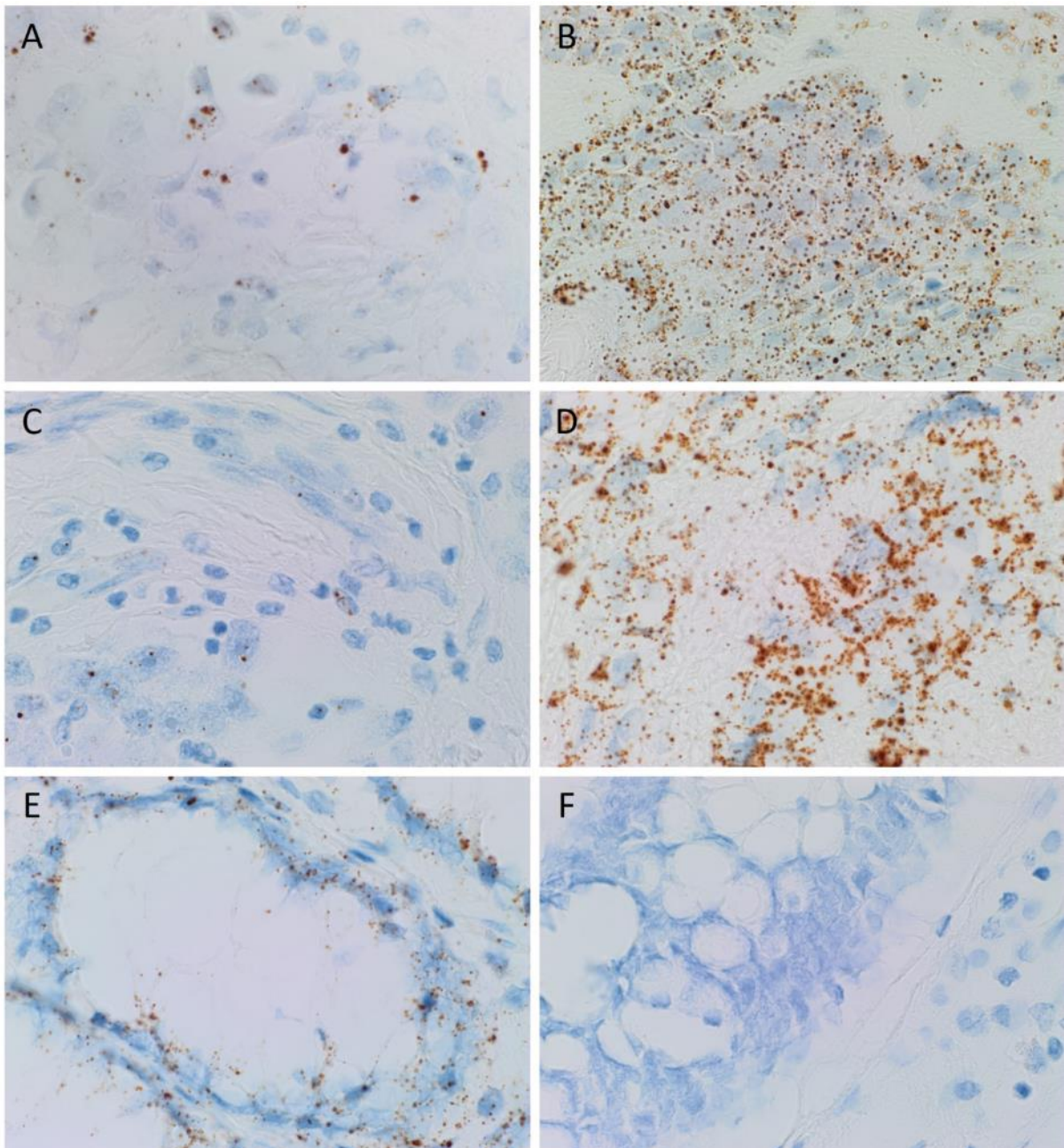
Images showing the negative control slides from IHC staining of OCT4 (A), SOX2 (B), NANOG (C), KLF4 (D), c-MYC (E), EpCAM (F), CD133 (G) and LGR5 (H). Negative controls for IHC were produced by using an isotype control antibody. Cell nuclei were counter-stained with 4'6-diamino-2-phenylindole (DAPI; blue). Original magnification: 400X.

Appendix Figure A4.4: iPSC marker IF staining negative controls.



Images showing the negative control slides from IF staining of OCT4/KLF4 (A), OCT4/NANOG (B), OCT4/SOX2 (C), OCT4/c-MYC (D) and OCT4/CD133 (E). Negative controls for IF were produced by omitting the primary antibody. Cell nuclei were counter-stained with 4'6-diamino-2-phenylindole (DAPI; blue). Original magnification: 400X.

Appendix Figure A4.5: *In situ* hybridisation controls.



In-situ hybridisation positive human control tissues for OCT4 (A, brown), NANOG (B, brown) and KLF4 (C, brown) on seminoma; SOX2 (D, brown) on normal skin, and c-MYC (E, brown) on normal colon. Negative control (F) to confirm the specificity of probes were produced using a probe for *Bacillus* (NM_L38424). Original magnification: 1000x.

Appendix Table A4.1: Supplementary data for iPSC marker RT-qPCR in tissues

		ΔCT (tumour)	Average ΔCT (normal)	$\Delta\Delta CT$ (T-N)	$2^{-(\Delta\Delta CT)}$
OCT4	LG1	-10.34	-10.27	-0.07	0.95
		-11.21		-0.94	0.52
		-10.71		-0.44	0.74
	LG2	-9.18	-9.71	0.53	1.44
		-9.05		0.66	1.58
		-9.06		0.65	1.57
	LG3	-12.47	-13.3	0.83	1.77
		-12.30		0.95	1.93
		-12.67		0.63	1.54
	LG4	-10.68	-9.78	-0.90	0.54
		-10.85		-1.07	0.48
		-10.86		-1.08	0.47
	LG5	-10.59	-9.86	-0.74	0.60
		-10.79		-0.94	0.52
		-11.27		-1.42	0.37
	LG6	-10.32	-8.86	-1.46	0.36
		-10.67		-1.81	0.28
		-10.87		-2.01	0.25
	HG1	-16.58	-14.29	-2.29	0.20
		-16.75		-2.46	0.18
		-15.92		-1.63	0.32
	HG2	-8.73	-10.16	1.43	2.69
		-9.02		1.14	2.20
		-9.02		1.14	2.20
	HG3	-15.16	-13.59	-1.57	0.34
		-15.03		-1.44	0.37
		-15.40		-1.81	0.28
	HG4	-11.09	0	-11.09	∞
		-11.31		-11.31	∞
		-11.09		-11.10	∞
	HG5	-8.97	-10.13	1.16	2.23
		-9.28		0.85	1.80
		-9.19		0.94	1.92
	HG6	-10.09	-9.08	-1.01	0.50
		-9.87		-0.79	0.58
		-10.08		-1	0.50
	LG1	-5.31	-4.67	-0.64	0.64
		-5.40		-0.73	0.60
		-5.35		-0.68	0.62
	LG2	-14.77	-14.67	-0.10	0.94
		-14.93		-0.26	0.84
		-14.56		0.11	1.08
	LG3	-12.09	-7.70	-4.40	0.05
		-12.12		-4.43	0.05
		-12.28		-4.59	0.04
	LG4	-11.00	-4.04	-6.96	0.01
		-11.13		-7.09	0.01
		-11.07		-7.03	0.01
	LG5	24.75	-15.32	40.07	0
		24.75		40.07	0
		24.75		40.07	0

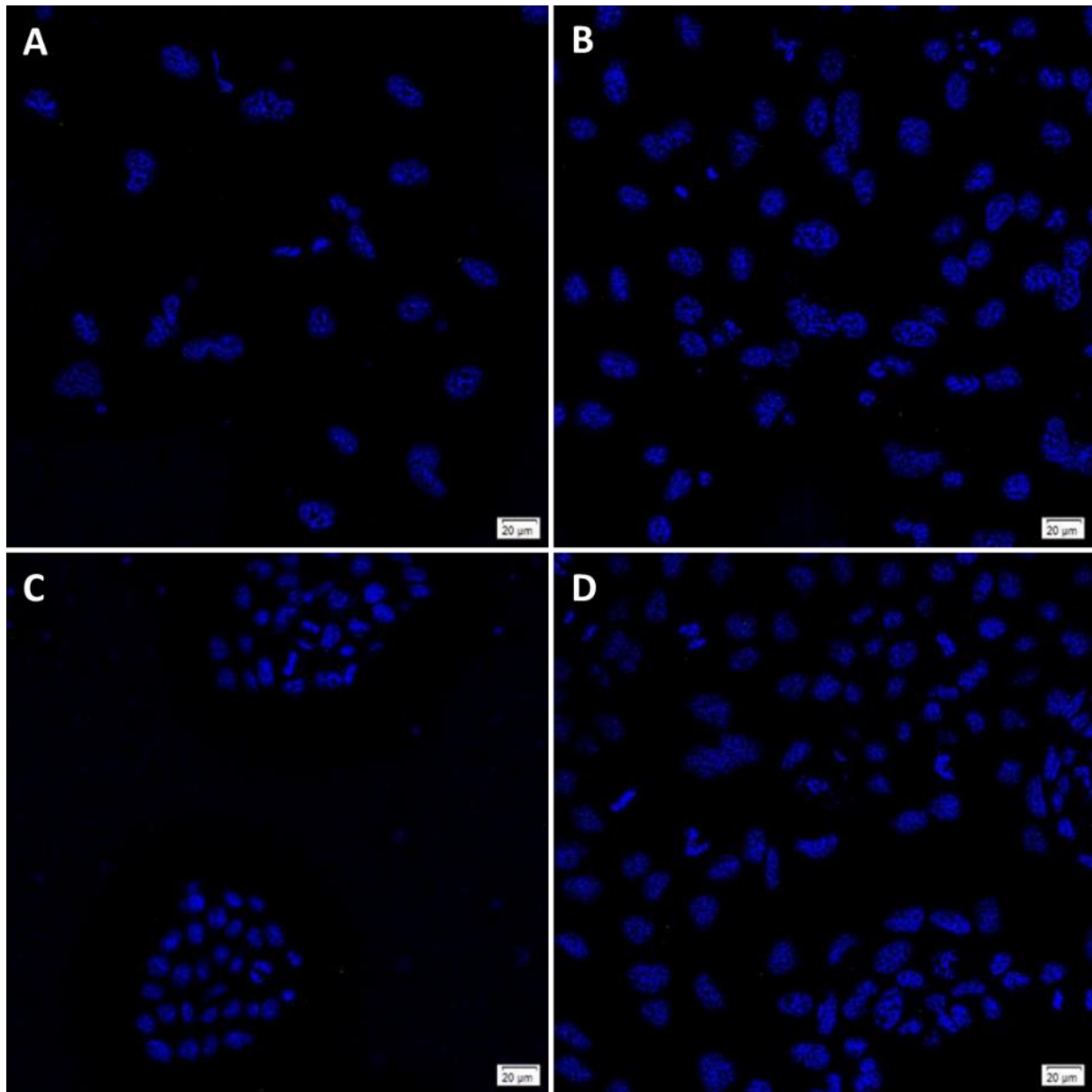
SOX2	LG6	22.19	24.78	-2.60	0
		22.19		-2.60	0
		22.19		-2.60	0
	HG1	-10.16	-9.38	-0.78	0.58
		-11.06		-1.68	0.31
		-10.07		-0.69	0.62
	HG2	21.75	22.94	-1.19	0
		21.75		-1.19	0
		21.75		-1.19	0
NANOG	HG3	-8.48	-11.46	2.98	7.89
		-8.12		3.34	10.13
		-8.3		3.16	8.94
	HG4	-16.17	0	-16.17	∞
		-16.33		-16.33	∞
		-16.73		-16.73	∞
	HG5	24.25	9.26	14.99	0
		24.25		14.99	0
		24.25		14.99	0
	HG6	24.23	0	24.23	0
		24.23		24.23	0
		24.23		24.23	0
	LG1	-10.53	-13.56	3.03	8.17
		-11.01		2.55	5.86
		-11.27		2.29	4.89
	LG2	-10.76	-11.39	0.62	1.54
		-10.67		0.71	1.64
		-10.83		0.55	1.47
	LG3	-12.63	-14.17	1.54	2.91
		-13.18		0.99	1.99
		-12.64		1.53	2.89
	LG4	-8.78	-10.10	1.31	2.91
		-9.13		0.96	1.99
		-9.01		1.08	2.89
	LG5	-13.68	-11.09	-2.59	0.17
		-13.50		-2.41	0.19
		-13.57		-2.48	0.18
	LG6	-11.42	-9.74	-1.68	0.31
		-11.34		-1.60	0.33
		-11.79		-2.05	0.24
	HG1	-14.29	-11.83	-2.46	0.18
		-14.77		-2.94	0.13
		-14.97		-3.14	0.11
	HG2	-8.57	-8.69	0.12	1.09
		-8.50		0.19	1.14
		-8.68		0.01	1.01
	HG3	-14.24	-14.52	0.28	1.21
		-14.30		0.22	1.16
		-14.08		0.44	1.36
	HG4	-9.88	-8.09	-1.79	0.29
		-9.99		-1.90	0.27
		-10.20		-2.11	0.23
	HG5	-8.33	-9.07	0.73	1.66
		-8.36		0.70	1.63
		-8.25		0.81	1.76
	HG6	-8.49	-8.59	0.10	1.07
		-8.52		0.07	1.05

		-8.53		0.06	1.04
KLF4	LG1	-5.98	-8.74	2.76	6.79
		-5.66		3.08	8.48
		-5.50		3.24	9.47
	LG2	-11.04	-8.42	-2.62	0.16
		-10.62		-2.20	0.22
		-10.45		-2.03	0.24
	LG3	-11.45	-11.01	-0.45	0.73
		-10.91		0.09	1.07
		-11.43		-0.43	0.74
	LG4	-11.06	-8.90	-2.17	0.22
		-10.66		-1.77	0.29
		-10.70		-1.81	0.29
	LG5	-8.99	-7.05	-1.94	0.26
		-8.72		-1.67	0.31
		-8.80		-1.75	0.30
c-MYC	LG6	-8.45	-5.56	-2.90	0.13
		-8.44		-2.89	0.14
		-8.35		-2.80	0.14
	HG1	-12.58	-8.74	-3.84	0.07
		-12.39		-3.65	0.08
		-12.87		-4.13	0.06
	HG2	-10.85	-9.34	-1.50	0.35
		-11.24		-1.89	0.27
		-10.94		-1.59	0.33
	HG3	-12.25	-11.77	-0.48	0.72
		-12.10		-0.33	0.80
		-12.22		-0.45	0.73
	HG4	-10.37	-7.64	-2.73	0.15
		-10.66		-3.02	0.12
		-10.70		-3.06	0.12
	HG5	-9.11	-11.80	2.69	6.45
		-8.71		3.09	8.51
		-9.97		1.83	3.56
	HG6	-9.56	-8.00	-1.55	0.34
		-9.72		-1.71	0.31
		-9.92		-1.91	0.27
c-MYC	LG1	-6.30	-11.71	5.41	42.62
		-6.02		5.69	51.74
		-5.99		5.72	52.83
	LG2	-4.49	-5.75	1.26	2.39
		-4.68		1.07	2.10
		-4.02		1.73	3.32
	LG3	-8.01	-7.72	-0.29	0.82
		-7.97		-0.25	0.84
		-7.66		0.06	1.04
	LG4	-4.32	-3.31	-1.01	0.50
		-3.94		-0.63	0.65
		-4.01		-0.70	0.61
	LG5	-8.52	-4.91	-3.62	0.08
		-6.67		-1.77	0.29
		-8.08		-3.18	0.11
	LG6	-6.70	-9.25	2.55	5.84
		-7.15		2.10	4.28
		-6.54		2.71	6.52
	HG1	-10.14	-8.35	-1.80	0.29

		-10.14		-1.79	0.29
		-10.52		-2.17	0.22
	HG2	-2.16	-5.52	3.36	10.29
		-2.32		3.20	9.21
		-2.48		3.04	8.24
	HG3	-8.09	-8.81	0.72	1.65
		-8.03		0.78	1.72
		-8.01		0.80	1.74
	HG4	-4.85	-7.08	2.24	4.71
		-4.68		2.41	5.30
		-5.06		2.03	4.07
	HG5	-4.34	-6.23	1.88	3.69
		-4.47		1.75	3.37
		-4.95		1.27	2.42
	HG6	-4.79	-5.49	0.70	1.63
		-4.94		0.55	1.47
		-5.18		0.31	1.24

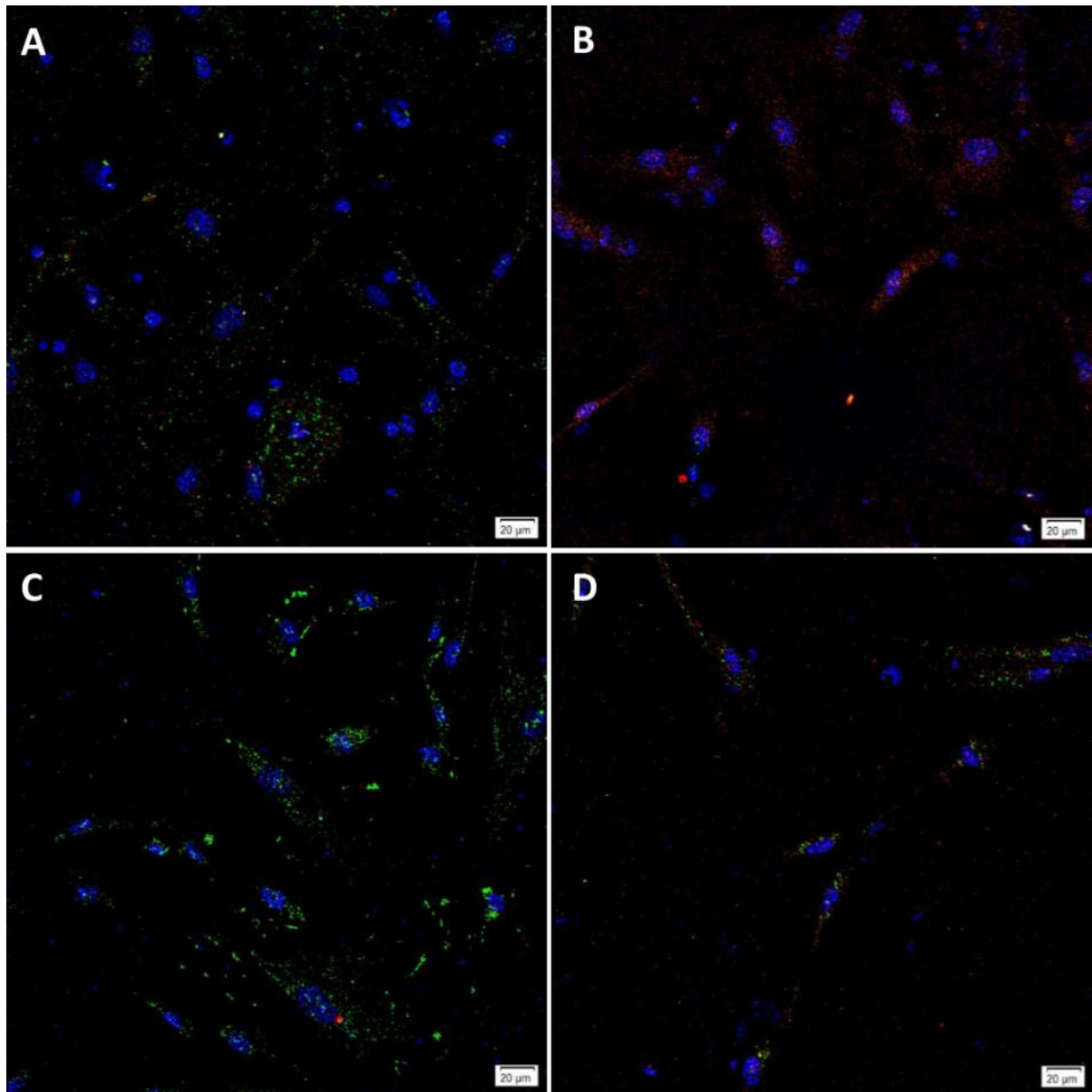
RT-qPCR data showing expression of iPSC markers *OCT4*, *SOX2*, *NANOG*, *KLF4* and *c-MYC*. ΔCT values calculated by comparing the gene of interest to housekeeper *GAPDH*, and $\Delta\Delta CT$ values by comparing high-grade (HG) and low-grade (LG) tumours to their patient-matched normal colon samples. $\Delta\Delta CT$ values used to calculate fold changes using the equation $2^{(\Delta\Delta CT)}$.

Appendix Figure A5.1: Representative images of ICC negative controls.



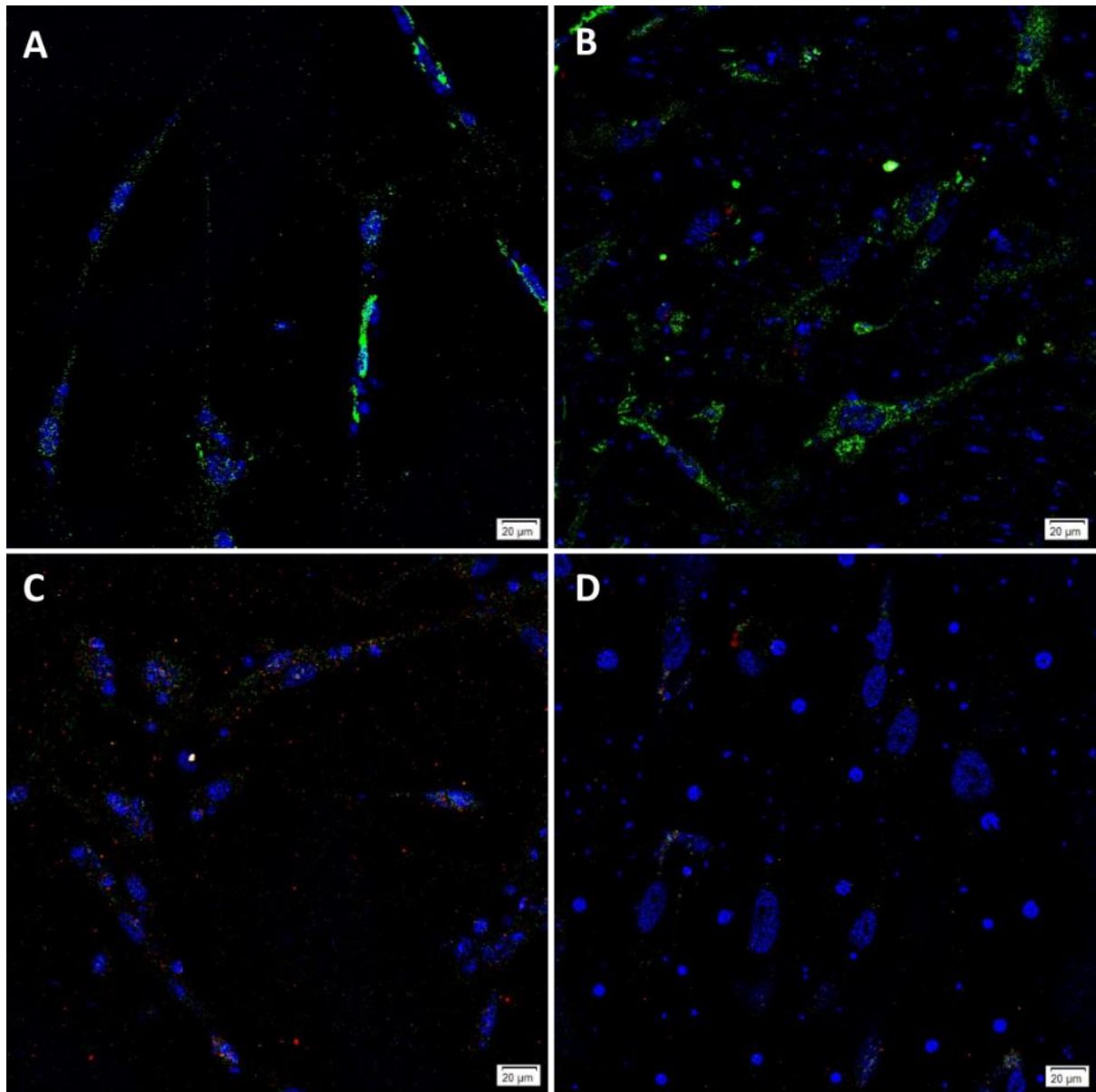
Negative controls were run for the control cells by omitting the primary antibodies for SSEA4/OCT4 (A) and SOX2/TRA-1-60 (B) in NTERA-2 cells, and for SSEA4/OCT4 (C) and SOX2/TRA-1-60 (D) in CaCo2 cells. Original magnification: 400x; scale bar = 20 µm.

Appendix Figure A5.2: Representative images of ICC staining for sample LGCA1.



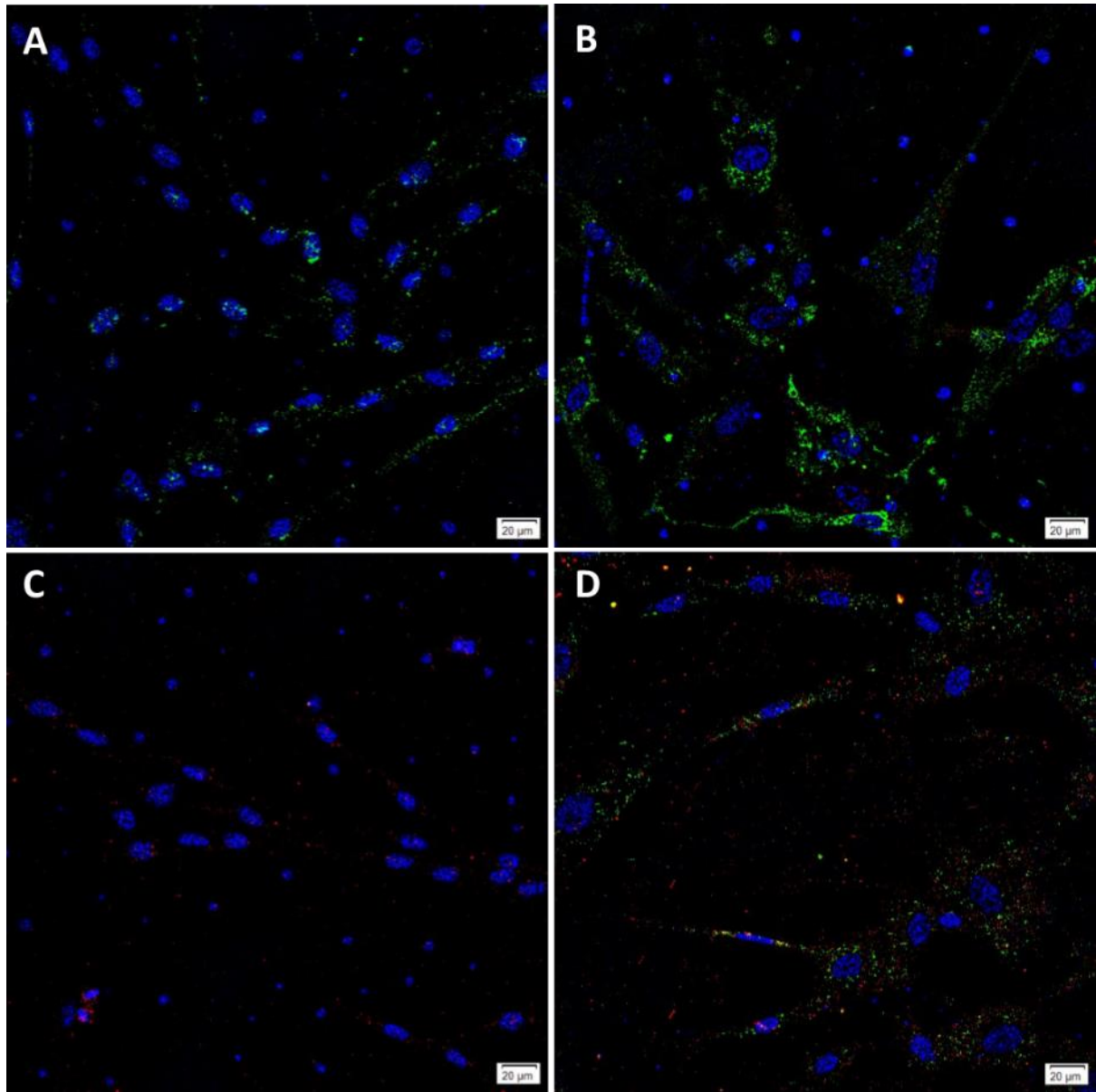
EpCAM^{Low} cells from sample LGCA1 were stained for SSEA4 (green; A) and OCT4 (red; A), and for SOX2 (green; C) and TRA-1-60 (red; C). EpCAM^{High} cells from LGCA1 were stained for SSEA4 (green; B) and OCT4 (red; B), and for SOX2 (green; D) and TRA-1-60 (red; D). Original magnification: 400x; scale bar = 20 μm.

Appendix Figure A5.3: Representative images of ICC staining for sample LGCA2.



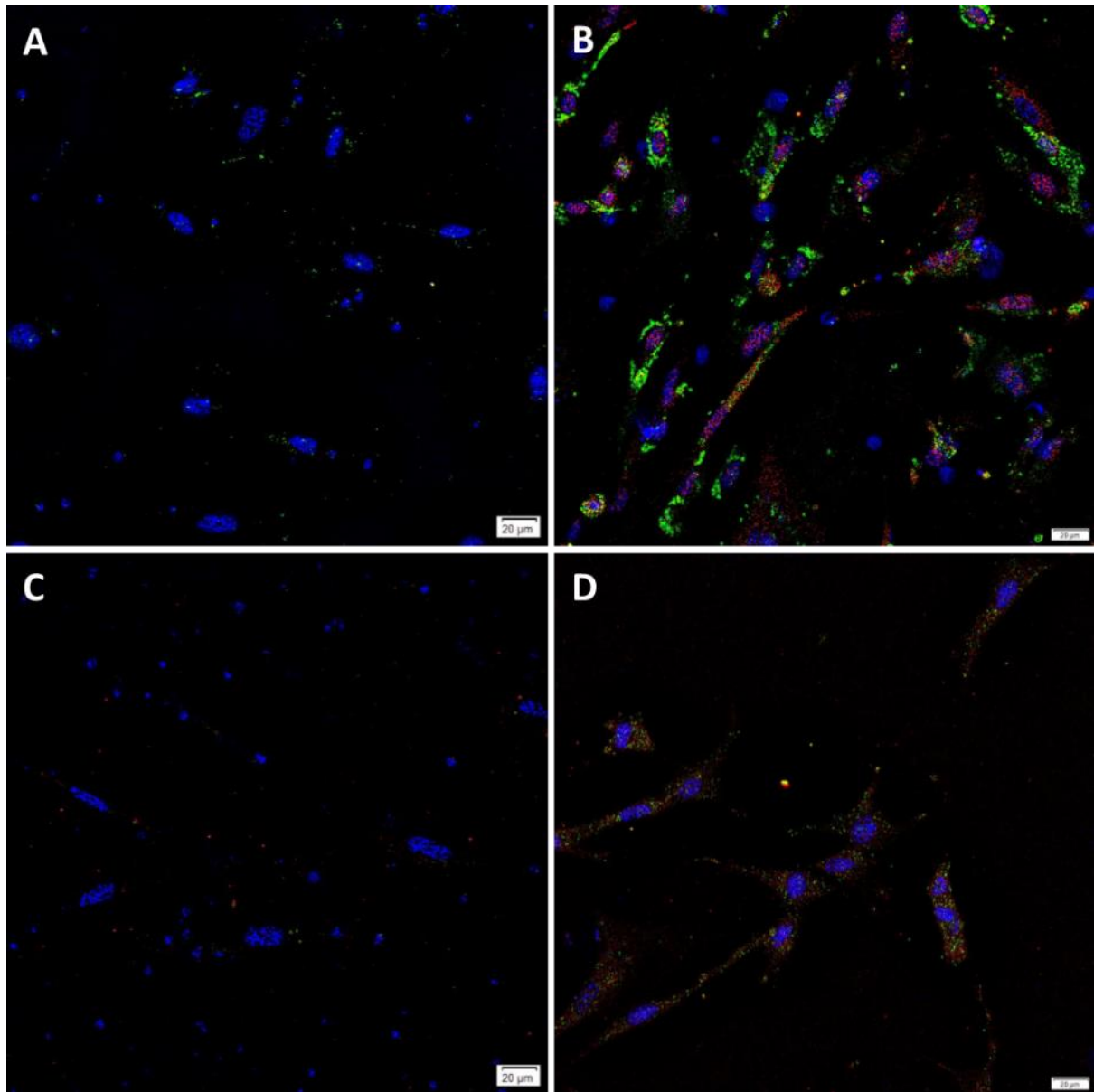
EpCAM^{Low} cells from sample LGCA2 were stained for SSEA4 (green; A) and OCT4 (red; A), and for SOX2 (green; C) and TRA-1-60 (red; C). EpCAM^{High} cells from LGCA2 were stained for SSEA4 (green; B) and OCT4 (red; B), and for SOX2 (green; D) and TRA-1-60 (red; D). Original magnification: 400x; scale bar = 20 μ m.

Appendix Figure A5.4: Representative images of ICC staining for sample LGCA3.



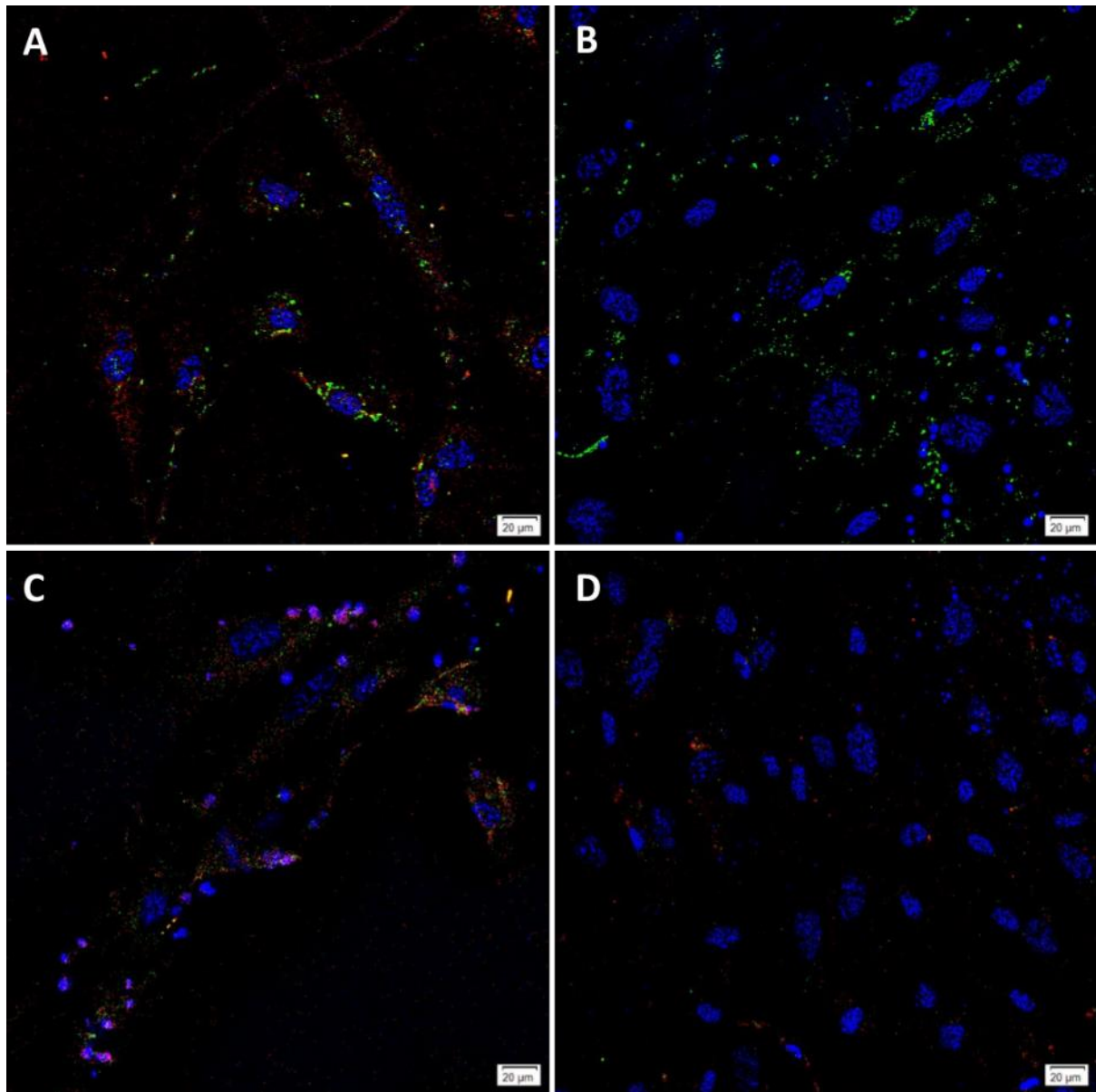
EpCAM^{Low} cells from sample LGCA3 were stained for SSEA4 (green; A) and OCT4 (red; A), and for SOX2 (green; C) and TRA-1-60 (red; C). EpCAM^{High} cells from LGCA3 were stained for SSEA4 (green; B) and OCT4 (red; B), and for SOX2 (green; D) and TRA-1-60 (red; D). Original magnification: 400x; scale bar = 20 μm.

Appendix Figure A5.5: Representative images of ICC staining for sample HGCA1.



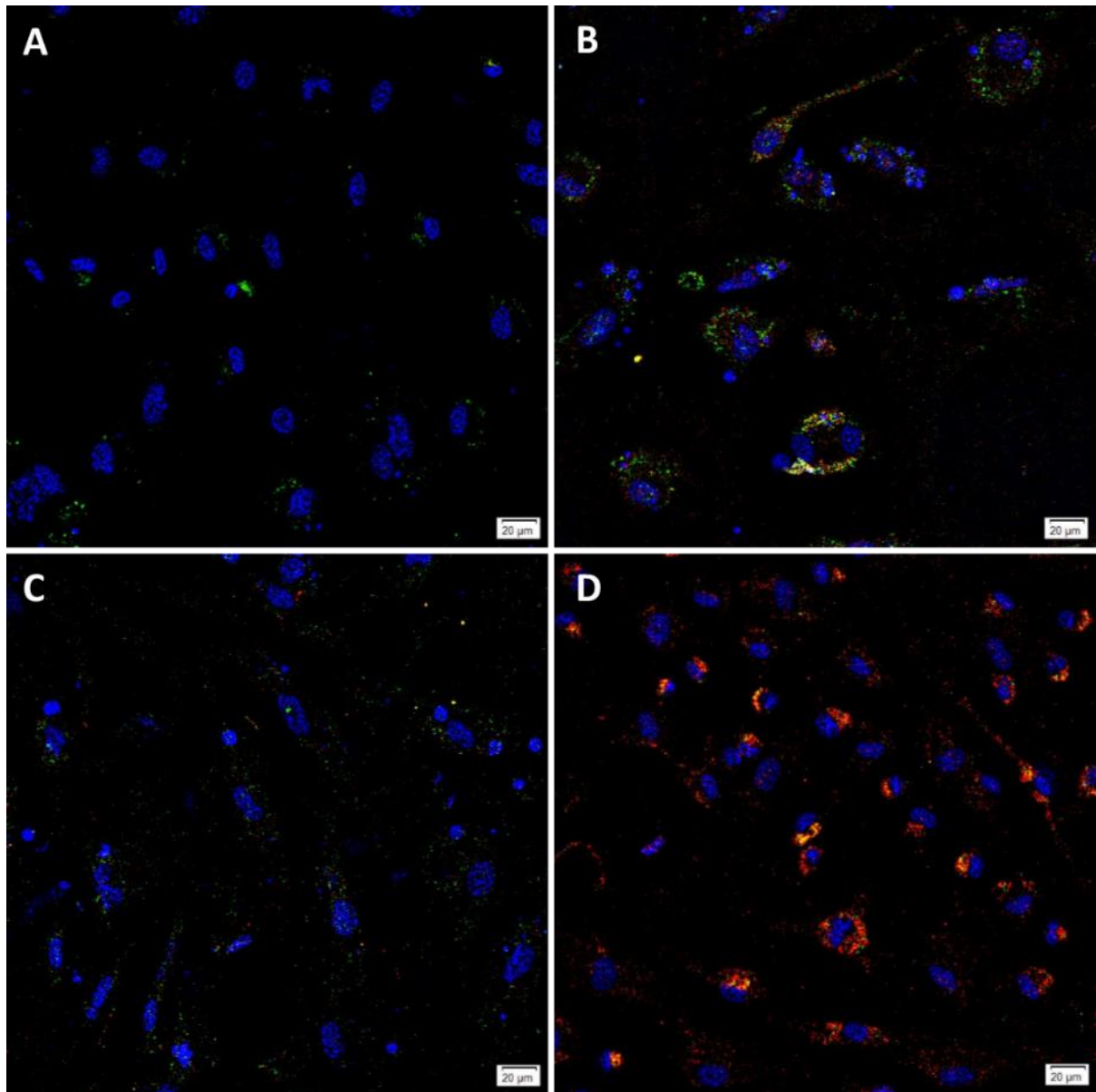
EpCAM^{Low} cells from sample HGCA1 were stained for SSEA4 (green; A) and OCT4 (red; A), and for SOX2 (green; C) and TRA-1-60 (red; C). EpCAM^{High} cells from HGCA1 were stained for SSEA4 (green; B) and OCT4 (red; B), and for SOX2 (green; D) and TRA-1-60 (red; D). Original magnification: 400x; scale bar = 20 μm.

Appendix Figure A5.6: Representative images of ICC staining for sample HGCA2.



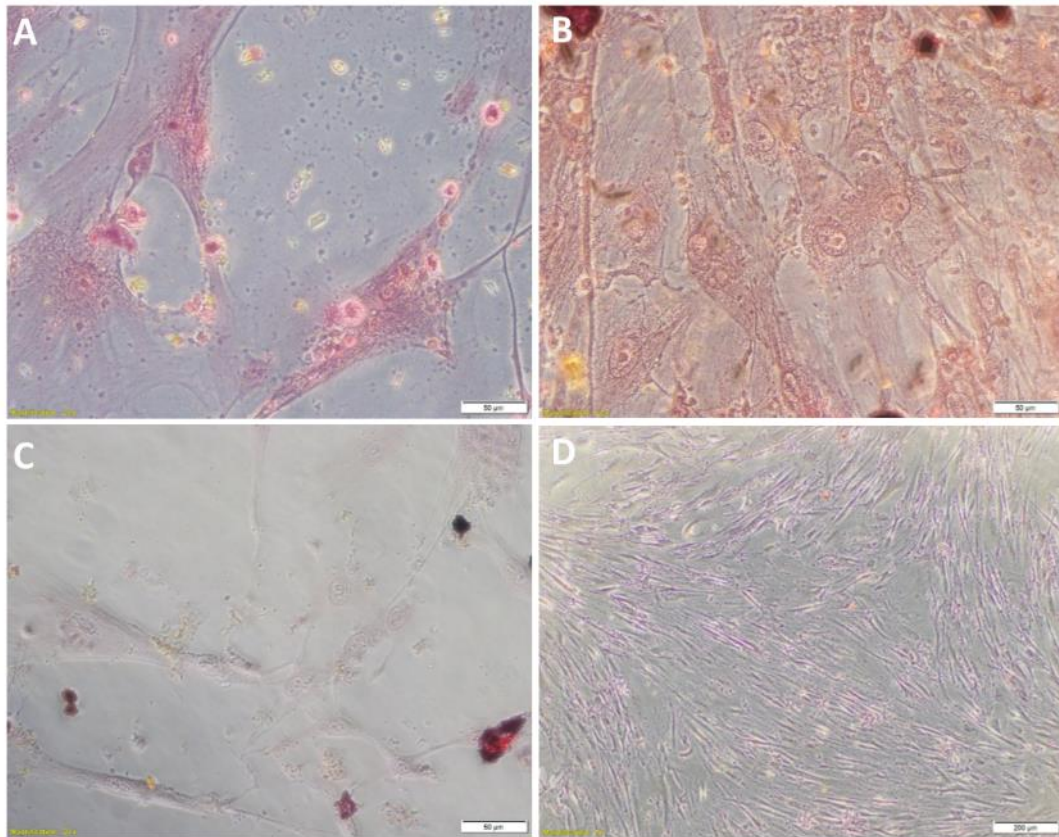
EpCAM^{Low} cells from sample HGCA2 were stained for SSEA4 (green; A) and OCT4 (red; A), and for SOX2 (green; C) and TRA-1-60 (red; C). EpCAM^{High} cells from HGCA2 were stained for SSEA4 (green; B) and OCT4 (red; B), and for SOX2 (green; D) and TRA-1-60 (red; D). Original magnification: 400x; scale bar = 20 μ m.

Appendix Figure A5.7: Representative images of ICC staining for sample HGCA3.



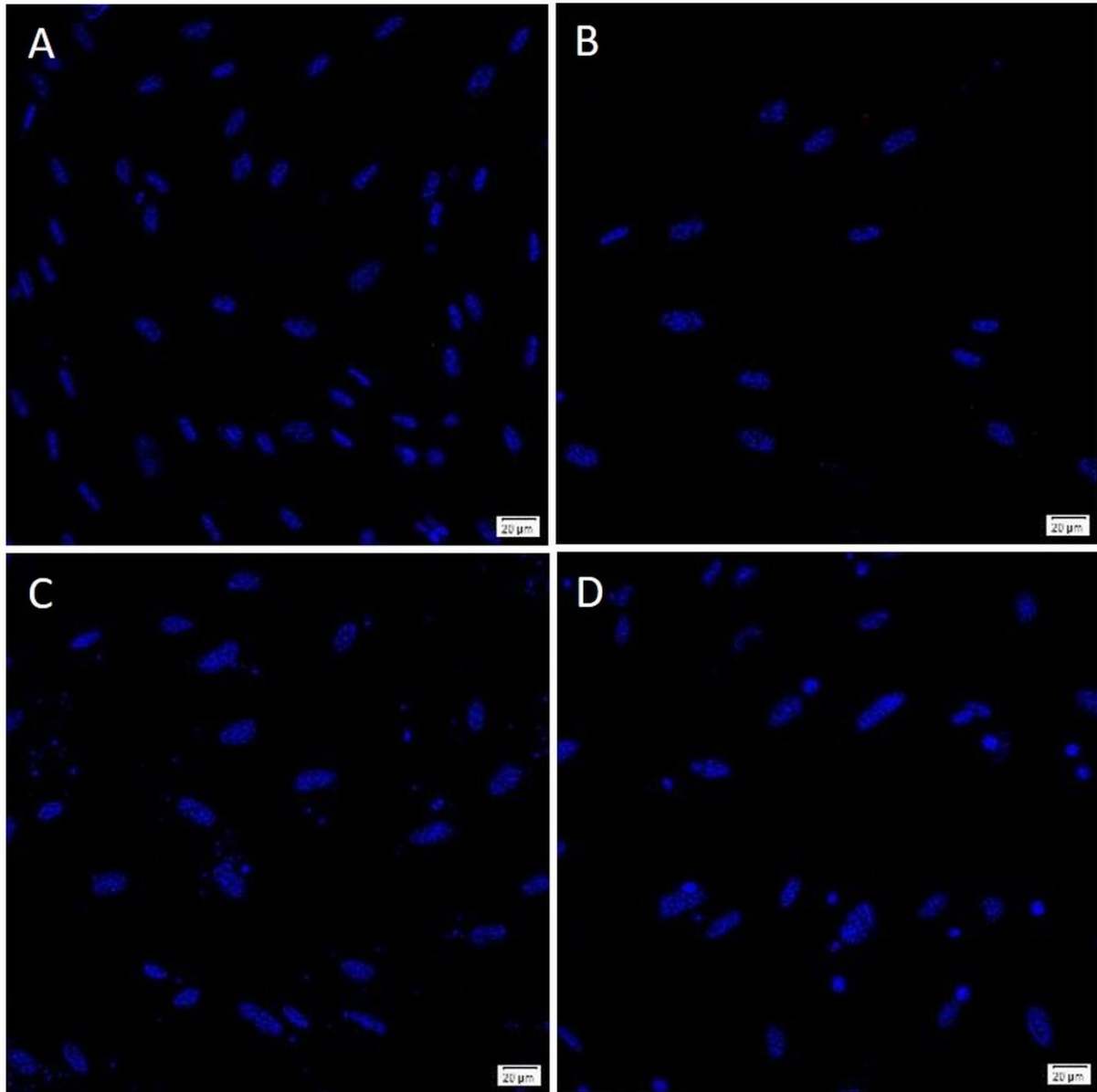
EpCAM^{Low} cells from sample HGCA3 were stained for SSEA4 (green; A) and OCT4 (red; A), and for SOX2 (green; C) and TRA-1-60 (red; C). EpCAM^{High} cells from HGCA3 were stained for SSEA4 (green; B) and OCT4 (red; B), and for SOX2 (green; D) and TRA-1-60 (red; D). Original magnification: 400x; scale bar = 20 μ m.

Appendix Figure A5.10: Mesodermal differentiation negative controls.



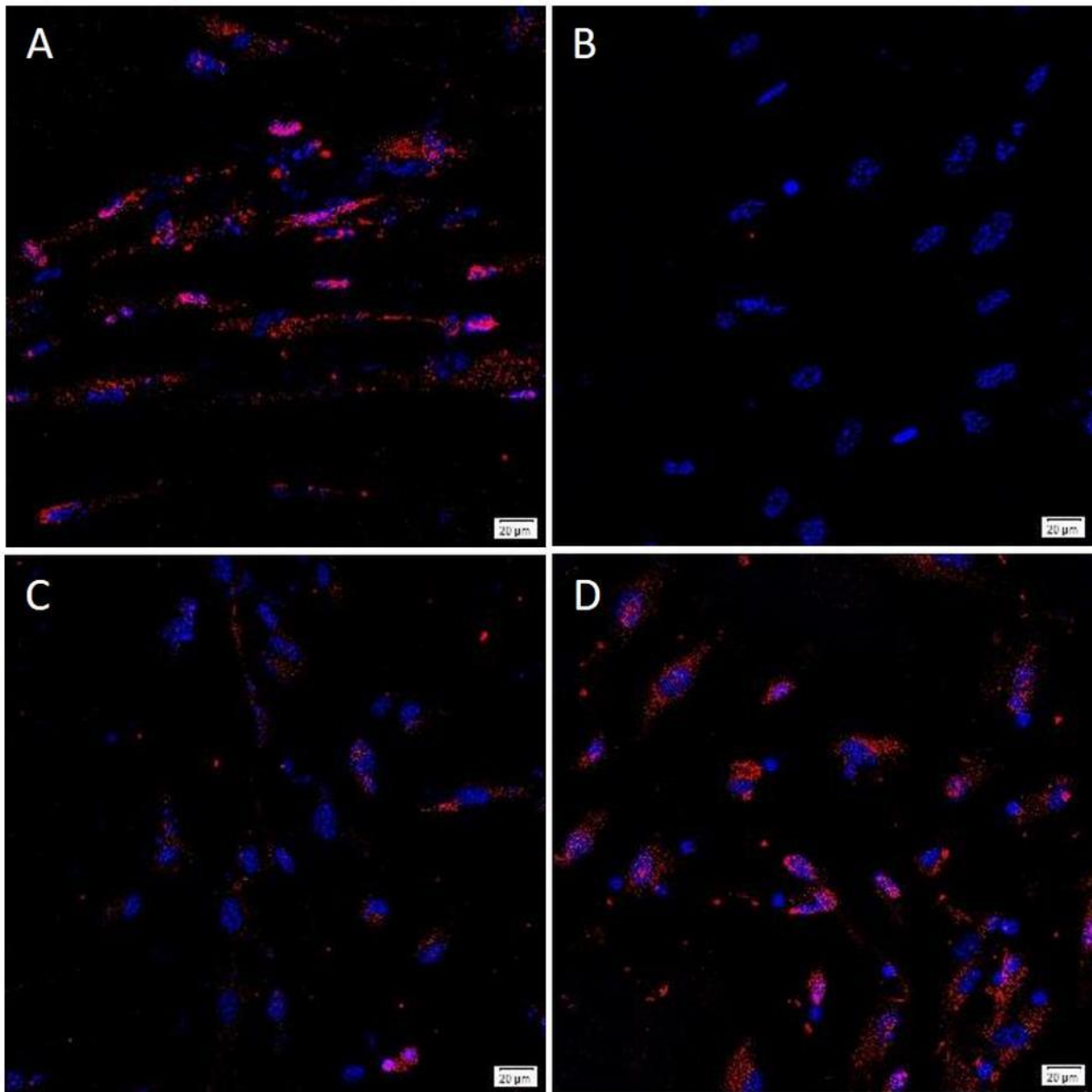
As a negative control, EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA-derived cells and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA-derived cells were grown in regular culture media rather than differentiation media before being exposed to Alizarin Red (pH 4.2). Original magnification: 40x; scale bar = 50 μm.

Appendix Figure A5.11: Endodermal differentiation negative controls.



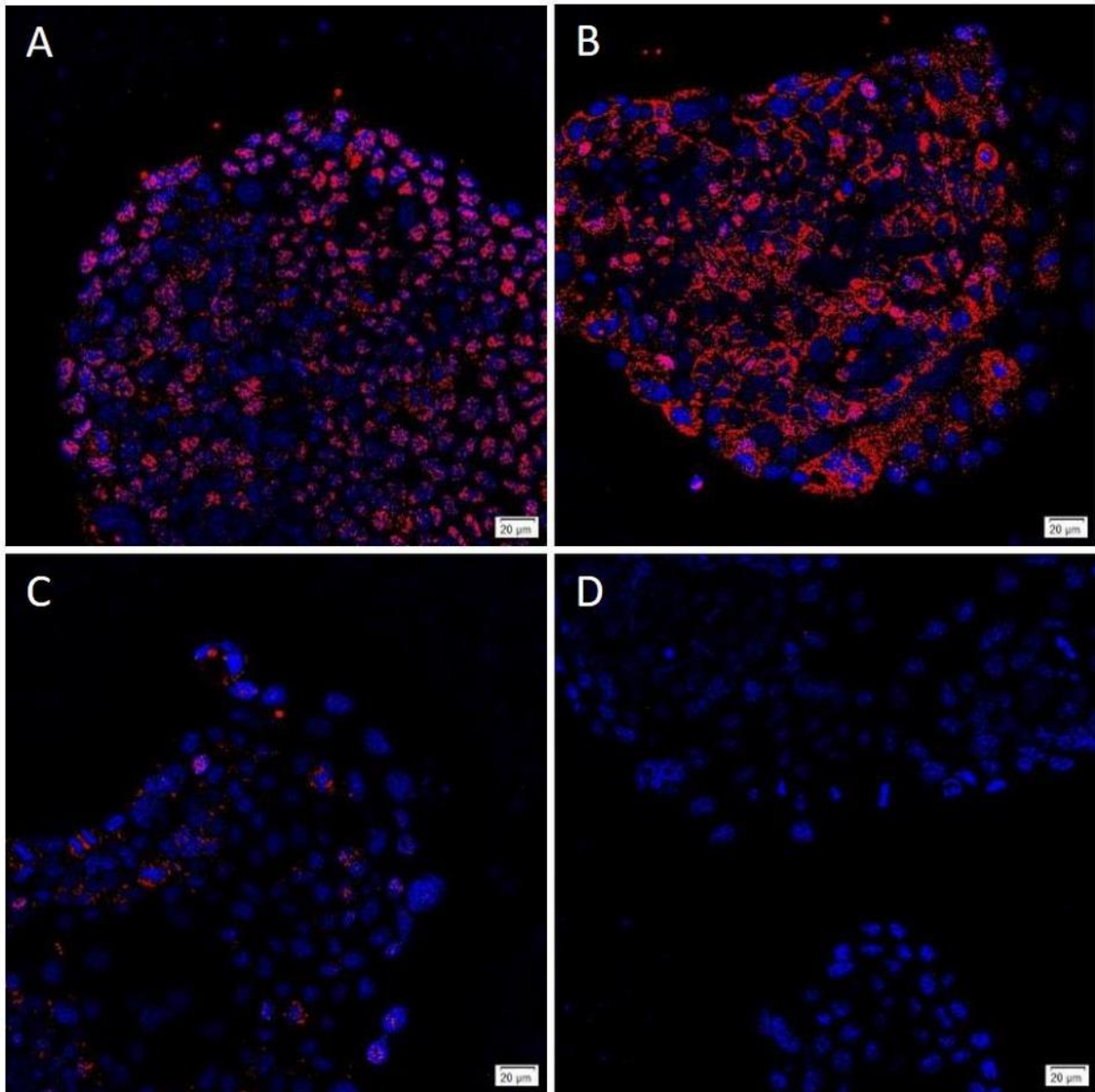
Negative controls were performed for EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA-derived cells and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA-derived cells by omitting the anti-SOX17 primary antibody. Original magnification: 400x; scale bar = 20 µm.

Appendix Figure A5.12: Endodermal differentiation negative controls.



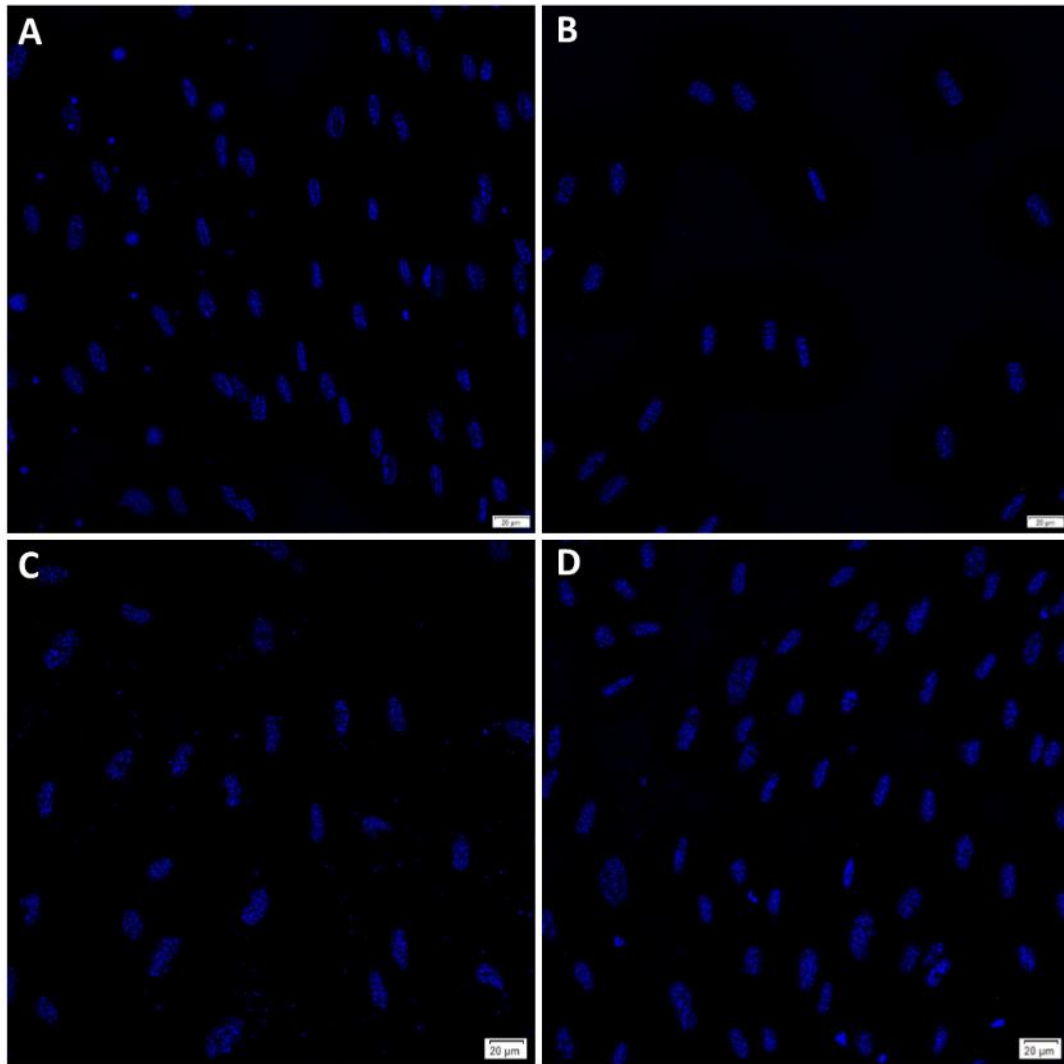
Negative controls were performed for EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA-derived cells and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA-derived cells by growing them in regular culture media rather than differentiation media before being exposed to anti-SOX17. Original magnification: 400x; scale bar = 20 µm.

Appendix Figure A5.13: Endodermal differentiation control cell line.



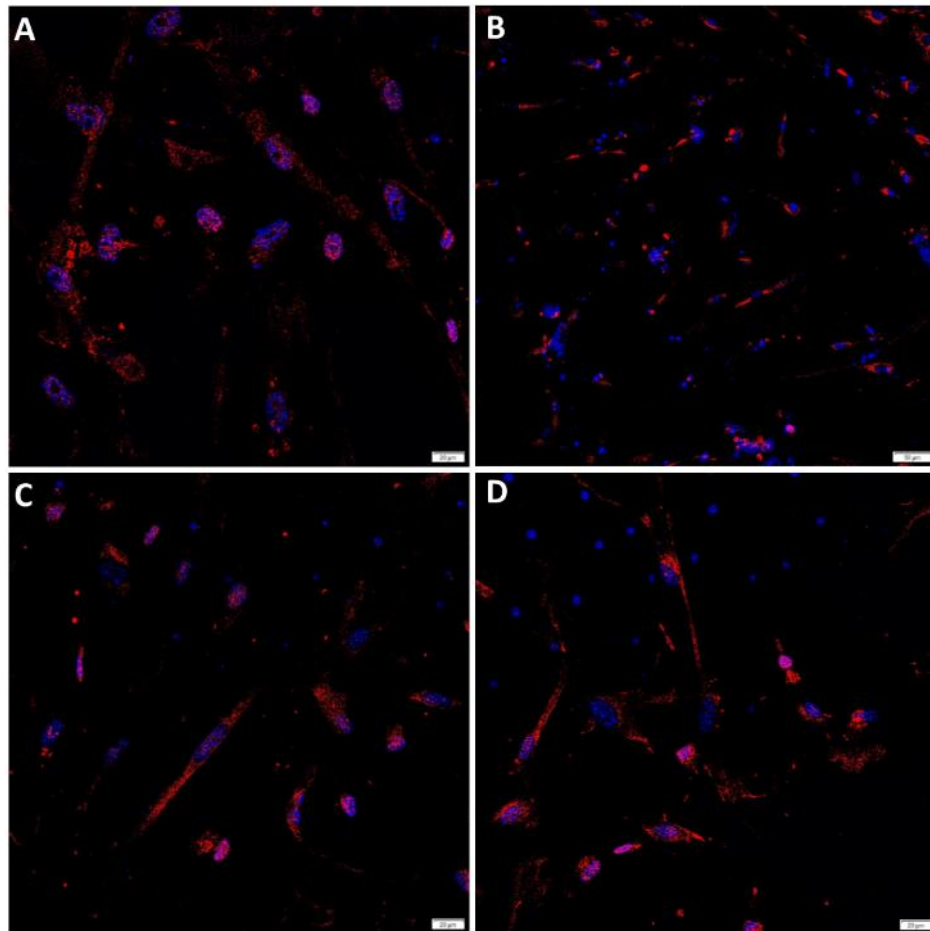
CaCo2 cells were used as a positive control for endodermal differentiation (A). CaCo2 cells were also grown in regular culture media as a control (B, C). A negative control was performed by omitting the anti-SOX17 primary antibody from CaCo2 cell grown in differentiation media (D). Original magnification: 400x; scale bar = 20 µm.

Appendix Figure A5.14: Ectodermal differentiation negative controls.



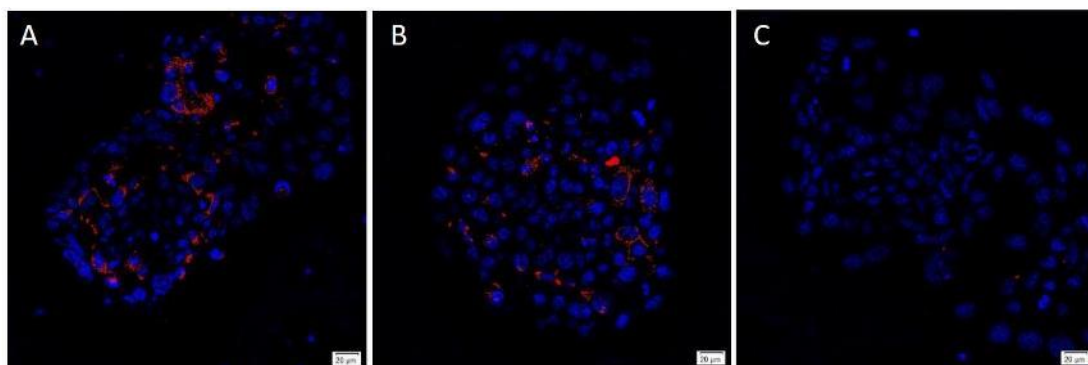
Negative controls were performed for EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA-derived cells and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA-derived cells by omitting the anti-Otx2 primary antibody. Original magnification: 400x; scale bar = 20 µm.

Appendix Figure A5.15: Ectodermal differentiation negative controls.



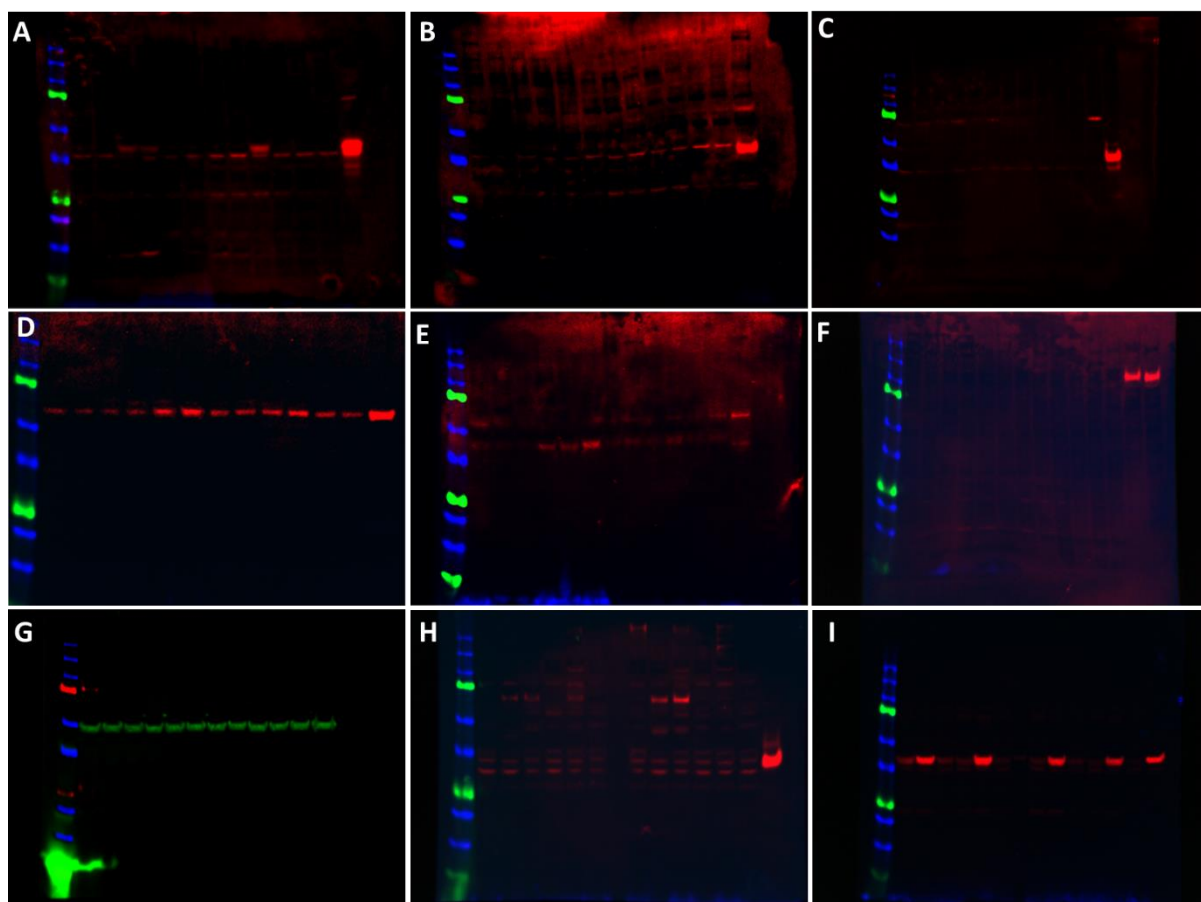
Negative controls were performed for EpCAM^{Low} (A) and EpCAM^{High} (B) LGCA-derived cells and EpCAM^{Low} (C) and EpCAM^{High} (D) HGCA-derived cells by growing them in regular culture media rather than differentiation media before being exposed to anti-Otx2. Original magnification: 400x; scale bar = 20 μ m.

Appendix Figure A5.16: Ectodermal differentiation control cell line.



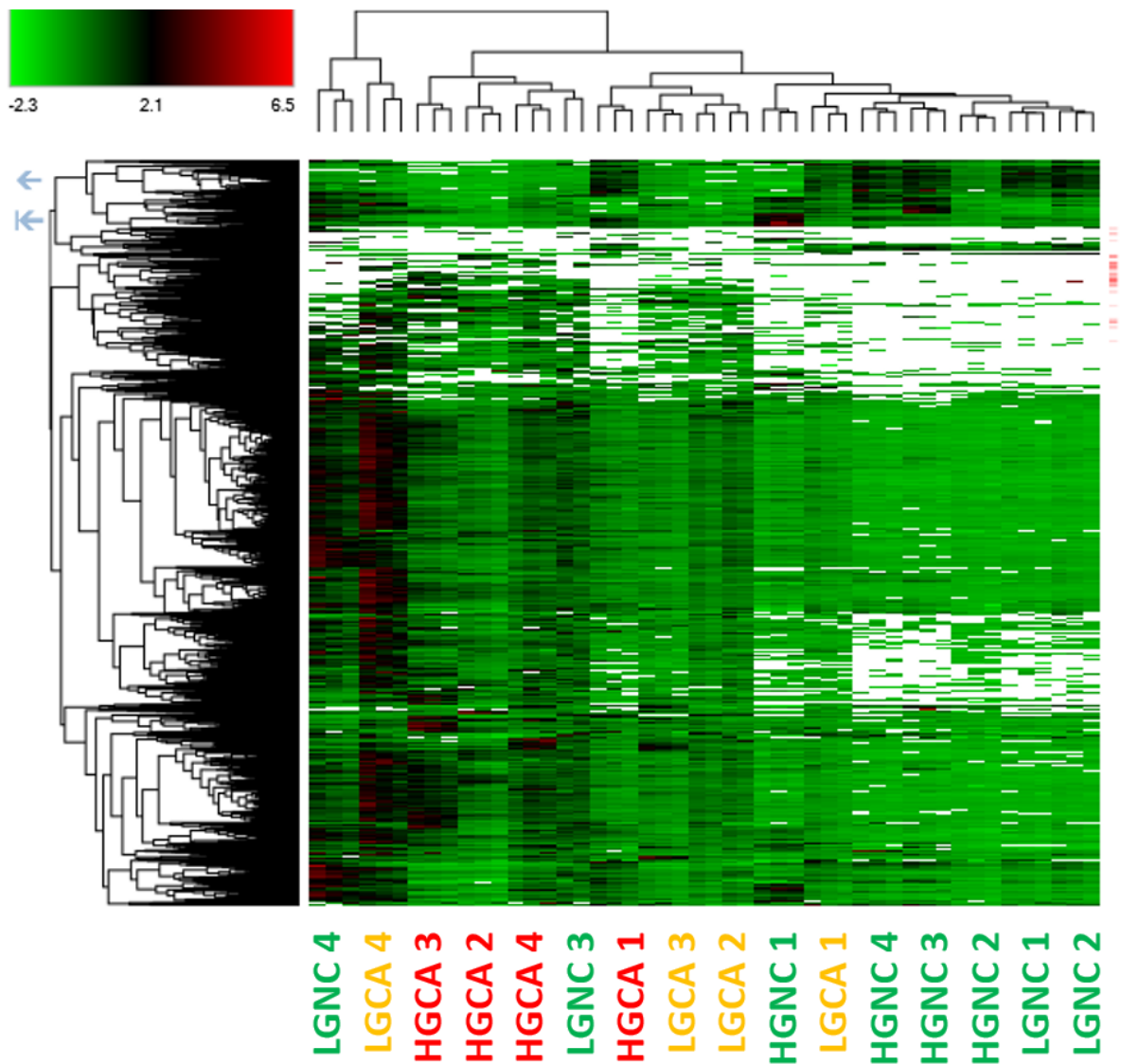
CaCo2 cells were used as a positive control for ectodermal differentiation (A). CaCo2 cells were also grown in regular culture media as a control (B). A negative control was performed by omitting the anti-Otx2 primary antibody from CaCo2 cell grown in differentiation media (D). Original magnification: 400x; scale bar = 20 μ m.

Appendix Figure A5.17: Unmodified western blot images.



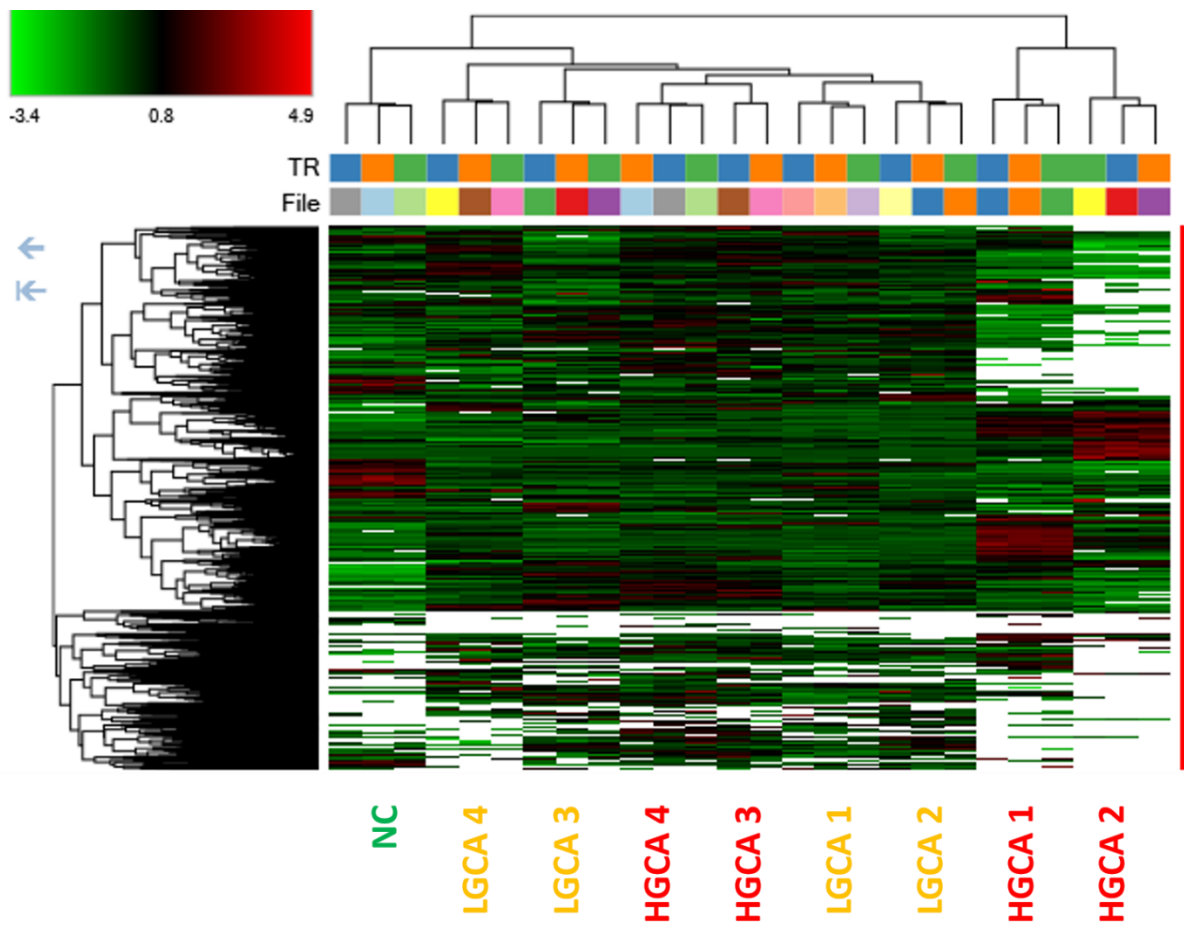
Uncropped images of western blotting membranes before adjusting for background fluorescence, showing OCT4 (A), SOX2 (B), NANOG (C), KLF4 (D), c-MYC (E), CD133 (F), α -tubulin (G), EpCAM (H) and α -SMA (I).

Appendix Fig A6.1: Heat map for individual tissue samples.



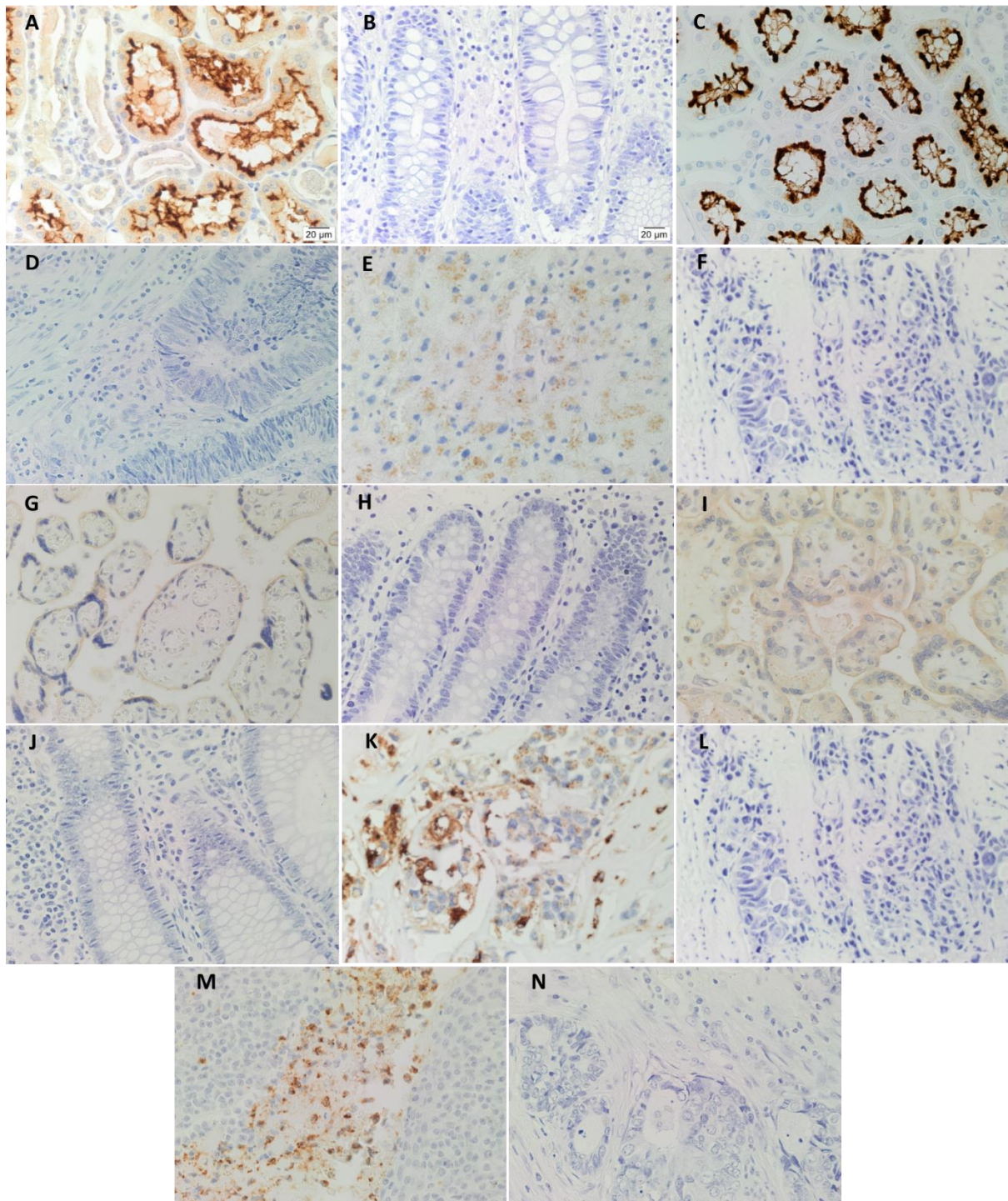
X-axis shows: NC tissues = green, LGCA tissues = yellow, HGCA tissues = red. Within the heat map, relative abundances are shown as: Green = low, Black = moderate, Red = high, White = not detected

Appendix Fig A6.2: Heat map for individual cell lines.



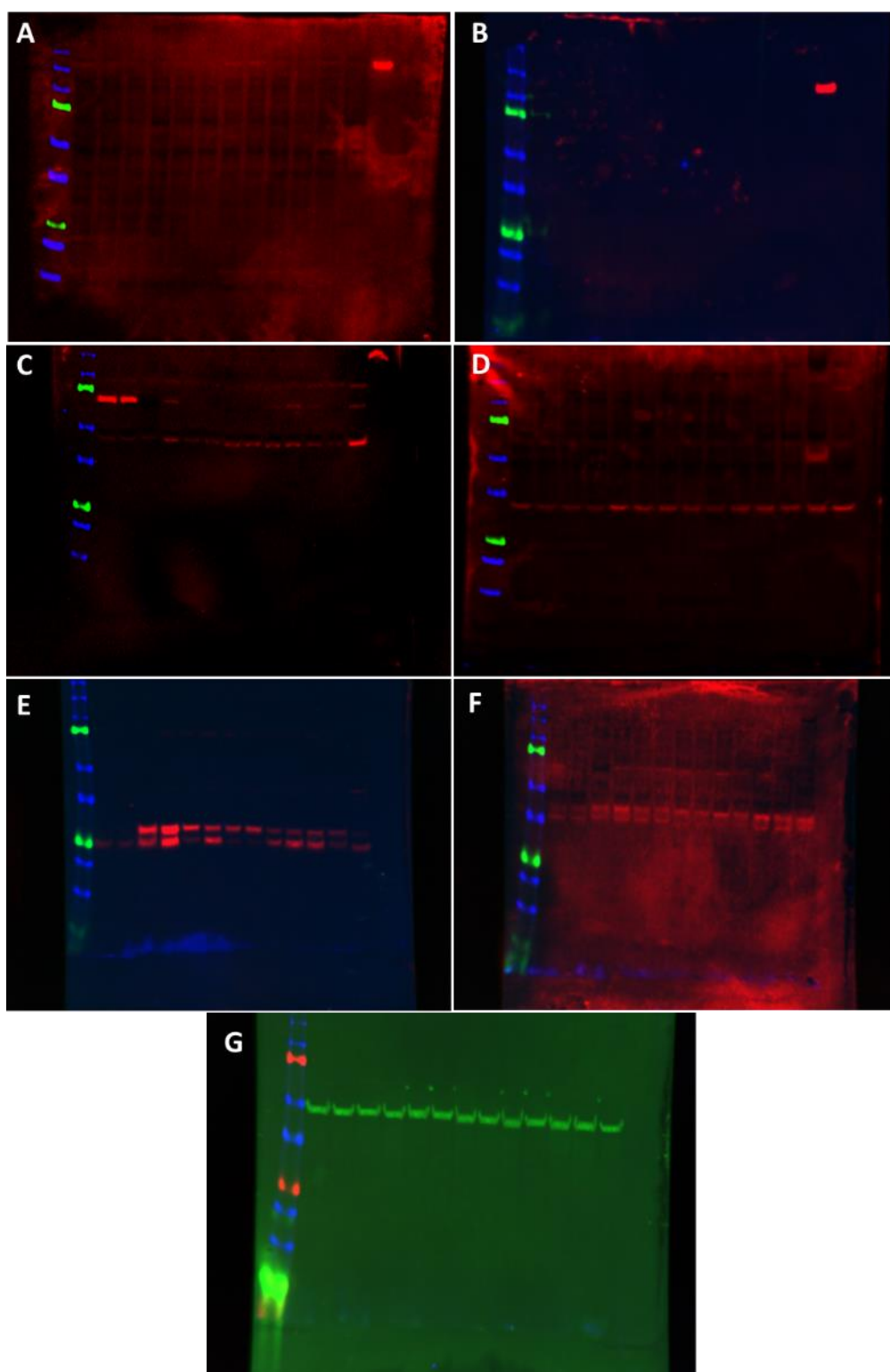
X-axis shows: NC cell line = green, LGCA-derived cell lines = yellow, HGCA-derived cell lines = red. Within the heat map, relative abundances are shown as: Green = low, Black = moderate, Red = high, White = not detected

Appendix Figure A7.1: IHC staining positive and negative controls.



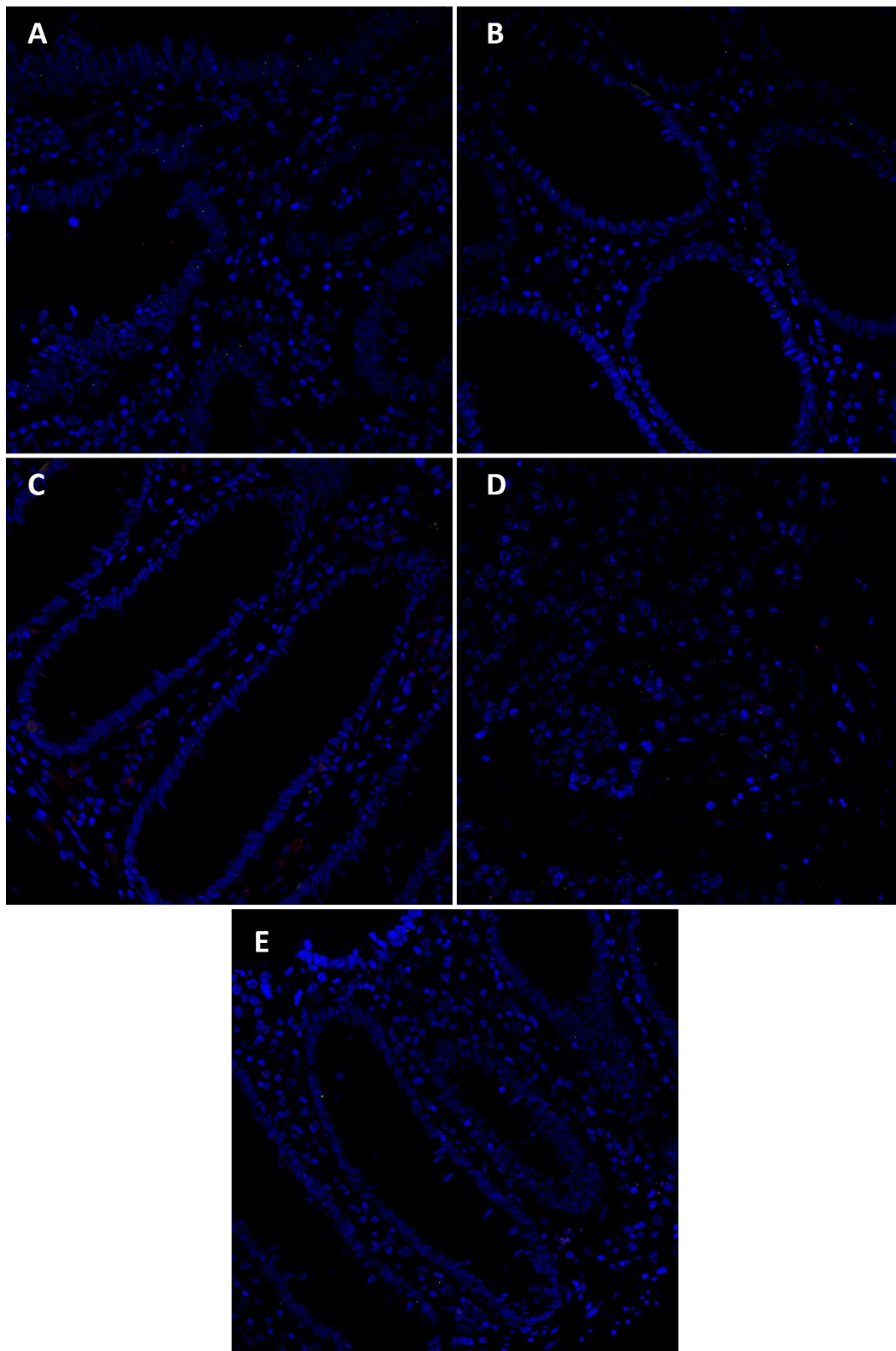
Positive controls for ACE (A; kidney), ACE2 (C; kidney), AT₂R (E; kidney), PRR (G; placenta), Cathepsin B (I; placenta), Cathepsin D (K; breast), and Cathepsin G (M; tonsil), and negative controls for ACE (B), ACE2 (D), AT₂R (F), PRR (H), cathepsin B (J), cathepsin D (L), and cathepsin G (N). Original magnification = 400x.

Appendix Figure A7.2: Unmodified western blot images.



Uncropped images of western blotting membranes before adjusting for background fluorescence for ACE (A), ACE2 (B), AT₂R (C), PRR (D), cathepsin B (E), cathepsin D (F) and α -tubulin (G).

Appendix Figure A7.3: iPSC marker IF staining negative controls.



Images showing the negative control slides from IF staining of OCT4/AT₂R (A), OCT4/cathepsin B (B), OCT4/cathepsin D (C), NANOG/ACE2 (D) and NANOG/AT₂R (E). Negative controls for IF were produced by omitting the primary antibody. Cell nuclei were counter-stained with DAPI (blue). Original magnification: 400x.

Appendix Table A7.1: Supplementary data for RAS component RT-qPCR in tissues and cells

Tissues		ΔCT (tumour)	Average ΔCT (normal)	$\Delta\Delta CT$ (T-N)	$2^{-(\Delta\Delta CT)}$
ACE	LG1	-10.99	-11.06	0.07	1.05
		-11.32		-0.26	0.83
		-10.69		0.37	1.29
	LG2	-10.53	-10.14	-0.39	0.76
		-10.35		-0.21	0.86
		-10.38		-0.24	0.84
	LG3	-12.49	-13.35	0.86	1.82
		-12.47		0.88	1.84
		-12.03		1.32	2.50
	LG4	-10.50	-11.47	0.96	1.95
		-10.21		1.25	2.38
		-10.02		1.44	2.72
	LG5	-9.28	-11.92	2.64	6.23
		-8.76		3.16	8.94
		-8.98		2.94	7.67
	LG6	-9.37	-8.01	-1.37	0.39
		-9.90		-1.90	0.27
		-9.58		-1.58	0.34
	HG1	-14.10	-13.78	-0.32	0.80
		-13.35		0.43	1.35
		-13.60		0.18	1.13
	HG2	-10.40	-11.03	0.64	1.55
		-9.94		1.10	2.14
		-10.48		0.56	1.47
	HG3	-15.25	-14.74	-0.51	0.70
		-15.29		-0.55	0.68
		-15.40		-0.66	0.63
	HG4	-10.48	-9.05	-1.43	0.37
		-11.25		-2.20	0.22
		-10.95		-1.90	0.27
	HG5	-8.05	-9.22	1.17	2.24
		-8.22		1.00	2.00
		-8.42		0.80	1.74
	HG6	-7.59	-8.80	1.21	2.31
		-7.35		1.45	2.73
		-7.28		1.52	2.87
	LG1	-10.98	-11.88	0.90	1.87
		-11.21		0.67	1.59
		-11.44		0.44	1.36
	LG2	-	-9.02	-	0
		-		-	0
		-14.28		-5.26	0.03
	LG3	-	-11.94	-	0
		-		-	0
		-		-	0
	LG4	-	-11.18	-	0
		-14.75		-3.57	0.08
		-		-	0
	LG5	-	-11.56	-	0
		-		-	0
		-		-	0

AT ₂ R	LG6	- - -	-	- - -	- - -
	HG1	- - -	-	- - -	- - -
	HG2	- -13.55 -11.73	-12.31	- -1.24 0.58	0 0.42 1.49
	HG3	-11.58 -11.35 -11.16	-	-33.05 -32.82 -32.63	∞ ∞ ∞
	HG4	-11.96 -12.15 -12.05	-	-40.67 -40.86 -40.76	∞ ∞ ∞
	HG5	-11.63 -10.64 -10.27	-13.32	1.69 2.68 3.05	3.22 6.39 8.26
	HG6	-9.56 -9.98 25.87	-	-35.33 -35.75 ‡	∞ ∞ -
PRR	LG1	-3.62 -3.33 -3.39	-3.55	-0.07 0.22 0.16	0.95 1.16 1.12
	LG2	-3.92 -3.95 -3.69	-2.18	-1.74 -1.77 -1.51	0.30 0.29 0.35
	LG3	-7.89 -7.77 -7.42	-6.07	-1.83 -1.71 -1.36	0.28 0.31 0.39
	LG4	-4.64 -4.70 -4.63	-3.19	-1.45 -1.51 -1.44	0.37 0.35 0.37
	LG5	1.39 1.14 1.42	-5.16	6.55 6.30 6.58	93.70 78.79 95.67
	LG6	-3.28 -3.09 -3.28	-2.70	-0.59 -0.40 -0.59	0.67 0.76 0.67
	HG1	-10.56 -10.58 -10.73	-9.25	-1.31 -1.33 -1.48	0.40 0.40 0.36
	HG2	-2.37 -2.56 -2.27	-1.73	-0.64 -0.83 -0.54	0.64 0.56 0.69
	HG3	-8.34 -7.94 -8.01	-9.73	1.39 1.79 1.72	2.63 3.47 3.30
	HG4	-2.84 -2.81 -2.80	-3.42	0.59 0.62 0.63	1.50 1.53 1.54
	HG5	-2.73 -2.52 -2.85	-2.31	-0.42 -0.21 -0.54	0.75 0.86 0.69
	HG6	-1.41 -1.62	-3.88	2.47 2.26	5.55 4.80

		-1.68		2.20	4.61
Cells		ΔCT (tumour)	Average ΔCT (normal)	$\Delta\Delta CT$ (T-N)	$2^{\Delta(\Delta CT)}$
ACE	LG1+	-7.38	-4.65	-2.73	0.15
		-7.42		-2.77	0.15
		-7.19		-2.54	0.17
	LG1-	-6.11	-3.24	-2.87	0.14
		-6.15		-2.91	0.13
		-6.19		-2.95	0.13
	LG2+	-7.87	-4.65	-3.22	0.11
		-8.09		-3.44	0.09
		-7.90		-3.25	0.11
	LG2-	-7.09	-3.24	-3.85	0.07
		-7.21		-3.97	0.06
		-7.22		-3.98	0.06
	LG3+	-6.70	-4.65	-2.05	0.24
		-6.53		-1.88	0.27
		-6.80		-2.15	0.23
	LG3-	-5.13	-3.24	-1.89	0.27
		-4.97		-1.73	0.30
		-4.76		-1.52	0.35
	HG1+	-6.47	-4.51	-1.95	0.26
		-6.45		-1.93	0.26
		-6.51		-1.99	0.25
	HG1-	-6.01	-3.98	-2.03	0.25
		-6.15		-2.17	0.22
		-6.13		-2.15	0.23
	HG2+	-7.62	-4.51	-3.11	0.12
		-7.57		-3.06	0.12
		-7.56		-3.05	0.12
	HG2-	-6.69	-3.98	-2.71	0.15
		-6.62		-2.64	0.16
		-6.41		-2.43	0.19
	HG3+	-5.87	-4.51	-1.36	0.39
		-5.78		-1.27	0.42
		-5.89		-1.38	0.38
	HG3-	-5.84	-3.98	-1.86	0.28
		-6.01		-2.03	0.24
		-5.96		-1.98	0.25
	LG1+	-	-5.87	-	0
		-		-	0
		-		-	0
	LG1-	-13.30	-5.29	-8.02	0.00
		-		-	0
		-		-	0
	LG2+	-	-5.87	-0.29	-
		-11.68		-0.25	0.02
		-		0.06	-
	LG2-	-12.90	-5.29	-7.62	0.01
		-		-	0
		-12.09		-6.81	0.01
	LG3+	-	-5.87	-	0
		-		-	0
		-15.16		-9.28	0.00
	LG3-	-	-5.29	-	0
		-		-	0

ACE2		-		-	0
		-	-5.59	-	0
		-		-	0
	HG1+	-		-	0
		-	-5.82	3.36	10.29
		-		3.20	9.21
	HG1-	-		3.04	8.24
		-12.28	-5.59	-6.69	0.01
AT ₂ R		-		-	0
		-		-	0
	HG2+	-		-	0
		-	-5.82	-	0
		-		-	0
	HG2-	-13.80		-7.98	0.00
		-	-5.59	-	0
		-		-	0
	HG3+	-		-	0
		-	-5.82	-	0
		-		-	0
	HG3-	-		-	0
		-	-9.22	-	0
		-		-	0
	LG1+	-		-	0
		-	-9.17	-	0
		-		-	0
	LG1-	-		-	0
		-	-9.22	-	0
		-		-	0
	LG2+	-		-	0
		-	-9.17	-	0
		-		-	0
	LG2-	-		-	0
		-	-9.22	-	0
		-		-	0
	LG3+	-		-	0
		-	-9.17	-	0
		-		-	0
	LG3-	-		-	0
		-	-9.26	-	0
		-		-	0
	HG1+	-		-	0
		-	-8.88	-	0
		-		-	0
	HG1-	-		-	0
		-	-9.26	-	0
		-		-	0
	HG2+	-		-	0
		-	-8.88	-	0
		-		-	0
	HG2-	-		-	0
		-	-9.26	-	0
		-		-	0
	HG3+	-		-	0
		-	-8.88	-	0
		-		-	0
	HG3-	-		-	0
	LG1+	0.97	-1.317	2.28	4.86

PRR		0.50 0.75		1.82 2.07	3.53 4.20
	LG1-	0.59 0.52 0.67	-0.983	1.57 1.50 1.65	2.98 2.83 3.14
	LG2+	1.19 1.15 1.25	-1.317	2.50 2.46 2.56	5.66 5.51 5.90
	LG2-	1.16 0.87 1.04	-0.983	2.14 1.85 2.02	4.42 3.62 4.07
	LG3+	0.94 0.83 0.87	-1.317	2.26 2.15 2.19	4.78 4.43 4.55
	LG3-	1.28 1.10 1.33	-0.983	2.26 2.08 2.31	4.78 4.22 4.95
	HG1+	1.93 1.73 2.01	-0.202	2.13 1.93 2.21	4.38 3.81 4.63
	HG1-	2.06 1.88 1.91	-0.173	2.23 2.05 2.08	4.68 4.13 4.22
	HG2+	1.67 1.69 1.45	-0.202	1.88 1.90 1.66	3.67 3.72 3.15
	HG2-	1.47 1.33 1.31	-0.173	1.64 1.50 1.48	3.13 2.84 2.80
	HG3+	1.49 1.43 1.48	-0.202	1.69 1.63 1.68	3.23 3.10 3.21
	HG3-	1.73 1.45 1.28	-0.173	1.90 1.62 1.45	3.74 3.08 2.73
CTSB	LG1+	0.28 0.44 0.28	-1.20	1.48 1.64 1.48	2.79 3.12 2.79
	LG1-	-0.20 -0.10 -0.19	-1.89	1.69 1.79 1.70	3.22 3.45 3.24
	LG2+	1.84 1.75 2.05	-1.20	3.04 2.95 3.25	8.20 7.70 9.48
	LG2-	1.80 1.77 1.52	-1.89	3.69 3.66 3.41	12.90 12.63 10.62
	LG3+	2.50 2.24 2.22	-1.20	3.70 3.44 3.42	12.98 10.84 10.69
	LG3-	2.98 2.94 2.61	-1.89	4.87 4.83 4.50	29.15 28.36 22.56
	HG1+	1.01 1.10 1.16	-1.16	2.17 2.26 2.32	4.49 4.78 4.98

CTSD	HG1-	0.76	-1.48	2.25	4.74
		0.79		2.28	4.84
		0.88		2.37	5.15
	HG2+	1.27	-1.16	2.43	5.40
		1.17		2.33	5.03
		1.31		2.47	5.55
	HG2-	0.81	-1.48	2.29	4.91
		0.79		2.27	4.84
		0.94		2.42	5.37
	HG3+	1.61	-1.16	2.77	6.81
		1.39		2.55	5.85
		1.32		2.48	5.57
	HG3-	1.40	-1.48	2.88	7.36
		1.36		2.84	7.15
		1.48		2.96	7.78
CTSD	LG1+	-2.46	-3.40	0.94	1.92
		-2.36		1.04	2.06
		-2.33		1.07	2.10
	LG1-	-1.91	-2.36	0.45	1.36
		-1.66		0.70	1.62
		-1.47		0.89	1.85
	LG2+	-0.23	-3.40	3.17	9.01
		-0.35		3.05	8.29
		0.02		3.42	10.72
	LG2-	0.81	-2.36	3.17	8.97
		0.62		2.98	7.86
		0.74		3.10	8.54
	LG3+	-1.99	-3.40	1.42	2.67
		-2.00		1.41	2.65
		-2.00		1.41	2.65
	LG3-	-2.10	-2.36	0.25	1.19
		-2.16		0.19	1.14
		-2.18		0.17	1.13
	HG1+	-2.05	-2.74	0.68	1.61
		-2.20		0.53	1.45
		-2.13		0.60	1.52
	HG1-	-2.47	-2.87	0.40	1.32
		-2.54		0.33	1.26
		-2.24		0.63	1.55
	HG2+	-1.93	-2.74	0.80	1.74
		-1.94		0.79	1.73
		-1.89		0.84	1.79
	HG2-	-	-2.87	-	0
		-2.33		0.55	1.46
		-2.33		0.55	1.46
	HG3+	-2.45	-2.74	0.29	1.22
		-2.21		0.53	1.44
		-2.41		0.33	1.26
	HG3-	-2.07	-2.87	0.81	1.75
		-2.26		0.62	1.53
		-2.23		0.65	1.57

RT-qPCR data showing expression of iPSC markers *ACE*, *AT₂R* and *PRR* in CA tissues, and *ACE*, *ACE2*, *AT₂R*, *PRR*, *CTSB* (cathepsin B) and *CTSD* (cathepsin D) in CA-derived cells. Δ CT values calculated by comparing the gene of interest to 2 housekeeper genes (*UBC* and *PUM1*). $\Delta\Delta$ CT values calculated by

comparing CA tissues to their patient-matched NC samples, and CA-derived cells to the pooled data from 4 NC tissues. $\Delta\Delta CT$ values used to calculate fold changes using the equation $2^{(\Delta\Delta CT)}$.

Appendix B: List of Publications

The following manuscripts were published during the course of this project. They have been updated as new data came to light, modified with further editing, and formatted to fit within the thesis as a whole. Permissions for use of the review paper published in *J Clin Path* (2017) have been obtained and attached to the Statement of Authorship and Thesis Length for Doctoral Candidates. The two papers published in *PLoS One* (2019 & 2020) are open-access and do not require further permissions to use in this thesis.

Munro, M.J., Wickremesekera, S.K., Peng, L., Tan, S.T., & Itinteang, T. (2017). Cancer stem cells in colorectal cancer: a review. *J Clin Path*, 71(2), 110-116. doi:10.1136/jclinpath-2017-204739. Epub 2017 Sep 23.

Munro, M.J., Wickremesekera, S.K., Peng, L., Marsh, R.W., Itinteang, T., & Tan, S.T. (2019). Cancer stem cell subpopulations in primary colon adenocarcinoma. *PLoS One*, 14(9). doi:10.1371/journal.pone.0221963. 2019 Sep 6.

Munro, M.J., Peng, L., Wickremesekera, S.K., & Tan, S.T. (2020). Colon adenocarcinoma-derived cells that express induced-pluripotent stem cell markers possess stem cell function. *PLoS One*, 15(5). doi:10.1371/journal.pone.0232934. 2020 May 19.