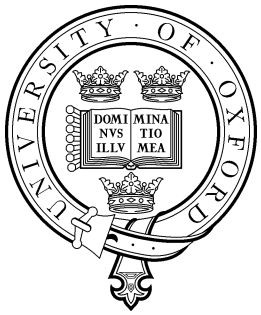
Driver Genes for Metastasis of Colorectal Cancer



Lennard Lee
Queen's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hiliary 2017

This thesis is dedicated
to Ta Ku

Acknowledgements

I very am thankful to Professor Ian Tomlinson for welcoming me into his lab and for giving me the time and freedom to make this thesis into what it is.

I am also extremely grateful for my mum, dad and Julian for their ongoing support and encouragement. I would like to thank Connor, Sujata, Annie, Luke, Raquel, Hayley, Chiara for sharing their time, valuable techniques and skills. A thesis is never possible without people with kind words and a keen ear, and for this Tom, Danny, Anna, Katie, Alicia and David have been most helpful.

A final special mention must go to Chris for his continual support and encouragement.

Abstract

Cancer metastasis is the principal cause of death in patients with solid organ tumours. Once cancer has spread, it is generally considered incurable. This spread of cells from the primary tumour to distant sites (metastasis) is a complex process and this DPhil focuses on colorectal cancer (CRC) metastasis. It has three aims; the development of a new CRC metastasis model, the completion of a whole-genome screen for drivers of metastasis and the development of a greater understanding of the molecular drivers of metastasis through use of human lymph node metastasis (LNM) samples.

The new CRC metastasis model is a mouse model utilising endoscope-guided orthotopic transplantation of tumour cells in combination with in-vivo imaging. This model has allowed the analysis of the role of the differentiation factor, Serum/Glucocorticoid regulated Kinase 1 (*SGK1*) in preventing metastasis and tumorigenesis. I provide preliminary data suggesting that the inhibition of metastasis by re-expression of *SGK1* is associated with the down-regulation of the transcription factor Avian Myelocytomatosis Viral Oncogene Homolog (*c-MYC*), and low *SGK1* is a poor prognosis biomarker identifying CRC patients with high risk of disease recurrence/death.

A genome-wide clustered regularly interspaced short palindromic repeats (CRISPR)/cas9 screen was then performed to identify other genes that play important roles in CRC metastasis. The screen involved the orthotopic transplantation of the MC38 line in a mouse model of metastasis and subsequent analysis using targeted next generation sequencing. This approach identified the pioneer transcription factor, Forkhead Box F1 (*FOXF1*) as an important regulator of metastasis. I demonstrate that *FOXF1* is down-regulated in human metastatic lesions, including LNM, liver and lung metastases. Metastasis driven by low *FOXF1* is mediated through Mammalian Target of Rapamycin Complex (MTORC) signalling through transcriptional control of Regulatory Associated Protein of MTOR Complex 1 (RAPTOR).

Finally, I performed an analysis of 69 stage III paired CRC tumour-LNM human specimens. Metastasis in this cohort was mediated through a num-

ber of important genes and gene networks, including C-X-C chemokine receptor type 4 (*CXCR4*) and Kirsten Rat Sarcoma Viral Oncogene (KRAS) signalling. Non-hierarchical clustering was performed and identified two distinct subsets of LNM, LNMS1 and LNMS2 each conferring different prognoses. I have also identified different immune profiles within the "*Lymphoid ecological niche*" using gene expression profiling. This has enabled me to demonstrate an association between a patient's immune profile, particularly activation of CD8 cells and KRAS signalling within the tumour, with improved disease free survival in stage III CRC patients.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the last few decades, there has been a shift in the demographic of the population of the United Kingdom. The Office of National Statistics has highlighted the rise in the age of the UK population with the proportion of children falling from 25% in 1974 to less than 20% in 2004 and the proportion of those aged 90 years and older increasing by four-fold since the 1980s. The UK has an "ageing population" [2] which has a significant impact on UK healthcare. Despite only 357,000 cancers diagnosed in 2014, expenditure on cancers and tumours make up nearly a fifth of UK health expenditure and is the third largest healthcare disease expenditure with year on year increases[3].

A small number of cancer types account for the majority of cancer diagnoses and these include breast, prostate, lung and colorectal cancers (CRC). Since the late 1970s, incidence rates of CRC have increased by 14%, with a 19% rise observed in men[4]. Most CRC cases are diagnosed in people aged 70 years and over and many are still diagnosed at a late stage[5 6 7].

1.0.1 Current management strategies for CRC

For patients diagnosed with CRC treatment is dependent primarily on TNM staging. The most effective form of treatment remains surgical resection. There are five stages of CRC with stages I-II categorised by depth of tumour invasion into the bowel lining. Stages III/IV represent metastatic cancers, stage III when the cancer has metastasised to regional lymph nodes (regional metastasis) and stage IV when there is spread of the cancer to distant organs (distant metastasis). For patients with stage I-III tumours, surgical resection of the tumour gives the only chance of long-term remission. For patients with stage IV cancers, resection is offered for those presenting with bowel obstruction or those with solitary metastatic lesions (oligometastasis). Once multiple metastases have developed, in general, patients are treated palliatively.

Chemotherapy is given routinely to patients with stage III CRC in an adjuvant setting to prevent relapse/recurrence after surgery. It has remained largely unchanged since 2004 and consists of the use of Oxaliplatin, Leucovorin and 5-FU, and this adjuvant chemotherapy increases five-year survival from 64% to 71%. The role of chemotherapy for patients with stage 2 CRC is uncertain and even for the highest risk patients, 15-30 patients need to be treated to prevent one recurrence or death[8]. Chemotherapy for stage IV CRC is given with palliative intent to prolong survival.

The outlook of patients diagnosed with CRC is likely to improve in the future with the introduction of bowel cancer screening and earlier diagnosis and treatment[9]. However, improvements in the treatment of CRC may be achieved through a greater understanding of tumour biology, particularly in the area of cancer metastasis which remains the leading cause of death in CRC patients.

1.0.2 Precursor lesions to CRC

An adenoma/polyp is a benign, but pre-malignant growth of epithelial tissue in the colon. Macroscopically, they are distinguished by their growth pattern, with hyperplastic polyp forming a curved elevation on the colon lining and an adenoma forming a finger-like growth. Aberrant Wnt/β-catenin signalling following loss of the tumour suppressor Adenomatous Polyposis Coli (*APC)* initiates colon adenoma/polyp formation[10 11]. *APC* has been historically termed the gate-keeper gene for CRC. It is a protein that normally binds glycogen synthases kinase 3 (GSK3) and AXIN to form a destruction complex. This complex binds to β-catenin and targets the protein for degradation and thus prevents its function as a transcription factor driving cell proliferation.

Histologically, adenomas can be subclassified further into tubular, tubulovillous, villous or serrated adenomas. Villous and serrated adenomas have the highest risk of the development into CRC. The traditional theory asserts that as an adenoma progressed from early, to intermediate, to late stage lesions, mutations accumulate in a step-wise process, with *KRAS* occurring early and *SMAD4* mutations occurring late, the so termed *Vogelstein progression*. However, this progression model is likely to be an over-simplification. Different types of adenomas have different mutation spectrums; serrated lesions and hyperplastic lesions commonly have *BRAF* mutations[12 13], villous adenomas commonly have *KRAS* gene mutations[14], and *PIK3CA* mutations are commonly identified in tubular adenomas. These different premalignant lesions have different drivers and are likely to give rise to the different cancer subtypes.

1.0.3 CRC tumorigenesis

Tumorigenesis is the transformation of a benign precursor lesion to carcinoma. It is principally diagnosed on histological sections with the aid of cross-sectional imaging.

Specific phenotypic cellular changes are seen in carcinomas and severely dysplastic adenomas, including larger nuclei, prominent nucleoli and an increase of mitosis. Molecularly, carcinomas display certain cellular features and these have been termed the *"Hallmarks of cancer"*. They include sustaining proliferative capacity, evading growth suppression, resisting apoptosis, enabling replicative immortality, inducing angiogenesis, reprogramming energy metabolism, evading immune destruction, and activating invasion and metastasis[15].

One gene stands out as being of crucial importance in CRC tumorigenesis, Adenomatous polyposis coli (APC). It is the most commonly mutated gene, present in 81% of tumours. Other genes are also frequently mutated in CRC and large-scale sequencing efforts have identified 32 other frequently somatically mutated genes[16]. Many of these mutations are found preferentially in different subtypes of cancers, for example, *ACVR2A* is a receptor that is a transmembrane kinase protein that regulates cell proliferation and migration and is most frequently observed in hypermutated cancers (63%).

Mutation analysis is one method of understanding the features of CRC, however, there are a number of complementary molecular techniques that have aided our understanding. Very early studies have noted that CRC exhibits *aneuploidy*, the presence of an abnormal chromosomal structure. Approximately 70% of CRC exhibit this phenomenon and CRCs may be categorised as CIN +ve or CIN -ve[17]. These alterations in chromosomal structure lead to loss or gain of gene allele(s), leading to "somatic copy number alterations" (SCNA) and these have been hypothesised to drive cancer evolution through the loss or gain of oncogenic or tumour suppressor genes. An alternative method of categorising CRC is by the absence or presence of CpG island methylator phenotype (CIMP) or microsatellite instability (MSI), and

these categories overlap. CIMP is defined by hypermethylation of CpG islands of genes and MSI through alterations in size of nucleotide repeat sequences. MSI commonly arises through defects in mismatch repair proteins (e.g. *MSH2, MLH1, MSH6*) and in sporadic CRC, *MLH1* loss is frequently observed, arising from promoter hypermethylation[18 19]. Whilst these methods of categorising tumours are widespread and established techniques, a new method of classifying tumours has been recently developed and is increasing in popularity.

Gene expression profiling (GEP) allows a surrogate measurement of the activity of thousands of genes to create a global picture of cellular function through the quantification of mRNA transcripts. It enables the analysis of differentially expressed genes/gene networks, and has identified that CRCs have dysregulation of *WNT/MAPK/PI3K/TGFb/MYC* and *TP53* signalling. Alterations in Wnt signalling is found in 93% of all tumours, usually through bilallelic inactivation of APC. Nearly 100% of tumours have changes in *MYC* transcription[16].

In the most recent analysis of 4,151 CRC patients, GEP has identified five main CRC subtypes, each with different gene signalling drivers. CMS1 is characterised by immune infiltration and activation (CMS1-"MSI/immune"), CMS2 characterised by WNT and c-MYC activation (CMS2-"Canonical"), CMS3 characterised by metabolic deregulation (CMS3-"Metabolic"), CMS4 characterised by TGFb activation and stromal response (CMS4-"Mesenchymal") and CMS-unclassified with no special features[20]. An important point is that these subsets are not solely determined by the underlying tumour biology but also by the presence of host stromal (CMS4) or immune cells (CMS1).

A final layer of complexity to an understanding of CRC is that of intra-tumour het-

erogeneity. This is the observation that tumour clones exhibit different mutations and gene alterations based on their location within the tumour. Sequencing performed at different locations in renal tumours has identified contrasting and divergent mutations and GEP[21] and this is very likely to be observed in CRC too.

1.0.4 Metastasis of CRC

When cancers metastasise, they do so to specific organs; for example, breast and prostate cancers preferentially metastasise to bone, and colon cancer to the liver. The underlying reasons are unknown, however, the *"Seed and Soil"* theory remains prevalent. The premise is that it is challenging for a cancer cell to survive outside its region of origin, so it metastasises to locations with similar characteristics that are permissive to its growth. However, there are a number of other alternative theories, including the *"Filter and Flow"* theory that suggests that metastasis is determined by anatomic and mechanical routes. For example, in the case of CRC, the hepatic portal vein drains first to the liver and so this is the likely first point of CRC metastasis. In this theory, lung metastases arises later and are less common in CRC patients because they do not arise from the primary tumour, but from the metastatic deposits within the liver[22]. Newer molecular biology theories to explain the metastatic sites include the *"Adhesion"* or *"Chemotaxis"* metastasis theories, where the site of metastasis is principally determined by adhesion molecules on the target organ site or the ability of the cancer to up-regulate and respond to chemotaxis signals[23].

Distant metastasis is present in 20% of patients with newly diagnosed with CRC and can also develop post-surgical resection of the colonic tumour, despite adequate surgical margins. Tumour recurrence is observed in 10-30% of patients with stage II disease, and 30-70% of patients with stage III disease[24]. The liver is the most frequent site of CRC metastasis (70%) with thorax (32%), peritoneum (21%), nervous

system and bone occurring less frequently. The pattern of metastasis depends on the subtype of CRC, rectal cancers are much more likely to spread to the thorax and bone compared to colon cancer, and mucinous and signet ring adenocarcinomas are more likely to develop peritoneal metastases[25].

Metastasis has historically been simplified into four-step process, each step obligatory for tumour metastasis.

To metastasise, malignant cells must have firstly fulfilled certain tumorigenic prerequisites. These include important cellular features such as the ability to proliferate indefinitely, resist apoptosis and maintain important tumour features such as an adequate blood supply. These features are not important per se for metastasis but act as crucial enablers for the emergence of metastasis-capable cells. Multiple genes have been implicated in this stage, for example, tissue inhibitors of metalloproteinases (*TIMP1,2,3*) and *HIF1a* are important in the induction of vascular endothelial growth factors (*VEGF*) and the process of angiogenesis[26].

The second step requires the development of metastasis-initiating attributes. These metastatic initiating attributes include the ability of cells to resist anoikis (the process of cell death when there is no attachment to other cells or the extracellular matrix) and the ability to migrate and invade through the extracellular matrix of tumour and normal tissues.

Some of the genes involved in these processes have been identified through genetic screens. Using shRNAs to screen 151 colon cancer candidate genes, it was demonstrated that the signalling pathway JNK is important for anchorage-independent growth[27]. An RNAi genetic screen was used to study 159 commonly mutated genes

in CRC to determine which of these were important for cellular invasion across a extracellular matrix (ECM) gel on a transwell plate. This identified *ADAMTS18* as an important gene that allows actin cytoskeleton rearrangement in CRC cells[28].

Cellular migration and invasion are clearly important metastasis initiating features and may be best observed at the tumour-host interface of human CRC specimens, the *invasive front*. This invasive front may form a very clear demarcation between the tumour and the host, a so-called *pushing margin*; conversely, the clear margin may be lost resulting in the development of an *infiltrative tumour border* and this represents a tumour that has developed the ability to migrate and invade through the normal tissue. In these tumours, small clusters of cancer cells are observed up to five cells ahead of the invasive front and this phenomenon has been termed *tumour budding*. *Tumour budding* and *invasive front* are histological markers of poor prognosis that are linked to tumour recurrence/metastasis. 5-year survival for patients with stage III CRC is 81.8% for the infiltrative tumour border group compared to 92.9% for the pushing margin group ($p<0.02$)[29]. The infiltrating cells are particularly interesting because they have many phenotypic differences from the primary tumour. They have reduced intercellular contacts and reorganisation of their cytoskeleton into lamellipodia or cytoplasmic protrusion and this characteristic is similar to a process observed during embryonic development, *Epithelial Mesenchymal Transition* (EMT). Cancer cells that have undergone EMT express mesenchymal markers such as Vimentin and alpha-SMA with decreases in epithelium-specific markers such as Cytokeratin and E-cadherin. In CRC, like many other cancers, the signals to induce EMT partly originates from the tumour-associated stroma cells and include *HGF, EGF, PDGF* and *TGF-b*. This results in the induction of a number of transcription factors associated with EMT including *SNAIL, SLUG, TWIST* and *ZEB1*[30]. However, it is important to note that transformation is not a one-way process and when these cells establish

at metastatic sites, they lose their mesenchymal phenotype and a reverse process of mesenchymal-epithelial transition (MET) occurs.

As well as demonstrating EMT, the cells at the invasive front demonstrate a number of other features that enable them to invade the host stroma. *CXCL12*, a factor involved in chemotaxis and angiogenesis and its receptor *CXCR4* are strongly up-regulated at the tumour-host interface[31]. There is over-expression of matrix metalloproteinase proteins, proteases involved in the breakdown of extracellular matrix. In human CRC, MMP-1,-2,-7,-9 and -13 are expressed, conferring poor outcomes and the importance of these MMPs in mediating invasion has been confirmed using Boyden transwell assays[32 33].

Metastatic propagation is the collective term for the third group of steps that are required for tumours to colonize new organs. This enables tumour cells to enter the vasculature (intravasation) and become circulating tumour cells (CTC). Practically all tumour cells that enter the circulation are unable to survive the stresses of dissemination and undergo senescence or apoptosis. CTC are challenging to detect and usually low in number with less than 5-10 cells in 7.5mls of blood for late stage tumours. However, the amount of CTC are dependent on CRC stage and 5% of patients with stage I disease have detectable CTC and this increases to 18.8% of patients with stage IV disease. The detection of CTC is a bad prognostic marker, particularly for clinically non-metastatic CRC patients, with a hazard ratio of 5.5 for recurrence/metastasis[34]. CTC have a number of features that distinguish them from primary tumour cells. They display hallmarks of stem cells with increased expression of the stem cell markers (CD26, CD44 and ALDH1A1) often with the development of new mutations such as the BRAF V600E mutation and up-regulation of genes involved with metabolic activity[35].

The final step, metastatic virulence attributes, allows the activation of dormant lesions and promotes metastasis outgrowth. Metastasis virulence attributes enable a tumour cell to survive in the circulation, mediate attachment to capillary walls in the vasculature of target organs, transmigrate across capillary walls or breach them to enter the parenchyma of target organs. They allow a tumour to develop in a new permissive environment, develop a blood supply and permit further dissemination. Often this will involve the co-option of both tumour-intrinsic and tumour-extrinsic features and many genes/gene networks have been identified. For example, in breast cancer metastasis, tumour extravasation from the lung capillaries is dependent on *COX2, MMP1* and *EREG*.

The origin of metastatic cells remains an area of much debate. In haematological malignancies, it is accepted that the majority of leukaemic cells are not able to establish leukaemia progeny cells[36]. The small percentage of cells that retain this capacity have been termed *Cancer Stem Cells* (CSC), and are defined as subpopulations of tumour cells that have the ability to form tumours specifically in mouse models. In haematological malignancies, CSCs express similar markers to normal stem cells, expressing markers such as CD33, CD90 and CD123[37]. Cancer stem cells have since been identified in numerous tumour solid organ tumour types including colon[38], prostate[39],[40] and breast cancer[41].

One hypothesis derived from the *Cancer Stem Cell theory* is that only CSC are able to establish metastases. These cells do not need to undergo epithelial-mesenchymal-transition phenotypic changes because CSCs have many intrinsic features that fulfil the steps in the metastatic cascade. They are able to migrate from one location to another, proliferate indefinitely and survive anoikis. In patients with CRC, high

expression of CSC markers CD133, CD44, CD166 identifies patients at high risk of metastasis[42]. Furthermore, in a very specific xenograft model, it was demonstrated that only CSC (defined by CD44/CD166/EpCAM) were able to establish tumours[43].

The definitive proof of the exclusivity of CSC in establishing metastasis may rely on the in-depth analysis of circulating tumour cells (CTC). Unfortunately, there is no evidence that CTCs exclusively express markers characteristic of CSC (CD133, CD44)[35] and it is likely that metastasic cells may derive from a combination of CSC, normal tumour cells that had undergone EMT or tumour cells that activated other hithero undescribed metastasis gene programs.

Leucine-rich-repeat-containing G-protein coupled receptor 5 (LGR5), is a downstream effector protein of WNT signalling and a particularly important stem cell marker. The landmark paper identifying the gene as a stem cell marker used an expression array-based analysis of the cells in the intestinal crypt[44]. Subsequent work from this group identified that Lgr5 was expressed in the stem cell compartment at the +4 position in the small intestine. This was confirmed using in-situ hybridization but also through the generation of a *Lgr5-lacZ* reporter mouse. Lineage tracing over a 60-day period demonstrated that these stem cells could give rise to all epithelial cells of the small intestine and colon of mice[45]. Further work confirming the importance of the protein to the self-renewing ability of the intestinal epithelium was performed using the intestinal organoid system. Lgr5+ stem cells were sorted and it was demonstrated that 6% of these cells were able to grow out into organoids and these were indistinguishable from organoids derived from whole crypts. The group claimed that these results identified that the intestinal crypt-villus units may be completely built from a single Lgr5 stem cell in the absence of a non-epithelial cellular niche[46].

The prognostic value of LGR5 expression in clinical cohorts have been confirmed

by a recent meta-analysis. Patients that overexpressed the protein were at higher risk of both lymph node metastasis (OR 0.45, p=0.003) and distant metastasis (OR 0.37, p<0.001)[47].

Cancer stem cells are different to intestinal stem cells, however, two papers have recently been published that highlight the importance of LGR5+ CSCs in tumours and their potential role in mediating metastasis. In the first paper, the authors established a colorectal tumour organoid library and inserted a LGR5 expression label. This was achieved by using CRISPR-Cas9 vectors to insert a GFP label downstream of the *LGR5* locus. These cells were then transplanted into mice and lineage tracing was performed to demonstrate that these cell maintain long-term self-renewal capacity. LGR5+ stem cells were then ablated using a new knock-in construct consisting of LGR5 linked to Caspase 9. Upon LGR5 ablation, there was a transient reduction in tumour size, but tumour eventually regrew and the tumour re-acquired LGR5 expression. The molecular mechanism of reversion was not identified, however, it was suggested that these experiments represented further proof of LGR5 cancer CSCs[48].

The second paper involved the use of the *Apc (Min/+), Kras (G12D/+), Vil1(CRE)* model of intestinal tumourigenesis. This mouse was crossed with a *Lgr5(DTR/eGFP)* line. This meant that Lgr5+ cells could be visualised and subsequently selectively ablated through the administration of diphtheria toxin (DT). Organoids were derived from these mice and two further mutations were introduced, *Trp53* and *Smad4* using CRISPR-Cas9 genome editing. This multiply mutated line was then transplanted into mice and following tumour engraftment, it was possible to demonstrate that Lgr5+ cells were necessary for tumour initiation. However, in contrast to the human organoid model, depletion of the Lgr5+ CSC in the mouse organoids was only evident during administration of DT and following treatment discontinuation, there

was rapid re-initiation of tumour growth. To monitor the effects on metastasis, these organoids were transduced with a lentivirus co-expressing a luciferase protein to allow bioluminescent imaging. FACS analysis revealed that whilst not all metastatic cells were Lgr5+, there was enrichment of the Lgr5 cells in micro-metastasis. There was no enrichment observed in macro-metastases. Selective ablation of Lgr5+ cells was then performed using DT treatment and this demonstrated that in this model, ablation of these cells was able to reduce, but not completely ablate tumour burden in the liver. The authors conclude that Lgr5+ CSC were key to tumour initiation, growth and metastasis in CRC[49].

There is good evidence that host factors are co-opted to aid this process. Exosomes secreted by tumours initiate pre-metastatic niche formation and in pancreatic cancer this is achieved through *TGFb* and fibronectin expression in liver Kupffer cells[50]. Within the circulation, platelets are recruited to CTC to form a protective layer, disguising the cells, and in mouse models with platelet dysfunction, metastasis rates are decreased[51 52]. Similarly, adhesion of tumour cells to endothelial cells is crucial for organ colonisation. There is evidence that tumour cells induce an increase of E-selectins on the surface of endothelial cells that facilitates further tumour adhesion[53]. Outgrowth at metastatic sites is also dependent on tissue macrophages that are recruited to areas of hypoxia within tumours and are stimulated to release growth factors such as *VEGF* that lead to angiogenesis and enable further tumour spread[54].

Intrinsic changes in tumour biology are crucial in determining the success of metastasis outgrowth. Expression of these genes tend to be tissue specific, for example, when chemokine *CXCR3* was over-expressed in a colon cancer cell line and implanted into the rectum of nude mice, *CXCR3* was only sufficient to induce para-aortic lymph node metastasis, but did not promote rates of distant metastases[55].

c-MYC signalling plays an important role in mediating metastasis outgrowth and is a transcription factor that is activated in response to mitogenic signals and results in cell proliferation, migration and growth. Gene amplification of *c-MYC* is a very common finding in liver lesions, being present in 59% of cases compared to 4% of controls[56]. Tumours with overexpression of c-MYC are more likely to have lymph node metastasis with a much poorer prognosis[57] and c-MYC promotes metastasis by directly regulating EMT through a *LIN28B/let-7/HMGA2* cascade, regulating cell-cell-matrix interactions through *LGALS1* and by the activation of RhoA, a protein involved in cytoskeleton organisation and motility[58].

Mutation analysis, using techniques such as Next Generation Sequencing, have been important in helping to identify how the mutation spectrum of the tumour varies between the primary and metastatic sites. Currently, there are no known germ-line SNPs/variants that predispose patients to a predominantly-metastatic tumour. It is generally believed that somatic mutations (mutations that arise during the course of tumurigenesis) play a much more important role. The largest somatic sequencing project of colorectal cancer to date is *The Cancer Genome Atlas*, and this has identified 32 somatic recurrently mutated genes, including *APC, TP53, KRAS, PIK3CA, FBXW7, SMAD4, TCF7L2* and *NRAS*[16]. In this cohort, the presence of metastasis was not analysed specifically, but they did define an aggressive tumour as those with lymph node metastasis, distant metastasis, higher tumour grade or demonstrating the presence of vascular invasion. These tumours demonstrated an association with somatic mutations in *APC, TP53, PIK3CA, BRAF* and *FBXW7*. The genotypes of smaller cohorts have been studied using targeted sequencing and these studies have identified that *KRAS* mutated tumours are associated with increased lungs, brain and bone metastases[59], and *JAZF1* is frequently mutated and identified in lung metastases[60]. There is also limited evidence that metastasis outgrowth may

be driven through CIN. (Sub-)chromosomal gains and losses leads to SCNA and papers have identified more SCNA in liver metastases than the primary tumour[61].

Large scale sequencing of metastatic lesions in colorectal cancers has not yet been performed, though it is likely that the development of metastasis will involve a complex interplay between somatic mutations, gene expression changes in the tumour, the tumour microenvironment and the effect of host immune control.

The mechanistic target of rapamycin (mTOR) is a protein that is an important mediator of metastasis. The protein is a protein kinase, and a member of the phosphatidylinositol 3-kinase related family and binds with several key proteins to form two distinct protein complexes, mTOR complex 1 and mTOR complex 2. Overall, mTOR functions to regulate cellular growth, proliferation and motility in response to growth factors, nutrition abundance, oxygen and insulin. The mTOR complex 1 consists of a complex consisting of RPTOR and DEPTOR and controls protein synthesis in response to cellular nutrition status. The mTORC omplex 2 acts principally to regulate the actin cytoskeleton through its stimulation of RhoA/Rac1 and regulates changes in celllular shape and motility. Activation of these protein complexes have been demonstrated to be important for cellular invasion and metastasis in several cancer types, including kidney, lung and CRC[62].

Finally, it is also generally accepted that the host immune surveillance acts as a important mechanism that may promote or restrain carcinogenesis, and a dynamic interaction of host and tumour cells exists. Transformed and metastatic capable cells may be eliminated during tumour progression. Immune selection results in tumour cells that are able to avoid the immune system with decreased expression in MHC antigens and upregulaton of proteins that are able to inactivate the immune response

(PDL1, CTLA-4). This process has been termed *Cancer Immunoediting*. Experimental evidence for this phenomenon has been demonstrated using transgenic mice with depletion of natural killer (NK) and macrophage cells, and these mice demonstrate markedly different tumour growth characteristics than control mice[63].

In a similar fashion, mice with homozygous mutation of the *Rag2* gene resulting in loss of mature B/T lymphocyte function, demonstrate a greater frequency of tumour formation in a carcinogen-induced sarcoma model[64].

Our lab has identified that the degree of intratumoral T-cell response is strongly correlated with the amount of "neopeptides", or new protein sequences that arise within tumours with a high mutation rate[65] and this may determine immune response and subsequent degree of immune editing.

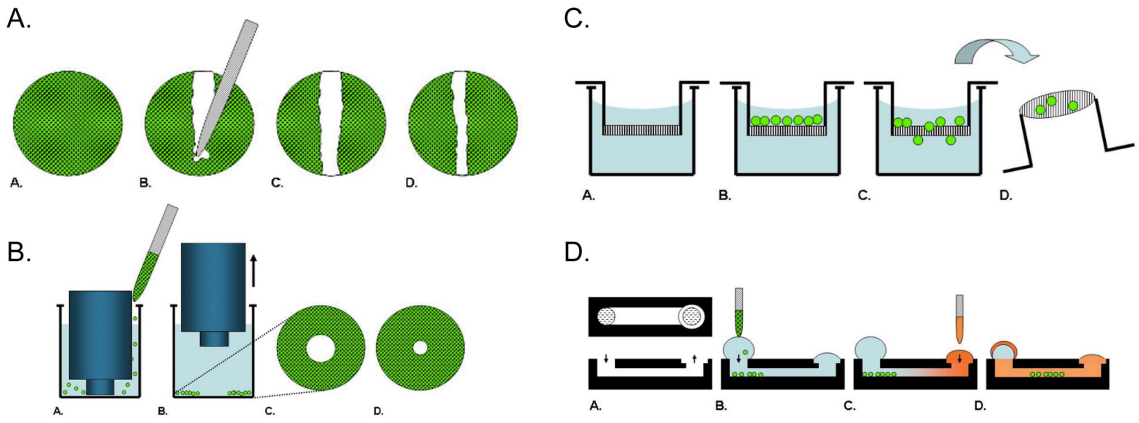
1.0.5 Models of CRC metastasis

Robust models of metastasis are useful to enable a greater understanding of molecular determinants of metastasis. In-vitro techniques are able to recreate many of the prerequisites for metastasis, particularly metastasis initiating attributes, however, modelling later metastasis attributes requires the use of in-vivo models. In-vivo models allow an analysis of the effects of host immunology, endothelial barriers and vasculature upon metastatic cells.

In-vitro assays are a reductionist approach to study the early stages of metastasis. They are numerous and their benefit is that they are comparatively easy, inexpensive and quick. Broadly, they can be used to assess cell proliferation, cell migration/invasion and anchorage independent growth.

Proliferation assays measure the number of cells or the change in proportion of cells that are dividing and this can be measured indirectly by measuring DNA synthesis,

Figure 1.1: In vitro models of metastasis. A. Scratch assay. B. Gap closure assay. C. Boyden Transwell assay. D. Microfluidic assay



metabolic activity, antigens associated with cell proliferation or ATP concentration. Cellular migration and invasion of tumour cells can be analysed using simple cell exclusion assays such as a scratch assay or a gap-closure assay. More complicated assays use three-dimensional models such as Boyden chamber transwells, biological or micropatterned scaffolds or the use of a microfluidic assay that allows the analysis of single cells (Fig 1.1). Cellular invasion is a similar process to and includes cell migration, however, cells must migrate through the ECM by degradation and proteolysis. The previously mentioned techniques for measuring cellular migration can be adapted to analyse invasion by the introduction of a layer of ECM gel through which cells must invade, however, the semblance to normal tissue architecture is open to debate.

Anchorage independent growth is another important metastasis prerequisite feature as normal cells maintain tight connections between neighbouring cells as these connections provide essential signals for growth and survival. When cells become detached, the loss of these normal cell-matrix interactions results in anoikis or programmed cell death. In-vitro assays involve isolating single cells in a three-dimensional

media, such as soft agar matrix and assessing the ability of these single cells to proliferate and generate cell colonies. In this assay, normal cells form very few or no colonies, however, numerous colonies are formed by metastatic-capable cells.

Modelling the later stages of metastasis requires the use of more complex in-vivo mouse models. These models recreate and allow analysis of the highly complex processes involving the release of tumour cells into the circulation (intravasation), the ability of tumour cells to exit at a target organ (extravascularisation) and their adaptation and outgrowth at the new target organ (tissue colonization). There are two categories of in-vivo models that can be used to understand the metastasis process, Genetically Engineered Mice (GEM) and Transplantation Mouse Models (TMMs).

GEM models allows the introduction of specific mutations or gene expression changes at a specific time point, in a specific tissue, in a specific temporal order. This can be achieved through models with inducible tissue specific recombinases (Cre or FLP), but also genome engineering tools (CRISPR). Recent advances in GEM models involve the introduction of a label in mutated cells, usually fluorescent/luciferase labels, to allow visualisation and/or lineage tracing. Labelling permits the analysis of the entire metastatic process from tumour cell invasion, tumour intravasation and tumour lodging and extravascularisation. GEM mice are powerful tools, however, they have limitations including cost and that few CRC GEM models develop metastasis[66].

Transplantation Mouse models (TMMs) remain the workhorse of metastasis research. The basic premise is that cells are labelled, transplanted and the resulting growth and metastasis analysed. TMMs are versatile and allow an analysis of specific mutations in specific cell types, the use of genetic screens and the role of the immune system. The implanted cells can be derived from a number of sources, including cell

culture (which allow good reproducibility and homogeneity), tumours or primary cell cultures (allowing use of a mixture of cell types/heterogeneous tumour admixtures). Labelling of cells prior to transplantation allows for accurate tracking and these may then be introduced into a variety of host mice. Immunocompromised mice (NOD scid gamma, RAG, Nude) allow the implantation of human cells without risk of rejection, whereas transgenic/wild type mice allow analysis of murine tumour development on a specific genetic background. The simplest models involve cells transplanted onto the dorsal flank of the mouse allowing high reproducibility and ease of access. More complicated models introduce cells into the correct anatomical sites (orthotopic TMMs), and allow the analysis in the correct tumour microenvironment.

One of the fist orthotopic colorectal transplantation models was described in 2005. HT29 colorectal cancer cells were transplanted into the caecum of an anesthethised mouse which had undergone a midline incision[67]. The tumour was able to establish in the caecum and grow, eventually occluding the colonic lumen. This technique has formed the basis of orthotopic transplantation models in a large number of studies[68],[69]. However, more recently, less invasive techniques have been developed that does not require major surgery on the animal and takes an endoscopic implantation approach. The MC38 murine CRC line was successfully implanted into the submucosa of the descending colon of C57BL/6 mice using endoscopy. All mice developed intestinal tumours with minimal unexpected adverse side effects[70].

It is clear the microenvironment of implanted cells is extremely important, and in colorectal cancer orthotopic implantation significantly increases rates of metastasis[68].

Table 1.1: Orthotopic colon cancer mouse models using different cell lines. Table demonstrates metastatic potential, predominance to LN, liver or lung metastases, overall propensity (no. mouse affected), duration and distribution

Cell line	Potential	LN	Liver	Lung	Propensity	Duration	Distribution	Ref
Caco2	+	+	-	-	25%	25 weeks	Predominantly LN	Flatmark(2004)[71]
WiDr	+	++	-	-	50%	28 weeks	Predominantly LN	Flatmark(2004)[71]
HT29	++	+++	-	-	83%	12 weeks	Predominantly LN	Flatmark(2004)[71]
KM20L2	+++	+++	-	-	75%	9 weeks	Predominantly LN	Flatmark(2004)[71]
HCC2998	+	+++	-	-	86%	13 weeks	Predominantly LN	Flatmark(2004)[71]
Co205	+	+++	-	-	50%	31 weeks	Predominantly LN	Flatmark(2004)[71]
HCT116	+++	+	-	+++	70%	9 weeks	Predominantly Lung/LN	Rajput(2008)[68], Chen(2012)[69], Cespedes(2007)[72]
Col 15	+++	+++	-	-	56%	9 weeks	Predominantly LN	Flatmark(2004)[71]
HCT15	+++	++	-	-	69%	10 weeks	Predominantly LN	Flatmark(2004)[71]
SW480	+++	-	-	-	50%	9 weeks	Predominantly LN	Flatmark(2004)[71]
Sw620	+++	+++	-	++	11%	6 weeks	Predominantly Lung/LN	Cespedes(2007)[72]
DLD-1	+++	+++	+	+	29%	12 weeks	LN/Liver/Lung	Cespedes(2007)[72]
Colo320	-	-	-	-	17%	4 weeks	Not metastatic	Flatmark(2004)[71]

Table 1.2: Transgenic colorectal cancer mouse models which demonstrate overall metastasis potential, predominance to LN, liver or lung metastases, metastases propensity, duration to effect and distribution

	Potential	LN	Liver	Lung	Propensity	Duration	Distribution	Findings	Ref
APC580S/KRAS-	+	+	+	-	20%	20 weeks	LN/Liver	Wnt activation in mets	Hung(2010)[73]
SMAD3-	+	+	-	-	No info	24 weeks	LN		Zhu(1998)[74]
TGFb-/KRAS	+	+	-	+	15%	20 weeks	LN/Lung	Kras & Tgfb loss required	Trobridge(2009)[75]

The very latest models of colon cancer metastasis involve the use of cancer organoids. Organoids consist of a collection of cells that are derived from either the normal intestine or a colorectal tumour and are established and grown in-vitro. They are derived from stem cells and form a 3-dimensional structure that consist of the appropriate make-up of cells (eg. Paneth cells/goblet cells) with similar micro-anatomy and the formation of villi and crypts[76]. This system was first described in 2011[77] and required the addition of the Wnt growth factor, nicotinamide and inhibitors of Alk and p38 and these were sufficient to establish long-term culture conditions. These organoids can be transplanted orthotopically and have the capability of forming micrometastases[78] and also liver metastases (APC/KRASG12V/TP53 knockout organoids)[79].

1.0.6 Serum and Glucocorticoid regulated kinase 1

One gene of interest to our laboratory has been the gene Serum and glucocorticoid-regulated kinase 1, SGK1. The gene is a serine/threonine kinase that is activated in response to insulin, cellular stress and growth factors via PDK1. The kinase has marked sequence and target preferences to AKT and contributes to hormone release, inflammation, proliferation and the activity of a number of ion channels[80] The gene had originally been identified in our lab from the expression profiling of 63 mouse adenomas. It was found that SGK1 was significantly downregulated in adenomas, but was expressed in normal tissues at high levels in differentiated cells at the crypt top[81]. Our lab had also demonstrated that downregulation of SGK1 in tumours is partly mediated through promoter methylation, though this did not account fully for different expression levels between normal and colorectal tumour tissue[82].

In a transgenic model, where a *Sgk1* knockout mouse was crossed with an *Apc (min/+)* mouse, it has been observed that adenoma burden was significantly lower and this was thought to be mediated through decreased expression of β-catenin expression[83].

There have been no studies into the effect of SGK1 on other features of tumorigenesis, such as cellular invasion or metastasis.

1.0.7 Aims and objectives of this thesis

Chapter 3

Aim 1. Understand the expression of SGK1 in normal human large intestine. I hypothesise that expression of SGK1 will mimic findings in mouse small bowel and I will be able to identify SGK1 expression principally in differentiated cells where it could act as a differentiation factor/marker.

Aim 2. Determine the expression of SGK1 in CRC. Work in our lab has already demonstrated that SGK1 expression is lower in mouse adenomas compared to normal mouse intestine, and so I would hypothesise that expression of SGK1 in CRC would be decreased. Secondly, in keeping with our observation that SGK1 was only observed in differentiated cells in the mouse intestine, I also hypothesise that there would be a relationship between tumour differentiation status and the degree of SGK1 staining.

Aim 3. Determine if the expression of SGK1 might confer differences in prognoses from CRC. Work performed in SGK1 knockdown mice has demonstrated that these mice are more prone to tumourigenesis, and I would hypothesise that the extent of SGK1 down-regulation might be associated with tumour aggressiveness.

Aim 4. To understand the effects of SGK1 expression on tumour invasiveness and metastasis in a murine model and gain greater understanding of the underlying mechanism(s).

Chapter 4

Aim 1. Perform a whole genome CRISPR/Cas9 screen to identify genes that are important for driving metastasis. I hypothesise that by taking a whole genome approach, it would be possible to identify a small subset of genes, which when silenced would enable a pro-metastasis cellular phenotype.

Aim 2. To validate the findings from the screen through validation in a in-vivo model.

Aim 3. To gain a greater mechanistic insight of the effects of these genes through RNA sequencing.

Aim 4. To determine the relevance of the hits of the CRISPR/Cas9 screen and the prognosis of patients with CRC. Previous genome screens in mouse models have been highly relevant to the study of human cancers[84], and I envisage that some of these genes identified in the screen could act to identify patients at high risk of disease relapse following surgery.

Chapter 5

Aim 1. To assemble a cohort of stage III colorectal cancer patients and perform RNA sequencing. By taking a hypothesis-free approach, I hope to understand the relative importance of the different signalling pathways in promoting metastasis to neighbouring lymph nodes.

Aim 2. To understand if there are groups or patterns of gene expression profiles within the lymph nodes of patients with stage III colorectal cancer patients that confer different prognoses. "Consensus molecular subtype" signatures may be derived from the transcriptome of primary tumours and these are associated with different

patient outcomes. I hypothesise that the signatures in metastatic lesions in the lymph nodes could be of greater prognostic relevance than the signatures derived from the primary tumours.

Aim 3. To derive a means to determine a patient's immune response immediately adjacent to the tumour. I hypothesise that the degree to which a patient is able to mount an anti-tumour response is likely to be associated with likelihood of relapse following surgery for stage III CRC, and this could be determined by RNA sequencing of the normal lymph node adjacent to the tumour.

Chapter 2

Materials and Methods

2.1 Laboratory methods

2.1.1 Plasmid manipulation

2.1.1.1 DNA cloning

Plasmid vectors and/or Polymerase Chain Reaction (PCR) products were digested using the appropriate restriction endonuclease to form the insert and vectors required for cloning. Endonucleases were purchased from *New England Biolabs* and used according to manufacturer's protocol. In cloning steps where standard DNA cloning was challenging, the vector was treated with 1μL of Calf Intestinal Alkaline Phosphatase (CIP) (*New England Biolabs*) to prevent re-circularisation of the vector. Digested inserts and vectors were excised from a gel using a blue light transilluminator *Clare Chemical Transilluminator DR-46B*. They were purified by gel or PCR extraction using the *QIAquick Gel Extraction Kit* or the *QIAquick PCR purification Kit* (*QIAGEN*). Nucleic acid concentration was determined by Nanodrop spectrophotometry. A mixture of insert and vector at a molecular ratio of 3:1 was ligated with a T4 DNA Ligase *New England Biolabs* at room temperature overnight.

2.1.1.2 Ligation-free DNA cloning

The method of ligation-free DNA cloning I have used in this thesis is *Gibson Assembly*. This technique relies on the creation of overlapping DNA fragments between the vector and insert and allows the assembly of multiple DNA inserts at high efficiency. The technique relies on an exonuclease to create 3' overhangs, a polymerase to fill in the gaps within each fragment and a ligase to assemble the full vector. The insert with the addition of overhangs was created by a PCR reaction using primers designed using the *NEBuilder Assembly Tool* (*New England Biolabs*) introducing between 15-20 nucleotides that overlapped the vector. Inserts were combined with vectors at a molecular ratio of 3:1 and the Gibson Assembly master mix added and incubated at 50°C for 1 hour.

2.1.1.3 Transformation

Plasmids were transformed into XL-10 Gold UltraCompetent cells (*Stratagem*). Briefly, 50μL of the cell suspension was transferred into a 1.5ml eppendorf and 2μL of ligation or Gibson assembly mix was added to the cells. This cell suspension-vector mix was incubated on ice for 30 minutes, heat-shocked at 42°C for 30s, before recovering on ice for 2 minutes. The E.Coli cells were plated out onto LB agar plates with the appropriate antibiotic for selection. Plasmid amplification was achieved by selecting single colonies from the agar plate and incubating in a 14ml BD Falcon polypropylene round-bottomed tube with LB broth for 24 hours. DNA extraction was performed using a mini-prep kit from *Macherey-Nagel* or *Qiagen*. Both kits are based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high salt on a spin column followed by washing and elution of plasmid DNA. Where required, positive clones were amplified again by culturing in 200ml of LB broth and DNA was extracted using the Endotoxin free Maxiprep kit (*Qiagen*) as per manufacturer's instructions.

2.1.1.4 Primers used in thesis

2.1.1.5 Gene synthesis-aided cloning

Where the DNA cloning inserts could not generated by PCR or from another vector, gene synthesis was performed. Nucleic acid sequences were generated in-silico and ordered from either *Integrated DNA technologies* or *Life technologies*.

2.1.1.6 CRISPR cloning

Clustered regularly interspaced short palindromic repeats (CRISPR) requires the generation of plasmids with the addition of a 20 nucleotide sequence to generate a synthetic guide RNA (gRNA) that allows targeting of the cas9 nuclease enzyme. CRISPR plasmids were purchased from *Addgene*. Inserts were designed in silico to minimise off-targeting effects and to enhance targeting to the genomic region of interest. This was achieved by use of the *CRISPR Design Tool* (*MIT*) or *CRISPR gRNA Design Tool* (*DNA 2.0*) or *Cas9-SAM sgRNA Design Tool*. The inserts were purchased as single strand oligos from *Integrated DNA technologies*. Inserts were generated by annealing the inserts and phosphorylating the oligos with T4 PNK (*New England Biolabs*). Vectors were linearised with BsmBI (pLenti-CRISPR) (*New England Biolabs*) and dephosphorylated with CIP (*New England Biolabs*). Inserts and vectors were combined at a molecular ratio of 3:1 and ligated with a T4 ligase (*New England Biolabs*) for 3 hours at room temperature.

2.1.1.7 CRISPR efficacy

To determine CRISPR efficacy, TIDE analysis was performed. This is a technique that allows the spectrum and frequency of targeted mutations generated by CRISPR in a pool of cells to be determined. DNA was harvested from the pool of control and CRISPR/Cas9 targeted cells. PCR reactions were performed using primers to amplify

Table 2.1: Primer sequences

Primer name	Sequence	Experiment
710-fwd	cgacggtatcggttatccag	3.2.4
710-rev	tctttcccctgcactgtacc	3.2.4
v2Adaptor_F	aatggactatcatatgcttaccgtaacttgaaagtatttcg	4.2.5
v2Adaptor_R	tctactattctttcccctgcactgttgtgggcgatgtgcgctctg	4.2.5
Rbl2-aOligo1	caccgCCCCATGATTAGCGATGATC	4.2.14
Rbl2-bOligo1	caccgCTAATCTTTCAGTACGTTCT	4.2.14
Rbl2-cOligo1	caccgACCTCTGCTCCGAACAGCGA	4.2.14
Mdc1-aOligo1	caccgGCTGTCGTTCCGTTAGCTTC	4.2.14
Mdc1-bOligo1	caccgTGCTGCAGGTGACCCGCTTT	4.2.14
Mdc1-cOligo1	caccgTTACGATTTGAGTGCCATTG	4.2.14
Foxf1-aOligo1	caccgGTACCCGCACCACGACAGCT	4.2.14
Foxf1-bOligo1	caccgGCGACTGTGAGTGATACCGA	4.2.14
Foxf1-cOligo1	caccgGTGCGGCTCCATGACTCCGC	4.2.14
Rbl2-aOligo2	aaacGATCATCGCTAATCATGGGGc	4.2.14
Rbl2-bOligo2	aaacAGAACGTACTGAAAGATTAGc	4.2.14
Rbl2-cOligo2	aaacTCGCTGTTCGGAGCAGAGGTc	4.2.14
Mdc1-aOligo2	aaacGAAGCTAACGGAACGACAGCc	4.2.14
Mdc1-bOligo2	aaacAAAGCGGGTCACCTGCAGCAc	4.2.14
Mdc1-cOligo2	aaacCAATGGCACTCAAATCGTAAc	4.2.14
Foxf1-aOligo2	aaacAGCTGTCGTGGTGCGGGTACc	4.2.14
Foxf1-bOligo2	aaacTCGGTATCACTCACAGTCGCc	4.2.14
Foxf1-cOligo2	aaacGCGGAGTCATGGAGCCGCACc	4.2.14
tide-fox-1-f	gtgcttcatcaaactgccca	4.2.15
tide-fox-1-r	CCCGTTGTGACTGTTTTGGT	4.2.15
tide-fox2-f	AGTAAAGCAGTTTGTGGCGG	4.2.15
tide-fox2-r	ccccttccgcatcaaagaag	4.2.15
tide-fox3-f	CAAGATCGGTTCACAGTGGC	4.2.15
tide-fox3-rr	tccggctagcgagtttatgt	4.2.15
tide-RBL2-1-f	tggtataatGGTGTCTGGTGTCA	4.2.15
tide-RBL2-1-r	aagcctgcatctggggatac	4.2.15
tide-RBL2-2-f	aaacaagcagagaagtcccg	4.2.15
tide-RBL2-2-r	ggcattcacttgttctggca	4.2.15
tide-RBL2-3-f	aaaggaaaaccacgactcaga	4.2.15
tide-RBL2-3-r	tccagatgagtgcaggaagt	4.2.15
tide-mdc1-1-f	atctgtctcactgtccgcaa	4.2.15
tide-mdc1-1-r	ctttggTTTGGGCAGTGACA	4.2.15
tide-mdc1-2-f	gcaacacttaccatcacagca	4.2.15
tide-mdc1-2-r	agctgttcaggaaggctcat	4.2.15
tide-mdc1-3-f	gcagccaacaaaactctgga	4.2.15
tide-mdc1-3-r	cccctctaccttggcaagaa	4.2.15
SAMFoxf1-aOligo1	caccgGAGGAGGAGGAAGAGGACGA	4.2.18
SAMFoxf1-bOligo1	caccgGCGGGCGGGAGCGCGGGGGC	4.2.18
SAMFoxf1-cOligo1	caccgGGGGGCTGCGCGCGCGGGGC	4.2.18

a 700 bp region around the targeted site. Sanger sequencing was performed on the control and experimental cells. Sequencing files were uploaded to https://tide.nki.nl and the web tool was used to reconstruct the spectrum of indels from the sequence traces and generate an estimate of CRISPR efficacy.

2.1.2 DNA/RNA manipulation

2.1.2.1 DNA extraction from cell lines

DNA was extracted from cell lines using the *DNeasy blood and tissue kit* (*Qiagen*) following manufacturer's protocols.

2.1.2.2 DNA extraction using NaCl salt precipitation

Salt precipitation was performed on whole organs that had been homogenised and digested with Proteinase K. Equal volumes of lysis solutions A and B were added to the specimens (Solution A, 75mM NaCl, 25mM EDTA and Solution B, 10mM Tris, 10mM EDTA, 1% EDTA). The solution was centrifuged at 13,000rpm for 20 minutes, and the supernatant transferred to a new tube. Precipitation of DNA was performed using isopranolol precipitation.

2.1.2.3 DNA quantification

Nanodrop Spectophotometer : DNA concentration and quality were analysed using the UV-visible spectrophotometer (*Nanodrop 2000*, ThermoScientific). 1μL of DNA was loaded onto the optical pedestal and the absorbed light from the sample was measured at a wavelength of 260 nm. The ratio of absorbance at 260 nm and 280 nm provided an estimate of DNA purity.

Qubit Fluorometric Quantification : Where precise quantification of DNA concentration was required, *Qubit dsDNA Fluorometric Quantification 3.0 assay* (*Thermo*

Scientific) was performed. The assay uses a dye that emits fluorescence only when specifically bound to DNA and is more sensitive than UV absorbance. Qubit assays were performed according to manufacturer's instructions.

TapeStation : For medium-throughput quantification of DNA, the *Tape station 4200, Agilent* was utilised as per manufacturer's instructions. Samples were mixed with an intercalating dye, inserted into a gel matrix and imaged. The process allows a rapid, accurate quantification of DNA molecular weight.

Picogreen : For high-throughput quantification of DNA specimens from 96 well plates, *Quant-iT PicoGreen dsDNA Assay kit* (*ThermoFisher*) was utilised. The assay uses an ultra-sensitive fluorescent nucleic acid stain for detecting very low quantities of dsDNA. The kit was utilised as per manufacturer's instructions and imaged on a *Tecan SpectraFluor* fluorescence imaging instrument.

2.1.2.4 DNA sequencing

Sanger sequencing A BigDye Terminator sequencing reaction was performed on DNA that was to be Sanger sequenced (*BigDye Terminator v1.1 cycle sequencing kit, Thermofisher*). The DNA was purified and the dyes removed using *DyeEx 2.0 Spin Kit* (*Qiagen*), dehydrated using a *Savant ISS110 SpeedVac Concentrator* (*ThermoFisher*) and sent for sequencing on the *ABI 3730xl* machine (*ThermoFisher*). For later experiments in this thesis, Sanger sequencing was performed using an overnight service provided by *Eurofin Genomics*.

2.1.2.5 Polymerase Chain Reaction (PCR)

All polymerase Chain Reactions were performed according to manufacturer's instructions. PCR reactions were carried out on the G-Storm Thermal Cycler Systems (*Life Science Research*). A typical 25μL reaction consisted of 10-100ng of template DNA,

0.2 pmoles of each primer, 2.5μl of 10x buffer, 0.5μl of 10mM dNTPs, 0.125μL of polymerase with nuclease free water to make up a 25μL reaction. The number of cycles performed varied from 30-35 depending on amount of DNA required and the need to minimise PCR errors. Temperature of annealing was calculated using *Tm calculator* (*NEB*).

2.1.2.6 RNA extraction

RNA extraction from cell lines RNA was extracted from cell lines using the *RNeasy Mini kit* (*Qiagen*) following manufacturer's protocols. For histological specimens, samples were deparaffinised and RNA was extracted using the *High Pure FFPET RNA Isolation Kit* (*Roche*).

2.1.2.7 RNA quantification

RNA concentration and quality were analysed using the UV-visible spectrophotometer (*Nanodrop 2000*, ThermoScientific). Where more accurate RNA concentration and quality were required, the *Tape station 4200, Agilent* was utilised as per manufacturer's instructions. RNA quality was assessed by analysing both RNA Integrity Score (RIN score) and the percentage of RNA fragments >200 bps (DV200).

2.1.2.8 RNA sequencing

RNA sequencing using the SMARTer protocol was performed at the *Wellcome Trust Centre for Human genetics*. RNA samples were processed using the *SMARTer Ultra Low Input RNA kit*(*Clontech*) using non-stranded polyA enrichment for mRNA amplification. Library preparation was performed using *NEBNext*, (*New England Biolabs*). Next generation sequencing was performed on the *Illumina Hiseq 4000 platform* with 75 paired end reads.

3' mRNA sequencing was performed using the *Quantseq 3' mRNA-seq kit* (*Lexogen*). This approach was taken when there was low quality, quantity, FFPE specimens or a large number of samples. This kit uses total RNA as input with no poly(A) enrichment and starts with an oligodT primer to initiate first strand synthesis. RNA was removed and second strand synthesis was initiated by random primer binding. A magnetic bead-based purification step was used to allow size selection and for nucleic acid purification. Linker sequences for illumina platforms were attached during the library amplification step. Next generation sequencing was performed on the *Illumina Hiseq 4000 platform* with 75 paired end reads.

2.1.2.9 Reverse transcription of RNA into cDNA

DNA was removed from extracted RNA using the DNase I kit (*Thermo Scientific*) as per manufacturer's instructions. Complementary DNA (cDNA) was reverse transcribed from extracted RNA using the High Capacity cDNA Reverse Transcription Kit (*Applied Biosystems*) following manufacturer's instructions. Sample concentrations were determined using the NanoDrop Spectrophotometer.

2.1.3 Cell culture

2.1.3.1 Maintenance of cell lines

All cell culture experiments performed in this thesis were carried out in a class II laminar flow hood. Cell lines were grown from frozen stocks, having thawed the cells rapidly at 37°C, cells were pelleted and resuspended in appropriate tissue culture media. In general, human cell lines were maintained in DMEM media plus L-Glutamine (*Lonza*) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Mouse cell lines were maintained in RPMI media (*Lonza*) also supplemented with the same concentration of FBS and penicillin-streptomycin. Cells were grown at 37°C with 5% CO_2 and were passaged as they reached confluence.

To subculture cells, they were detached with trypsin-EDTA (*Gibco*), neutralised with media and passaged. Cells that were not required for experiments were stored in liquid nitrogen. Cells being prepared for storage were harvested in the log phase of growth, resuspended in freezing media (90% FBS, 10% Dimethylsulfoxide-DMSO).

2.1.3.2 Quantification of cell counts

Cell quantity was counted by two methods. Cell counts during the first half of the thesis were performed using a *FastRead102* disposable haemocytometer (*Immune systems*). The number of cells per ml was calculated using the following calculation: (Number of cells / No of 4x4 grids counted) x 10^4 x sample dilution. Cell counts performed during the second half of the thesis were performed using the *Countess II* automated cell counter (*Life Technologies*).

2.1.4 Lentivirus Usage

2.1.4.1 Lentivirus generation

Lentivirus cell culture was carried out in a designated class II laminar flow hood and usage was carried out according to standard group operating procedures, specifically with regards to personal protective clothing and disposal of viral waste. HEK293T cells were maintained in a T175 flask in Iscove's Modified Dulbecco's medium (IMDM) (*Gibco*) supplemented with 10% FBS and 1% penicillin-streptomycin. The cells were cultured until they reached 50-70% confluence. The media was changed prior to transfection. For generation of lentiviruses, I used the Calcium Chloride transfection method to generate 2nd generation lentiviruses. The following plasmids were required for viral packaging; an envelope glycoprotein-encoding plasmids (pMD2.VSVG) which contained the *Env* gene and the packaging plasmid (pCMV.dR8.74) which contained the *Gag, Pol, Rev, Tat* genes. The following plasmids were combined in a 50ml falcon tube; 16.25ug of pCMV.dR8.74, 9ug of VSV-g, 32ug of transfer vector (or

25ug of transfer vector if the transgene and promoter was less than 1500bp). These plasmids were combined with 750μL of 0.1% sterile filtered TE, 375μL of deionised water and 125μL of calcium chloride at a concentration of 2.5M. The falcon tube was vortexed and 1250μL of HBSS was added in a drop-wise fashion to form a DNA-calcium chloride co-precipitate that has the ability to adhere to the cell surface of the HEK293T and is taken by the cells by endocytosis. This solution was added to the media of the T175 flask containing the HEK293T cells. Cells were incubated in this solution for 16 hours before the media was removed and fresh IMDM added. At 72 hours, the virus was harvested. This was achieved by removing the IMDM media that contains the virus particles. This was centrifuged at 1000g for 5 minutes to remove stray cells and filtered with a 0.45uM filter. Virus particles were stored in 2ml screw top tubes and stored at -80°C. If super-concentrated virus particles were required (>300x concentration), an ultracentrifugation step was performed. The supernatant was ultracentrifuged (*Sorvall*) at 24,000rpm for 2 hours at 4°C and resuspended overnight at room temperature with PBS.

2.1.4.2 Lentivirus quantification

Lentivirus supernatant titers were quantified using the *Lenti-X p24 Rapid Titer kit* (*Clontech*) following manufacturer's instructions. This kit is an ELISA-based protocol that assesses p24 capsid protein content, an HIV-1 virus core capsid protein by anti-p24 coated wells.

2.1.4.3 Lentivirus transduction

Cells due to be transduced with lentivirus were plated between 50-100,000 cells in a 24 well plate. When the cells had adhered to the plate, the media was supplemented with polybrene (Hexadimethrine bromide) at a concentration of 8ug/ml. Polybrene increases retrovirus gene transfer by enhancing receptor-independent virus adsorption

on target cell membranes. Transductions were performed with a virus titration of between 10-400μL of virus supernatant. Target cells were spinoculated to further increase virus adsorption. Spinoculation involves centrifuging cells at 1000g at room temperature for 1 hour. Following spinoculation, the virus supernatant was removed and cells maintained in normal cell media. After 72 hours following transduction, cells either were utilised for experiments, or underwent antibiotic selection by adding the appropriate antibiotic, or visualised for the presence of fluorescent markers.

2.1.5 Gene expression assays

2.1.5.1 Real time quantitative PCR (rt-qPCR)

Absolute real time-PCR (rt-qPCR) was performed using the *ABI 7900HT* cycler (*A*pplied Biosystems). Taqman gene expression assays were purchased from *Integrated DNA technologies* and consisted of a pre-prepared assay that contains primer pairs and probes. A reporter dye (FAM dye) is linked to the 5' end of the probe and a non-fluorescent quencher is at the 3' end of the probe. Taqman probe efficiency was determined through the use of serial dilutions of input DNA and probes with efficiencies of >98% were used in this thesis. The rt-qPCR reaction consisted of 10ng of purified cDNA, 0.5μL of gene expression assay, 5μL of Taqman FAST Universal Mastermix and dH2O to a total volume of 10μL. The samples were run on the *ABI 7900HT* cycler with the following conditions: 95°C for 20 s followed by 40 cycles of denaturation at 95°C for 1 s and simultaneous annealing and extension at 60°C for 20 s. Reactions were performed at least in duplicate, and in triplicate whenever possible. Data was analysed using standard protocols to calculate relative expression with the dCT method with GAPDH serving as an endogenous control. Percentage knockdown of gene expression was calculated by 100 multiplied by (1-(2^{-ddCT})).

2.1.5.2 Western Blotting

Protein extraction from tissue samples was performed using either Tris-HCl/SDS for cytoplasmic proteins or using RIPA buffer (*ThermoFisher Scientific*) for the extraction of cytoplasmic/nuclear and membrane proteins. Tris-HCl/SDS lysis was performed with 1 part SDS (10%) and 2 parts water and subsequently, samples underwent lysis by sonication at 20% power for 10 seconds. Protein extract was clarified by centrifugation at 13,000g for 10 minutes. Quantification of protein lysates was performed using the *Pierce Bicinchoninic acid assay* (BCA). This technique is based on the buret reaction (reduction of copper by protein in an alkaline medium) and allows colorimetric detection and quantification of protein content. A standard curve was created by diluting a known protein (bovine serum albumin) to defined dilutions and the absorbance was measured at 562nm following addition of the kit reagent (*Novaspec Plus, Amersham Biosciences*). This allows the protein content of unknown lysates to be calculated. 20g of protein per sample was denatured at 95°C for 5 minutes and loaded on the *NuPAGE Gel System* (*Invitrogen*) according to the manufacturer's protocol. Proteins were run alongside a protein ladder, and the gel was transferred onto a PVDF membrane (*Immobilon P, Millipore*) in a semi-dry tank. Blocking was performed by incubation with PBS containing 5% milk. The membranes were incubated overnight in the appropriate primary antibody in PBS with 5% milk (Table 2.2). After washing, the membranes were incubated with the appropriate secondary antibody. After further washes, the blots were incubated in ECL reagents *GE healthcare* and chemiluminescence was detected with a developer (*Compact X4, Xograph*).

2.1.5.3 Immunohistochemistry

Formalin-fixed, paraffin-embedded tissue sections were de-waxed in xylene and rehydrated through graded concentrations of alcohol to water for 5 minutes each. En-

Table 2.2: Antibodies used for western blotting

Primary Antibody	Manufacturer	Dilution	Block	Secondary
SGK1 (07-315)	EMD Millipore	1 in 300	Goat	Goat anti-rabbit
AKT (4691L)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
c-MYC (5605T)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
p44/42 MAPK (4695S)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
Phos-MYC- T58 (ab185655)	Abcam	1 in 1000	Goat	Goat anti-rabbit
Phos-MYC- S62 (ab185656)	Abcam	1 in 1000	Goat	Goat anti-rabbit
TIMP (ab211926)	Abcam	1 in 1000	Goat	Goat anti-rabbit
CYTH1 (ab154847)	Abcam	1 in 1000	Goat	Goat anti-rabbit
Snail (3879P)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
Slug (9585P)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
RPTOR (2280t)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
Debtor (ab11816)	Abcam	1 in 1000	Goat	Goat anti-rabbit
FOXF1 (ab23194)	Abcam	1 in 1000	Goat	Goat anti-rabbit
4EBP1 (9644T)	Cell signaling	1 in 1000	Goat	Goat anti-rabbit
S6K (ab32529)	Abcam	1 in 1000	Goat	Goat anti-rabbit

dogenous peroxidase was blocked using a 1.6% Hydrogen peroxidase block for 20 minutes. Sections were boiled in a pressure cooker in 10mM citrate buffer (pH 6.0) for 5 minutes for antigen retrieval. Sections were blocked with 5% goat serum for 30 minutes. Slides were incubated with the primary antibody overnight at 4°C before washing with PBS prior to the application of the secondary antibody (1 hour at room temperature). Sections were washed in PBS and incubated in ABC reagent that contains avidin and biotinylated horseradish peroxidase for 30 minutes prior to washing. DAB solution was applied for between 2-5 minutes depending on the development of the colour reaction that was monitored microscopically. Slides were washed in PBS, counterstained with Haematoxylin, dehydrated, and mounted with a coverslip using PDX.

2.1.6 Co-Immunoprecipitation

Cells were grown to confluence on T175 flasks, and washed once with PBS at room temperature. 500μl of Bezonase lysis buffer (20mM HEPES, 40mM KCl, 2mM MgCl2,

Table 2.3: Antibodies used for immunohistochemistry

Primary Antibody	Manufacturer	Dilution	Block	Secondary
SGK1 (07-315)	EMD Millipore	1 in 100	Goat	Goat anti-rabbit
S6K (ab32529)	Abcam	1 in 200	Goat	Goat anti-rabbit
MTOR (ab32028)	Abcam	1 in 50	Goat	Goat anti-rabbit
p-AKT (9275)	Cell signaling	1 in 500	Goat	Goat anti-rabbit
SNAIL (3879)	Cell signaling	1 in 300	Goat	Goat anti-rabbit
SLUG (9585)	Cell signaling	1 in 300	Goat	Goat anti-rabbit
TIMP1 (ab211926)	Abcam	1 in 1000	Goat	Goat anti-rabbit
CYTH (ab154847)	Abcam	1 in 1000	Goat	Goat anti-rabbit

10% Glycerol, 0.5% NP-40, 2x Roche EDTA-Protease inhibitor, 0.5x Sigma phosphatase inhibitor P0044/P5726, 50U/ml Benzonase) was added to the flask and cells were removed using cell scrapers. The cell lysates were incubated on ice for 15 minutes, KCl was added to bring the final concentration to 450mM. Protein lysates were incubated on a rotating wheel (10rpm) for 30 minutes at 4°C. The lysate was clarified by spinning at 13000g at 4°C for 15 minutes. Co-IP protein A Dynabeads (*ThermoFisher*) were prepared. 20μL of the beads were used per reaction and washed three times with wash buffer (20mM HEPES, 100mM KCl, 10% Glycerol, 0.5mM DTT, 1x Roche EDTA-free-Protease inhibitor) and retrieved by binding to a magnetic plate. The beads were resuspended in 20μL of wash buffer and 1ug of antibody was added. The beads were incubated at room temperature with the antibody, with agitation every 2 minutes for a total of 30 minutes. After the antibody was bound to the beads, the magnetic beads were washed 3 times with wash buffer at room temperature. 2mg of the protein extract was used to perform the co-immunoprecipitation protocol. Protein extracts were diluted to 150mM of salt by adding 2x volume of no-salt Equilibration Buffer (20mM HEPES, 10% Glycerol, 0.5mM DTT, 1x Roche EDTA-free protease inhibitor, 0.5mM EDTA, 0.5x Sigma phosphatase inhibitor cocktail P0044/P5726). The low salt solution containing the protein lysates was mixed and equilibrated on ice for 15 minutes. 20μL of the bead-antibody conjugates were

added to the solution and incubated on a rotating wheel at 4 °C for 2 hours. The beads containing the bound target protein and protein-protein complexes were washed twice in wash buffer, and transferred into a new eppendorf. This was to prevent the detection of any proteins that were bound to the eppendorf. The magnetic beads were washed a further two more times, before finally being resuspended in 40μL of 1x SDS loading buffer. The beads suspended in 1x SDS loading buffer were boiled for 3 minutes at 95°C to elute off the denatured proteins. The beads were removed from the solution using the magnetic plate and a western performed as described above.

2.1.7 Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) allows an analysis of the interactions between proteins and DNA in the cell. Cells were grown to confluence on a T175 flask. Cross-linking of cellular proteins to DNA was achieved through incubation at 10 minutes at room temperature using 1% formaldehyde. The fixation reaction was quenched using 1.5mls of 2.5M glycine and cells were collected in cold PBS. Cells were lysed using 1% SDS, 10mM EDTA and 50mM Tris-HCL (10 minutes) and underwent sonication to generate fragments of 1000-1500bps (15 secs x 15). Protease inhibitor, phenylmethylsulfonyl fluoride and *cOmplete mini tablets* (*Roche*) were used to prevent protein degradation. The DNA was diluted and protein-DNA complexes were incubated with 5μL of antibody overnight at 4°C. Immunoprecipitation was performed using Protein A Dynabeads with a 4 hour incubation at 4°C. The beads were washed with TSE 1 (1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, 150mM NaCl, 0.1% SDS), TSE II (1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, 50mM NaCl, 0.1% SDS) and wash Buffer III (0.25M LiCl, 1mM EDTA, 10mM Tris-HCL, 1% NP-40, 1% deoxycholate). The immunoprecipitated DNA was extracted by the use of DNA extraction solution (1% SDS, 0.1M NaHCO3) and an overnight incubation at 65°C overnight. Purification of DNA was performed using the Qiaquick spin PCR

purification Kit (*Qiagen*). rt-qPCR was performed using SYBR Green real time PCR master mix (*Thermo Fisher*). SYBR green probes had been optimised through an analysis of the dissociation curve to ensure homogeneity and specificity of the PCR products.

2.1.8 Flow cytometry

Fluorescence activated cell sorting (FACS) was performed on trypsinised cells that had been filtered through an 80um filter to select for single cells. FACS was performed on a BD Aria III flow cytometer. Cells were selected for the highest 5% of Zsgreen using the Alexa 488 filter.

2.1.9 Functional assays

2.1.9.1 Cell proliferation assay

Cell proliferation is a core cancer prerequisites attribute required for a tumour to develop. Proliferative assays measure the number of cells or the change in proportion of cells that are dividing. Cell proliferation may be determined by measuring DNA synthesis, metabolic activity, antigens associated with cell proliferation or ATP concentration. However, my preferred method is using the *CellTitre-Glo Luminescent cell assay* (*Promega*), an assay where the amount of ATP present, and thus the amount of cells present is proportional to the luminescent signal generated in the assay. Cells were detached with trypsin-EDTA (*Gibco*) and then centrifuged at 1000rpm for 5 minutes in a tabletop centrifuge to collect cells. They were plated out in a 24 well plate at a concentration such that at 72 hours, they are not more than 80% confluent, and in practice this was between 20,000 and 100,000 cells depending the cell line and mutation of interest. At 72 hours, 300μL of the *CellTiter-Glo* reagent was added to each well and luminescence was recorded on the *FLUOstar Omega* (*BMG Labtech*) 10

minutes later. Quantification was performed using the *MARS data analysis software* (*BMG Labtech*).

2.1.9.2 Migration assay- Boyden Chamber Migration Assay

The ability of cells to gain the ability to migrate is a multistep process that plays an important role in the progression of cancer. The migration assay that I have utilised involves the use of the *Boyden Chamber Migration Assay*. Cells were detached with trypsin-EDTA (*Gibco*) and the centrifuged at 1000rpm for 5 minutes in a tabletop centrifuge to collect cells. Cells were washed once and resuspended in FBS free media at a concentration of 100,000 cells/100μL. Transwell permeable supports with an 8.0um polycarbonate membrane were placed in a 24 well plate (*Costar*) and 100μL of cell/FBS free media was placed on the top of a transwell plate. 600μL of normal cell media was placed on the bottom of the well. 16 hours later, transwell permeable supports were removed and cells from the upper chamber were removed by mechanical detachment. Cells that were able to migrate through the transwell permeable supports to the lower surface were stained using Quik-diff staining solution (*Thermo*). Cells were visualised on a Nikon TE2000u inverted wide-field microscope and cell counts were obtained by using the automatic feature counting software within the *NIS-Elements imaging software* (*Nikon*).

2.1.10 Animal models

2.1.10.1 Breeding and maintenance

Animals were maintained in individual ventilated cages under specific pathogen free conditions and food and water were supplied ad libitum. Housing and all procedures involving animals were performed according to protocols approved by the United Kingdom Home Office PPL 30/3311, in compliance with institutional and national guidelines (Animals (Scientific Procedures) Act/ home office guidelines) on animal

welfare.

All experiments were designed to adhere to the principles of the 3Rs (Replacement, Reduction, Refinement). Wherever possible, I attempted to utilise methods which avoided or replaced the use of animals. I also performed pilot experiments to gain an idea of the effect size and thus minimise the number of animals used per experiment. Finally, I kept logs, audits of my experiments and my reflections and this allowed me to develop new methods that enabled me to minimise suffering and thus improve animal welfare.

2.1.10.2 Orthotopic transplantation model- (OTM)

The Endoscopic guided Orthotopic mouse Transplantation Model (OTM) was performed in animals of at least 6 weeks of age.

Immunocompromised strains were selected for the experiments performed in this thesis NOD.CB17-Prkdc<scid>J (NOD scid) or NOD.Cg-Prkdc<scid>Il2rg<tm1Wjl>SzJ (NOD scid gamma).

Animals were restrained as per the methods outlined in the NC3R guidelines to encourage defecation of faeces and anaesthetised using 3% isofluorane. A mouse endoscope system was used to localise an area of the descending colon with good views and scarcity of blood vessels (*Karl Storz coloview system*). A 14g needle was introduced into the instrument channel of the endoscope and a submucosal injection of a cell suspension of volume 50μL was performed with the needle bevel faced down. A similar protocol has been published in the literature[70].

2.1.10.3 Tail vein transplantation

Tail vein transplantation was performed in animals of at least 6 weeks of age. Immunocompromised strains were selected for experiments (NOD scid or NOD scid gamma). Animals were restrained and a cell suspension of 50μl with between 1-2

million cells (depending on cell line) was introduced directly into the lateral tail vein. In-vivo imaging was performed 24 hours later to confirm success of the procedure.

Nod scid gamma mice carry two mutations, the first in the SCID gene and the second in the IL2 receptor common gamma chain. This renders the mice B and T cell deficient, but also NK deficient (as a result of impaired signalling through the IL2 receptor). These mice were used where feasible. Unfortunately, this cohort of mice were prone to high rates of death of neonates and infants, and so for later experiments, I also used Nod Scid mice. These mice have do not have the IL2 receptor mutation and whilst they are B and T cell deficient, they maintain NK function.

2.1.10.4 Harvest of single cells from organs for flow cytometry analysis

Organs were resected and macroscopically dissected into small pieces 1mm in size. Tissues were incubated with Dispase I for 1 hour at 37oC in an orbital shaker. Cells were filtered through an 80-micron filter to select for single cells.

2.1.10.5 In-vivo imaging

In-vivo imaging was performed using the *IVIS Spectrum in-vivo Optical Imaging system* (*Perkin Elmer*). In-vivo imaging was performed in animals of at least 6 weeks of age. An intraperitoneal injection of 100μL of D-luciferin (1.5g) was administered using a 20g needle. Animals were anaesthetised with 3% isofluorane and imaged as per the manufacturer's instructions. For colon and whole body imaging, the animals were imaged at 7.5 minutes. For lung imaging, the animals were imaged at 10 minutes. Following the acquisition of the image, animals were released into a recovery cage and monitored as per the project license protocols. To improve the accuracy of the lung metastasis readings, an endpoint ex-vivo image of the lungs was acquired. Briefly, following in-vivo imaging, animals underwent schedule 1 methods, then a midline incision was performed to remove the lungs. Lungs were imaged separately

at 10 minutes. Quantification of signal was performed using the *IVIS Lumina Living Image software* with the luminescence measured in radiance (p/sec/cm^2/sr).

2.1.10.6 In-vivo drug dosing

Following tumour transplantation and one day after the tumours were observed in the lungs of mice, chemotherapeutic agents or drugs were administered to mice. Sirolimus was diluted in DMSO and administered intraperitoneally (i.p) at a dose of 1.5mg/kg every 2 days (*Stratech*). LY294002 was diluted in DMSO and administered i.p at a dose of 50mg/kg twice a week (*Cambridge Bioscience*). Control mice received no chemotherapy, only DMSO vehicle. The mice were observed for 4 weeks and then underwent schedule 1 methods.

The volume of DMSO utilised was 50μL.

2.1.10.7 Histological preparation

Tissue specimens were placed in a histology cassette and processed using a *Histomaster machine* (*Bavimed*) using the following protocol: 70% ethanol for 45 minutes, 95% ethanol for 30 minutes, 95% ethanol for 45 minutes, 100% ethanol for 30 minutes, 100% ethanol for 45 minutes, Histoclear for 30 minutes, Histoclear for 30 minutes, Histoclear for 30 minutes, wax for 30 minutes, wax for 45 minutes, wax for 45 minutes. Processed samples were embedded in paraffin wax using a paraffin embedding station (*EG1150H, Leica*). Tissues embedded in paraffin blocks were sectioned using a microtome (*RM2255, Leica*). For immunohistochemistry, 4um slices were obtained. For DNA extraction, 10um slices were obtained.

2.1.10.8 Histological analysis

Haematoxylin and Eosin staining was performed on paraffin sections. Sections were dewaxed in xylene for 5 minutes, rehydrated through graded ethanol solutions (100%,

90%, 70%) for 5 minutes each followed by 2-minute wash in tap water. Sections were stained in Harris Haematoxylin for 30 seconds and washed in running water. The slides were briefly dipped in acid alcohol and left in water for 5 minutes. Sections were stained in Eosin for 3 minutes, followed by a brief water wash. Sections were dehydrated through graded ethanols (70%, 90%, 100%) for 2 minutes each prior to mounting using DPX.

2.1.11 CRISPR GeCKO genetic screen- (GECKmA)

2.1.11.1 Generation of lentiviral library

Generation of the GeCKOv2 mouse lentiviral library A (GECKmA) was performed as per the protocol described by Shalem[85 86] with a few modifications.

The mouse CRISPR knockout pooled library was purchased from Addgene (1000000052) and amplified by a collaborator (Dr Andrew Bassett) according to the protocol described above.

In brief, 10 T-175 flasks of HEK293T cells were each transfected with 32ug of the lentiCRISPR plasmid library, 9ug of VSV-g and 16.25ug of dr8.74 using Plus reagent and Lipofectamine 2000. After 72 hours, the media containing the viral supernatant was removed, filtered then ultracentrifuged at 24,000rpm for 2 hours at 4°C to obtain the concentrated viral stocks.

2.1.11.2 Transduction of cells

Cells were transduced with optimal virus volumes to establish a MOI of 0.3-0.5. This was achieved by plating 1 million cells in a 12 well plate, supplementing with 8ug/ml polybrene, titrating the virus amount and performing spinoculation. The next morning, cells were split with half receiving puromycin (4μl/ml) and half receiving normal media. Transduced cells were puromycin resistant as the lentiviral library contained a puromycin resistant tag, and this allows percentage transduction to be

calculated. The virus volume yielding a MOI closest to 0.5 was selected for large-scale screening. To maintain a 100x representation for large scale screening, 7 million cells were transduced with the lentiCRISPR library virus.

2.1.11.3 Extraction of DNA

The tissues from the experimental mice were placed in cryovials that were placed in liquid nitrogen and stored for a period at -80°C. To extract DNA from these tissues, cryopulverization was utilised. This method relies on the fact that as tissue has high water content, tissues becomes as brittle as glass at liquid nitrogen temperatures. We used the Biopulverizer (*BioSpec*) and DNA was extracted using sodium chloride salt precipitation (NaCl 75mM, EDTA 25mM, Tris 10mM, 1% SDS-page).

2.1.11.4 GeCKO targeted sequencing

Targeted DNA sequencing was performed using a semi-nested PCR strategy as described in the aforementioned protocols. First stage PCR was performed using the following lentiCRISPR gRNAs. F1 tcttgtggaaaggacgaaacaccg R1 tctactattctttcccctgcactgt . To maintain genomic DNA coverage at 300x, 13 separate 100μL reactions were performed with 10ug of genomic DNA using the Hercules II Fusion DNA polymerase (*Agilent*). The first phase PCR was performed for 18 cycles. The second phase PCR was a semi-nested PCR strategy and was used to insert Illumina sequencing barcodes, a sequence stagger and sequencing primers. This second phase was performed for 24 cycles. DNA was purified, normalised, pooled and quality of DNA assessed using Tapestation. Sequencing was performed using the HiSeq 2500 using 30% PhiX spike-in.

2.1.12 Human specimens and tissue

2.1.12.1 Tissue Microarrays

Tissue Microarrays were obtained from *Insight Biotechnology Ltd*, who are the UK suppliers of *US Biomax*. The tissue microarrays used were CO484a, CO485. These are histological sections of 48 colon cancer specimens with information on specimen TNM, clinical stage and pathological grade. Scoring for SGK1 was performed on IHC of Tissue microarrays and staining was scored on a scale of 1-10 for each specimen. Differentiation status of tumours were graded from 0-10, with >7 graded well differentiated, 3-7 moderately differentiated and 0-2 poorly differentiated.

2.1.12.2 Oxford Stage III CRC-Life expectancy, adverse Events and Relapse cohort- OSLER sample set

This cohort consists of biospecimens collected from newly diagnosed patients with stage III sporadic colon adenocarcinoma who had undergone surgical resection and had received no prior treatment for their disease at the Oxford University Hospitals between March 2013 to May 2015. All patients had given general consent for analysis of their specimens and this was performed under ethics reference 11/SC/0236. Cases were staged according to the European Society of Medical Oncology (ESMO) guidelines. Patients were excluded from this analysis if there was a previous history of CRC, any predisposing conditions for CRC (such as IBD, polyposis syndromes), patients presenting with synchronous CRC lesions or if there was evidence of inherited CRC predisposition syndromes. Specimens were processed and embedded in formalin-fixed paraffin-embedded (FFPE) blocks and stored at temperatures in the range of 17-22°C. Histological sections were reviewed by a consultant histopathologist to confirm that the tumour was consistent with a diagnosis of colon adenocarcinoma and representative regions were marked on the primary tissue (contained >60% tu-

mour foci). Histological sectioning of size 5um was performed by the *Oxford Centre for Histopathology Research*. RNA was extracted from the marked regions using the High Pure FFPET RNA Isolation Kit (Roche) as per the manufacturer's guidelines.

2.2 Statistical and Bioinformatics methods

2.2.1 Next generation sequencing and RNA sequencing

Next generation sequencing reads were obtained in fastq format. Quality control and trimming of sequencing reads was performed using QC reports from the automated analysis pipeline from the bioinformatics team at the Genomics Centre (*WTCHG*), using FASTQC. Sequencing reads were demultiplexed using the standard Illumina sequencing software. However, where demultiplexing was not possible (i.e. because multiplexing was performed on the forward as well as reverse primer), I performed this using Cutadapt version 1.9.1[87]. Sequence alignment of reads was performed by the bioinformatics team at the Genomics Centre (*WTCHG*), using HISAT2 version 2.0.4. For RNA sequencing, to perform feature counting, I used HTSeq version 0.6.1 with alignment to a GTF file with features obtained from Ensembl for human assembly version GRCh37. I performed differential comparison by using DESeq version 1.10.1[88]. The package performs normalisation of the counts based on size of sequencing and allows differential expression to be determined from count data through use of the negative binomial distribution and a shrinkage estimator for the distribution's variance.

2.2.2 Gene set enrichment analysis

Gene set enrichment is a technique that is able to identify groups of genes that are over-represented in RNA sequencing differential expression lists. It utilises a priori gene sets that have been grouped together based on features such as relevance to

particular signaling pathway, biological phenotypes or protein localisation. I used two gene set enrichment programs. Principally, I used javaGSEA version 2 for *Gene Set Enrichment Analysis*, (*Broad Institute, Harvard, USA*)[89] using gene sets from the Molecular Signatures Database (MSigDB)[90] and custom curated gene signatures generated from our group. GSEA were performed using pre-ranked gene differential expression lists. The second gene set enrichment program was the Database for Annotation, Visualisation and Integrated Discovery (DAVID) version 6.7/6.8[91]. DAVID uses a different algorithm to GSEA, looking at overlap between the supplied gene list and the curated database and uses Fisher's exact and Chi-squared tests. GSEA uses every datapoint in the RNA sequencing differential expression dataset, i.e. performing an entire gene-expression profile using Wilcoxon and Kolmogorov-Smirnov tests. GSEA contains up to date datasets from MSigDB including Biocarta, Kegg, Reactome, Go and immunologic signatures and was used where gene expression changes were available. DAVID was used where candidate gene lists had been generated.

2.2.3 Hierarchical clustering

Hierarchical clustering was performed using the R package, hclust. This calculation works by assigning each object (i.e. sample GEP) to its own cluster. The hierarchical clustering algorithm uses to *Complete Linkage Method* or *Ward's minimum variance method* to proceed iteratively to join two of the most similar clusters, continuing until there is just a single cluster. The distances between each cluster were calculated by using the Lance-Williams dissimilarity update formula. This formula calculates the differences between new clusters and existing points based on the differences that were present prior to the formation of the new cluster.

2.2.4 Principal component analysis

Principal component analysis (PCA) was performed using the R package, prcomp. PCA is a method that allows the extraction of important variables from a high dimensional data set, in this case, sample gene expression profile (GEP). The aim is to represent this high dimensional dataset as a two dimensional PCA plot. This technique enables a calculation of the first principal component (1st PCA) and plot it against the second principal component (2nd PCA). The 1st PCA is the linear combination of the original genes that captures the maximum variance in the dataset and the 2nd PCA is the linear combination of the original predictors that captures the 2nd most common variance. As each component is identified in an unsupervised fashion, the technique is an unsupervised approach.

2.2.5 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) genetic screen

The CRISPR genetic screen utilises a library of lentiviruses, each containing a synthetic guide RNA (gRNA) that had been designed to target a particular gene with high specificity. In order to determine how gRNA diversity changes under different selection pressures, targeted next generation sequencing was used to amplify the 20bp gRNA sequence. The GECKmA screen contains 67,405 gRNAs. In order to align the FASTQ reads back to the GECKmA gRNA sequences, I demultiplexed the files using Cutadapt version 1.9.1 to identify the barcodes I had assigned to each animal. I trimmed the barcodes so that only the gRNA sequence remained. These fastq files were piped into the MAGeCK Linux bioinformatics program (version 0.5.3). This program is a computational tool that converts the gRNA sequences to genes and performs summary counts for each gene[92].

2.2.6 Public repositories for high-throughput data

Gene expression data from a number of colorectal cancer cohorts is available within the public domain. I obtained the data from two sources. Expression data from primary and metastatic lesions were obtained from the Gene Expression Omnibus (GEO-GSE41258, GEO-GSE24551). I used the GEO2R web tool to download data as this allowed me to interrogate the original data file directly and not rely on the curated Dataset. Expression data from The Cancer Genome Atlas (TCGA) colorectal cancer cohort, was obtained directly from the Broad Institute website (https://gdac.broadinstitute.org). Normalised mRNA sequencing data was used throughout this thesis. The TCGA dataset allows gene expression to be compared between primary and normal tissue and this was performed directly using the Broad Institute Firebrowse web application (www.firebrowse.org).

2.2.7 Gene expression data from Connectivity Map (CMAP)

Connectivity Map (CMAP- *Broad Institute*) is a public repository of genome-wide transcriptional expression data from human cells treated with 1,309 compounds and consists of 7,000 expression profiles. The CMAP version 2.0 interface was used to interrogate similarities of gene expression profiles with those generated from chemicals and bioactive small molecules. The top 500 differentially over-expressed and top 500 differentially under-expressed genes from my experiments were used to generate a list of candidate chemicals and compounds with similar transcriptional expression alterations.

2.2.8 Statistical tests

The statistical tests used were selected to be appropriate for the dataset type and distribution of the data. Variables were classified into categorical, nominal, rank or

ordinal values for both my independent and dependent variables. A small number of pilot experiments were performed to determine if the results followed a Gaussian normal distribution and if this was observed, parametric tests were applied. The use of parametric tests implied that the samples had similar variance. If it was not clear, non-parametric tests were performed. During this thesis, two-tailed tests were performed as I did not make any assumptions about the direciton of deviation from the null hypothesis. For categorical parametric data, a two sided Chi-squared test was performed. For continuous parametrically distributed data, two sided t-tests were performed.

When I compared more than two sets of numerical data, a multiple group comparison test such as one-way analysis of variance (ANOVA) was performed. Where data were grouped, this was performed to enable equal data points to be placed into each category. In the specific cases of gene expression and survival data, all patients were used in the analysis and gene expression was expressed in quintiles. Data sets for survival trends were always considered to be non-parametric and time to event analyses were performed using the Wilcoxon or log-rank test. Survival analyses were performed within the statistical analysis software package SPSS version 23).

2.2.9 Time-to-event analysis

Survival analyses were performed within the Statistical Analysis Software Package (SPSS version 23) using the Kaplain Meier non-parametric statistic and Cox proportional hazards.

2.2.10 Figures and diagrams

Figures were principally generated within GraphPad Prism 7.0a or within Microsoft Excel version 14.2.2. Chord diagrams were generated using the Circos software pack-

age (*BC Cancer Agency, Canada*) using the interface *Circos Table Viewer* version 0.63-9[93].

Chapter 3

Investigation of *Serine/Glucocorticoid Regulated Kinase 1* expression in normal intestinal homeostasis and colorectal tumorigenesis

3.1 Introduction

The unit structure of the large bowel is the crypt, the stem cells at the crypt base produce progeny that expand in number and differentiate as they pass to the lumen into which they are shed. Although the biology of the stem cell niche is relatively well understood, much less is known about the process of differentiation. The involvement of signalling pathways such as Wnt, TGFb/BMP/hedgehog, notch and ephrin is required for production of the main differentiated cell types (colonocyte, neuroendocrine, goblet, Paneth). However, especially for colonocytes, it is unclear to

what extent differentiation is a passive process, for example, resulting from absence of Wnt pathway activity, or an active process in which unknown pathways are activated to drive the expression of genes involved in the functioning of differentiated cells.

Serine/Glucocorticoid Regulated Kinase 1 (*SGK1*) is expressed at high levels in all tissues of the digestive tract including the oesophagus, stomach, small intestine, intestine, liver and pancreas[94]. Expression in normal intestinal epithelium is principally in the upper most part of the crypt, corresponding to the positions of differentiated colonocytes[81] making SGK1 a good candidate gene for the normal functioning of colonocytes.

SGK1 was first described in 1993 as a gene that was activated in response to glucocorticoids in mammary tumour epithelial cells[95]. It has high sequence similarity with another notable kinase, Protein Kinase B (also known as *AKT1*), with substantial overlap in substrate specificity[96]. Its activity is regulated both at a transcriptional and post-translational level. The *SGK1* promoter contains a number of transcription factor binding sites and expression is regulated by transcription factors. These transcription factors are up-regulated in response to glucocorticoids, growth factors, TP53 signalling, osmotic stress, FSH, heat shock and ischaemic injury[97]. There exist alternative translation initiation sites and four *SGK1* isoforms have been described with the long isoforms 47 and 49kDa most abundant and found in the endoplasmic reticulum. The shorter isoforms 42 and 45kDa are found in the cytoplasm and nucleus[98]. Phosphorylation of *SGK1* is required at a threonine in the catalytic domain (T256) and a serine in the hydrophobic motif (S422) for the protein to be fully active[99]. This is principally mediated by 3-phosphoinositide dependent protein kinase 1 (PDK1) in a fashion similar to *AKT*, but it can also be mediated by the mTORC2 complex[100] and p38 MAPK at Ser78.

Historically, the gene was thought to primarily play a role in the activation of ion channels including EnaC, SCN5A, TRPV4-6, ROMK, kv1.3, kv1.5, kv4.3, KNCE1, KCNQ1, KCNQ5, ASIC1,GlUR6, ClCKa/barttin, ClC2 and CFTR. Together these channels mediate diverse effects including hormone release, inflammation, cell proliferation and apoptosis[101]. *SGK1* has been shown also to phosphorylate a number other important proteins such as NEDD4-2[102], GSK3b, p27, MDM2[103], ERK[104], B-RAF[105] and the pro-apoptotic transcription factors FOXO3[106].

The function of *SGK1* in cancer has yet to be clearly defined. *SGK1* is down-regulated in many different cancer types, particularly in carcinomas of the breast, the GI tract, lung and prostate[107]. In breast cancer models, despite these observations, it is generally believed that *SGK1* functions as an oncogene. *SGK1* has been demonstrated in some models to activate the PI3/AKT/MTOR signalling pathway, a key oncogenic network for cell proliferation and growth, by phosphorylation and inactivation of TSC2[108]. *SGK1* also phosphorylates other proteins involved in cell cycle progression and apoptosis, including FOXO3[106] and p27[105 103]. In myeloma cell lines, growth rates are decreased by *SGK1* knockdown using shRNAs[109], and this effect may be observed in a proportion of AKT-inhibitor-resistant breast cancer cell lines[110].

SGK1 has contrasting effects in different cell models and this is best exemplified by its effects on the RAS-RAF-ERK pathway. On one hand, it phosphorylates and promotes ERK1/2 function[104], on the other it phosphorylates and inactivates B-RAF thus reducing cell proliferation driven by the RAS/RAF/ERK pathway[111]. The effects of *SGK1* on proliferation may not even be particularly marked as *SGK1* knockout mice have no tumour predisposition and are phenotypically normal, with

the exception of decreased urinary salt loss on a modified diet[112]. To summarise, *SGK1* targets several important oncogenic signalling networks, but, its effects appear to be dependent on the type and model of tumour involved.

Unpublished work in our laboratory to date has focused on the function of *SGK1* using CRC cell line models. This work has identified that agents that cause differentiation of CRC cell lines, such as sodium butyrate, induce expression of *SGK1*. Re-expression of *SGK1* results in the expression of proteins consistent with differentiation of CRC cell lines (Zo-1, β-catenin) and leads to decreased proliferation and a failure of these cell lines to survive anchorage-independent growth. These findings suggest that *SGK1* may act as a differentiation driver and a tumour suppressor gene in CRC.

There have been few studies into the effect of *SGK1* on other features of tumorigenesis, such as cellular invasion or metastasis. In a non-small cell lung cancer cell line, *SGK1* knockdown was reported to reduce metastasis rates in mouse models by modulating β-catenin signalling[113]. Constitutively active *SGK1* mutants in Caco-2 colon tumour cells and HEK293T embryonic kidney cells both demonstrated enhanced motility mediated by viniculin[114].

The aims of this chapter are to i) understand the role of *SGK1* in human CRC cohorts, ii) identify if there was any prognostic significance of the expression of *SGK1* in cancers, iii) identify the mechanism through which *SGK1* exert its tumour suppressor effect and iv) develop a new mouse model to investigate its effect on CRC metastasis.

3.2 Chapter Methods

3.2.1 Lentivirus Usage

3.2.1.1 Lentivirus generation

Lentivirus cell culture was carried out in a designated class II laminar flow hood and usage was carried out according to standard group operating procedures, specifically with regards to personal protective clothing and disposal of viral waste. HEK293T cells were maintained in a T175 flask in Iscove's Modified Dulbecco's medium (IMDM) (*Gibco*) supplemented with 10% FBS and 1% penicillin-streptomycin. The cells were cultured until they reached 50-70% confluence. The media was changed prior to transfection. For generation of lentiviruses, I used the Calcium Chloride transfection method to generate 2nd generation lentiviruses. The following plasmids were required for viral packaging; an envelope glycoprotein-encoding plasmids (pMD2.VSVG) which contained the *Env* gene and the packaging plasmid (pCMV.dR8.74) which contained the *Gag, Pol, Rev, Tat* genes. The following plasmids were combined in a 50ml falcon tube; 16.25ug of pCMV.dR8.74, 9ug of VSV-g, 32ug of transfer vector (or 25ug of transfer vector if the transgene and promoter was less than 1500bp). These plasmids were combined with 750μL of 0.1% sterile filtered TE, 375μL of deionised water and 125μL of calcium chloride at a concentration of 2.5M. The falcon tube was vortexed and 1250μL of HBSS was added in a drop-wise fashion to form a DNA-calcium chloride co-precipitate that has the ability to adhere to the cell surface of the HEK293T and is taken by the cells by endocytosis. This solution was added to the media of the T175 flask containing the HEK293T cells. Cells were incubated in this solution for 16 hours before the media was removed and fresh IMDM added. At 72 hours, the virus was harvested. This was achieved by removing the IMDM media that contains the virus particles. This was centrifuged at 1000g for 5 minutes to remove stray cells and filtered with a 0.45uM filter. Virus particles were stored in 2ml screw

top tubes and stored at -80°C. If super-concentrated virus particles were required (>300x concentration), an ultracentrifugation step was performed. The supernatant was ultracentrifuged (*Sorvall*) at 24,000rpm for 2 hours at 4°C and resuspended overnight at room temperature with PBS.

3.2.1.2 Lentivirus quantification

Lentivirus supernatant titers were quantified using the *Lenti-X p24 Rapid Titer kit* (*Clontech*) following manufacturer's instructions. This kit is an ELISA-based protocol that assesses p24 capsid protein content, an HIV-1 virus core capsid protein by anti-p24 coated wells.

3.2.1.3 Lentivirus transduction

Cells due to be transduced with lentivirus were plated between 50-100,000 cells in a 24 well plate. When the cells had adhered to the plate, the media was supplemented with polybrene (Hexadimethrine bromide) at a concentration of 8ug/ml. Polybrene increases retrovirus gene transfer by enhancing receptor-independent virus adsorption on target cell membranes. Transductions were performed with a virus titration of between 10-400μL of virus supernatant. Target cells were spinoculated to further increase virus adsorption. Spinoculation involves centrifuging cells at 1000g at room temperature for 1 hour. Following spinoculation, the virus supernatant was removed and cells maintained in normal cell media. After 72 hours following transduction, cells either were utilised for experiments, or underwent antibiotic selection by adding the appropriate antibiotic, or visualised for the presence of fluorescent markers.

3.2.1.4 Orthotopic transplantation model- (OTM)

The Endoscopic guided Orthotopic mouse Transplantation Model (OTM) was performed in animals of at least 6 weeks of age. Immunocompromised strains were se-

lected for the experiments performed in this thesis (NOD scid or NOD scid gamma). Animals were restrained as per the methods outlined in the NC3R guidelines to encourage defecation of faeces and anaesthetised using 3% isofluorane. A mouse endoscope system was used to localise an area of the descending colon with good views and scarcity of blood vessels (*Karl Storz coloview system*). A 14g needle was introduced into the instrument channel of the endoscope and a submucosal injection of a cell suspension of volume 50μL was performed with the needle bevel faced down. A similar protocol has been published in the literature[70].

3.2.1.5 Migration assay- Boyden Chamber Migration Assay

The ability of cells to gain the ability to migrate is a multistep process that plays an important role in the progression of cancer. The migration assay that I have utilised involves the use of the *Boyden Chamber Migration Assay*. Cells were detached with trypsin-EDTA (*Gibco*) and the centrifuged at 1000rpm for 5 minutes in a tabletop centrifuge to collect cells. Cells were washed once and resuspended in FBS free media at a concentration of 100,000 cells/100μL. Transwell permeable supports with an 8.0um polycarbonate membrane were placed in a 24 well plate (*Costar*) and 100μL of cell/FBS free media was placed on the top of a transwell plate. 600μL of normal cell media was placed on the bottom of the well. 16 hours later, transwell permeable supports were removed and cells from the upper chamber were removed by mechanical detachment. Cells that were able to migrate through the transwell permeable supports to the lower surface were stained using Quik-diff staining solution (*Thermo*). Cells were visualised on a Nikon TE2000u inverted wide-field microscope and cell counts were obtained by using the automatic feature counting software within the *NIS-Elements imaging software* (*Nikon*).

3.2.2 Human specimens and tissue

3.2.2.1 Tissue Microarrays

Tissue Microarrays were obtained from *Insight Biotechnology Ltd*, who are the UK suppliers of *US Biomax*. The tissue microarrays used were CO484a, CO485. These are histological sections of 48 colon cancer specimens with information on specimen TNM, clinical stage and pathological grade. Scoring for SGK1 was performed on IHC of Tissue microarrays and staining was scored on a scale of 1-10 for each specimen. Differentiation status of tumours were graded from 0-10, with >7 graded well differentiated, 3-7 moderately differentiated and 0-2 poorly differentiated.

3.2.2.2 Oxford Stage III CRC-Life expectancy, adverse Events and Relapse cohort- OSLER sample set

This cohort consists of biospecimens collected from newly diagnosed patients with stage III sporadic colon adenocarcinoma who had undergone surgical resection and had received no prior treatment for their disease at the Oxford University Hospitals between March 2013 to May 2015. All patients had given general consent for analysis of their specimens and this was performed under ethics reference 11/SC/0236. Cases were staged according to the European Society of Medical Oncology (ESMO) guidelines. Patients were excluded from this analysis if there was a previous history of CRC, any predisposing conditions for CRC (such as IBD, polyposis syndromes), patients presenting with synchronous CRC lesions or if there was evidence of inherited CRC predisposition syndromes. Specimens were processed and embedded in formalin-fixed paraffin-embedded (FFPE) blocks and stored at temperatures in the range of 17-22°C. Histological sections were reviewed by a consultant histopathologist to confirm that the tumour was consistent with a diagnosis of colon adenocarcinoma and representative regions were marked on the primary tissue (contained >60% tu-

mour foci). Histological sectioning of size 5um was performed by the *Oxford Centre for Histopathology Research*. RNA was extracted from the marked regions using the High Pure FFPET RNA Isolation Kit (Roche) as per the manufacturer's guidelines.

3.3 Results

3.3.1 *SGK1* expression in normal specimens

Previous work performed by our lab had identified that *SGK1* was predominantly expressed by differentiated cells at the top of the mouse intestinal crypt. I validated these findings using human CRC specimens. I obtained and analysed human colonic tissue employing a number of methods. Firstly, I performed immunohistochemistry. The images confirmed that there was a gradient from the crypt base to the crypt tops (Fig 3.1). To quantify this gradient, the expression of *SGK1* expression in colonic specimens was measured by rt-qPCR. Fresh biopsy specimens of colonic tissues were obtained from 5 patients and the crypts tops were dissociated from the crypt bottoms by laser capture dissection and RNA was extracted. rt-qPCR was performed on these specimens. The *SGK1* Taqman efficiency was determined to be 99.1% (Fig 3.4) denoting a probe with a linear standard curve and good accuracy. There was on average a 20-fold increase in *SGK1* mRNA expression in the crypt tops compared to the crypt bottoms (two-sided t test, p=0.037). This expression pattern supports our previous findings that *SGK1* is a gene that functions as a marker of differentiation in colonocytes.

Figure 3.1: Representative expression of *SGK1* determined by immunohistochemistry in normal tissue. Representative figures from Human Protein Atlas and my sample set demonstrating increasing expression from bottom to top of crypt.

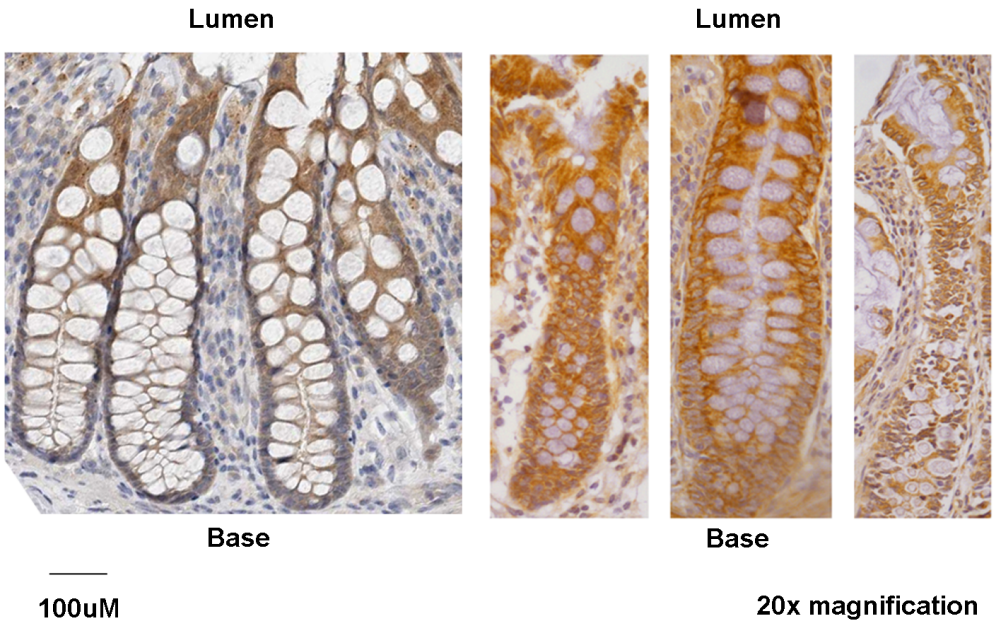
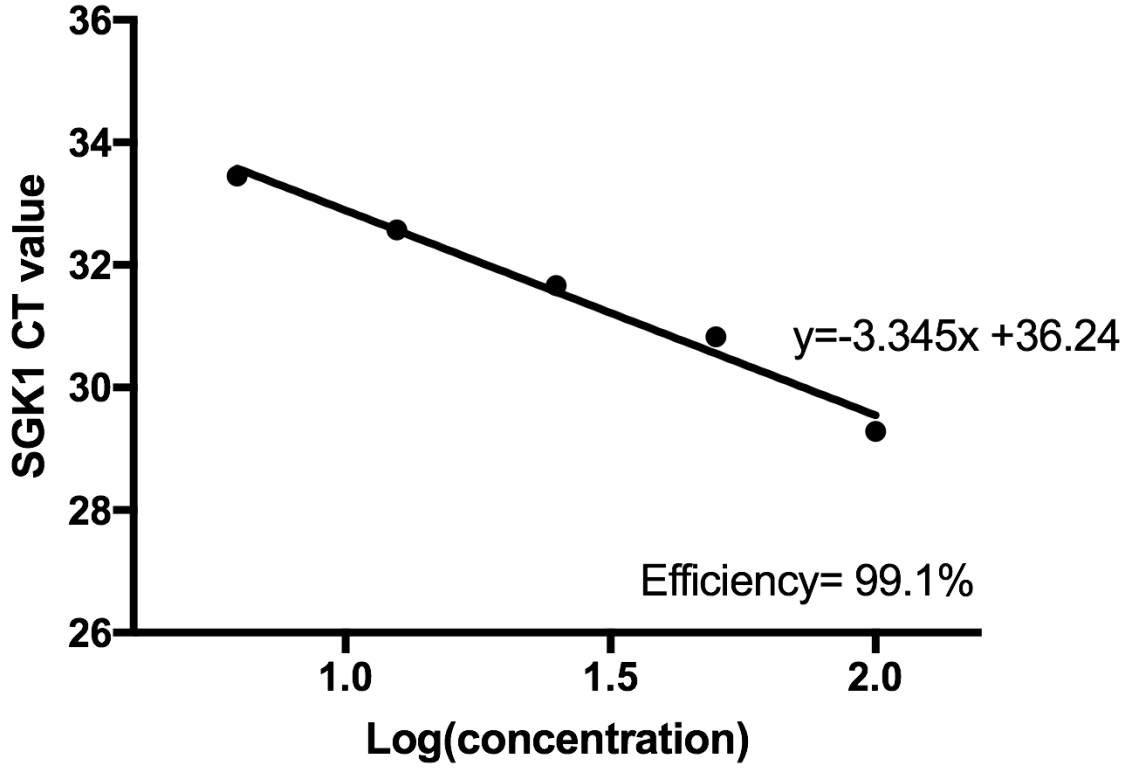


Figure 3.2: Efficiency of *SGK1* taqman probes was determined by 8 serial dilutions of cDNA and use of the CT slope method



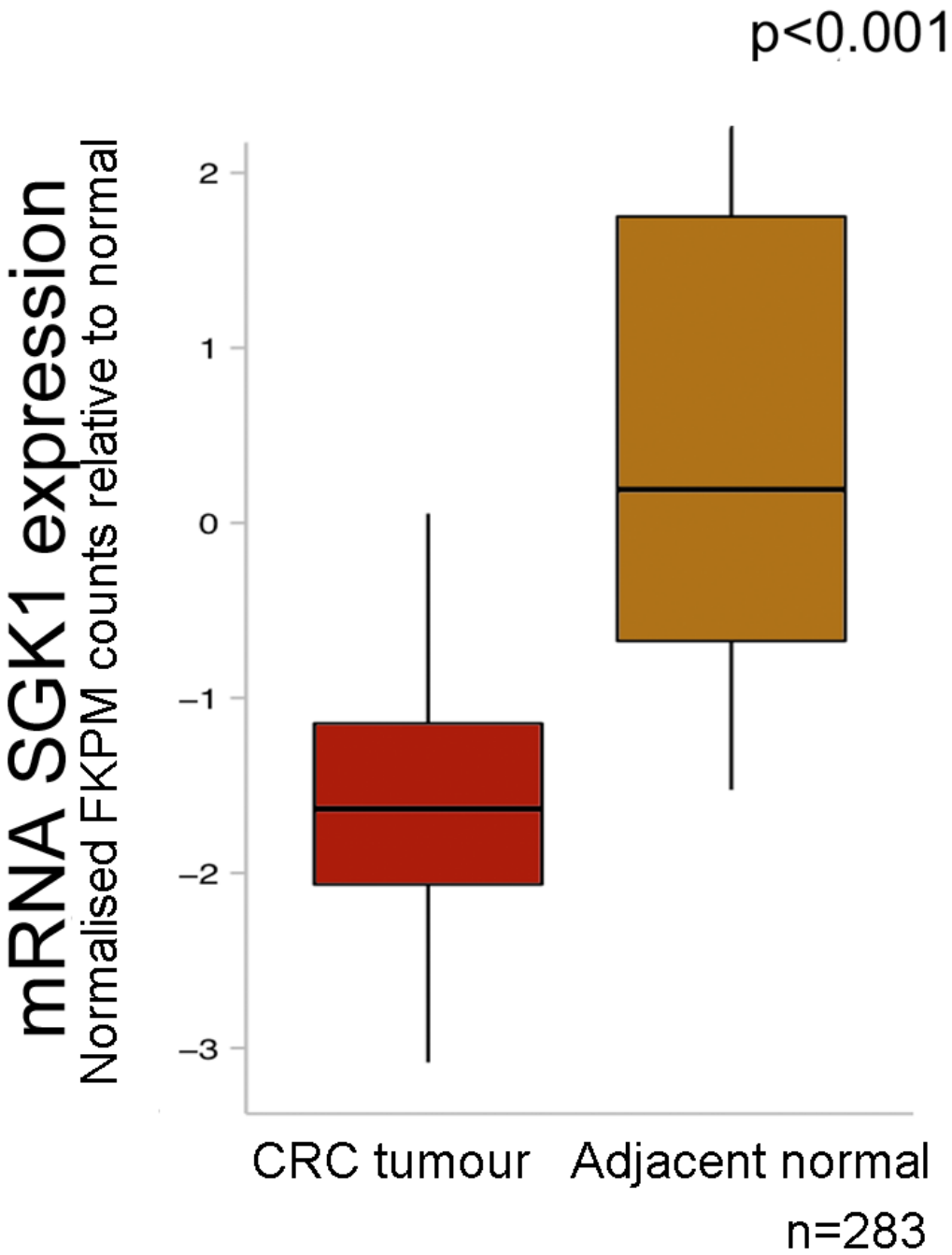
3.3.2 *SGK1* expression in CRC

Tumour cells demonstrate many features similar to undifferentiated cells, including the ability to migrate, invade and proliferate. I therefore set out to identify if *SGK1* expression might be down-regulated in tumour specimens.

To gain an understanding of the differences in expression of *SGK1* in tumour compared with normal specimens, I took two complementary approaches. Firstly, I obtained a tissue microarray (TMA) of 86 primary CRC specimens of various stages and grades (US Biomax, C0484/485) and compared these with normal controls using immunohistochemistry. There was significantly decreased expression of SGK1 protein

in tumours vs. normals (IHC score 4.2 vs. 9.0, see methods for scoring methodology, two-sided t test p=0.04). I also analysed CRC *SGK1* mRNA expression data from the TCGA cohort. This is an American cohort of 283 patients from whom RNA sequencing data is publicly available. This confirmed that there was a 3 fold decreased expression ($p<0.001$) of *SGK1* in tumours compared to normal tissues (Fig 3.3).

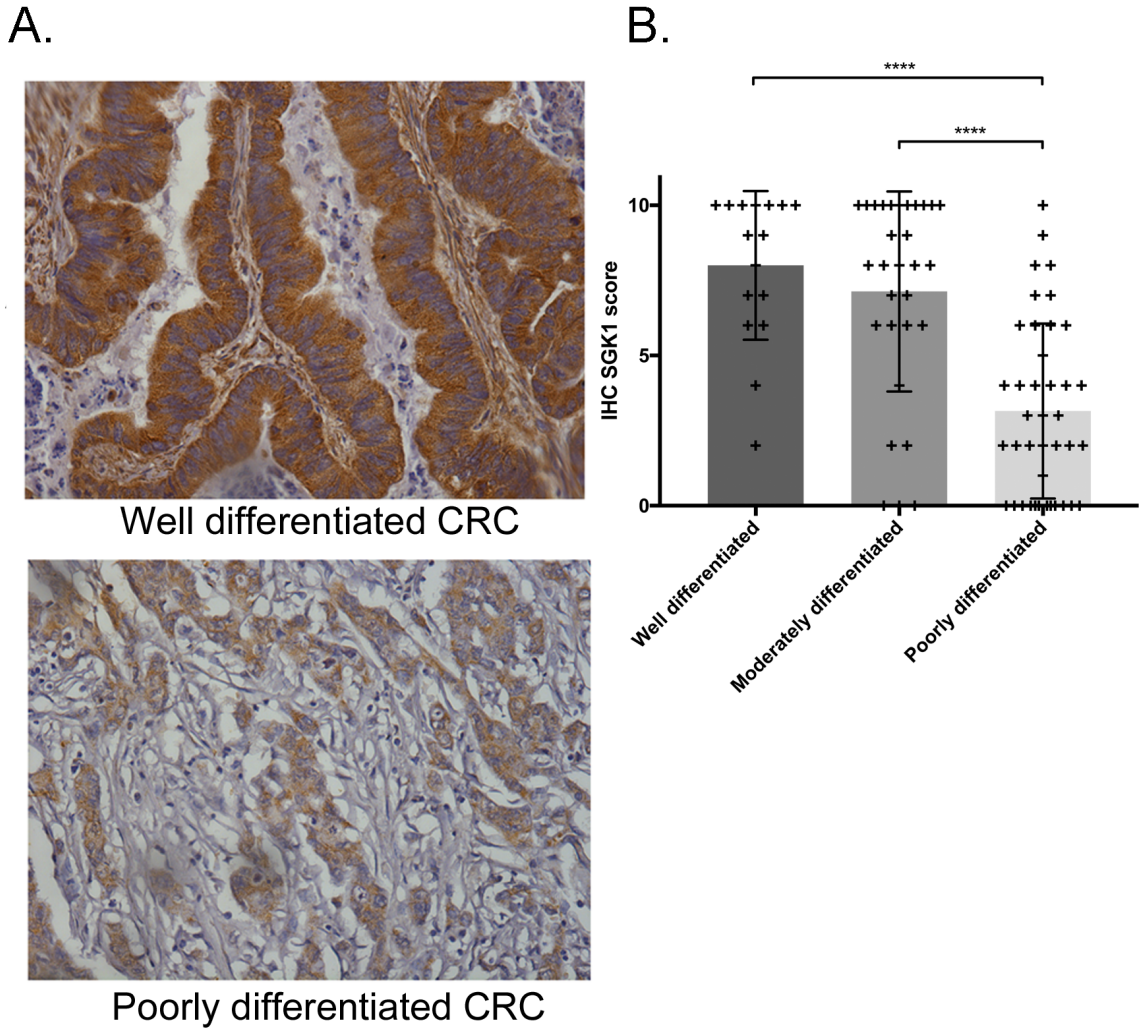
Figure 3.3: Expression of *SGK1* in tumour vs. normal tissue (n=283). Data from mRNA sequencing from The Cancer Genome Atlas. Error bars denote standard deviation



Analysis of the TMA demonstrated that there was very different expression of *SGK1* in tumours. I explored the relationship between *SGK1* expression and histo-

logical features available. I observed an association with the level of differentiation of the tumour and *SGK1* expression. There was a statistically significant difference in SGK1 protein expression between poorly differentiated tumours and well differentiated tumours (two sided t test unequal variance, $p<0.01$) and between poorly differentiated tumours and moderately differentiated tumours (two sided t test unequal variance, $p<0.01$) (Fig 3.4a,b).

Figure 3.4: A. Representative expression of *SGK1* determined by immunohistochemistry in CRC, well differentiated vs. undifferentiated. B. *SGK1* IHC intensity in CRC of different differentiation status. Data from Tissue Microarray- US Biobax C0485, C0484. Error bars denote standard deviation



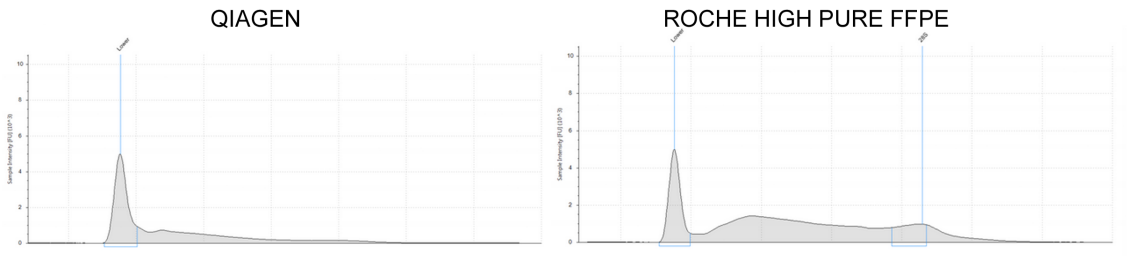
3.3.3 The effect of *SGK1* on cancer outcomes- OSLER sample set

Having identified *SGK1* as a gene that was both down-regulated in cancer specimens and down-regulated in poorly differentiated tumours, I sought to understand if *SGK1* identified patients with cancers of poor prognosis. To achieve this, I assembled a cohort of sporadic stage III CRC patients from the Oxford University Hospitals trust, the *OSLER sample set*. This was performed under the research ethics reference 11/SC/0236 and consisted of a retrospective analysis of patients treated between 2013-15.

FFPE primary CRC and lymph node metastases (LNM) histological sections were available from 53 patients. I optimised RNA extraction from FFPE specimens (Fig 3.5) and expression of *SGK1* in these tissues was analysed using rt-qPCR.

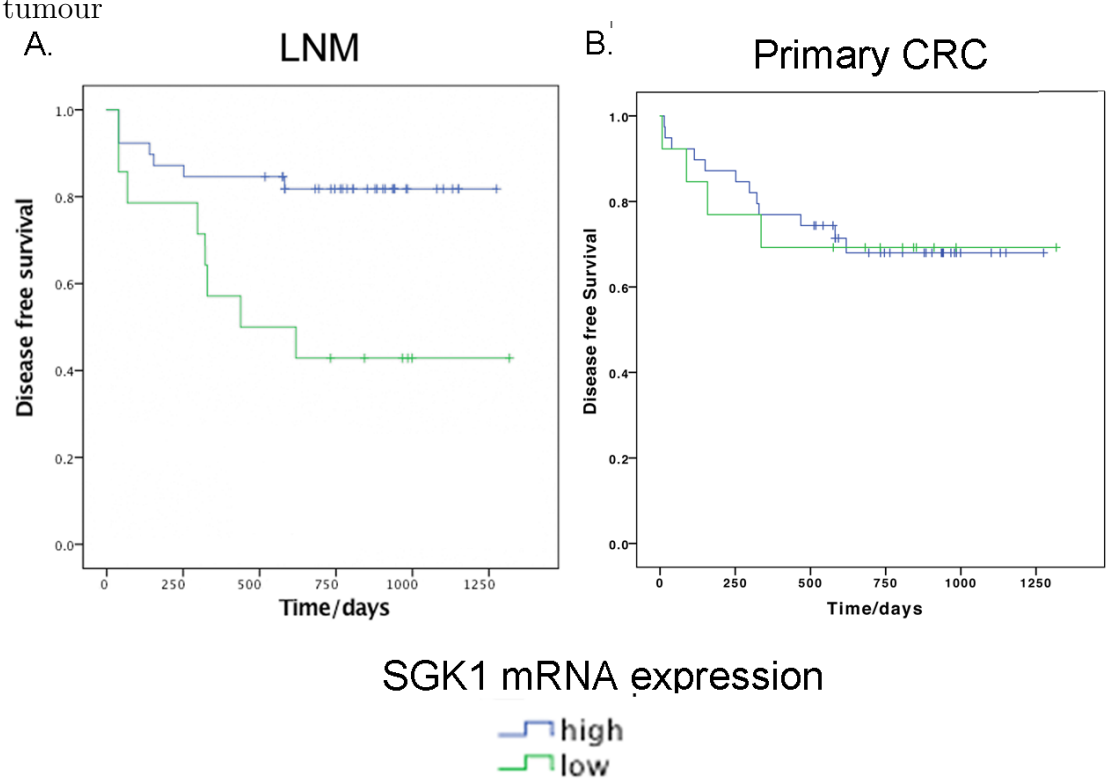
Figure 3.5: Optimisation of RNA extraction. Comparison of three RNA extraction kits using a single FFPE specimen. (Qiagen RNeasy FFPE, Roche High Pure RNA Kit, Roche FFPE High Pure RNA kit). RNA yield and quality as judged by RNA integrity number and DV200 and were highest with Roche FFPE High Pure RNA kit

	Conc/ng/uL			
Extraction type	Total	<200bp	RIN	DV200
QIAGEN	20.7	10.5	2.3	50.37
ROCHE HIGH PURE	27	7.73	2.6	28.49
ROCHE HIGH PURE FFPE	75.4	18.3	2.7	24.21



I observed that within the lymph node metastasis (LNM) samples, patients with the lowest *SGK1* expression (lowest 20% vs. remainder) had significantly poorer disease free survival (Hazard ratio= 3.68, 718 days vs. 1076 days, log rank p=0.006)

Figure 3.6: A. Efficiency of *SGK1* Taqman probes was determined by 8 serial dilutions of cDNA and use of the Ct slope method. The high efficiency denotes a reliable rt-qPCR probe. B. *SGK1* mRNA expression within LNs and disease free survival in OSLER sample set demonstrating that low mRNA of *SGK1* was associated with poorer disease free survival. C. *SGK1* mRNA expression within primary CRC and disease free survival in OSLER sample set. No association was observed in the primary tumour



(Fig 3.6b). This effect was not observed from the mRNA *SGK1* expression in the primary CRC (Fig 3.6c).

An analysis of the histological features of these tumours at diagnosis was performed to identify if there were any associations linked to low *SGK1* expression. *SGK1* expression in the primary tumour was not significantly associated with any histological features, although admittedly the number of patients in this sample set are relatively small and so it is challenging to exclude associations (Table 3.1).

Having identified *SGK1* as a putative survival biomarker using the OSLER sample

Table 3.1: Analysis of patient OSLER sample set- analysis of tumour expression of *SGK1* and baseline histological features. No significant differences were identified.

	High SGK1 (n=41)	Low SGK1 (n=11)
p3 stage (%)	19 (46.3)	5 (45.5)
p4 stage (%)	17 (41.5)	6 (54.5)
N1 stage (%)	27 (65.9)	5 (45.5)
N2 stage (%)	14 (56.1)	6 (54.5)
High budding (%)	27 (56.1)	3 (27.3)
V1 (%)	23 (56.1)	8 (72.3)
L1 (%)	23 (56.1)	6 (54.5)
MSI+ve (%)	7 (17.1)	4 (36.4)

set, I set out to validate these findings using other publicly available CRC expression datasets. I performed survival analyses based on expression of *SGK1* in tumours of patients from The Cancer Genome Atlas (TCGA) dataset. This is a cohort of north American patients, which began in 2009 and consists of 231 patients with CRC in whom RNA sequencing had been performed and survival data available. mRNA expression of *SGK1* was divided into quintiles and patient prognosis analysed using survival curves. Patients with the lowest 20% mRNA expression of *SGK1* demonstrated marked poorer prognosis with an overall survival of 127 months vs. 166 months (Hazard ratio 3.1, Log Rank test, p=0.001). I performed a separate analysis, splitting the TCGA cohort into different stages of tumours, and this difference in overall survival was present for all stages of colorectal tumours (Fig 3.7a,b).

To be more confident of *SGK1* mRNA expression as a biomarker of poor prognosis, I found and analysed a second cohort, a Norwegian CRC cohort. Gene expression was determined by microarrays (Affymetrix Human Exon 1.0 array) in 160 patients with stage II and stage III CRC (GSE24551)[115]. Similar to the TCGA cohort, patients with

Figure 3.7: A. *SGK1* mRNA expression in primary tissue and overall survival in TCGA. B. *SGK1* expression and overall survival by stage of disease

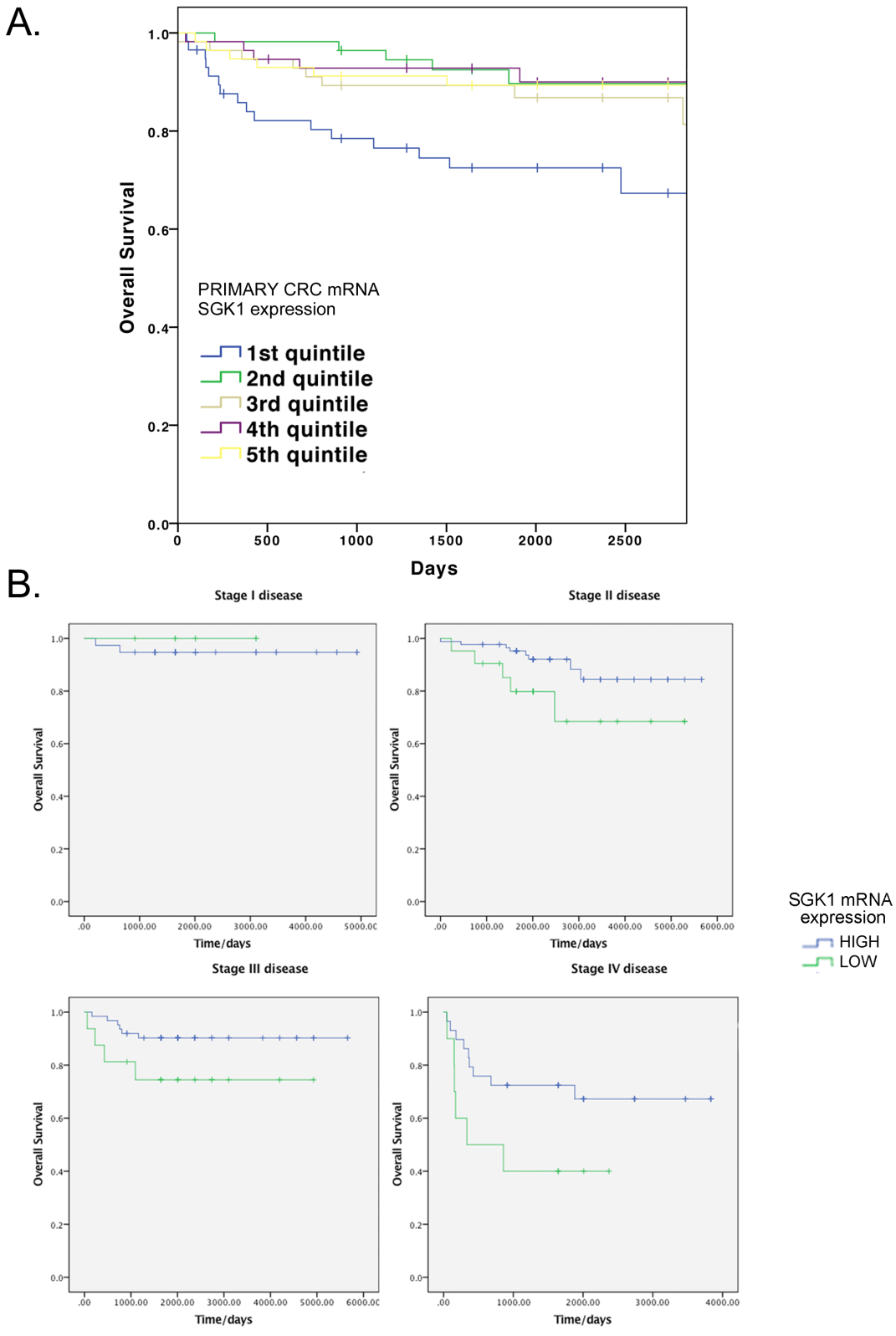


Figure 3.8: A. *SGK1* mRNA expression in primary tissue and overall survival in GSE24551. B. *SGK1* expression and overall survival by stage of disease

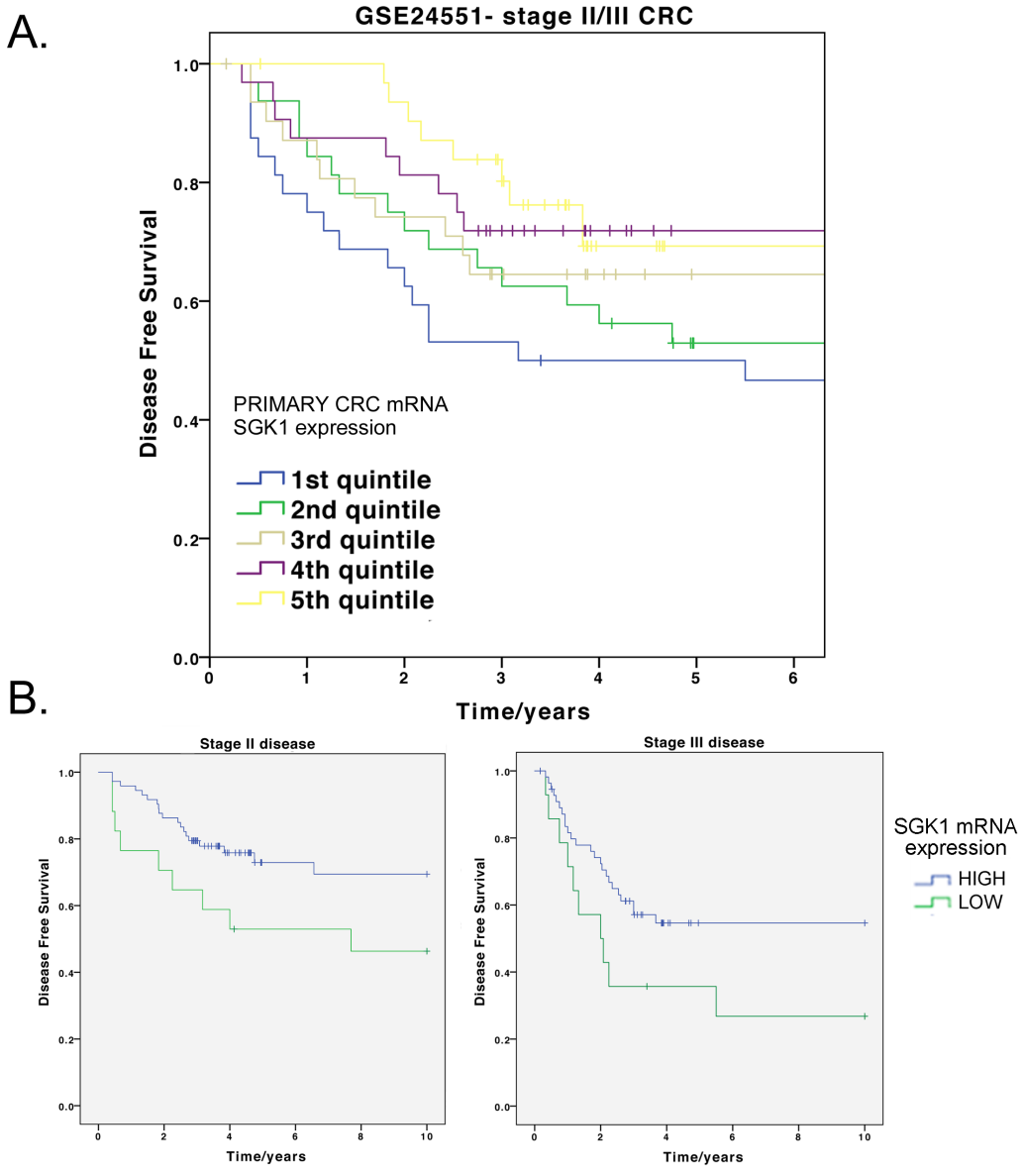
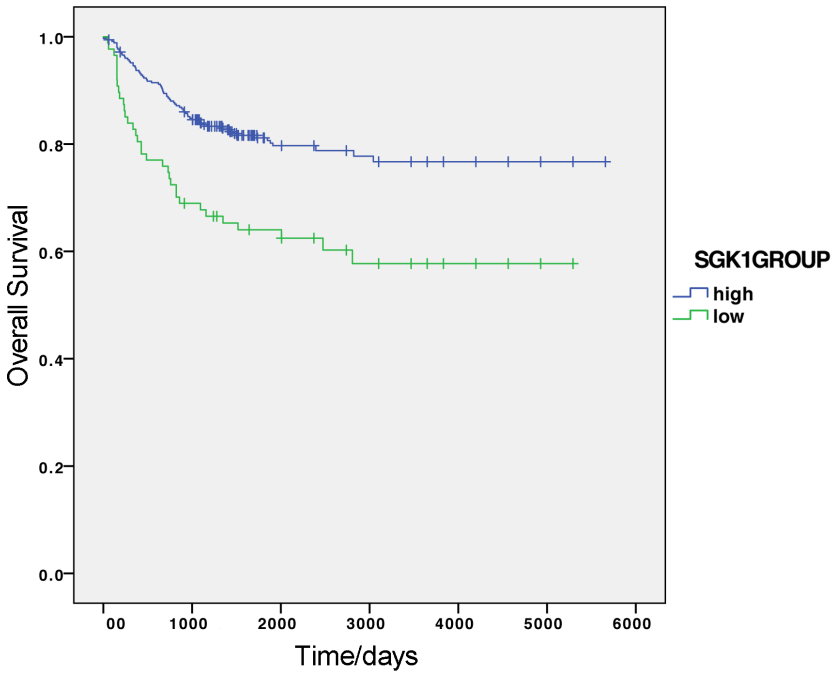


Figure 3.9: Meta-analysis of patients from TCGA and GSE24551. Association between SGK1 mRNA expression and overall survival. n=391



the lowest 20% *SGK1* expression of *SGK1* in this Norwegian cohort demonstrated poorer prognosis with a disease free survival of 5.4 years vs. 7.0 years (Hazard ratio 1.79, Log Rank test, p=0.03). A separate analysis was performed splitting this cohort into stage II and stage III tumours, and this difference in disease free survival was present for both stage II/III disease (Fig 3.8a,b).

I then combined these two cohorts together to perform a meta-analysis. This combined the dataset from the TCGA (n=231) and the Norwegian cohort (n=160), giving a total cohort of 391 patients with stage II/III CRC. Patients with the lowest 20% expression of SGK1 demonstrated poorer prognosis with a hazard ratio 2.19, log rank test, p<0.001 (Fig 3.9).

These findings need to be interpreted with considerations to the limitations of these analyses. RNA was extracted from representative tumour histological regions,

and were marked by a histopathologist based on based on the presence of at least 60% tumour nuclei (TCGA/OSLER). However, these regions also contain fibroblasts/lymphocytes and some normal tissue and it challenging to adjust for the presence of these non-tumour cells.

However, the analysis from these three CRC cohorts suggest that low mRNA expression of *SGK1* in primary tumours or LNs resection specimens are more likely to be associated with poor disease free survival or overall survival.

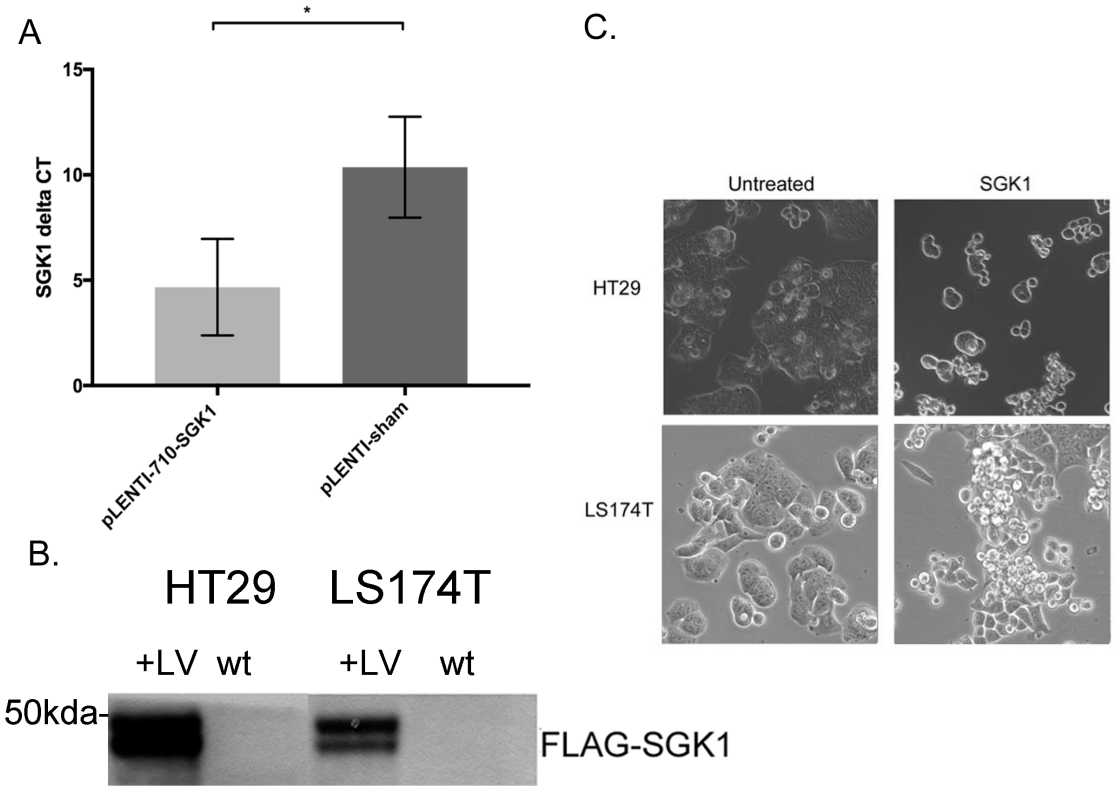
3.3.4 The development of a *SGK1* re-expression vector

To gain a greater functional understanding of the effects of *SGK1*, a strategy was developed to re-express *SGK1* in CRC cell lines. CRC cell lines, like human CRCs, have low expression of *SGK1*. A lentiviral delivery strategy was developed by Dr S.Segredtas (postdoc) and the construct was obtained by cloning the coding sequence of *SGK1* with a FLAG-tag into the over-expression vector pLENTI-710. I transduced cells with either this pLENTI-710-*SGK1* or an empty lentivirus. I confirmed that this strategy enabled re-expression of *SGK1* within cell lines by rt-qPCR (Fig 3.10a). Western blotting were performed by Dr S.Segredtas (Fig 3.10b).

3.3.5 The phenotypic effects of *SGK1* re-expression

Following re-expression of *SGK1* in CRC cell lines, it was possible to observe certain morphological changes in cell lines. LS174T and HT29 are derived from adenocarcinomas with the capacity to differentiate, and upon re-expression of *SGK1*, they displayed increased cell aggregation and a decreased tendency to grow in monolayers (Fig 3.10c). Increased aggregation or clustering of cells is a feature of cells with changes in cell adhesion properties, particularly increased expression of cadherins[116]. Increased cellular adhesion is also a feature of differentiated colonocytes.

Figure 3.10: A. Increased expression of *SGK1* demonstrated by rt-qPCR following transduction with pLENTI-710-*SGK1* (two sided t test. $p<0.01$). B. Western blot demonstrating expression of FLAG-tagged SGK1. Performed by S.Segredtas. C. Changes in cellular shape of HT29 and LS174T following transduction with pLENTI-710-*SGK1* demonstrating increased cell aggregation. Photograph of HT29-SGK1 taken by S.Segredtas.



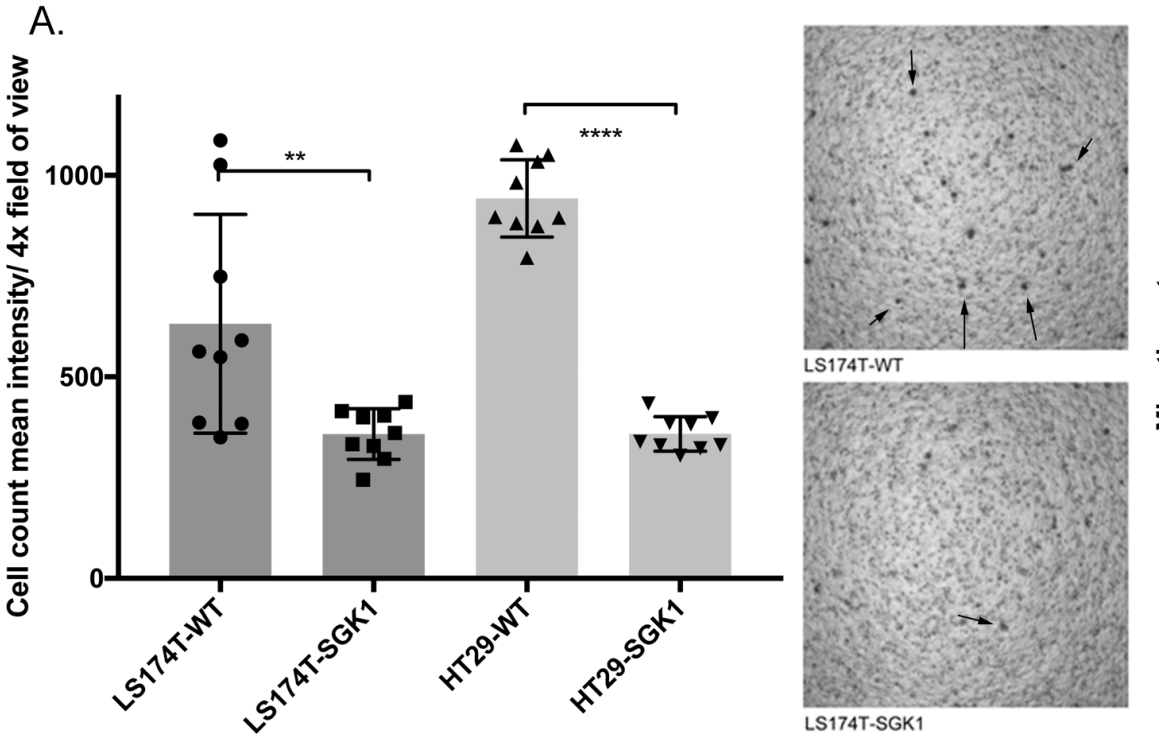
For cells to migrate, they must dissociate from neighbouring cells, and so I investigated if cellular migration was altered in these *SGK1* re-expressing cell lines. Cellular migration was assessed using a Boyden transwell assay. This consists of a transwell suspended in a cell culture plate. Single cells in suspension are placed on the top of the transwell and they are induced to migrate across a growth factor gradient to the bottom of the transwell plate where they can be counted using automated counting software. Re-expression of *SGK1* resulted in a significant reduction in rates of migration (change of migration rate, LS174T: -43% and HT29: -62%, two-sided t.test, $p<0.01$, $p<0.0001$ respectively) compared with cells that had been transduced with an empty lentiviral construct (Fig 3.11a).

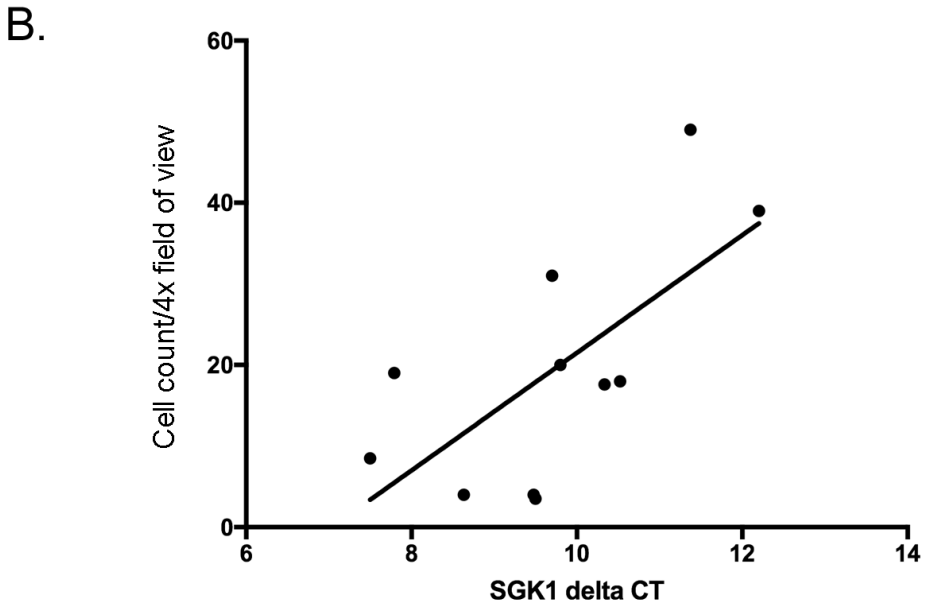
I then used a complementary approach to validate this putative link between low *SGK1* expression and cellular migration. I utilised our laboratory's bank of CRC cell lines (Colo205, GP2D, HCT116, HT29, LS174T, LS180, SNU-C4, SW1222, SW48, SW480, T84). I determined their baseline migration rates in the Boyden Transwell assay and then analysed *SGK1* expression in these cells, as determined by rt-qPCR. I was able to observe that cell lines with low *SGK1* expression, such as LS174T and SW480, exhibited the highest migration rate. Cells with high *SGK1* (Colo205, SNU-C4, SW1222) demonstrated low cellular migration. There was a significant association between a cell line's baseline *SGK1* expression determined by rt-qPCR and migration capacity (Pearson correlation 0.68, $p= 0.02$) (Fig 3.11b).

3.3.6 Transcriptome analysis of *SGK1* re-expression

To gain a more comprehensive understanding of the effects of *SGK1* re-expression, whole transcriptome profiling using 3' RNA sequencing was performed on 7 cell lines

Figure 3.11: A. Effect of *SGK1* re-expression and migration rates SGK1 re-expressing lines demonstrate lower migration rates. Error bars denote standard deviation B. *SGK1* expression was determined by rt-qPCR in 11 CRC lines and migration rates were determined. High SGK1 expressing cell lines demonstrating lower migration.





(GP2D, HT29, HCT116, LS174T, SW1222, SW480 and T84 cell lines) following *SGK1* re-expression or transduction with an empty lentivirus.

Specimens were sequenced at a depth of 5 million reads and mapped to human assembly GRCh37 to obtain gene count data. Gene Set Enrichment Analysis (GSEA) was performed and this identified enrichment for gene sets of the *Protein Kinase Cascade*, *Stress Activated Protein Kinase Signalling*, *Generation of Cell Polarity*, *Regulation of Cell Differentiation*, *Cell Adhesion* and *Epithelial Morphogenesis*. Enrichment of these gene sets are consistent with *SGK1*'s described role as a protein kinase and consistent with the work performed in our group which has identified *SGK1* as a marker and also a driver of cellular differentiation (Table 3.2, Fig 3.12).

Cancer signalling pathway enrichment was also analysed and I observed that following re-expression of *SGK1*, there was a negative enrichment of genes within pro-oncogenic pathways such as MYC/mTOR and Rb/E2F compared to wild type cells. This effect was most pronounced for c-MYC targets (NES -2.61, $p<0.001$) (Table 3.2).

Figure 3.12: GSEA of *SGK1* over-expressing cell lines- effect on Myc targets, cell differentiation and cellular polarity. In cells that re-express SGK1, there is an enrichment of genes involved in cell differentiation and genes involved in the "Establishment and Maintenance of cell polarity", but a negative enrichment of Myc gene targets

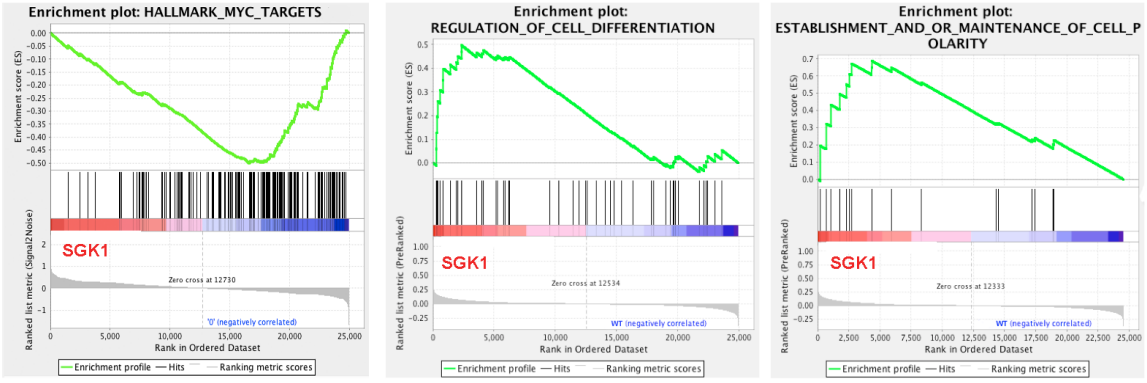


Table 3.2: Biological properties changed by SGK1 over-expression. There was significant down-regulation of MYC targets and enrichment for cell differentiation. Full MSigDB gene set in appendix A.1 on page 205

Cellular phenotypes	NES	p-value
Protein Kinase cascade	1.34	0.004
Cell polarity	1.70	0.010
Regulation of cell differentiation	1.54	0.014
Cell substrate adhesion	1.45	0.031
Stress activated Protein Kinase signalling pathway	1.39	0.040
Cell matrix Adhesion	1.43	0.044
Regulation of cell adhesion	1.47	0.047
Epithelial morphogenesis	-1.38	0.005
Epithelial Mesenchymal transition	-1.69	0.013
MYC targets	-2.61	<0.001
mTORC1 signalling	-2.61	<0.001
Rb/E2F signalling	-1.56	<0.001

Table 3.3: Differentially expressed genes following SGK1 re-expression- Leading edge genes

Gene	Base mean	log2FC	pvalue
SGK1	139.52	1.80	3.00E-15
NEFM	35.67	1.54	7.90E-12
RNA18S5	7117.04	1.02	4.96E-07
RNU6ATAC	26.55	0.88	2.93E-05
RUFY3	42.48	0.58	0.00634
PPIAP22	133.04	0.57	0.0115
FAM115A	73.15	0.56	0.0099
RN7SL4P	7.41	0.55	0.00714
BEX4	14.74	0.55	0.00308
SIPA1L3	39.04	0.54	0.00584
HBB	6.45	0.53	0.00782
ZNF346	70.16	0.49	0.0207
ANKRD18B	2.89	0.49	0.00402
RP5-1085F17.3	27.43	0.49	0.0270

3.3.7 *SGK1* and MYC expression

It was notable that this GSEA analysis identified MYC signalling as significantly down-regulated following *SGK1* re-expression. c-MYC is an oncogene that lies at the crossroads of many growth-promoting signal transduction pathways. In CRC, defects in the Wnt pathway, directly drive tumorigenesis as a result of enhanced TCF7L2 activation of c-MYC[117]. *c-MYC* is also upstream of a number of growth signalling pathways, including the MAPK, TGF-beta, PI3K and mTORC pathways. *c-MYC* is a pro-metastatic transcription factor, regulating epithelial-to-mesenchymal transition, cell-cell matrix interactions and cellular invasion[58]. In breast cancer, this is partly mediated through the induction of *RHOA*, a GTPase protein that is associated with cellular cytoskeleton regulation[118]. In colorectal cancer, gene amplification of *c-MYC* is much more frequent in metastases compared to primary tumours[56] and is an independent poor prognostic in stage II and stage III CRC patients with a hazard ratio of 2.35[119].

To confirm the findings of the GSEA from a protein perspective, I performed western blots on three cell lines (HT29, LS174T, SW480). Compared to cells transduced with an empty lentivirus, following *SGK1* re-expression, normalised c-MYC protein expression was decreased in all three cell lines with a mean reduction of 56.4% (Fig 3.14a). Analysis of other upstream signalling proteins, including AKT1 and p44/42 MAPK (ERK1/2) was performed, but these were not consistently dysregulated in the *SGK1* re-expressing cell lines. AKT expression levels did appear to be dysregulated in two of the cell lines, but this was not observed in LS174T-*SGK1*, and this could have arisen through a known negative feedback loop[120].

To determine if down-regulation of c-MYC occurred through direct protein-protein interactions between *SGK1* and c-MYC, I performed a co-immunoprecipitation assay.

An *SGK1* antibody was used to capture SGK1 protein bound with protein substrates onto Protein A-coated magnetic dynabeads. Protein-protein interactions were maintained by the use of physiological lysis and wash buffers that were of low ionic strength and contained non-ionic detergents. All the proteins were then denatured and proteins that were bound to SGK1 were released. Using this technique, I was able to confirm that in all cell lines analysed (HCT116, LS174T, SW480), c-MYC was precipitated with *SGK1*, suggesting the presence of strong protein-protein interactions (Fig 3.13).

The c-MYC transcription factor is under tight control with expression regulated by post-translational modification (PTMs). PTMs occur on amino acid side chains or the C- or N- terminal and can include new functional groups such as phosphate, acetate, amide or methyl groups. *SGK1* is a member of the serine/threonine kinase that phosphorylates the OH group of serine or threonine amino acids and I hypothesised that *SGK1* might alter c-MYC levels through phosphorylation of c-MYC. There are number of important phosphorylation sites on c-MYC including Thr58, Ser62, Ser71, Thy74, Ser344, Ser373. However, of these Thr58 and Ser62 are particularly relevant, as mutations in these regions are common in a number of cancer types such as Burkitt's lymphoma and act to stabilise and activate c-MYC signalling expression[121]. Phosphorylation at Ser62 and Thr58 leads to ubiquination and subsequent proteasomal degradation of the protein by FBXW7[122][123].

I performed a western blot using phosphoprotein antibodies specific to Ser62 and Thr58, to determine if there was a relative increase in phosphorylation in these sites, which in term would lead to MYC degradation. In all cell lines, for both Ser62 and Thr58, there was an increase in MYC phosphorylation at these sites. On average, following *SGK1* re-expression, Ser62 phosphorylation was increased by 46.4% and

Thr58 phosphorylation was increased by 42.0% (Fig 3.14).

Figure 3.13: Co-immunoprecipitation of c-MYC with SGK1 demonstrating binding of SGK1 with c-MYC

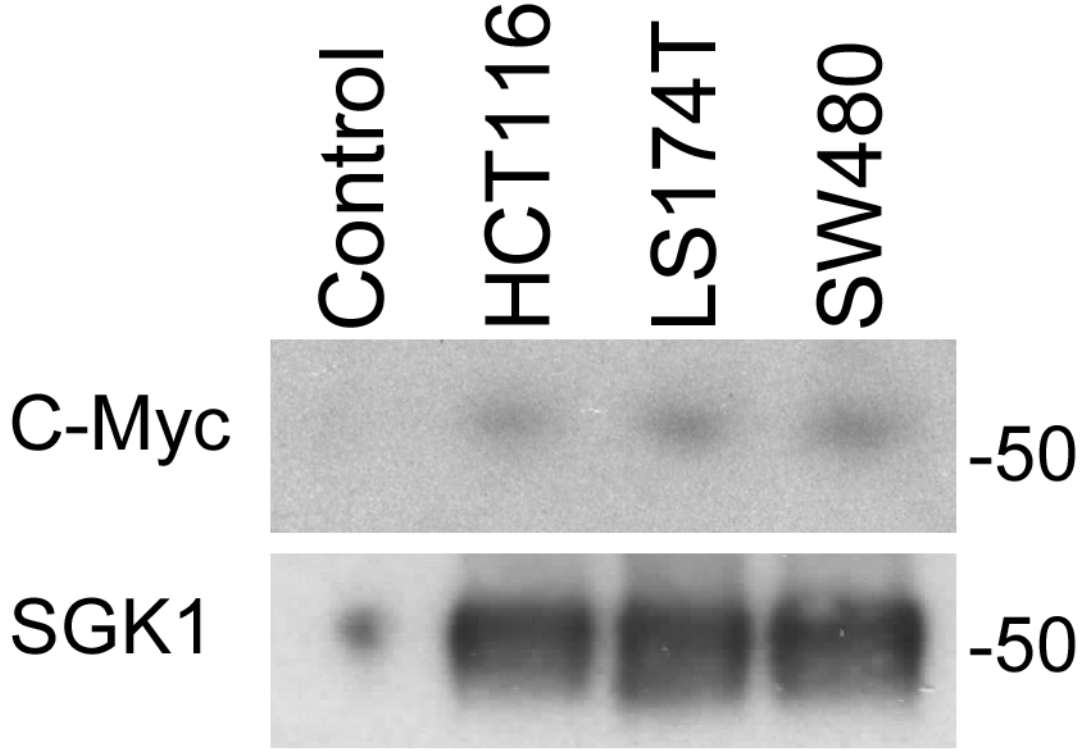
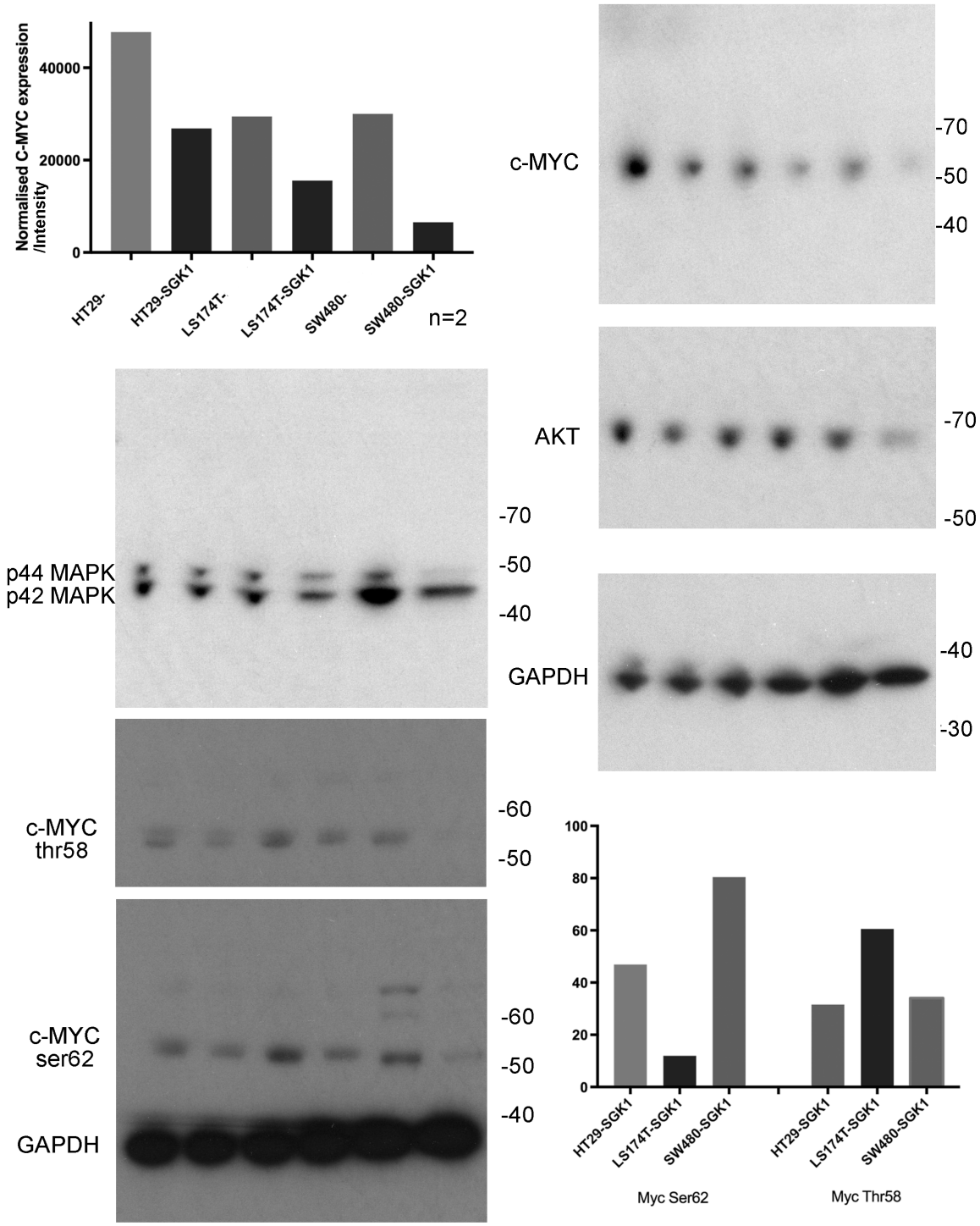


Figure 3.14: Representative blots demonstrating the effect of SGK1 re-expression and c-MYC expression and targets and phosphorylation sites. Decreased protein expression of c-MYC is observed following SGK1 re-expression



3.3.8 Development of a mouse model- OTM

These results suggest that *SGK1* has an important role in cellular migration and is linked to poor prognosis in human cohorts. To gain a greater understanding, I sought to develop a new in-vivo CRC model, an endoscopic-guided orthotopic mouse transplantation model (OTM). I hypothesised that *SGK1* might have effects on CRC metastasis rates.

Orthotopic transplantation of cells has been shown to have significant advantages over the use of transgenic mice or standard mouse dorsum transplantation models. Unlike orthotopic transplantation models, dorsum transplantation models rarely develop distant metastases as they are unable to recapitulate the rate limiting steps of local invasion[68 72]. Furthermore, only a few transgenic models develop metastasis; CRC Smad3 mutant mice develop only local LN metastases[74]; the Kras(G12V)-Tgf-b1(-/-) mice only develops lung metastases in 10% of cases[75]; and a Kras/Apc ko animal with viral delivery of Adeno-CRE rectally only developed liver lesions in 20% of cases[73]. A further issue with the aforementioned transgenic models is that they require several months/years before the development of metastases is detectable.

Mouse endoscopy is a technique that is used to monitor colonic tumour burden over time[124]. I sought to modify this technique to allow it to be used as a tool for orthotopic transplantation of CRC cell lines. At the time, endoscopic orthotopic transplantation had only been successfully achieved by one group in Israel, however, they had demonstrated convincing submucosal growth of tumours[70].

An 8-inch long flexible stainless steel 30-gauge microinjection syringe and luer lock were purchased from *Injecta*, Germany and I developed my own protocol. I used NOD scid gamma mice and performed microinjection of both human and mouse cell lines.

A high rate of submucosal tumours was achieved by subtly modifying the published protocol. I pointed the needle bevel downwards to minimise retrograde leakage of cells and introduced a second operator. Having a second operator allowed increased stability of the instrument during administration of the cells and a higher likelihood of success (Fig 3.15a). To enable accurate tracking and quantification of the tumour cells as they grew and metastasised, I set out to combine my OTM with in-vivo imaging to allow real-time detection, quantification and localisation of labelled cells in an animal over a period of time.

Optimisation of in-vivo imaging was performed. The MC38 cell line was used and cells were labelled with either a near-infrared fluorophore (pLENTI-RFP670, pLENTI-RFP702, pLENTI-RFP720) or luciferase construct (pHIV-Zsgreen-Luc2), transplanted into mice and then imaged using the *IVIS Spectrum*. I established that labelling tumour cells with the luciferase construct was superior and allowed detection of as few as 100 cells (Fig 3.15b,c).

Bioluminescence assays are notable for their a linear-range and its use during in-vivo imaging has been demonstrated to very accurately evaluate tumour growth[125].

Albino mice were used to increase the sensitivity of the technique and luciferin kinetic studies were performed to determine the optimal time for imaging following administration of luciferin (Fig 3.16b).

The pilot studies demonstrated that there were no fatalities from the procedure. In mice in which the procedure was deemed successful at 16 hours, mice developed colonic tumours that became obstructive by week 4. The most common reason for failure of the procedure was intraluminal delivery of the tumour cells or more rarely, intraperitoneal spread. Orthotopic injection of tumour cells into the colon frequently gave rise to lung metastasis, with liver metastases an infrequent event (Fig 3.16a).

Figure 3.15: A. Image 1:- Endoscopic view of the descending colon of mouse, insufflated with air, with faeces in the top right. Image 2:- Micro-syringe introduced into the colon in bottom left. Image 3:- Micro-syringe used to puncture surface of colonic epithelium. Image 4&5:-Introduction of cells beneath colonic epithelium. Image 6:- removal of syringe and demonstration of colonic bleb on epithelium containing cells. B. In-vivo imaging using near-IR imaging vs. bioluminescent imaging. Optimisation was performed using CRC cells labelled with either RFP702, RFP720 or Luciferase. Luciferase and RFP were detected using the IVIS spectrum, measurements in radiant efficiency using automated detection . C. Limit of detection using bioluminescent imaging. Between 100-10,000 cells were administered subcutaneously. In-vivo imaging was able to detect as few as 100 cells.

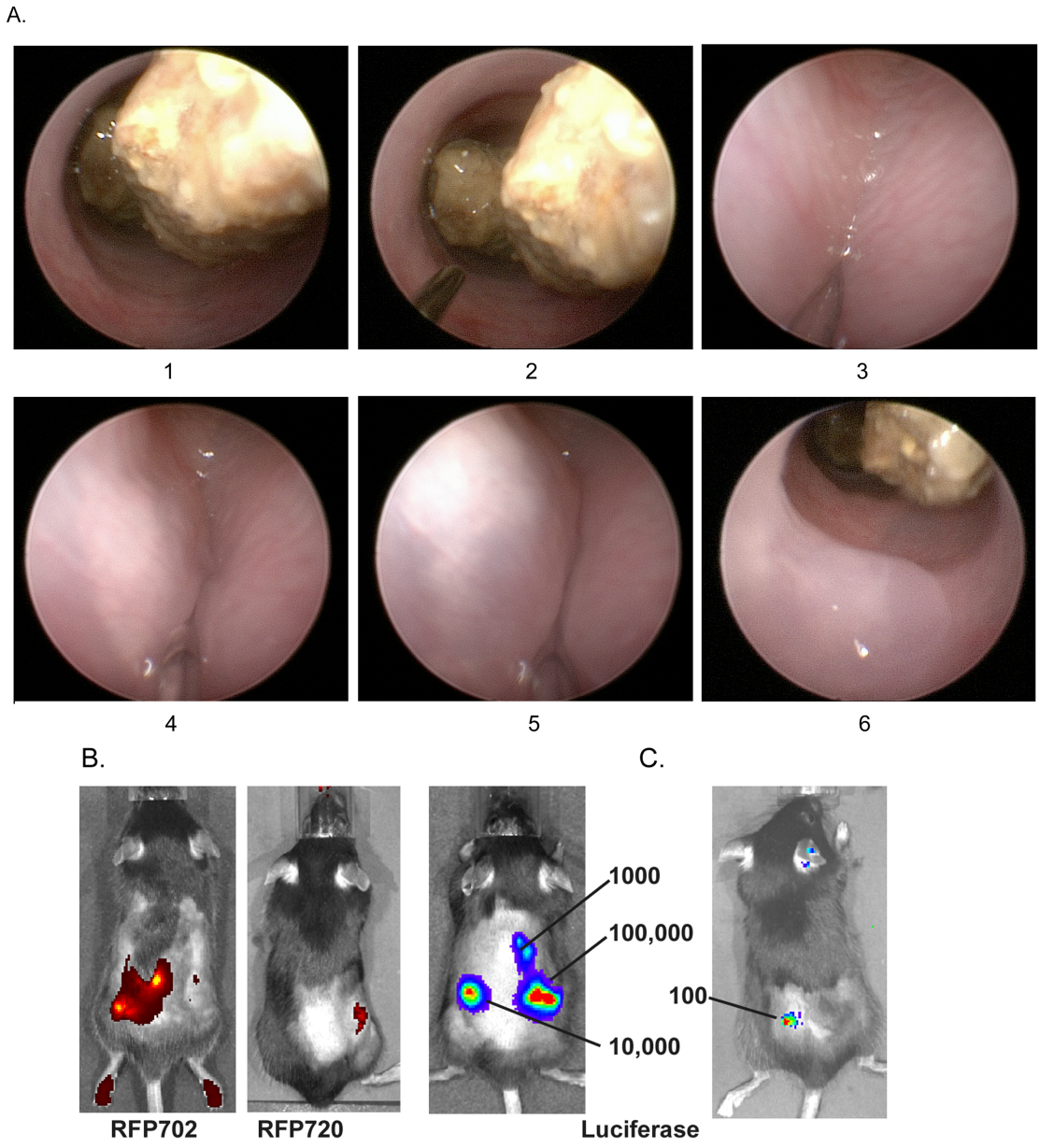
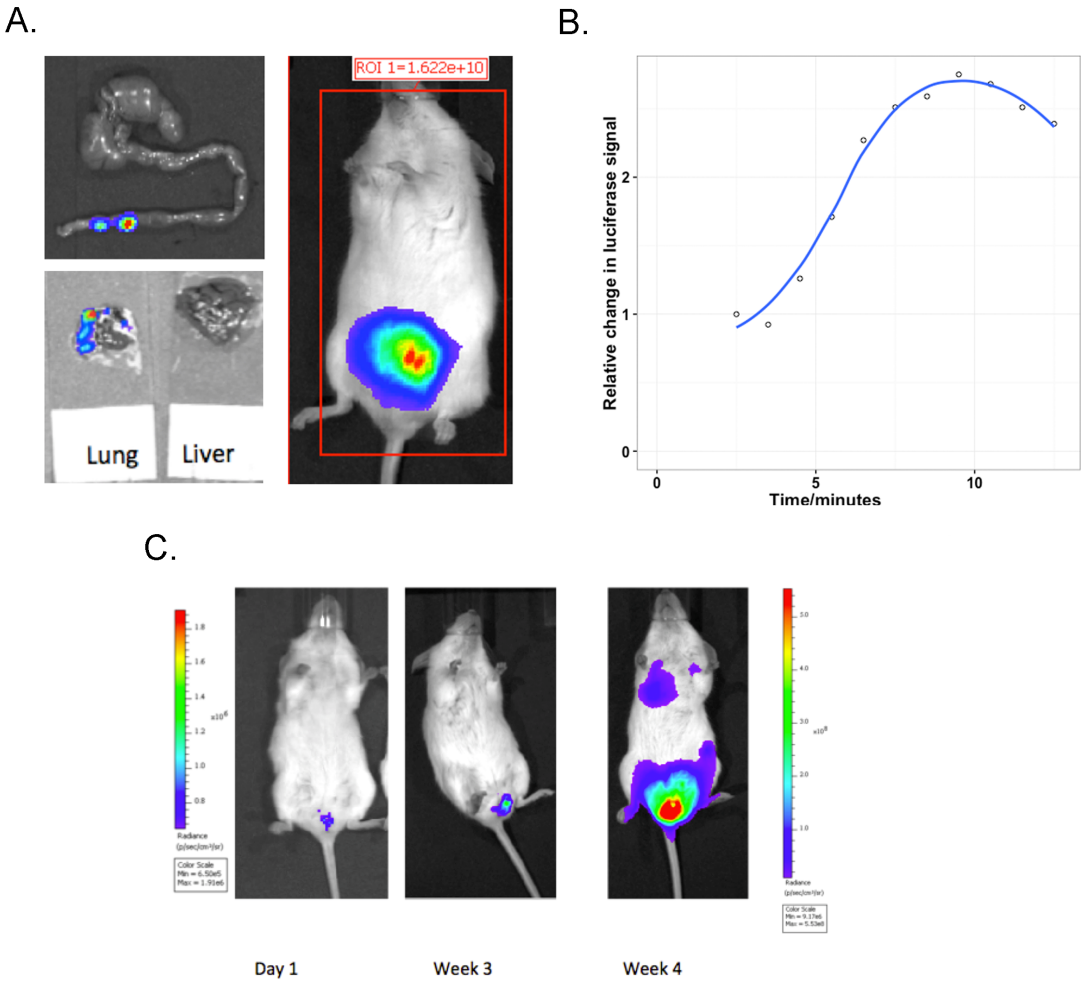


Figure 3.16: A. Ex-vivo imaging using bioluminescent imaging, demonstrating establishment in the descending colon with lung metastasis. Whole body imaging demonstrates colonic growth of tumour cells. B. Bioluminescent kinetic assay optimisation. Luciferin was administered to mice intraperitoneally and luminescence whole body signal was quantified over time. C. Growth of CRC cell lines and metastasis over 4 weeks, following administration of 0.5 million CRC cells. Metastatic outgrowths were detectable by week 4.



3.3.9 *SGK1* orthotopic mouse transplantation experiment

The OTM was then performed using the *SGK1* re-expression cell line to determine effects on tumour growth kinetics. Luciferase labelled LS174T cells were transduced with either the pLENTI-710-*SGK1* re-expression virus or an empty lentivirus. These cell lines were injected into two cohorts of 16 mice. 1 million cells were transplanted

into the descending colon of NOD scid gamma mice and the success of this procedure was determined by in-vivo imaging at 16 hours. Imaging of mice was then performed at weekly intervals or until animals became symptomatic. The primary growth of the tumour was analysed using in-vivo imaging.

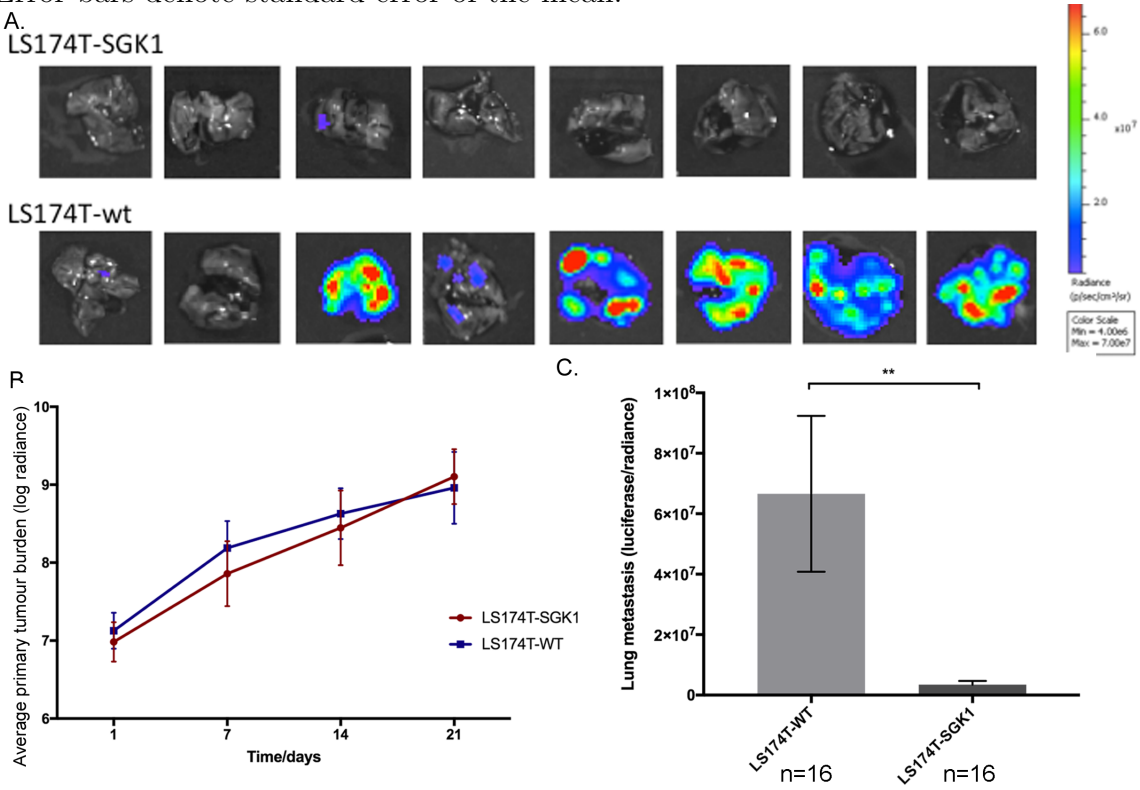
In-vivo imaging identified no differences in rates of primary tumour growth between SGK1 or empty vector transduced LS174T cells (Fig 3.17b). However, analysis of lung metastasis burden demonstrated metastasis was almost 20 times lower in the animals that received LS174T-*SGK1* than LS174T-empty ($3.4 \text{x} 10^6$ vs. $66.6 \text{ x} 10^6$ rad). This increase in lung metastasis rates was significant and observed in both cohorts of experimental mice (two sided t-test, p=0.002) (Fig 3.17c).

There was no difference in lung weight. Unfortunately, I did not find a way to determine if the differences in lung metastasis burden represented a growth delay in the establishment of lung lesions or decreased growth. It is possible that future experiments would require me to aim to modify the number of cells administered to enable me to reach the same tumour seeding burden at day 1. This would then enable me to specifically analyse growth potential.

3.3.10 *SGK1* mouse tail vein experiment

The OTM models the entire metastatic process including growth of the primary tumour, intravasation in the circulation, extravascularisation and tumour outgrowth in the lung. However, to determine if *SGK1* re-expression was more important at a late, than early stage, I performed a mouse tail vein metastasis assay. In this experimental model, the tumour cells are introduced directly into the circulation through the lateral tail vein and the model assesses the ability of cells to lodge and grow in the lungs. 1 million cells were administered into each mouse (n=7). All mice developed lung metastatic lesions. The LS174T-*SGK1* cells formed significantly fewer

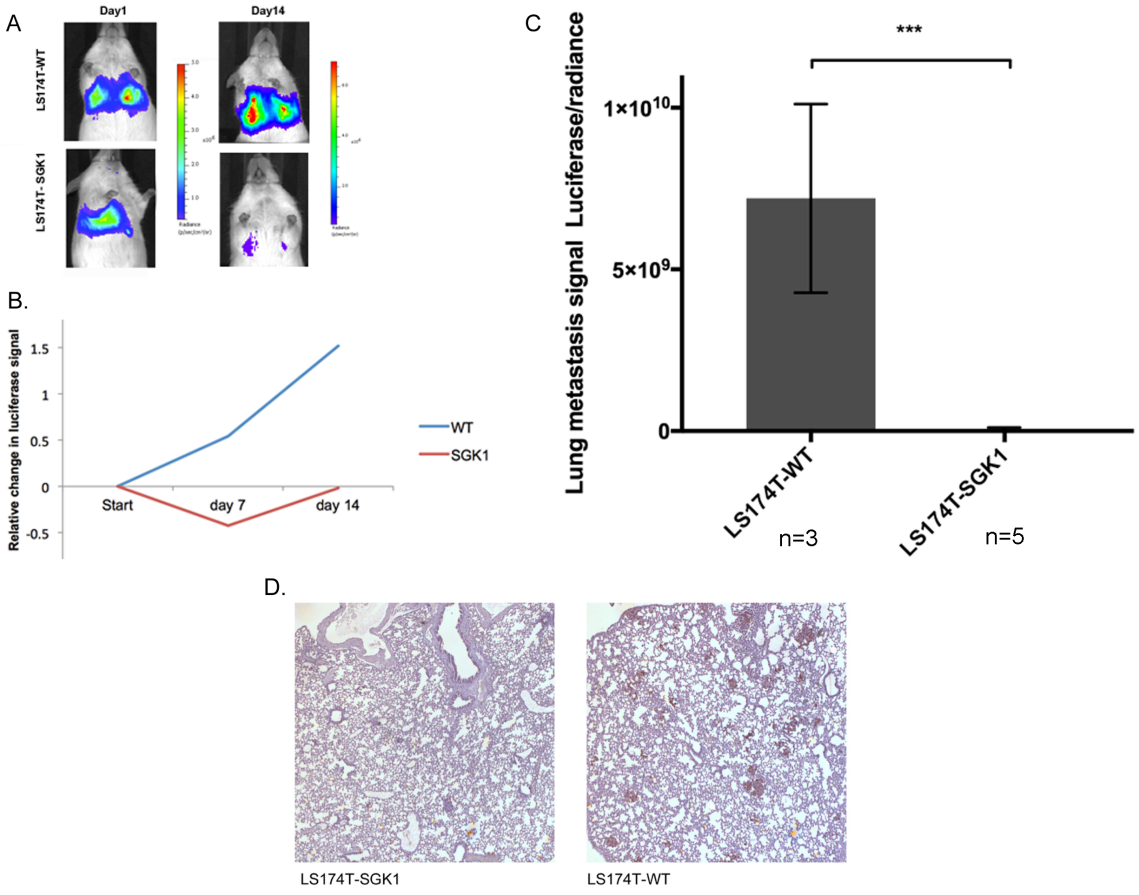
Figure 3.17: A. Effect of *SGK1* re-expression and lung metastasis. Ex-vivo lung bioluminescent imaging demonstrating that mice that received LS174T-WT cells had high metastasis burden whereas those that received LS174T-SGK1 had lower metastasis burden. B. Effect of *SGK1* re-expression and primary tumour growth. Tumour growth was quantified using in-vivo imaging at multiple time points. Average primary tumour burden did not differ between the groups. C. Bar chart demonstrating lung metastasis burden was significantly lower following *SGK1* re-expression ($p<0.01$). Error bars denote standard error of the mean.



lung metastatic lesions than the LS174T-WT cells, reflecting the findings from the orthotopic transplantation experiment. Metastasis burden was several hundred times lower in mice that received the LS174T-*SGK1* cells(Fig 3.18a,b,c). These results demonstrate that *SGK1* had a negative impact on the later stages of metastasis of LS174T (i.e. survive in the circulation, establish and proliferate in the lungs).

The data suggests SGK1 re-expression is likely to affect the early phases of metastasis. Following administration of the cells into the lateral tail vein, cells must establish and lodge in the lungs. At 24 hours, there was a trend towards a decrease in lung colonisation. I believe that SGK1 affects the survival of cells in the context of anchorage independent growth, i.e. whilst in the circulation. The rate of growth of the tumours at day 7 was 12.8 log2 FC for the LS174T-wt vs 5.4 log2 FC for the LS174T-*SGK1* mice. This was not statistically different. It will be challenging to prove if this difference also represents decreased growth as well as a growth delay in-vivo.

Figure 3.18: A. Tail vein injection assay. *SGK1* re-expression results in lower lung metastasis burden at day 14. B. Lung metastasis burden over time. WT cells demonstrate increases in lung metastasis burden over time. SGK1 re-expressing cells demonstrate a reduction in metastasis burden at day 7 before demonstrating a marginal increase by day 14. C. Bar chart demonstrating effect of *SGK1* on lung metastasis burden ($p<0.001$). Error bars represent standard of the mean. D. Histological sections using antibody for human TP53 demonstrating higher tumour deposits in mice that received LS174T-WT cells compared to those receiving LS174T-SGK1 cells . LS174T cells do not have mutations in TP53.



3.3.11 *SGK1* expression in primary and metastatic colonies

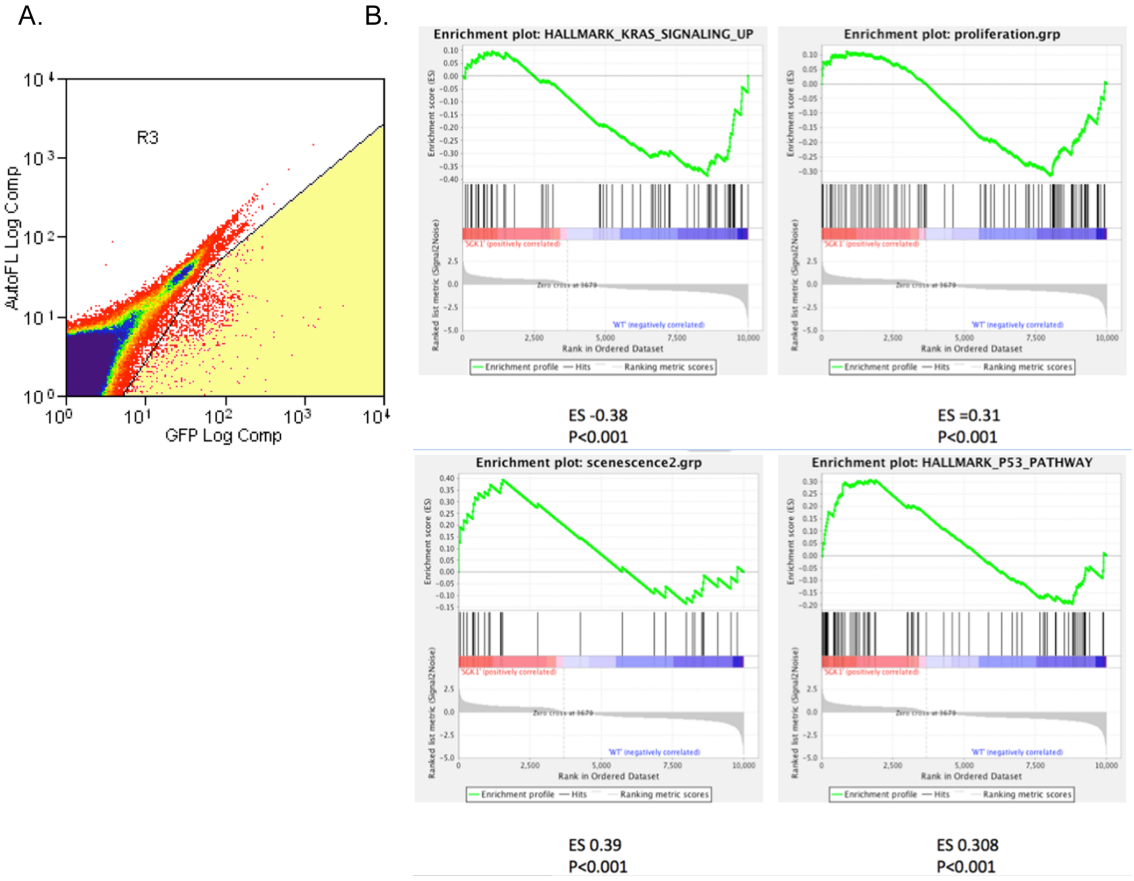
I proceeded to analyse the metastatic lesions of these mice models in more detail. Unfortunately, standard techniques such as IHC were not possible from the lung resections from the LS174T-*SGK1* mice because lesions were so few and small in size. However, transcriptome profiling of metastasis cells was feasible as the cellular in-

put requirements are smaller. The OTM was performed with the LS174T-WT and LS174T-SGK1 cell lines and the primary tumour and lungs containing the metastatic lesions were harvested. Metastatic cells and the transplanted primary tumour cells were isolated from the mouse by digesting tissues using Dispase I (2 units, incubation for 1 hour at 37°C) to generate a cell suspension and then Fluorescence Activated Cellular Sorting (FACS) for the Zsgreen-tag that was present on the transplanted LS174T cells (Fig 3.19a). RNA sequencing was performed on 6 experimental mice (3 from mice that received LS174T-WT cells and 3 from mice that received LS174T-SGK1 cells).

The sequencing reads obtained demonstrated that some reads from the primary and metastatic lesions aligned to the mouse genome, i.e. there was admixture of the human CRC cell line and mouse host mRNA. Mouse transcript admixture varied from specimen to specimen and varied from 12-71%. Sequencing reads that mapped to the mouse genome were filtered out and only sequencing reads that mapped to the human genome were used for feature counting and differential expression.

Confirming the success of this analysis, both the primary tumour and the lung metastases from the mice that received the LS174T-*SGK1* cell line transplants expressed higher counts of *SGK1* than those that received LS174T-SHAM cells (4.3x-primary tumour, 1.6x - lung metastases).

Figure 3.19: A. Fluorescence activated cellular sorting for GFP to sort for metastatic outgrowths. Highest expressing GFP cells were selected, identified in yellow section. B. Results from gene set enrichment analysis showing that SGK1 metastatic lesions had negative enrichment for KRAS signalling and cell proliferation, but positive enrichment for the senescence gene set



3.3.12 Metastatic cells- GSEA expression

Gene set enrichment of the metastatic and primary lesions demonstrated a number of interesting features. In metastatic lesions, *SGK1* re-expression was associated with enrichment for genes involved in apoptosis and senescence with a negative enrichment for proliferation compared to the LS174T-WT. Furthermore, pathway analysis demonstrated an up-regulation in the p53 pathway with a down-regulation of the KRAS pathway (Fig 3.19b, Table 3.4). These features were not identified in the primary tumour or in the cell line transcriptome dataset. The LS174T line does not

Table 3.4: GSEA results of *SGK1* re-expressing cell lines vs. sham cell lines from lung metastatic cells

	Metastatic cell		Primary CRC		Control Cell lines	
	NES	p-value	NES	p-value	NES	p-value
TP53	1.25	<0.01	-1.12	0.22	-0.74	0.84
KRAS	-1.63	<0.01	-0.94	0.61	-0.79	0.98
MYC	1.25	0.21	-0.66	0.64	-1.94	0.02
MTORC	1.12	0.18	-0.70	0.91	-1.65	0.03
NOTCH	1.52	<0.01	0.96	0.40	0.82	0.71
Wnt	0.99	0.50	-1.06	0.30	0.51	0.92
Apoptosis	1.63	<0.01	-0.90	0.70	-0.85	0.75
Proliferation	-1.51	<0.01	-0.94	0.69	1.16	0.10
Senescence	1.35	<0.01	-1.27	0.09	1.08	0.29
EMT	0.84	0.707	0.83	0.71	-0.88	0.653

contain any mutations in TP53 and its up-regulation specifically in metastasis might underpin the differential effects of SGK1 in primary tumour vs. lung metastasis lesions. The relative inability of the LS174T-*SGK1* to activate the KRAS pathway is particularly interesting as mutations causing a constitutively active KRAS in humans are a known driver for lung metastasis[126].

3.4 Discussion

Tumour differentiation (or grade) is an important histological feature that is associated with poor prognosis from colorectal cancer. Patients with a poorly differentiated tumour are more likely to develop metastasis than those without[127], however, the mechanisms through which differentiation status of a tumour affects prognosis has been overlooked and the mechanisms an enigma.

Previous work in our group had identified *SGK1* as a gene that was expressed primarily in differentiated colonocytes and significantly down-regulated in adenomas[81]. *SGK1* is a serine/threonine kinase that is part of the AGC group of kinases that shares

homology with the AKT family. To date, the role of *SGK1* has been most clearly elucidated for the regulation of ion channels[128], regulation of neuronal function[129] and its role in T helper cell differentiation[130]. However, more recently, there has been a growing realization that *SGK1* lies at a key juncture of several oncogenic signaling networks. The effects of *SGK1* vary depending on the cell lines analysed and the models used, for example, a number of studies have highlighted the pro-proliferative effects of *SGK1* in a hepatocyte model[104], breast cancer model[108] and in a murine colitis model[131]. However, the effect appears to be cancer and cell line dependent and use of the human embryonic kidney cells has demonstrated contrasting findings. In this model, *SGK1* inhibited the Ras-MAPK kinase pathway through phosphorylation of B-RAF[111].

To date there have been no studies looking at the role of *SGK1* in a CRC model. Certain paradoxes are present in the intestine, principally that differentiated colonocytes are found at the tops of colonic crypts and these cells have little or no proliferative potential, however, this is also where *SGK1* is predominantly expressed.

To understand the mechanisms of action of *SGK1*, I performed whole transcriptome analysis on several CRC cell lines following *SGK1* re-expression, representing the different subsets of colorectal cancer. A consistent signature from these cell lines was the down-regulation of a key oncogenic signalling target, c-MYC.

SGK1 re-expression results in the down regulation of c-MYC and I provide pilot evidence that one possible mechanism is through protein-protein interactions that is associated with increased phosphorylation of c-MYC at Ser62 and Thr58. Phosphorylation of c-MYC at these sites would then allow rapid proteolysis through ubiquination by FBXW7[122].

During normal intestinal homeostasis, within differentiated cells, *SGK1* has a role

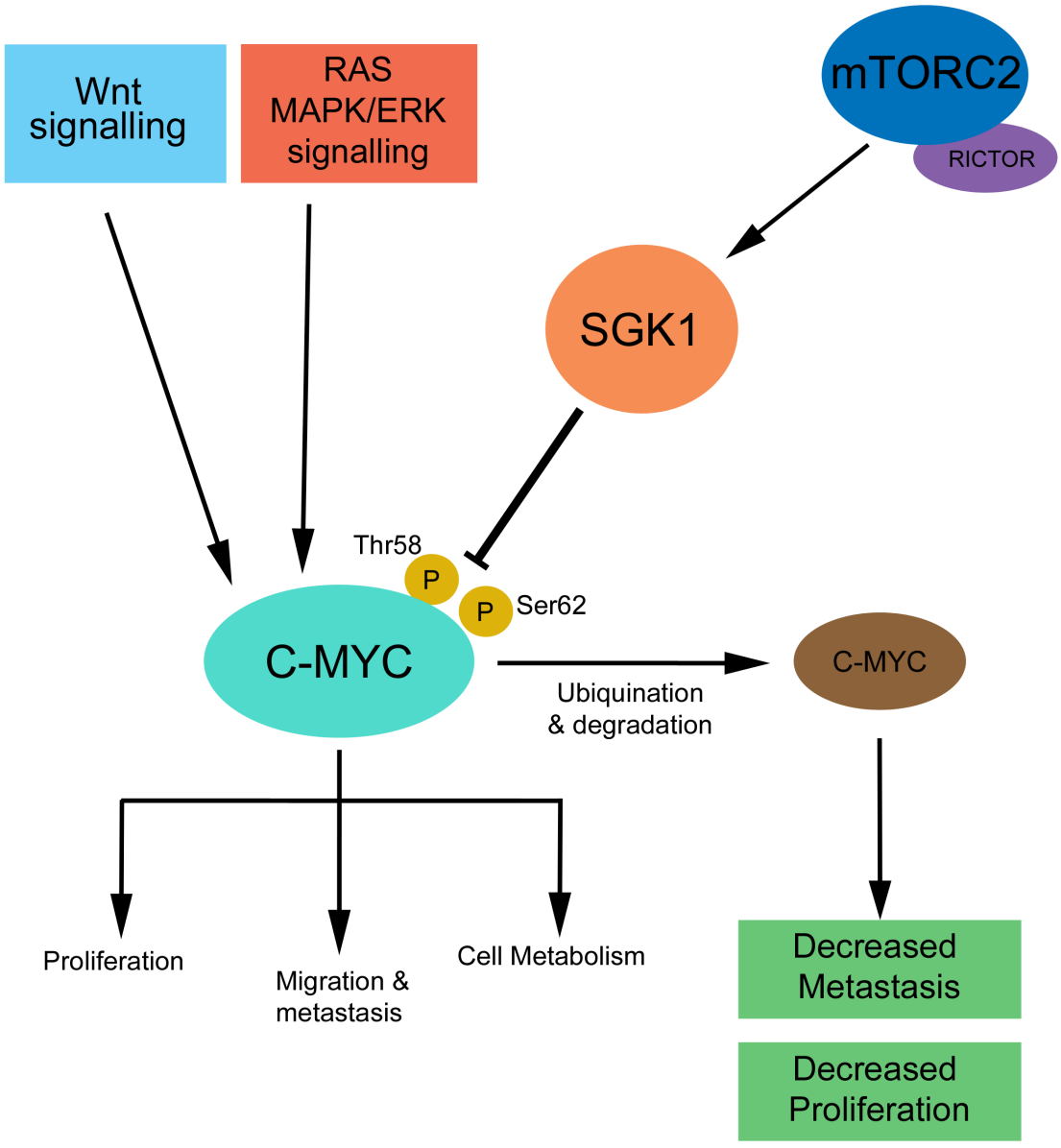
in driving cell polarity and promotes cell-cell and cell-matrix adhesion and the formation of ion channels. In these differentiated intestinal cells, it is important that pathways involved in cell proliferation and migration are suppressed. My data is consistent with *SGK1* playing an important role in this process by phosphorylating and inducing the degradation c-MYC (Fig 3.20).

This function of *SGK1* means that in addition to induction of differentiation of colonocytes, I am able to show pilot evidence that *SGK1* could act as a barrier to tumour development. Without the ability to activate genes downstream of the c-MYC transcription factor such as *BRCA1, CDC25A, cyclin A, p27, TK and E1F4E*, tumour proliferation is limited. This would explain why *SGK1* down-regulation is so pervasive for CRC and why low expression confer significantly poorer prognosis.

The effect of *SGK1* re-expression in CRC appears to be most marked on tumour metastasis. In my animal models, both the orthotopic transplantation model and the tail vein model, *SGK1* re-expressing cells have significantly reduced metastatic capabilities. The reasons for this are likely to be multifactorial; *SGK1* re-expressing cells have a much poorer ability to migrate, demonstrate more cell-cell adhesion and have lower c-MYC expression. The effect of c-MYC expression is likely to be particularly important and the metastasis promoting effects of this gene have been covered extensively already[58]. In addition it lies downstream of the KRAS oncogenic signalling pathway, a known key driver of CRC metastasis.

Work will need to be performed to identify the mechanism by which SGK1 is down regulated in tumours. Previous work in our lab using cancer cell lines have suggested that it does not arise through hypermethylation of the promoter region of SGK1[82], analysis of the TCGA demonstrates that there is no evidence that the SGK1

Figure 3.20: Proposed mechanism of how *SGK1* re-expression affects MYC function



loci on Chr6-q23 is prone to focal deletions or amplification, furthermore, mutation within the gene is rare (1-2% in the TCGA) with a truncating lesion observed in only 1 patient. It is therefore likely that transcriptional control may be exerted by enhancer elements within the promoter. The top transcription factor enhancer site for SGK1, GH06F134286, has binding sites for several transcription factors implicated in tumourigenesis and metastasis including MYC, BRCA1, ZEB1, TCF7L2 and LEF1 and it is feasible that any one of these could lead to down regulation of SGK1 in CRC.

Finally, my work suggests the possibility of targeting *SGK1* or cancer differentiation status to specifically treat metastatic lesions. Whilst *SGK1* re-expression had little effect on overall primary tumour burden, it did specifically reduce lung metastatic burden. Cancer differentiation agents, such as histone acetylation inhibitors (HDAC) vorinostat have failed early phase clinical studies due to poor efficacy[132], however, based on the results of these models, it is possible that these treatments may be much more efficacious at halting metastatic growth.

It is clear that other aspects of *SGK1*'s actions have not been fully elucidated. For example, why the *SGK1*-driven degradation of c-MYC is not more prominent in the other tumour types, and the underlying (epi)genetic features that modulate its effect. However, my data demonstrates that in CRC, *SGK1* has an important role in promoting cellular differentiation and perhaps acts as key kinase barrier that might inhibit the metastatic abilities of CRC.

Chapter 4

A genetic screen investigating regulators of CRC metastasis

4.1 Introduction

Transplantation Mouse models (TMM) remain the workhorse of metastasis research. The basic premise is that cells are labelled, implanted into a mouse and the resulting growth and metastasis analysed. Using such techniques, we have a greater understanding of the genes involved in driving colon cancer metastasis. Numerous genes have been implicated in the process, including *LIN28, S100P, IL8, Galectin 8 and NM23-H1* [133 134 135 136 137]. Metastatic cells often up-regulate expression of genes that are involved in MYC, MAPK, NF-kB and TOR signalling pathways. However, limitations of taking a candidate gene approach are that it is challenging to fully comprehend the relative importance of these genes and they do not screen the metastatic potential of every gene in the genome.

Genetic screening techniques allow a high-throughput, hypothesis-free approach to the identification of candidate metastasis driver genes. Commonly utilised meth-

ods of altering gene function include RNA interference (shRNA, siRNA), lentiviral cDNA expression or by the use of genome engineering tools like Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR).

Numerous screens have been performed in cancer research, principally using breast cancer models. A limited screen targeting 1,500 kinases/phosphatases in a breast cancer cell line was used in combination with imaging to determine the effect on cell motility. Thirty genes were identified that affected migration speed with *SRKP1* highly associated with metastasis in both human and in-vivo validation experiments[138]. A Boyden chamber assay was utilised in a screen for genes responsible for migration of a minimally migratory breast cancer cell line. 11,000 genes were screened using lentiviral delivery of shRNA and this approach identified 31 genes, with cAMP signaling emerging as a key regulator of cellular migration in breast cancer[139]. These techniques have also been employed in mouse breast cancer metastasis models and a cDNA screen identified *ERp5* as an important breast cancer metastasis gene[140], and *KLF17* in a breast cancer RNAi metastasis screen[141].

To date, the most comprehensive metastasis genetic screen has utilised the genome-engineering tool, CRISPR/Cas9. CRISPR/Cas9 is a prokaryotic immune system that was derived from Streptococcus pyogenes and has been repurposed to edit genomes. It consists of a Cas9 nuclease complexed with a synthetic guide RNA (gRNA), and it enables a cell's genome to be cut at a desired location. A lentiviral CRISPR library that targeted 20,611 genes and 1,175 microRNA precursors in the mouse genome was generated to allow a genome-wide screen to analyse gene function at each step of a lung cancer metastasis model. A mouse non-small cell lung cancer line was transduced with this CRISPR library and was injected subcutaneously onto the dorsal surface of mice. Only mice that received cells transduced with the CRISPR lentivi-

ral library developed lung metastasis. The primary tumour was sampled at both an early and late time point; blood and lung tissue containing metastatic cells were harvested. This approach meant that it was theoretically possible to determine drivers genes for each metastatic step. This genome-wide approach identified 24 genes that were over-represented in late stage tumours (metastasis initiating events causing increased growth), of which 4 genes (*NF2, PTEN, TRIM72, CDKN2A*) were validated as important for intravascularisation and 8 genes for metastatic outgrowth in the lung (metastatic virulence events). The majority of the genes correlated with expression changes observed in metastasis in human lung cancer specimens and greatly increased the understanding of genes at critical nodes of lung cancer metastasis[84].

Given that there have been no colon cancer metastasis screens performed to date, I aimed to utilise the OTM described in chapter 3 to determine if it could be used in conjunction with a CRISPR/Cas9 genetic screen to identify important driver genes involved in colorectal cancer metastasis and to subsequently determine their relevance to human cohorts.

4.2 Chapter Methods

4.2.1 CRISPR GeCKO genetic screen- (GECKmA)

4.2.1.1 Generation of lentiviral library

Generation of the GeCKOv2 mouse lentiviral library A (GECKmA) was performed as per the protocol described by Shalem[85,86] with a few modifications. In brief, 10 T-175 flasks of HEK293T cells were each transfected with 32ug of the lentiCRISPR plasmid library, 9ug of VSV-g and 16.25ug of dr8.74 using Plus reagent and Lipofectamine 2000. After 72 hours, the media containing the viral supernatant was removed, filtered

then ultracentrifuged at 24,000rpm for 2 hours at 4°C to obtain the concentrated viral stocks.

4.2.1.2 Transduction of cells

Cells were transduced with optimal virus volumes to establish a MOI of 0.3-0.5. This was achieved by plating 1 million cells in a 12 well plate, supplementing with 8ug/ml polybrene, titrating the virus amount and performing spinoculation. The next morning, cells were split with half receiving puromycin (4μl/ml) and half receiving normal media. Transduced cells were puromycin resistant as the lentiviral library contained a puromycin resistant tag, and this allows percentage transduction to be calculated. The virus volume yielding a MOI closest to 0.5 was selected for large-scale screening. To maintain a 100x representation for large scale screening, 7 million cells were transduced with the lentiCRISPR library virus.

4.2.1.3 Extraction of DNA

The tissues from the experimental mice were placed in cryovials that were placed in liquid nitrogen and stored for a period at -80°C. To extract DNA from these tissues, cryopulverization was utilised. This method relies on the fact that as tissue has high water content, tissues becomes as brittle as glass at liquid nitrogen temperatures. We used the Biopulverizer (*BioSpec*) and DNA was extracted using sodium chloride salt precipitation (NaCl 75mM, EDTA 25mM, Tris 10mM, 1% SDS-page).

4.2.1.4 GeCKO targeted sequencing

Targeted DNA sequencing was performed using a semi-nested PCR strategy as described in the aforementioned protocols. First stage PCR was performed using the following lentiCRISPR gRNAs. F1 tcttgtggaaaggacgaaacaccg R1 tctactattctttcccctgcactgt . To maintain genomic DNA coverage at 300x, 13 separate 100μL reactions

were performed with 10ug of genomic DNA using the Hercules II Fusion DNA polymerase (*Agilent*). The first phase PCR was performed for 18 cycles. The second phase PCR was a semi-nested PCR strategy and was used to insert Illumina sequencing barcodes, a sequence stagger and sequencing primers. This second phase was performed for 24 cycles. DNA was purified, normalised, pooled and quality of DNA assessed using Tapestation. Sequencing was performed using the HiSeq 2500 using 30% PhiX spike-in.

4.2.2 Gene set enrichment analysis

Gene set enrichment is a technique that is able to identify groups of genes that are over-represented in RNA sequencing differential expression lists. It utilises a priori gene sets that have been grouped together based on features such as relevance to particular signaling pathway, biological phenotypes or protein localisation. I used two gene set enrichment programs. Principally, I used javaGSEA version 2 for *Gene Set Enrichment Analysis*, (*Broad Institute, Harvard, USA*)[89] using gene sets from the Molecular Signatures Database (MSigDB)[90] and custom curated gene signatures generated from our group. GSEA were performed using pre-ranked gene differential expression lists. The second gene set enrichment program was the Database for Annotation, Visualisation and Integrated Discovery (DAVID) version 6.7/6.8[91]. DAVID uses a different algorithm to GSEA, looking at overlap between the supplied gene list and the curated database and uses Fisher's exact and Chi-squared tests. GSEA uses every datapoint in the RNA sequencing differential expression dataset, i.e. performing an entire gene-expression profile using Wilcoxon and Kolmogorov-Smirnov tests. GSEA contains up to date datasets from MSigDB including Biocarta, Kegg, Reactome, Go and immunologic signatures and was used where gene expression changes were available. DAVID was used where candidate gene lists had been generated.

4.2.3 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) genetic screen

The CRISPR genetic screen utilises a library of lentiviruses, each containing a synthetic guide RNA (gRNA) that had been designed to target a particular gene with high specificity. In order to determine how gRNA diversity changes under different selection pressures, targeted next generation sequencing was used to amplify the 20bp gRNA sequence. The GECKmA screen contains 67,405 gRNAs. In order to align the FASTQ reads back to the GECKmA gRNA sequences, I demultiplexed the files using Cutadapt version 1.9.1 to identify the barcodes I had assigned to each animal. I trimmed the barcodes so that only the gRNA sequence remained. These fastq files were piped into the MAGeCK Linux bioinformatics program (version 0.5.3). This program is a computational tool that converts the gRNA sequences to genes and performs summary counts for each gene[92].

4.2.4 Gene expression data from Connectivity Map (CMAP)

Connectivity Map (CMAP- *Broad Institute*) is a public repository of genome-wide transcriptional expression data from human cells treated with 1,309 compounds and consists of 7,000 expression profiles. The CMAP version 2.0 interface was used to interrogate similarities of gene expression profiles with those generated from chemicals and bioactive small molecules. The top 500 differentially over-expressed and top 500 differentially under-expressed genes from my experiments were used to generate a list of candidate chemicals and compounds with similar transcriptional expression alterations.

4.3 Results

4.3.1 Selection and characterisation of the MC38 cell line

There are a number of principles that are essential to ensure a successful whole genome genetic screen. One important feature was that any proposed genetic screen must have a low number of false positive results. To this end, it was necessary to identify a cell line that when introduced orthotopically into mice demonstrated low levels of metastasis.

Human CRC cell lines, such as LS174T (described in detail in Chapter 3), HT29 and HCT116 have high rates of metastasis in the NOD scid gamma (NSG) mice and were deemed not suitable. In contrast, I identified that the mouse CRC line, MC38 demonstrated barely detectable metastasis burden, on average $>3x10^3$ lower than human CRC lines. MC38 is a well-established and characterised CRC cell line that was derived from a C57BL6 mouse that had been subjected to subcutaneous injection of dimethyl-hydrazine[142].

The MC38 is a good model for canonical human CRC lesions. It has a similar mutation spectrum to CRC with an intronic mutation in *Apc*, a truncating mutation in *Axin2*, multiple mutations in *Sox9* (which regulates the Wnt pathway) as well as mutations in *Tp53, Pten* and *Braf*. It does not have any mutations in the *Kras* gene[143].

To identify the best host background, the MC38 cell line was orthotopically transplanted into an immunocompromised mouse (NSG) and an immunocompetent mouse (C57BL6) to understand if the presence of an immune response in C57BL6 might suppress MC38 growth following viral transduction. Overall, growth rates of MC38

were significantly lower in C57BL6 mice, but more importantly, there was greater mouse-to-mouse variability in tumour size when C57BL6 mice were used.

Five C57BL6 mice were injected with the MC38 cell line. The line was able to establish in 2 mice, had minimal growth in 1 mouse and had no growth in the remainder.

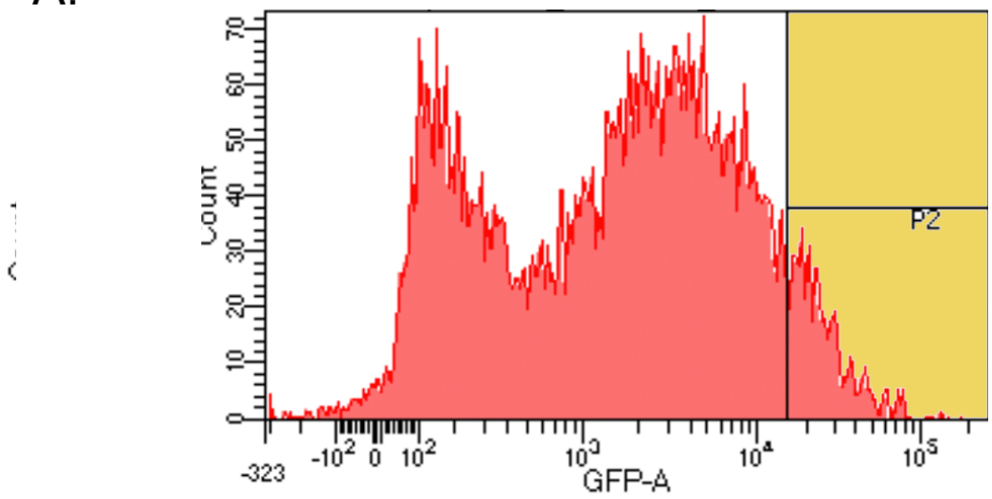
Although sacrificing the realism of an intact immune system, to minimise biological variability NSG mice were selected as the most appropriate model. In order to detect the MC38 tumours and metastatic lesions using in-vivo imaging, the cells were labelled with luciferase using a luciferase expressing lentivirus (pLenti-Zsgreen-Luc2) and then sorted for the top 5% of transduced MC38 cell line by expression of Zsgreen (Fig 4.1a).

4.3.2 Karyotyping of the MC38 cell line

Human colorectal tumours are characterised into those with chromosomal instability (CIN+ve) and these make up approximately 70% of cases, and CIN-ve. In order to be confident that the MC38 line represented a good model for human CRC, the MC38 cell lines underwent karyotype analysis by MFISH to analyse variability in chromosome number. MC38 demonstrated CIN with a modal number of 92 chromosomes (range 56-213), with some recurring structural abnormalities. These included, a deletion in chromosome 4, translocations between chromosomes 2 & 13 and 14 & 5, dicentric chromosomes between 4 & 13 and Robertsonian fusions between chromosomes 6 & 15 and 2 & 16 (Fig 4.2). Targeted sequencing of the MC38 cell line identified the presence of TP53 mutations (G239C and S255I).

Figure 4.1: A. Fluorescence activated cell sorting of MC38-Zsgreen-Luc2. Virally transduced cells have two peaks corresponding to cellular autofluorescence and Zsgreen labelled cells. Cells were sorted from highlighted region B. Endoscopic orthotopic injection, demonstrating injection technique and orthotopic tumour growth at day 7.

A.



B.

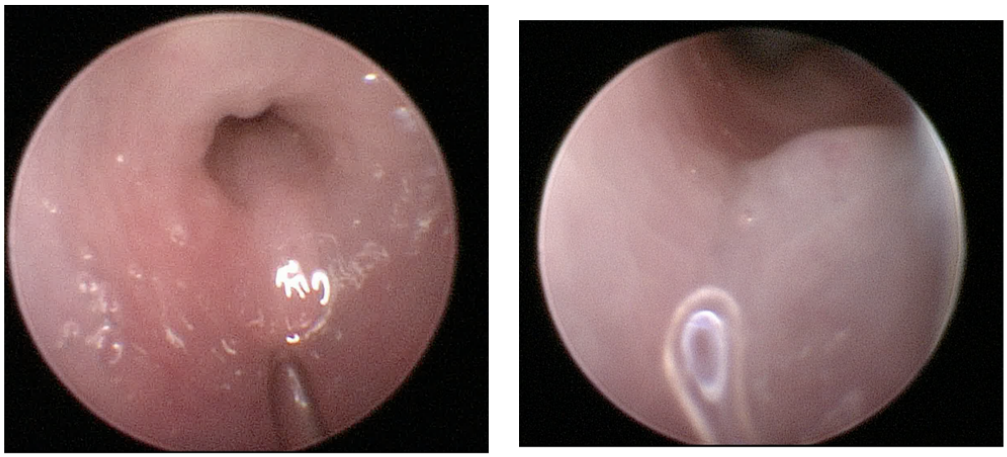
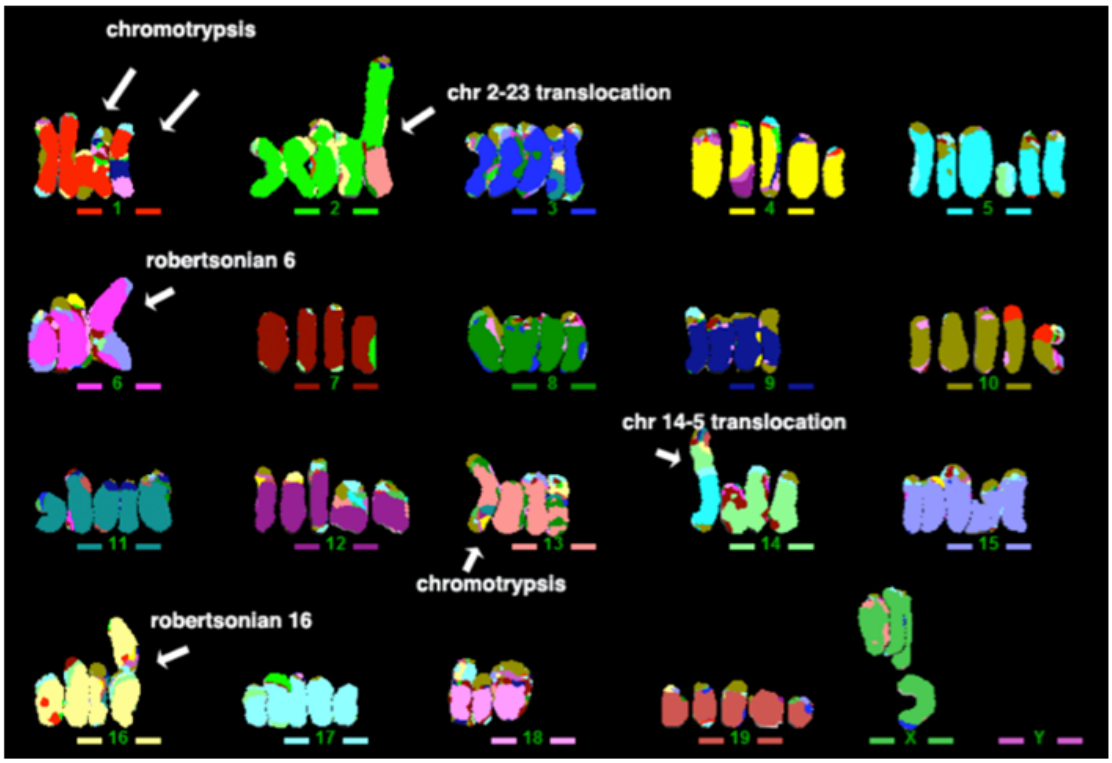


Figure 4.2: MFISH characterising of MC38 cell line demonstrating frequent translocations and structural abnormalities



4.3.3 GeCKO genetic screen in the OTM

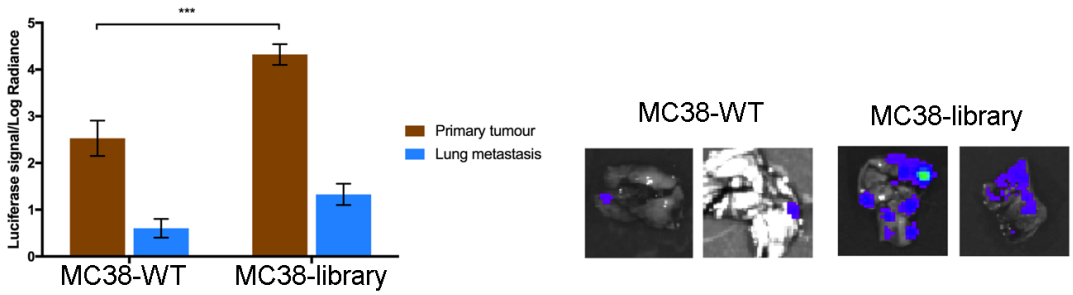
The genetic screen used in this experiment was the GeCKO knockout plasmid library A (GECKmA). It contains a pool of 67,405 gRNAs targeting 20,611 protein-coding genes and 1,175 microRNA precursors in the mouse genome with 1,000 control non-targeting gRNA. The gRNAs were obtained through array DNA oligonucleotide library synthesis and had been cloned into a lentivirus backbone containing a puromycin selection cassette. I used this plasmid library to generate the GECKmA lentiviruses as described in Chapter 2. A good viral yield was obtained with a concentration of 11.8ng/uL. The MC38-Zsgreen-Luc2 cell line was then transduced with the GECKmA library virus (6uL/well) and I obtained a low multiplicity of infection (MOI) of 0.44[85 86]. Low MOI is required to ensure that each cell would on average

receive a single library gRNA, enabling one targeted gene per cell.

Seven million MC38-Zsgreen-Luc2 cells were transduced with the GECKmA library in two independent infection replicate experiments. This meant that each replicate had a library representation (i.e. cells per lentivirus CRISPR construct) of greater than 100x, i.e. high library representation was maintained. After in-vitro culture for 3 days, 1 x 10^6 of the transduced MC38-Zsgreen-Luc2 cells were orthotopically injected into the descending colon of NSG mice using the OTM (Fig 4.1b). A control cohort of mice received 1 x 10^6 of the untransduced MC38-Zsgreen-Luc2 cells. Both GECKmA transduced and untransduced MC38-Zsgreen-Luc2 cells established colonic tumours and mice became symptomatic by day 14 as a result of intestinal obstruction.

The primary tumour size and metastasis rates were assessed by in-vivo imaging. At day 14, colonic tumour and metastasis burden were assessed by bioluminescent imaging. Colonic tumours induced by GECKmA induced cells had grown larger than untransduced cells (paired two-tailed t-test, p=0.009). Ex-vivo analysis of the lung metastasis signal, demonstrated that lung metastasis rates remained low but a higher rate of lung metastasis was observed in the GECKmA transduced cells (paired two-tailed t-test, p=0.05) (Fig 4.3). Mice were sacrificed at day 14 and examined for metastasis in various organs using ex-vivo imaging (imaging following organ resection). Metastases were observed in the lungs in 100% of the GECKmA mice (8/8) and 100% of the untransduced tumour mice (3/3). Liver metastases were less frequent and only detected in 37.5% of the GECKmA mice (3/8) and 33% of the untransduced tumour mice (1/3). It was not possible to detect metastases in the bone, brain, spinal cord, spleen or kidneys.

Figure 4.3: Figure demonstrating larger tumour size and metastasis burden following transduction with GECKmA screen



4.3.4 Optimisation of DNA extraction for GeCKO screen

Targeted next generation deep sequencing was used to investigate gRNA representation in the primary colon CRC and metastatic lesions, and thereby identify genes that led to a proliferative or metastatic phenotype. This is possible because lentiviruses integrate into the host genome, and in the case of the GECKmA viruses, there is integration of the gRNA sequence into the MC38 genome.

It was necessary to isolate the integrated CRISPR/CAS9 gRNA sequences from the primary tumour and metastatic lesions. There are two approaches for achieving this; firstly using metastasis enrichment strategies, such as fluorescence-activated cell sorting/magnetic-activated cell sorting or secondly using whole-organ DNA extraction methods. I was concerned that metastasis enrichment strategies would result in loss of a significant amount of metastatic colonies (and hence affect gRNA representation) and so I elected for the later approach to maintain sensitivity.

I also considered if the lung metastatic lesions should be excised whole, or if lobes or individual lesions should be excised. Lesions were so numerous as to make macro/micro-dissection infeasible. Other papers, including the CRISPR/Cas9 screen in a lung cancer metastasis model[84] had taken individual lobes. I elected to use whole

Table 4.1: Comparison of different methods of DNA extraction

Method	Concentration	260/280
Trizol	16.4ng/ul	1.67
Qiagen kit	1277ng/ul	1.9
Salt precipitation	4181ng/ul	1.75

lungs, as this would enable me to maximise signal to noise ratio of the screen.

At the experimental end point, resected tissues (primary tumour and lung) were snap-frozen and stored at -80°C. To enable DNA extraction from whole organs, they were crypopulverized and digested with protease K. A comparison was made of different methods for whole-organ DNA extraction, including Qiagen blood & tissue kit (*Qiagen*), Trizol extraction, and NaCl salt precipitation. The use of the Qiagen kit resulted in the highest quality DNA as measured by nanodrop 260/280 reading (Table 4.1). However, the Qiagen kit utilised a spin column protocol and was limited by the amount of DNA that could be stored in the columns and this significantly increased experimental cost. Sodium Chloride salt precipitation was selected to extract DNA from the lungs of the experimental mice. DNA was extracted from the primary tumour and lungs from each of the 8 GECKmA mice. At the same time DNA was extracted from the pre-orthotopic transplantation GECKmA transduced MC38 cells and this formed the experimental control.

4.3.5 Optimisation of genomic screen PCR reactions

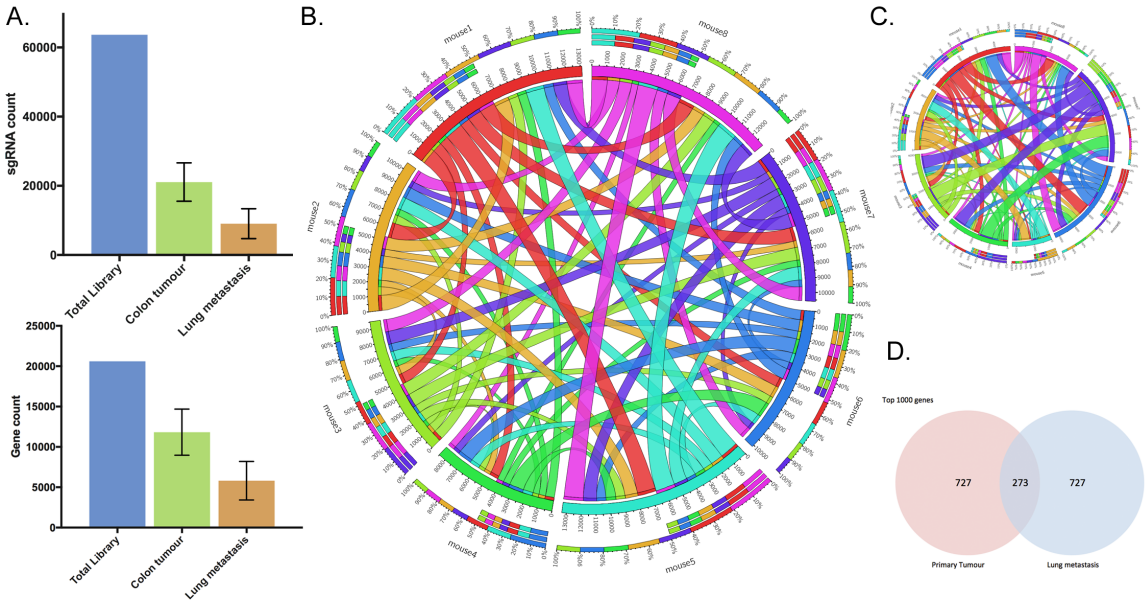
Lentiviral gRNA sequences were amplified from the DNA extracted from the organs and sequencing adapters and barcodes attached. This was achieved through a two-step PCR strategy as described in Chapter 2. To ensure there were no distortion of the gRNA counts/distribution, for both the first and second stage nested PCR reactions,

amplicons were obtained from the linear and not plateau phase of the PCR reaction. This was achieved through lengthy iterative optimisation experiments and I established that 18 cycles for the first PCR reaction (amplification of the lentiviral gRNA sequences) and 24 cycles for the second PCR reaction were sufficient to ensure that the gRNA representation was maintained. The second PCR reaction allowed the addition of a barcode to identify the experimental animal, Illumina sequencing primers and a stagger to minimise mono-template sequencing problems. Mono-template sequencing issues arise from where there is very low complexity in a sequencing run. This was highly likely to be the case during sequencing of the GECKmA transduced cells as the sequences around the gRNA are identical. The stagger introduced into the forwards and reverse primers enabled an increase in amplicon diversity and allowed successful sequencing.

4.3.6 gRNA representation

The gRNA representation of the pre-implantation GECKmA transduced MC38 cells demonstrated that the pre-orthotopic implantation library representation was extremely good. I obtained a total gRNA coverage of 63686/65960 (96.6%). gRNA library distribution was investigated in the different sample types (pre-orthotopic transplantation cells, primary tumour, lung metastasis). The primary tumours at day 14, the experimental end point, retained on average 33.1% of the gRNAs found in the transplanted cell population. This made up an average gene diversity of 57.4%. The gRNA diversity in the lung was markedly smaller with an average gene diversity of 13.7% (median gene count of 2549 genes per mouse) (Fig 4.4a).

Figure 4.4: A. Representation of gecko library-by gene and by gRNA. B. Chord diagram illustrating genes enriched in the lung metastases in each mouse. C. Chord diagram illustrating genes enriched in the primary tumour of each mouse. D. Venn diagram illustrating genes enriched between primary tumour and lung metastases



4.3.7 Analysis of results by cohort

The experiment was performed as two replicative cohorts separated in time, consisting of a total of 8 mice, with independent transductions of cells with the GECKmA virus library. In order to analyse for similarities between genes enriched in metastatic lesions between individual mice or cohorts of mice, I analysed the top 2000 genes that were present in the lung metastasis lesions for the whole cohort and analysed how often these were observed in the top 2000 genes in a single experimental mouse.

This analysis demonstrated that on average, of the top 2000 genes observed in lung metastasis lesions across the cohort, 47.0% of these genes were present in a single mouse (n=8, range 37.3-57.8%). This represents a relatively high level of reproducibility. A chord diagram was generated to illustrate similarities of enriched genes between animals of different cohorts. The segments of the graph denote ex-

perimental mice from different cohorts (mouse 1-4: cohort 1, mouse 5-8: cohort 2). The size of the segments in the diagram demonstrates how much similarity each experimental mice has to the entire cohort. The segments are joined by chords, whose size represents similarity between experimental mice. In this figure, the chords are of similar size and connect each animal with each other. This illustrates the high level of concordance between the mice and cohorts (Fig 4.4b).

I also analysed the genes enriched in the primary tumour. Of the top 2000 genes that were present in the primary tumours for the whole cohort, on average 25.8% of these genes could be identified in a single experimental mice (n=8, range 20.8-31.1%). This was lower than observed in the lung metastases (25.8% vs 47.0% two sided t test, $p<0.001$). One possible explanation is that the growth of a tumour in the colon might involve a much more complex interplay between the microenvironment and the genotype-phenotype association. For example, a gene that conferred a cell a high growth advantage, might be completely negated by being located in a hypoxic area of the tumour, or being on the luminal surface where growth is restricted. A chord diagram was generated to illustrate the genes enriched in the primary tumour of each mouse and demonstrates the greater degree of variability between experimental animals, but still a high degree of overlap (Fig 4.4c). There were some animals that shared a high proportion of genes, such as mouse 1 and mouse 3, and mouse 7 and mouse 4, however, the latter pair were from mice of different experimental cohorts making experimental bias unlikely.

One possible limitation of the experimental design was that it is possible that metastasis in this model could be a passive phenomenon, principally dictated by gRNA representation of the primary tumour. If this were the case, the enriched genes in the lung metastatic lesions would simply reflect those that were enriched

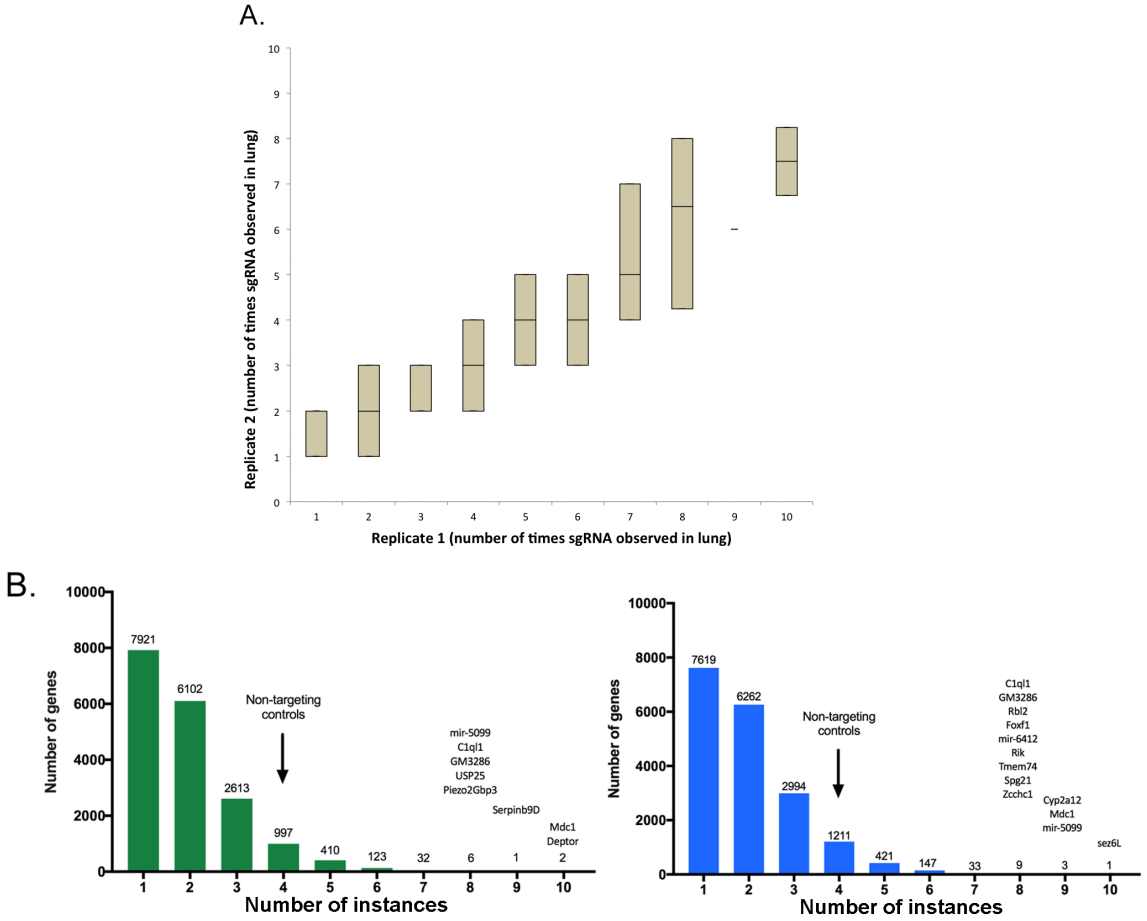
in the primary tumour. A comparison was performed of the top 1000 genes that were enriched in the primary tumours and compared to the top 1000 genes that were enriched in the lung metastasis. Only 27.3% of genes had overlap between the sites (Fig 4.4d). This result suggests that the gRNA diversity was different in the lung metastasis compared to the primary tissues, suggesting contrasting selection pressures.

4.3.8 Enriched genes in the lung metastases from individual cohorts

The GECKmA library is designed such that there are three different gRNAs targeting each gene. This allows the identification of enriched genes in the cohort, and if a gene was highly enriched, it would be identified in all four mice of a cohort, with three different gRNA each, giving a maximum number of instances of 12.

The top hits in the first cohort of mice (n=4 mice) were the genes *Mdc1* and *Deptor* that had 10 instances each. In the second cohort (n=4 mice), *Sez6L* was the top hit with 10 instances (Fig 4.5b). The results from each cohort were very similar. A figure was generated to illustrate how likely a gene was to be enriched in cohort 1 vs cohort 2. The figure demonstrates a high level of concordance, with genes highly enriched in cohort 1 also highly enriched in cohort 2 (Fig 4.5a).

Figure 4.5: A. Number of times genes were observed as enriched in replicate 1 vs replicate 2. B. Genes enriched from replicate 1 vs replicate 2



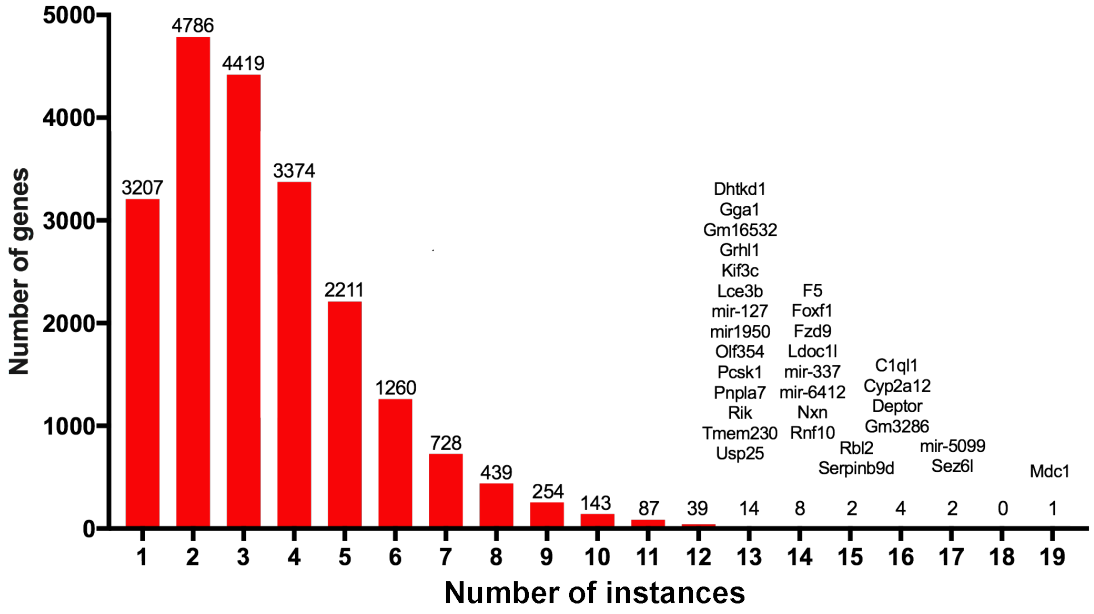
4.3.9 Identification of putative metastasis suppressor genes

The results from the two cohorts were combined to generate a list of genes with effects on metastatic potential (Fig 4.6). If a gene was highly enriched, it would be identified in each of the eight mice of the two cohorts, with three different gRNA each, giving a maximum number of instances of 24. In the metastatic lung lesions, the average gene was targeted in 2 instances. The mean number of times a non-targeting control sgRNA was targeted was 1.07 (maximum of 7). Very few genes were targeted with more than 10 instances.

Mdc1 had the highest number of gene instances in the lung metastasis lesions (19 instances). *Mdc1* is a DNA checkpoint protein and implicated in breast cancer progression. Knockout of this gene has been demonstrated to increase tumour growth in mouse studies[144], however, a role in metastasis has not yet been defined.

The screen identified a number of further interesting hits. The second most enriched gene, *Sez6l*, is a protein encoding a cell adhesion domain and therefore a biologically plausible metastasis regulator (17 instances). *Sez6L* is a protein frequently deleted in lung cancer[145,146] and linked to the development of bone marrow micrometastases in breast cancer[147]. *Deptor* (16 instances) is relevant in the field of CRC research as the gene is a crucial component of the MTORC pathway[62] and expression of this gene has been linked to histological CRC grade and to lymph node metastasis[148]. The genes *Rbl2* (15 instances) and *Usp25* (13 instances) are putative tumour suppressor genes[149,150]. Finally within the top hits were a number of genes linked to Wnt signalling, an important gene for CRC progression, including the Wnt receptor *FZD9* (14 instances), *C10orf11* (16 instances) and the upstream regulator of Wnt, *Nxn* (14 instances).

Figure 4.6: Bar chart identifying genes regulating lung metastasis from all experimental mice. Mdc1 was the top hit with other top hits listed. Most genes were observed in just two instances



To understand if these genes were frequently truncated or mutated somatically in CRC, I analysed data from the TCGA. Of the top 18 genes, the gene was mutated on average in 2.3% of CRC tumours and in 11.8% of the time, the mutations were truncating lesions. (Table 4.2)

4.3.10 Gene network analyses- biological pathways

To gain further insights into the biological pathways that were identified from this knockout screen, gene set enrichment analysis were performed. *Regulation of Heterotypic Cell Adhesion* was found to be significantly targeted (NES 1.78, p<0.001), and this might suggest a link between the ability of a cell to down-regulate cell adhesion and increased rates of metastasis, a phenomenon also noted in chapter 3.

Also enriched were genes involved in *Regulation of Histone Deacetylation* (NES 1.71, p=0.006) with genes like *Trps1, Ski* and *Akap8* being targeted. Histone acety-

Table 4.2: Genes enriched from CRISPR/Cas9 screen and likelihood that the gene had a mutation (data from TCGA) and the likelihood that the mutation would be a truncating mutation

Gene	% mutated tumours	% truncating mutations
MDC1	4.35	20
SEZ6L	3.95	9
DEPTOR	3.7	22
CYP2A7	1.8	7
C1QL1	1.25	14
RBL2	2.95	26
SERPINB9	1.2	14
RNF10	1.65	0
NXN	1.3	14
FZD9	0.95	08
FOXF1	1.2	11
F5	5.5	14
DHTKD1	3.1	12
GGA1	1	18
KIF3C	2.15	18
PCSK1	2.7	0
TMEM230	0.25	0
USP25	2.95	5

lation is an important mechanism for controlling global cell expression as acetylated histones have less positive charges, decreasing the interaction of the N termini of histones with DNA. Histone acetylation enables histones to decondense and result in increased gene transcription. *TRPS1* is an important upstream inducer of overall HDAC activity in cells[151] and *SKI* and *AKAP8* are components of the histone deacetylase complex[152]. The pro-metastatic effects of knockdown of histone deacetylation genes have not been widely described, though they have been previously observed in mouse experiments[153].

Gene Set Enrichment Analysis was then performed using the MSigDB hallmark signalling gene sets. The results from this screen identified that three signalling networks were enriched, *TGFbeta* (NES 1.38, p=0.047), *BMP2* (NES 1.27, p=0.051) and *Notch* (NES 1.51, p=0.007) (Table 4.3).

Transforming Growth factor beta (TGF-beta) is a cytokine that is secreted by many cell types including macrophages and lymphocytes. Its role is complex, but in early tumours without TGF-beta pathway mutations, it is anti-tumorigenic. TGF-beta acts to stop cell cycle progression at G1 and thus stop proliferation of cells[154,155], suppress angiogenesis[156], induce colonocyte differentiation[157] and promote apoptosis[158]. In keeping with this role, in this screen I found that genes from this pathway such as B-cell translocation gene (*Btg-1*) have been targeted in metastatic lesions. *Btg*-1 expression has been demonstrated to inhibit proliferation, migration and invasion in endothelial cells and gastric tumours[159,160].

BMPs are a group of growth factors that belong to the TGF-beta superfamily of proteins. Expression of BMP2 is known to be down-regulated in colon cancer compared to normal tissue, and a number of genes from the BMP2 pathway were

Table 4.3: GSEA/GO analyses of hits identified from CRISPR/Cas9 lung metastasis screen. Full hallmark MSigDB gene set results in appendix A.2 on page 207

	NES	p
Regulation of Cell adhesion	1.78	<0.01
Regulation of Histone Deacetylation	1.71	<0.01
Glycolysis	1.29	0.04
Apoptosis	1.29	0.06
Autophagy	1.11	0.3
EMT	0.97	0.53
Differentiation	0.88	0.67
Scenescence	0.83	0.74
Proliferation	0.61	0.99
TGFb signalling	1.38	0.05
BMP signalling	1.27	0.05
NOTCH signalling	1.51	<0.01
TP53 signalling	0.26	0.93
KRAS signalling	1.25	0.07
MTORC signalling	0.81	0.85
Myc targets	-1.08	0.17
Wnt signalling	-0.99	0.47

identified from this screen. These include *Gata2* and *Prkd1*, both known inhibitors of cellular migration[161 162].

The final enriched pathway, Notch, is generally believed to be an oncogenic pathway, being linked to tumour growth and angiogenesis and it was surprising that it was so strongly implicated in this screen. A number of genes from this pathway were identified, including *Notch3, Wnt4, Fzd5* and *Id2*. Whilst there is evidence that Notch may be pro-tumourigenic in-vitro, in animal studies, metastasis rates from neuroblastoma and breast cancer cell lines are actually increased by heterozygous loss of Notch 1 and from the use of the Notch inhibitor PF-03084014[163]. NOTCH signalling might therefore have differential effects on metastasis, depending on the tumour background.

4.3.11 Primary tumour- positive screen results

To investigate which gRNAs were enriched in the primary tumour, I selected the top 2000 most enriched genes in each mouse and performed *Gene Ontology* analysis. Few GO terms were enriched, though pathways involved in cell death and apoptosis were enriched. gRNA frequently targeted apoptotic genes, such as programmed cell death proteins such as *Pdcd10, Pdcd6, Pdcd7*. There was also enrichment for a number of genes targeting the BCL-2 family of regulator proteins that regulate apoptosis. This included proteins such as BCL2 binding component 3 (*Bbc3*), BCL2 modifying factor (*Bmf*), BCL2-associated transcription factor 1 (*Bclaf1*), BCL2-like 2 (*Bcl2l2*) and BCL2-like 14 (*Bcl2l14*). Also enriched in this screen, was *Btg2* a well described anti proliferative gene[164] and proteins involved in the caspase apoptosis signalling pathway (*Card9/Nod1*) (Table 4.4).

4.3.12 Primary tumour- negative screen results

A large number of gRNAs from the original population of cells were lost in the primary tumour. I selected the top 2000 genes that were most frequently lost from the primary tumour across the entire cohort and performed *Gene Ontology* analysis. The genes frequently lost in the tumour were those involved in fundamental cellular processes, such as *Regulation of Transcription*, *Regulation of RNA Metabolic Reactions*, *Protein Hydrolysis, Reduction, Transport and Localisation*. *Regulation of Transcription* was the GO term that was found in 12.4% of the lost gRNAs. These results demonstrate that there was selection against the loss of genes with key housekeeping functions in the primary tumour (Table 4.4).

Table 4.4: Gene set enrichment analysis of primary tissue- negative and positive screen results

GO TERM	% of genes
NEGATIVE SCREEN	
Regulation of transcription	12.4
DNA dependent regulation of transcription	8.4
Regulation of RNA metabolic process	8.4
Proteolysis	6
Oxidative reduction	4.3
Protein transport	3.9
Establishment of protein localisation	3.9
RNA processing	3
Negative regulation of gene expression	2.7
POSITIVE SCREEN	
Regulation of cell death	0.4
Regulation of apoptosis	0.4

4.3.13 Validation in-silico of the top enriched genes from lung metastases

In order to understand if the genes identified from this model system are important in regulating metastasis in human CRC cohorts, comparison with human metastatic specimens was necessary. Expression data from CRC metastasis specimens are challenging to find, though a literature review identified one study where data was publicly available. The authors had collected a total of 252 specimens (consisting of 186 primary tumours, 47 liver metastasis and 19 lung metastasis specimens) and had performed gene expression analysis using the Affymetrix U133A array[1]. The expression data was downloaded from the *Gene Expression Omnibus* and I performed analysis of the expression data to allow a comparison with the top 18 genes identified from my GECKmA screen.

Analysis of the human CRC lung metastasis specimens from this dataset, identified that 17 of the top 18 genes identified in this GECKmA knockout screen demonstrated

Table 4.5: Top 18 genes identified from GECKmA screen and expression in human metastasis[1]

	Lung metastasis		Liver metastasis	
	Expression Change	p-value	Expression Change	p-value
MDC1	-0.153	<0.001	0.060	0.099
SEZ6L	-0.090	<0.01	-0.019	0.476
DEPTOR	-0.084	0.39	0.065	0.381
CYP2A7	-0.058	0.22	0.220	0.003
C1QL1	-0.019	0.59	0.128	<0.001
RBL2	-0.062	0.24	0.090	0.02
SERPINB9	-0.131	0.10	-0.038	0.64
RNF10	-0.024	0.56	0.005	0.88
NXN	-0.007	0.93	-0.299	0.00
FZD9	0.041	0.37	-0.058	0.08
FOXF1	-0.419	<0.001	-0.606	<0.001
F5	-0.052	0.80	1.433	<0.01
DHTKD1	-0.219	<0.001	0.020	0.83
GGA1	-0.138	0.09	0.086	0.11
KIF3C	-0.098	0.04	0.034	0.39
PCSK1	-0.588	0.06	-0.580	0.08
TMEM230	-0.303	<0.001	-0.189	0.02
USP25	-0.044	0.41	0.147	<0.001

a corresponding decrease in expression in human lung metastasis compared to the primary tumours. The changes in expression of 6 of these 17 genes were significantly decreased ($p<0.05$). In the liver metastasis, expression of 7 of the 18 genes was lower than in the primary, with 1 gene significantly decreased (FOXF1, $p<0.001$) (Table 4.5).

In summary, there appeared to be an overlap between genes that had been identified from lung metastasis specimens from my GECKmA mouse CRISPR knockout screen and human lung metastasis samples. One of these genes, Forkhead box F1 (FOXF1), had decreased expression in both liver and lung human metastasis specimens suggesting the possibility that it might act as a pan-metastasis regulator.

4.3.14 Validation of GECKmA genes

My next aims were to validate selected results of the GECKmA screen and to gain a greater understanding of the biological effects. Three genes (FOXF1, MDC1, RBL2) were selected for validation experiments.

FOXF1 was selected for validation experiments as expression of this gene was significantly down-regulated in both liver and lung human metastasis. FOXF1 codes for a transcription factor that belongs to the forkhead family of proteins. These transcription factors are important during embryonic development where they play an important role in lung, vasculature and intestinal morphogenesis[165 166 167]. During intestinal development, FoxF1 has been identified as a target transcription factor for the hedgehog signalling pathway, and Foxf1 $^{+/-}$ mutant mice have perturbations in the crypt-villus axis with increased proliferation and Wnt signalling[168]. Control of expression of FOXF1 is primarily transcriptional with a role for promoter hypermethylation[169] and control by p53 response elements[170]. FOXF1 regulates the transcriptional activity of the cell adhesion proteins, E-cadherin (CDH1)[170], the integrin ITGB3, a receptor for cell-cell and cell-extracellular matrix interactions[171] and the cell cycle protein p21[172]. In breast and colon cancer, it has a role as a tumour suppressor gene enabling cell cycle G1 arrest, through inhibition of CDK2-Rb-E2F[169 173]. In prostate cancer, FOXF1 is proposed to promote tumour growth[174]. Its role in inducing EMT is unclear with different effects on differing cell lines[171 175].

Mediator of DNA damage checkpoint protein 1 (MDC1) was selected for validation experiments as it was the most significantly over-represented gene from the GECKmA screen. *MDC1* is a nuclear protein that plays a key role in the DNA damage response pathway and senses DNA damage by binding to y-H2AX where it facilitates retention of ATM. Mice with knockdown of Mdc1 are more radiosensitive and develop more

tumours than wild type mice and this was associated with increased rates of genomic instability[176]. In breast and lung cancer, it functions as a tumour suppressor[177,178], but functions as an oncogene in cervical cancer[179].

Retinoblastoma-like protein 2 (RBL2) is related to the tumour suppressor gene Rb. It functions as a key regulator of cell division. It targets histone methyltransferases leading to transcriptional repression and decreased cell proliferation[149]. It is an inhibitor of E2F and plays a role in the initiation of senescence[180]. Over-expression or restoration of RBL2 function suppresses the growth of glioma and lymphoma cell lines[181,182] and aberrant promoter methylation of this gene has been found in a high percentage of breast cancer tissues[183].

4.3.15 Generation of individual lenti-CRISPR viruses and efficiency

I cloned three gRNAs targeting each of these genes into the pLenti-CRISPR vector and generated a small lenti-CRISPR virus pool that targeted each gene (i.e. The *Foxf1* virus pool contained three gRNAs, targeting *Foxf1*). The MC38-Zsgreen-Luc2 cell line was then transduced with the gene virus pool (Table 4.6).

To determine the success of this strategy to knockdown gene expression, I took two approaches. Firstly, I performed TIDE analysis[184]. Double strand DNA breaks induced by the CRISPR/Cas9 gRNA complexes are repaired by cells using the error-prone non-homologous end-joining system and this results in the generation of indels at the repair loci.

These CRISPR/Cas9 induced indels are known to be predominantly heterozygous, but in a small percentage of cells, they may be homozygous. The degree of heterozygous to homozygous mutations was not tested directly in this thesis as the main

Table 4.6: Validation experiments, inserts for gRNA targeting Rbl2, Mdc1 and Foxf1

gRNA	Mouse
Rbl2-a	CCCCATGATTAGCGATGATC
Rbl2-b	CTAATCTTTCAGTACGTTCT
Rbl2-c	ACCTCTGCTCCGAACAGCGA
Mdc1-a	GCTGTCGTTCCGTTAGCTTC
Mdc1-b	TGCTGCAGGTGACCCGCTTT
Mdc1-c	TTACGATTTGAGTGCCATTG
Foxf1-a	GTACCCGCACCACGACAGCT
Foxf1-b	GCGACTGTGAGTGATACCGA
Foxf1-c	GTGCGGCTCCATGACTCCGC

Figure 4.7: A. DNA chromatograph following CRISPR-targeted gene knockdown, demonstrating indels and resulting frameshift. B. Tracking of Indels by Decomposition (TIDE) identifying regions of aberrant sequences identified in green. C. Indel frequency by position relative to gRNA target site

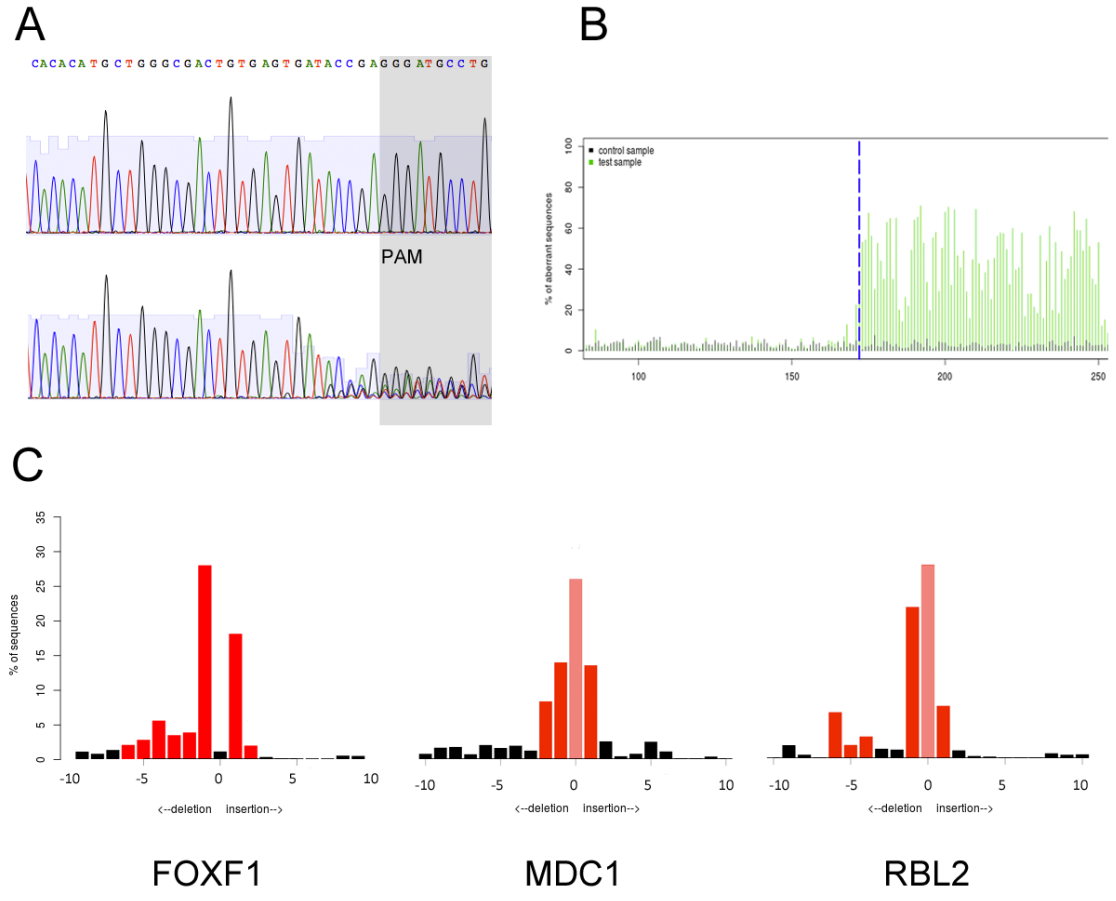
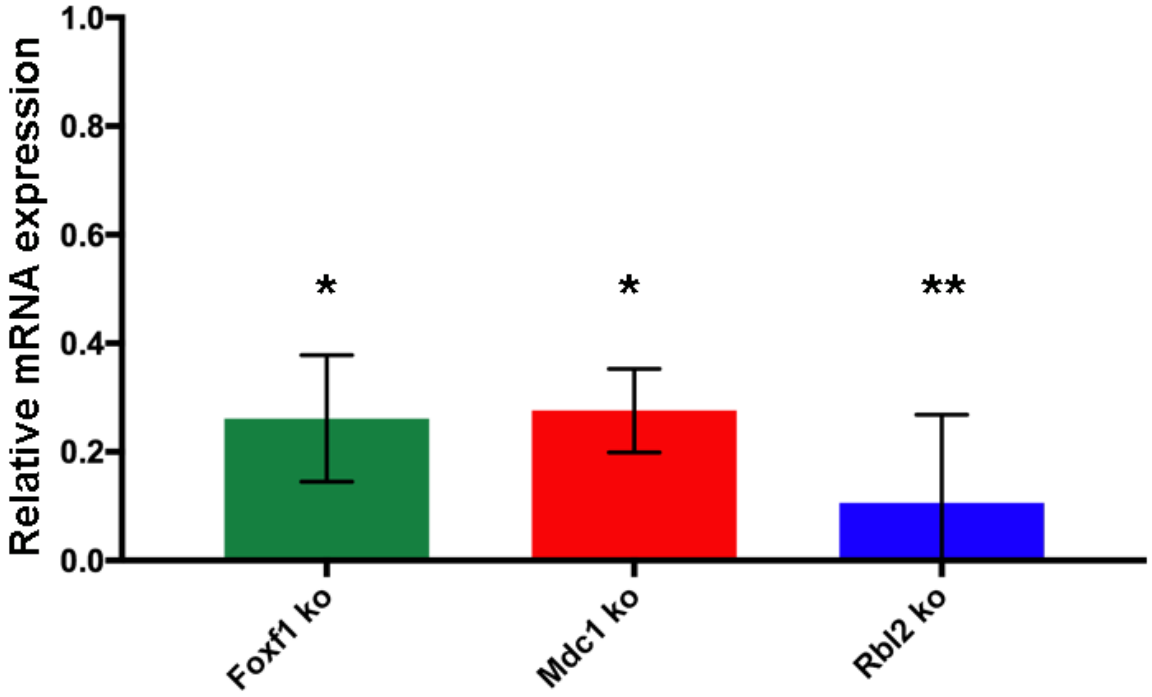


Figure 4.8: Relative CRISPR mediated knockdown of gene expression, determined by rt-qPCR. Expression was reduced to less than 30% for each of the genes relative to wild type cell lines (two sided t test $p<0.05$). Error bars denote standard deviation



objective was to down-regulate the expression of these genes in a cell line population, but were presumed to be heterozygous.

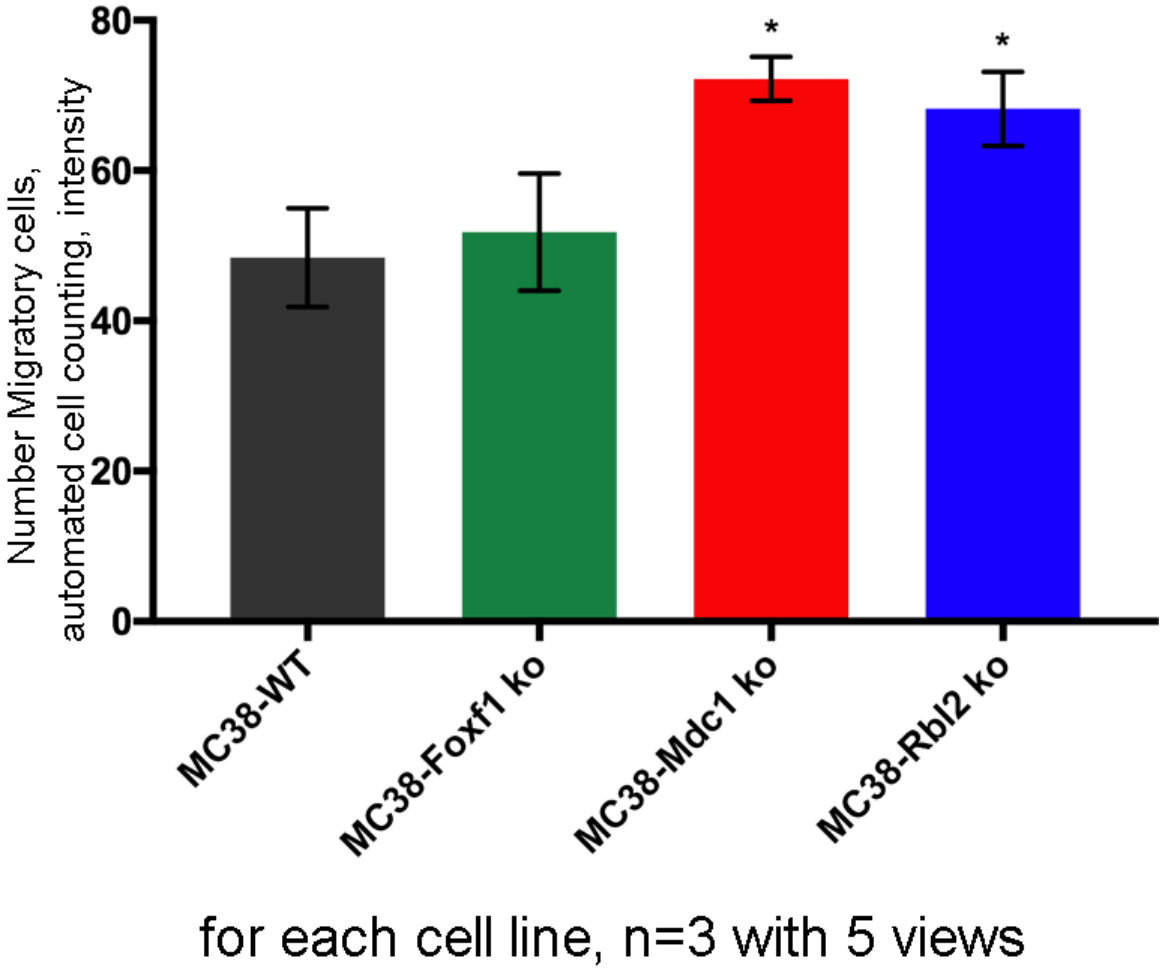
TIDE analysis was performed to determined efficiency of CRISPR/CAS9 through use of colony Sanger sequencing and the detection of these indels. As expected, there was a high frequency of indels around the target site with an average efficiency of 68.1% (45.6-85.2%). Insertions and or deletions were more common, and the majority of these happened in a 5 base pair region of the designed gRNA target site (Figure 4.7).

The presence of indels at a genomic locus disrupts protein function and expression. The mRNA transcripts generated are also liable to a cellular process of *Nonsense Mediated Decay*. This is a surveillance pathway in eukaryotes and results in the elimination of mRNA transcripts that contain premature stop codons. I looked at mRNA transcript expression at day 7 to confirm that they had been successfully targeted using the lenti-CRISPR system. The lenti-CRISPR system was highly effective and mRNA expression was decreased to less than 30% of baseline levels for each of the three genes analysed (Figure 4.8).

4.3.16 Analysis of migration using individual lenti-CRISPR transduced cells

An early prerequisite of metastasis is the development or increase in cellular motility. Having generated the three cell lines, MC38-Foxf1 ko, MC38-Mdc1 ko, MC38-Rbl2 ko, I analysed for a change in cellular motility using a Boyden transwell assay. Compared to the wild type MC38-Zsgreen-Luc2 cell lines, the number of migratory cells at 16 hours was significantly increased for Mdc1 and Rbl2 knockout cell lines with increases of 49.1% and 40.9% ($p=0.02$ and $p=0.04$ respectively). There was no significant change in migration rates for the FoxF1 knock down cell line (Figure 4.9).

Figure 4.9: Effect of gene knockdown and migration using in-vitro migration assay. Error bars represent standard error of the mean



4.3.17 Validation in-vivo using individual gRNAs

The lenti-CRISPR transduced cells (MC38-Mdc1 ko, MC38-Rbl2 ko, MC38-Foxf1 ko) and control cells (MC38-WT) were then transplanted orthotopically into the colon of immunocompromised mice (NSG). All transduced cell lines formed tumours in situ. At day 14, the primary tumour size was quantified using in-vivo imaging. There were no differences of the primary tumour size of any of the cell lines compared to the wild type cells. There was a trend for MC38-Mdc1 tumour to be smaller, however, this was not significant (p=0.067) (Figure 4.10a). This trend could be illustrative of the dual function of *MDC1* in cell cycle progression. It is a regulator of G2/M cell cycle checkpoint, but also has an anti-apoptotic role. MDC1 inhibits p53 driven apoptosis by binding to the n-terminus of p53 which blocks transactivation and through reduction of phosphorylation of p53 ser-15[185]. Knockout of *Mdc1* in these cell lines could have lead to increased apoptosis and smaller primary tumour growth.

However, the key result from this assay was that the lung metastasis burden was significantly increased for mice that received the MC38-Foxf1 ko and MC38-Mdc1 knockout cells (p=0.01, p=0.04 respectively), thus validating the results from the original GECKmA screen (Figure 4.10b,c).

The MC38-Rbl2 knockout line did not demonstrate any increase in lung metastasis burden and this was surprising. One possible reason for this difference was the design of the gRNA. The gRNAs I used in this experiment to target Rbl2 (Table 4.5) were not the same as those used the original GECKmA screen. The gRNAs used in these validation experiments were different from those in the initial GECKmA screen, designed to enable me to minimise bias from off-target effects from the original screen. My gRNAs were designed to target the 5' end of the mRNA transcript, whereas the gRNAs used in the GECKmA screen were selected solely on the basis

of specificity. There are likely to be differences in protein function depending on gRNA location, efficacy and off-target effects and this could be responsible for this contradictory finding.

To determine off-target effects, it would be necessary to perform whole genome analysis on the samples. This could not be analysed using my data set as targeted sequencing had been performed to understand the CRISPR/Cas9 library representation. The costs to perform whole genome sequencing on the several mouse lines had not been costed into the original thesis.

4.3.18 Analysis into the effects of FOXF1 expression

I set out to define the mechanisms by which decreased expression of *FOXF1* increased metastasis burden. Human CRC cell lines, like primary CRCs, express low levels of *FOXF1*, so I designed a lentivirus to allow me to re-express the gene in CRC cell lines. To achieve this I used two strategies, using CRISPR/Cas9 Synergistic Activation Mediators (SAM) and lentiviral expression of the *FOXF1* coding sequence. The aim of these two approaches were to give me physiological and super-normal expression of *FOXF1*.

I elected to investigate FOXF1 in more detail and not other targets because of the marked association between FOXF1 expression and liver and lung metastasis in human cohorts. I was also particularly interested in the ability of transcription factors to modulate metastatic potential. As FOXF1 was a relatively uninvestigated pioneer transcription factor, I selected this gene for investigation in more detail.

CRISPR-SAM consists of three components, a nucleolytically inactive Cas9-VP64 fusion, a gRNA incorporating MS2 RNA linkers, and a MS2-P65-HSF1 helper protein. This CRISPR-SAM complex induces gene expression through the presence of the three activation domains (VP64, P65 and HSF1) that are targeted to the *FOXF1*

Figure 4.10: A. No difference in primary tumour growth in orthotopic mouse transplantation model was observed following k/o of Foxf1, Mdc1 or Rbl2. B. Gene knock-down did result in increased lung metastasis burden with Foxf1 and Mdc1 k/o in orthotopic mouse transplantation model (two sided t test, $p<0.01$ and $p<0.05$ respectively). Error bars represent standard deviation. C. Ex-vivo lung bioluminescent images demonstrating increased lung metastasis burden following knock-down of FoxF1 and Mdc1

. Error bars represent standard deviation.

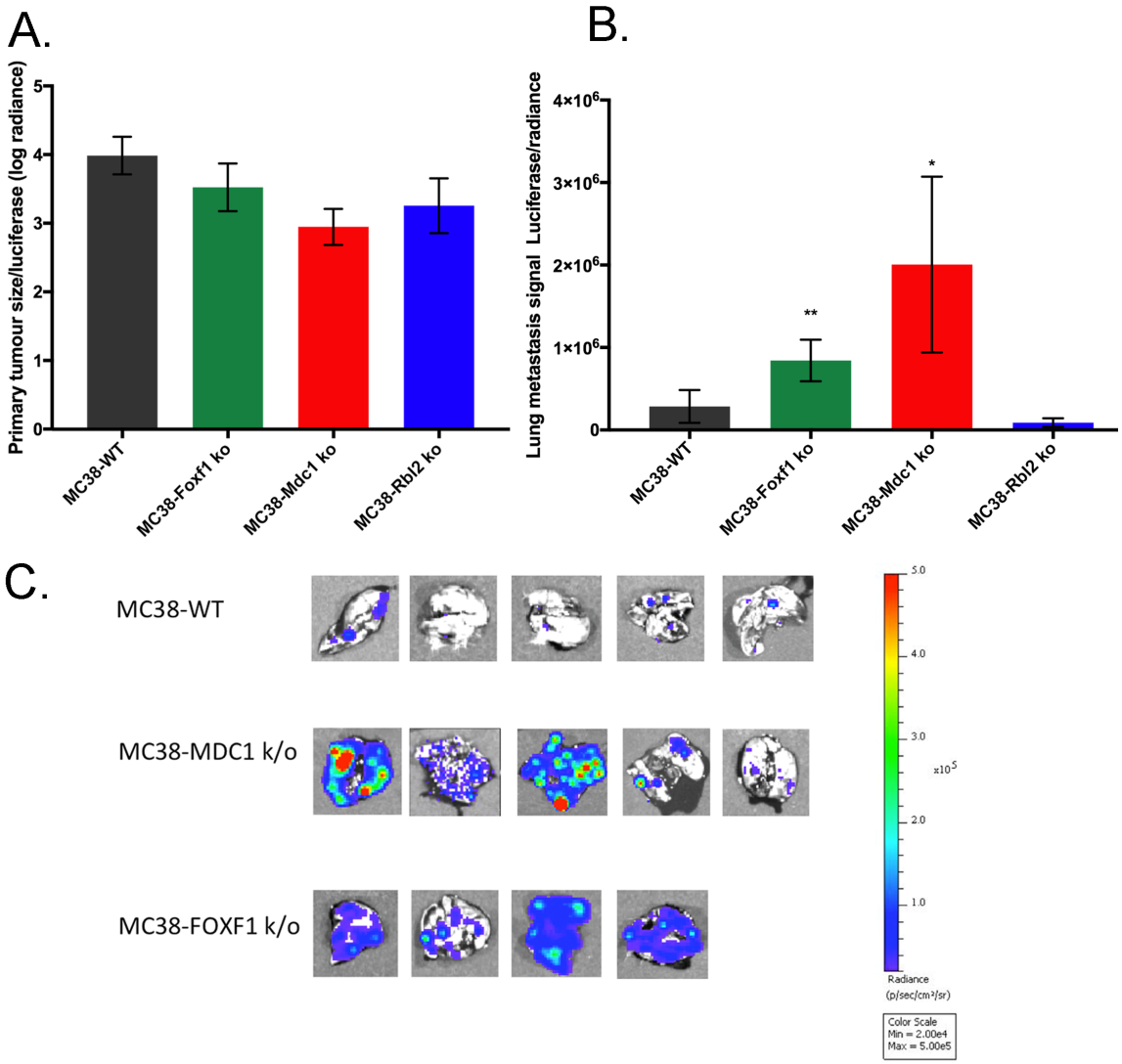


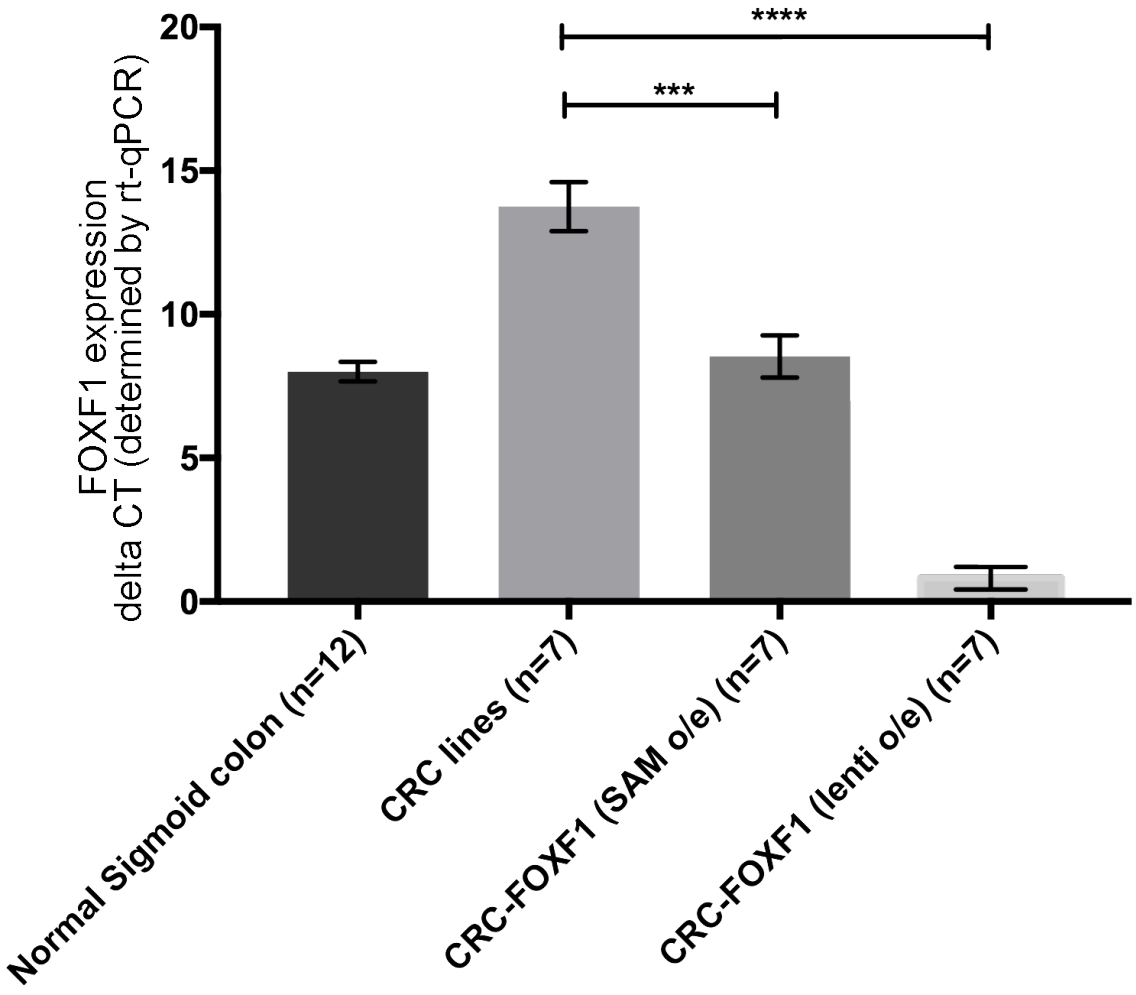
Table 4.7: SAM CRISPR/Cas9 activator FOXF1 gRNA sequences

	SAM gRNA
FOXF1 a	GAGGAGGAGGAAGAGGACGA
FOXF1 b	GCGGGCGGGAGCGCGGGGGC
FOXF1 c	GGGGGCTGCGCGCGCGGGGC
FOXF1 d	ATCCAGCAGCTCCTCGGGGA

loci and function to open the chromatin and allow initiation of transcription[186] (Table 4.7). Lentiviral expression of *FOXF1* was achieved by cloning the coding sequence of *FOXF1* into the lentivirus backbone, pLenti-710, generating the lentivirus, LENTI-710-FOXF1.

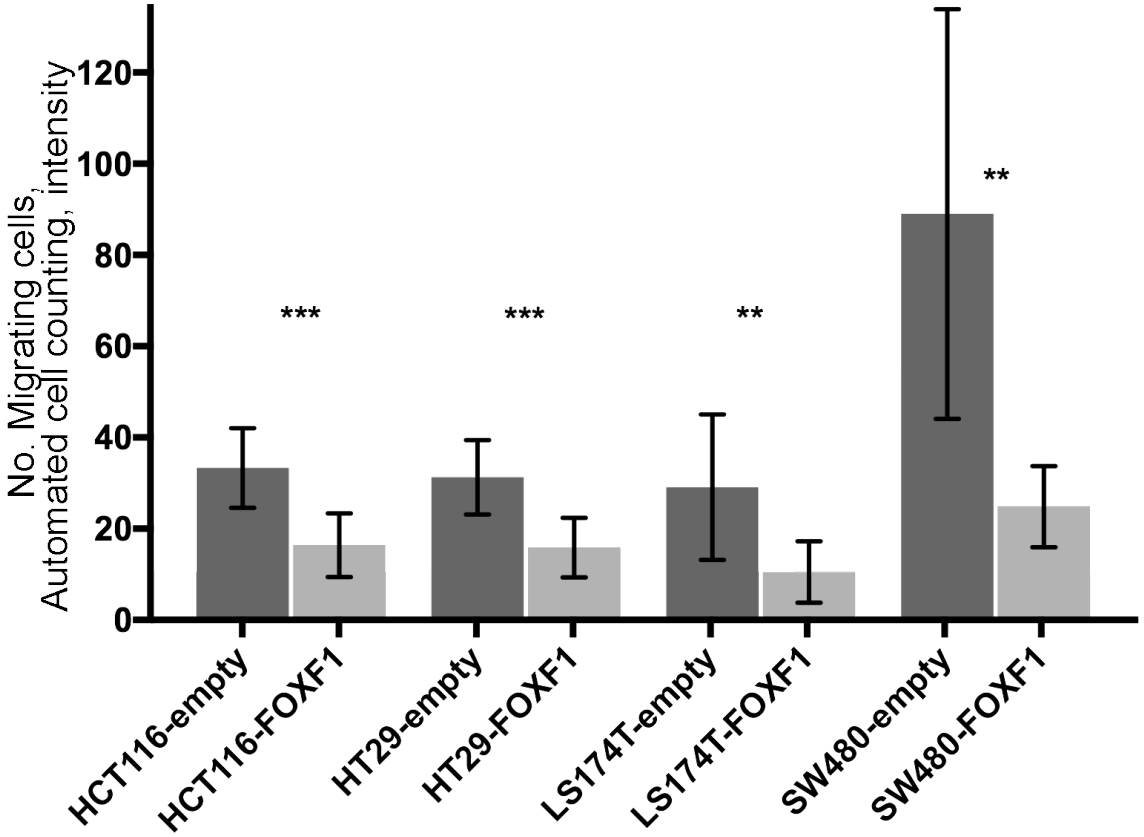
Seven CRC cell lines (GP2D, HT29, HCT116, LS174T, SW1222, SW480 and T84) were then transduced with these lentiviruses utilising these two complementary approaches. Control cell lines received empty lentivirus (lentivirus containing no cDNA). The success of these two strategies to re-express *FOXF1* was determined by rt-qPCR (Fig 4.11). The CRISPR-SAM strategy resulted in an increase of *FOXF1* to comparable levels as normal tissues. Expression of *FOXF1* using the LENTI-710-FOXF1 strategy resulted in greater increases in mRNA expression levels of *FOXF1* to between 50-1000 x the level found in normal tissues.

Figure 4.11: Overexpression of *FOXF1* using CRISPR-SAM strategy and Lentivirus strategy in CRC lines, compared to normal expression of *FOXF1* in sigmoid colon . Error bars represent standard deviation.



To investigate the effect of re-expression of *FOXF1* on cellular migration, a prerequisite to metastasis, a migration assay was performed using CRC lines over-expressing *FOXF1* by lentivirus strategy. I performed this assay in HCT116, HT29, LS174T, SW480 CRC lines as they have migratory capacity. Compared with cells that had been transduced with an empty lentivirus, *FOXF1* over-expressing cell lines had significantly lower rates of cellular migration (mean 38.4% of normal, $p<0.05$) (Fig 4.12).

Figure 4.12: Effect of FOXF1 over expression on cellular migration of CRC lines. Error bars represent standard deviation.



I performed 3' RNA sequencing on the *FOXF1* re-expressing cell lines, generated by either CRISPR-SAM or lentiviral transduction (GP2D, HT29, HCT116, LS174T, SW1222, SW480 and T84). Some of the perturbed gene sets were similarly altered whether *FOXF1* was re-expressed by either CRISPR-SAM or lentiviral transduction. For example, *FOX1* re-expressing cells had enrichment for "apoptosis" gene sets and negative enrichment for "Proliferation", "Glycolysis", "G2M checkpoint" and "MTORC1 signalling". However, other gene sets were only altered in the cell lines that had physiological re-expression of *FOXF1* using CRISPR-SAM ("P53 pathway", NES 1.32, p=0.02; "EMT", NES -1.37, p=0.02) and others only in cell lines that had higher expression of *FOXF1* using lentivirus transduction (Rb/E2F targets, -2.41, $p<0.01$; Myc targets, -2.76, $p<0.01$, Notch signalling, -1.55, p=0.02). These

Table 4.8: GSEA results of CRISPR-SAM and Lentivirus over expression of FOXF1. Full table in appendix A.3 on page 208

MsigDB	CRISPR-SAM NES	p-val	Lentivirus NES	p-val
MTORC1 SIGNALING	-1.47	0.00	-1.74	0.00
Rb/E2F TARGETS	1.06	0.30	-2.41	0.00
MYC TARGETS	-0.91	0.72	-2.76	0.00
NOTCH SIGNALING	0.91	0.61	-1.55	0.02
P53 PATHWAY	1.32	0.02	1.27	0.06
HEDGEHOG SIGNALING	1.09	0.33	-1.22	0.16
KRAS SIGNALING	-1.28	0.04	-0.96	0.60
WNT BETA CATENIN SIGNALING	0.96	0.53	0.87	0.68
GLYCOLYSIS	-1.74	0.00	-1.34	0.01
PROLIFERATION	-1.31	0.02	-2.08	0.00
APOPTOSIS	1.24	0.07	1.58	0.00
G2M CHECKPOINT	-1.21	0.09	-1.89	0.00
HYPOXIA	-2.01	0.00	1.42	0.00
EPITHELIAL MESENCHYMAL TRANSITION	-1.37	0.02	1.05	0.34

findings would be suggestive of the presence of a dose-effect (Table 4.8).

Some of these activated pathways reproduce findings from the literature, for example, the inverse relationship between *FOXF1* re-expression and *Rb/E2F* signalling has been demonstrated in breast cancer cell lines where *FOXF1* decreases phosphorylation of *CDK2*, an inhibitor of Rb[169]. Furthermore, the ability of *FOXF1* to effect EMT has been previously described in mammary epithelial cell lines and mesenchymal cells[175 171].

The benefit of whole transcriptome profiling is that it helped identify of a novel association that has not been previously described before, that of a negative enrichment of MTORC1 signalling. This was of interest because one of the original hits from the original GECKmA screen, DEPTOR, is a key component of MTOR signalling, and it was negatively enriched irrespective of the degree of re-expression of *FOXF1*.

4.3.19 FOXF1 and MTOR signalling pathway

The MTOR protein is a member of the phosphatidylinositol 3-kinase-related kinase family of protein kinases. It forms two distinct multi-protein complexes, MTORC1 and MTORC2, which in turn regulates distinct cellular processes. MTOR1 is induced in response to Akt/PKB, MAPK/ERK and Wnt signalling and enables activation of proteins that are important for cellular growth and proliferation through p70-S6 kinase 1 and 4E-BP1. MTOR2 is induced in response to insulin, growth factors and nutrient levels and regulates cell metabolism, the actin cytoskeleton and proliferation via IGF-IR, Akt/PKB and SGK1. In many types of cancer, MTOR signalling is dysregulated. This pathway and *PTEN* mutations are particularly important in tumour initiation of melanomas[187]. In CRC, the expression of MTOR signalling components, including MTOR, p70S6K and 4EBP1 are associated with higher grade tumours with increasing depth of invasion of CRC[188].

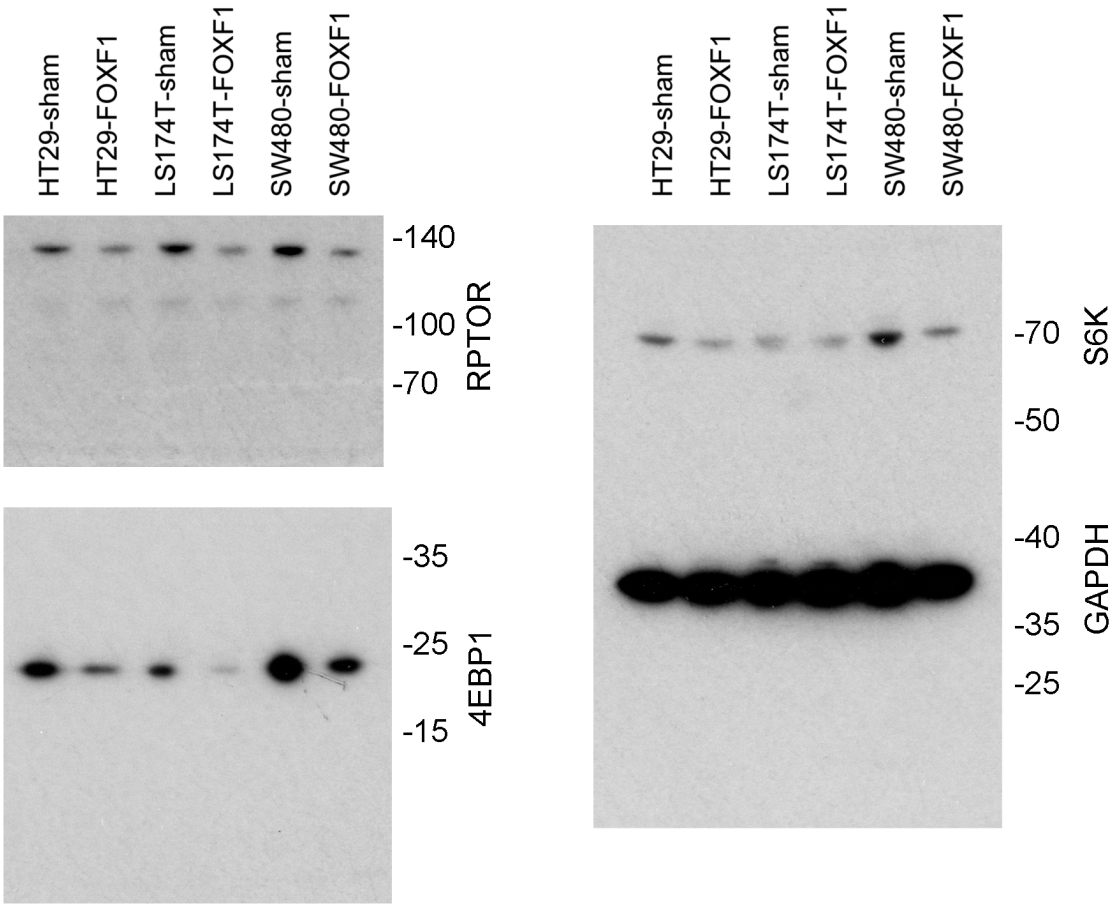
A more detailed analysis of the GSEA results demonstrated that FOXF1 re-expression lead to down-regulation of several oncogenes that are regulated by the MTOR signalling pathways, including *TFRC, MLLTII, IDH1, TCEA1* and *GMPS*. Other down-regulated downstream proteins of MTORC included genes involved in cell cycling (*PLK1, CDC25A*) and cell metabolism (*LDHA, IDI1, ACLY*).

To confirm these findings at a protein level, I analysed the expression of downstream effectors of MTOR signalling using western blotting. *FOXF1* was re-expressed in HT29, LS174T and SW480. Western blots were performed for S6K and 4EBP1. As suggested by the GSEA analysis, these genes were significantly down-regulated following FOXF1 re-expression (Fig 4.13b).

These findings pose interesting questions into the function of FOXF1. 4E-BP1

is a translation repressor protein and regulation of the protein is thought to be predominantly post-translational. Phosphorylation of 4E-BP1 is mediated by mTOR at Thr 37 and Thr 46, however, it can be phosphorylated by a number of other kinases including GSK3b, MAP kinase p38 and ERK. Phosphorylation of 4E-BP1 allows its dissociation from eIF4E and allows recruitment of eIF4G to allow translation to proceed[189]. What is particularly notable is that following FOXF1 re-expression, it is the total protein levels of 4EBP1, S6K and RPTORS that is altered. This hints at the possibility that FOXF1 may regulates expression at the promoter of these genes.

Figure 4.13: Representative western blot showing expression of S6K/4EBP1 and RPTOR following FOXF1 re-expression



4.3.20 In-silico prediction of FOXF1 transcription factor potential binding sites

FOXF1 codes for a transcription factor (TF) capable of altering the transcription of a number of genes. The binding sites of many TF have been determined by the ENCODE project consortium using ChIP-Seq[190], unfortunately, this did not include *FOXF1*. In order to identify putative targets relating to MTORC1 signalling, I performed an in-silico analysis of computationally predicted potential transcription binding sites (TFBS) of *FOXF1* using TRANSFAC and Multi-genome Analysis of Positions and Patterns of Elements of Relegation (MAPPER2)[191]. These are computational algorithms that use motif searches to predict likely target gene-transcription factor associations using known transcription factor binding motifs. This algorithm identified 1508 genes which contained potential TFBS of FOXF1. To gain a greater understanding of the function of these genes, I performed functional annotation using the MSigDB databases.

The hallmark MSigDB gene sets identified that a high number of genes with computationally generated FOXF1 potential TFBS overlapped with "Allograft Rejection". In support of this, total FOXF1 expression has been demonstrated to be associated with severe acute rejection pathology in a mouse lung transplantation model[192]. Other gene sets that demonstrated a high overlap with the computationally generated FOXF1 potential TFBS included "Apoptosis" ($p=1.15x10^{-7}$), "KRAS signalling" ($p=1.2x10^{-5}$), "MTORC signalling" ($p=2.3x10^{-5}$) and "Epithelial Mesenchymal Transition" ($p=6.0x10^{-5}$) (Table 4.9). It was particularly interesting to find that these gene sets were also identified from the RNA sequencing following *FOXF1* re-expresson.

As both RNA sequencing and in-silico prediction of FOXF1 TFBS pointed to the

Table 4.9: GSEA of in-silico FOXF1 potential TFBS genes from TRANSFAC. There was enrichment for apoptosis, EMT and MTORC signalling

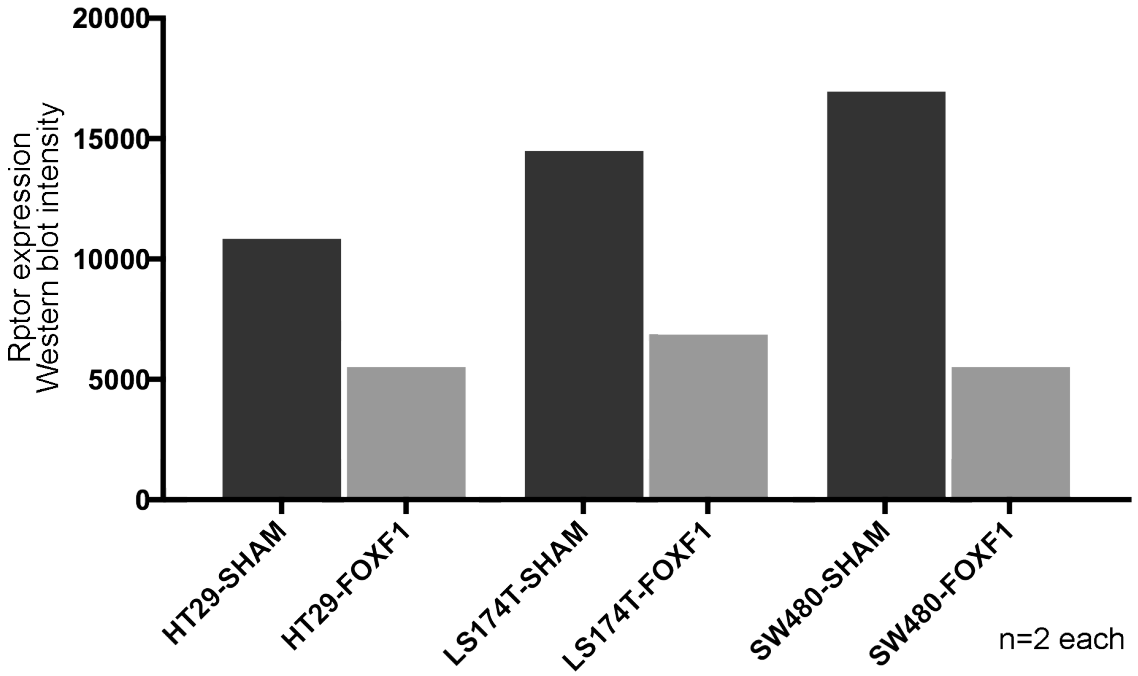
MSigDB	# Genes in Overlap (k)	p-value	FDR q-value
ALLOGRAFT_REJECTION	26	5.66 e-10	2.83 e-8
APOPTOSIS	22	4.59 e-9	1.15 e-7
ADIPOGENESIS	21	9.59 e-7	1.2 e-5
KRAS_SIGNALING	21	9.59 e-7	1.2 e-5
PROTEIN_SECRETION	14	1.28 e-6	1.28 e-5
PEROXISOME	14	3.41 e-6	2.29 e-5
HYPOXIA	20	3.67 e-6	2.29 e-5
MTORC1_SIGNALING	20	3.67 e-6	2.29 e-5
EMT	19	1.33 e-5	6.04 e-5
ESTROGEN_RESPONSE	19	1.33 e-5	6.04 e-5

importance of *FOXF1* with MTORC signalling, I re-analysed the *FOXF1* potential TFBS to determine which MTOR complex components were involved. I found that the *RPTOR* gene contained several TFBS for *FOXF1*. *RPTOR* normally forms a complex with MTOR and acts to activate the downstream effectors of MTORC1 signalling, S6K1[193] and 4E-BP1[194].

4.3.21 Chromatin Immunoprecipitation

To demonstrate that *FOXF1* exerted its effect through alterations in RPTOR protein expression, a western blot was performed in the *FOXF1* re-expressing cell lines. This demonstrated that RPTOR expression was reduced significantly in these cell lines by on average 56.4% (Fig 4.13a,Fig 4.14).

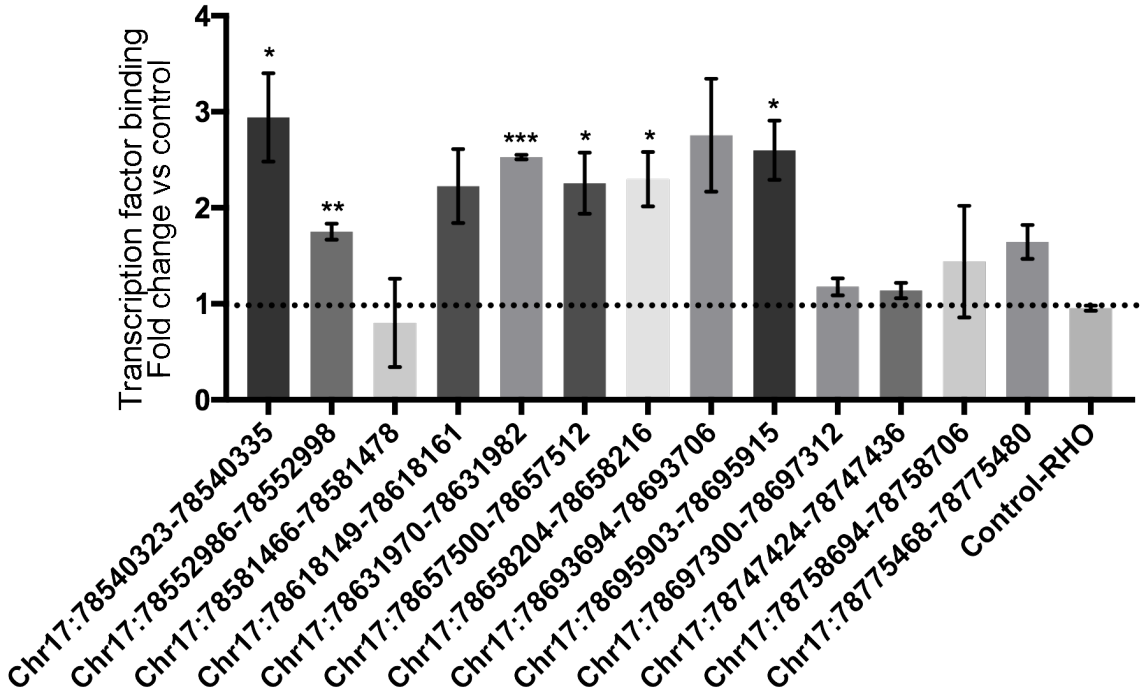
Figure 4.14: Summary bar chart demonstrating the effect of FOXF1 re-expression in three cell lines, HT29, LS174T and SW480. RPTER expression was determined by western blots and quantified by band intensity. Experiments were done in duplicates



Chromatin Immunoprecipitation (ChIP) is a technique that allows an analysis of binding between proteins and DNA in the cell. It determines if specific proteins are associated with certain genomic regions. Briefly, DNA and proteins are cross linked using formaldehyde, DNA-protein complexes are sheared through sonication, proteins of interest are selectively immunoprecipitated and DNA fragments are quantified. To demonstrate that FOXF1 binds to DNA within the *RPTOR* locus, I performed Chromatin Immunoprecipitation. The TRANSFAC/Mapper2 algorithm identified 21 potential TFBS of FOXF1 within the *RPTOR* locus. Two cell lines which expressed higher levels of FOXF1 were used for this assay (HCT116 and LS174T). FOXF1-chromatin-bound complexes were pulled down using immunoprecipitation (FOXF1 antibody, abcam, ab23194) with magnetic beads. The binding of FOXF1 to each of the predicted TFBS was quantified. It was possible to observe strongest binding of FOXF1 to RPTOR at intron 1 at a position 20.9 kb from the transcriptional start

site (Chr17:78540323-78540335) (Fig 4.15).

Figure 4.15: The relative binding of FOXF1 to different potential transcription factors was determined using Chromatin immunoprecipitation and two cell lines (HCT116 and LS174T). 13 potential TFBS within the RPTOR loci were selected based on the TRANSFAC/Mapper2 predictive algorithm and analysed. Error bars demonstrate standard deviation. Significantly increased binding was found at 6 TFBS

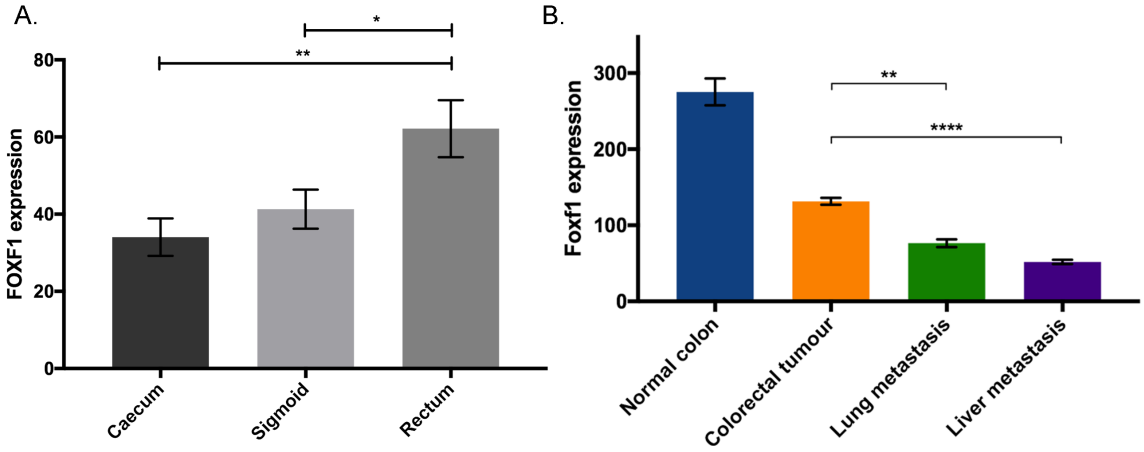


4.3.22 Analysis of effects of FOXF1 expression in human cohorts

I sought to gain a greater understanding of the relevance of FOXF1 to metastasis from human colon cancer cohorts.

I used two CRC datasets. I analysed unpublished data from our group, where RNA sequencing had been performed on 131 patients at different locations (Caecum, sigmoid, rectum). I also analysed the metastasis expression dataset from Sheffer et al[1].

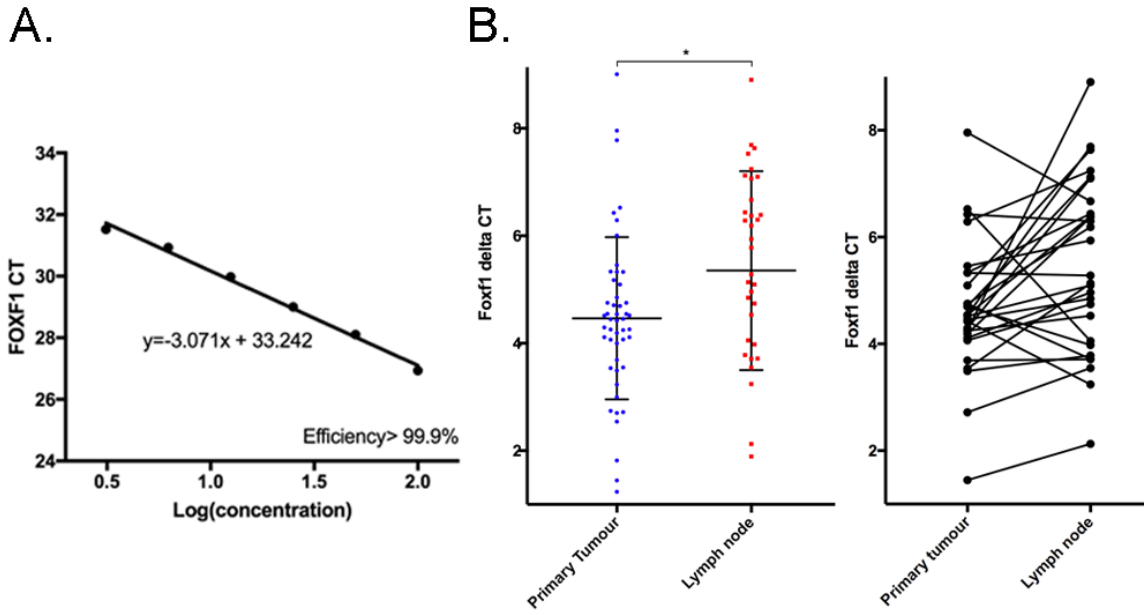
Figure 4.16: A. Expression of Foxf1 at different locations in the GI tract, demonstrating highest expression in the rectum. B. Expression of FOXF1 in tumours and metastasis, demonstrating decreased expression in tumours and metastasis compared to normal tissues



These datasets demonstrate that FOXF1 expression in normal tissue is highest in the sigmoid colon and rectum (Fig 4.16a). There is a stepwise statistical reduction in FOXF1 expression from normal human colon to colorectal tumours, with lung and liver metastasis showing very low levels of expression (CRC vs lung mets p=0.009, CRC vs liver mets p=<0.001, two-sided t test) (Fig 4.16b).

Secondly, an analysis of specimens from my OSLER sample set was performed. This was a cohort of sporadic stage III CRC patients from the Oxford University Trust. Baseline demographics can be found in Chapter 5 of this thesis. Paired LN-MTD and primary tissues were available for 41 patients. I determined the mRNA expression of FOXF1 in these tissues by rt-qPCR. The FOXF1 Taqman efficiency was >99.99% (Fig 4.17a). Expression of *FOXF1* in LNM compared to the primary tumour. Expression of *FOXF1* is lower in LNM compared to the primary tumour (two-sided t test, p=0.019) (Fig 4.17b).

Figure 4.17: A. FOXF1 Taqman efficiency determined 8 serial dilutions of cDNA and use of the Ct slope method denoting a probe with a linear standard curve and good accuracy. B. FOXF1 expression in lymph node metastasis vs primary tumour in paired specimens (y-axis is delta CT, the number of cycles to result in amplification of the signal, thus high delta CT means lower expression).



Together, these datasets suggest the association with decreased expression of *FOXF1* expression to both human CRC tumorigenesis and metastasis.

4.3.23 Identification of a therapeutic agent to increase FOXF1 expression

To identify if there were bioactive compounds or chemicals that had similar molecular effects to *FOXF1* re-expression, I used the *Connectivity Mapping, (CMAP)* algorithm[195]. This algorithm uses the results of expression changes derived from treating human cells with 1,309 compounds and by providing the expression changes observed from my *FOXF1* re-expression cell lines, it is possible to identify compounds that would exert similar effects.

Table 4.10: Connectivity mapping to identify drugs with a similar effect to FOXF1 re-expression

Rank	Drug	Enrichment	p
1	withaferin A	0.954	<0.0001
2	vorinostat	0.814	<0.0001
3	trichostatin A	0.806	<0.0001
4	LY-294002	0.521	<0.0001
5	sirolimus	0.5	<0.0001
6	valproic acid	0.234	<0.0001

This algorithm identified Withaferin A as the top hit. Withaferin A is a NOTCH, AKT and MTOR inhibitor[196]. The analysis also identified a number of HDAC inhibitors such as Trichostatin A, Vorinostat, Scriptaid and Valproic acid. Of particularly interest was Sirolimus and LY-294002, readily available drugs that are MTOR and PI3K inhibitors respectively (Table 4.10).

Sirolimus is a first generation MTOR inhibitor. It functions through the binding of the cytosolic protein FKBP12, which then directly interacts and inhibits MTORC1. It is widely used in clinical practice to prevent renal transplant rejection. A phase 1 clinical trial has been performed on 12 stage IV CRC patients who received a combined chemotherapeutic regime of Sirolimus, bevacizumab, 5-FU and irinotecan and it was reported that clinical benefit had been observed in the majority of cases[197].

LY-294002 is an inhibitor of AKT signalling, which is a pathway that is upstream of MTORC1. LY-294002 functions through the inhibition of PI3K, which is a signal transducer that activates AKT. There is some pre-clinical evidence that suppressor of AKT signalling does inhibit tumour progression[198][199] but there have been no clinical trials to date.

4.3.24 Mouse metastasis tail vein assay

I assessed the potential of these two drugs to inhibit metastasis using a mouse tail vein lung metastasis assay. In this assay, tumour cells are introduced directly into the circulation of mice by the lateral tail vein and an assessment can be performed of the ability of cells to lodge and grow in the lung. I selected this assay as it is more representative of early stage CRC patients, in whom the primary tissue would have been surgically resected, but are likely to have micrometastases in distant organs.

The NOD scid mouse was used as a model and human CRC cells (1M LS174T) were introduced into the tail vein. Mice then received Sirolimus, or LY294002, or vehicle (n=17).

In vitro culture experiments were not pursued because there is already a number of papers illustrating the growth inhibitory effect of the drugs in vitro, and secondly, because I did not feel that in-vitro metastasis models are able to recapitulate metastasis accurately.

Rapamycin has been demonstated to inhibit metastatic growth in several mouse cancer models. It inhibited liver outgrowth in a cholangiocarcinoma model[200], pulmonary outgrowth in a non-small cell lung cancer model[201], renal cancer model[202] and hepatocellular carcinoma model[203].

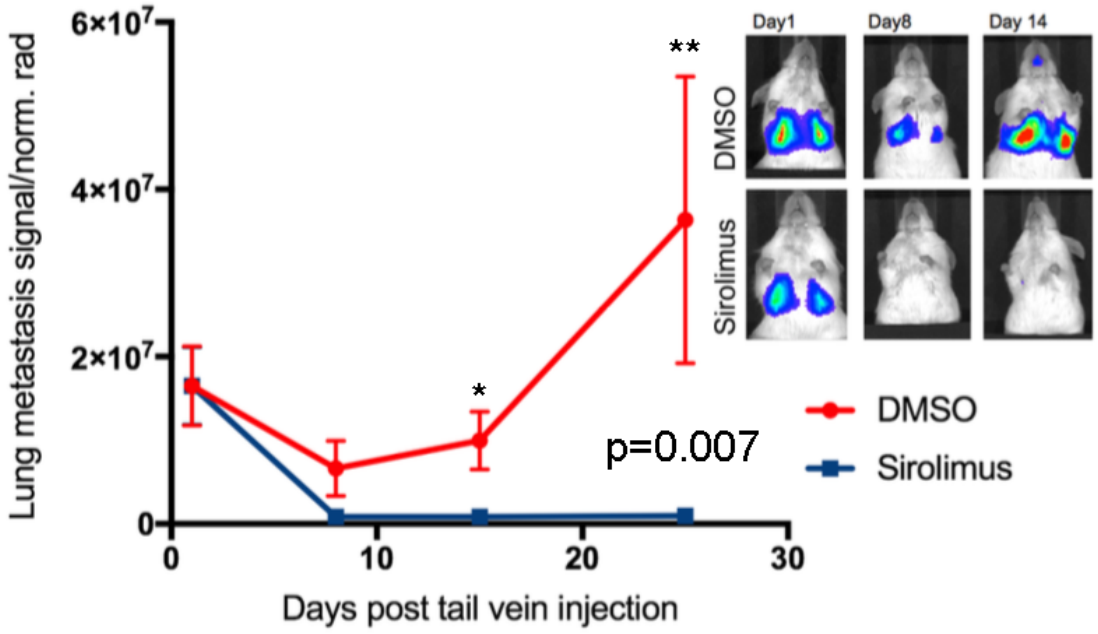
For the tail vein assay, doses of these drugs were used at the same level as used in other publications.

At day 1, the same amount of tumour cells were detected within the lung for all the experimental mice. However, over the course of three weeks, the amount of lung metastatic lesions diverged significantly. There was no difference between LY294002 and vehicle, however, Sirolimus treatment was extremely efficacious, significantly inhibiting the outgrowth of metastatic lesions. At day 25, the metastatic burden was 94% lower, confirming the importance of MTOR signalling in promoting the out-

growth of metastatic lesions (two-sided t test, p=0.007) (Fig 4.18).

The failure of the AKT inhibitors to inhibit metastatic outgrowth was surprising. There is a described a negative feedback loop, allowing activation of AKT following MTOR inhibition, however, the converse is not clear[204 205 206]. One possible explanation for the failure of LY294002 to inhibit metastasis could be an intrinsic resistance to the AKT inhibitor. LS174T has an activating mutation in PIK3CA and this could render it permanently active and resistant to the effects of LY-294002[207].

Figure 4.18: Effect of Sirolimus administration on the outgrowth of lung metastatic lesions demonstrating significantly decreased metastasis burden at day 14 and 21



4.4 Discussion

Cancer metastasis requires cells to gain the ability to proliferate, migrate, intravasate into the blood or lymphatic vessels, extravasate into the vessel of delivery and then grow in a distant organ. These resulting changes in cellular phenotype have been termed cellular *macro-evolution*, reflecting the wholescale phenotypic changes in-

volved. These phenotypes take a long time to develop and a number of genes have been implicated as important requirements for each step of this process, i.e. CXCR4, MMPs, and genes required for epithelial mesenchymal transition[208].

What I have attempted to do with this CRISPR/Cas9 genetic screen is to identify single gene alterations that could result in macro-evolution and result in distant metastatic lesions. A notable result from this screen was that many genes appeared to mediate this process. Within the lung metastatic lesions of mice, on average 2500 genes were identified. Many of these may have arisen through chance, but a large number were observed frequently across experimental mice and across the replicative cohorts. This suggests that unlike CRC tumorigenesis driver genes, there may be many genes involved in CRC metastasis.

Candidate genes identified from genetic screens are neither exhaustive nor exclusive and they require extensive follow up and validation. Of the genes I have identified, *FOXF1* is of particular interest. *FOXF1* is a member of the Forkhead Box proteins, evolutionary conserved transcription factors that are known to regulate the transcription of genes involved in cell cycle progression, differentiation and embryonic development. Their importance is that they are *Pioneer Factors*, transcription factors that are capable of directly decondensing chromatin, thus facilitating the binding of other sequence-specific transcription factors to target enhancers, repressors and promoters[209]. Like all Pioneer Factors, they are maintained under precise temporal and spatial gene control in normal cells. There is evidence that this control is lost in cancer, contributing to tumour initiation, progression and metastasis. For example, the *FOXO* proteins (*FOXO1, FOXO3A, FOXO4, FOXO6*) act as tumour suppressor, and deletions have been shown to induce lymphomas and haemangiomas in mice as well a promoting prostate tumorigenesis in humans[209 210 211]. Conversely, the *FOXM1*

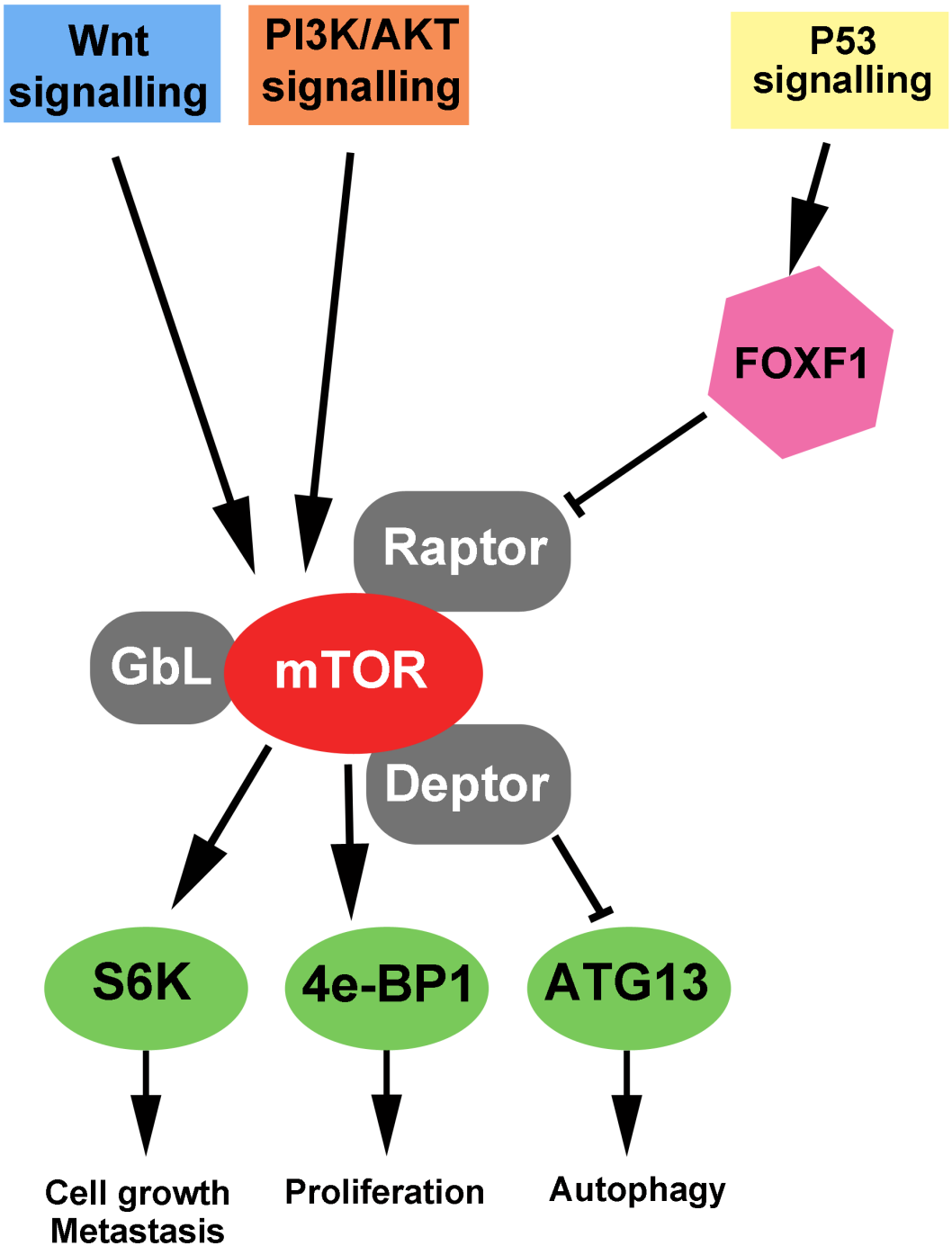
transcription factor acts as an oncogene in mouse models of lung and colon cancer through expression of cyclin proteins[212,213].

FOXF1 has been identified in breast and colon cancer studies as a tumour suppressor gene, enabling cell cycle G1 arrest through inhibition of CDK2-RB-E2F[169,173]. *FOXF1* suppresses cellular invasion of the HCT116 CRC cell line[170] and evidence from human CRC cohorts suggests there is expression of mislocalised/cytoplasmic FOXF1. Mislocalised FOXF1 is correlated with late stage CRC tumorigenesis[214]. Little more is known of the mechanisms or role of *FOXF1* in mediating human CRC metastasis.

My analysis of human cohorts, including my *OSLER sample set* has identified FOXF1 as a protein that is significantly and consistently down-regulated in tumours compared to normal tissues, and in CRC metastatic lesions (lymph node, lung and liver) compared to primary tumours. I have gone on to identify that FOXF1 exerts its CRC metastasis suppressor role in part through the inhibition of MTOR signalling. FOXF1 is able to bind to RPTOR at intron 1 and this is associated with decreased expression. RPTOR normally forms a multiprotein complex with MTORC1 and this allows phosphorylation of S6K1, a kinase that promotes cell motility and survival. RPTOR also recruits and enables inhibitory phosphorylation of 4E-BP1, allowing release of eIF4E and translation of oncogenic genes from the AKT and MAPK signalling pathway[193,194]. By reducing *RPTOR* expression, FOXF1 inhibits this process and acts to prevent AKT/MAPK and MTOR signalling (Figure 4.19). These pathways, particularly, MTOR are crucial to enable a cell to gain metastatic capabilities and would explain the high efficacy of the MTOR inhibitor, Sirolimus in my mouse metastasis model.

In summary, by performing a CRISPR/Cas9 whole-genome screen, it has enabled

Figure 4.19: Proposed mechanism of the effect of FOXF1 on MTORC1 signalling



me to identify a number of gene mediators that enable the macro-evolution of CRC clones. These mouse models and validation using human cohorts have identified the pioneer transcription factor FOXF1 as a likely inhibitor of CRC metastasis.

Chapter 5

Investigation of metastasis gene drivers in a human CRC cohort

5.1 Introduction

I felt my understanding of the molecular drivers of metastasis would not be adequate without the analysis of human CRC samples. One such approach involves the comparison of primary tumours that have given rise to metastatic deposits (stage III/IV CRC) to those that have not (stage I/II CRC) and this has formed the basis of a number of large CRC cohort studies[215,216,217,218]. However, these analyses do not necessarily enable a greater understanding of the process of metastasis, because they are based on the premise that stage III/IV tumours differ systematically from stage I/II tumours only in their ability to metastasise. There are numerous confounders including stage III/IV cancers being diagnosed later, are larger in size and often in a different anatomical location at diagnosis. In addition, this approach is fraught with difficulties, particularly because the ability to metastasise is not necessarily a phenotype that arises out of the whole primary tumour, but may be a phenotype that arises from a small proportion of cells or a small colony within the primary tumour.

A better approach to understanding the molecular events required to enable metastasis is through the comparison of primary tumour cells with cells from metastatic deposits. By understanding the changes that have been observed between these two samples, we can find salient features that define, enable and arise within metastatic cells. There are an increasing number of studies taking this approach, and they have begun to give us better insight into the molecular events required for metastasis.

Around 70% of CRC demonstrate a phenotype of chromosomal instability. This phenotype results in gene copy number changes as a result of chromosomal duplications and deletions, so called, somatic copy number alterations (SCNA). A number of papers have attempted to describe the SCNAs differences between primary tumours and their metastatic lesions. There is evidence to suggest that liver metastases have higher chromosomal aberrations compared to primary tumours[219] and that copy number changes in 6p are enriched in liver metastases[220]. However, contrasting results have also been observed with some groups demonstrating no allelic imbalances[221]. Lung metastases have also been analysed, and in 18 paired samples of primary and lung metastases, more chromosomal abnormalities were observed (12.6 vs. 5.7 p=0.024) with frequent gains in 20q/8q/13q/7q and more losses on 18q/8p/18p[222].

In terms of mutation analysis, there have been several studies looking at *KRAS, BRAF, PIK3CA* and *PTEN* to determine if they were differentially mutated at metastatic sites. The latest meta-analysis of 46 studies has noted that there is a high level of concordance of these mutations between primary tumour and metastatic sites, but there was some evidence that KRAS mutations are associated with lung metastases and that PIK3CA mutations are not detected in lymph node tumour metastatic deposits (LNM)[223]. In the largest cohort to date, targeted sequencing

of RAS, BRAF, PIK3CA and TP53 was performed in 323 primary tumours and 160 unmatched distant metastatic samples. This demonstrated that metastases often had fewer *BRAF* mutations (1.9% vs. 7.7%, p=0.01), but higher frequency of TP53 mutations (53.1% vs. 30.3%, p<0.001)[219].

Evidence of differences in chromosomal alterations or mutation profiles between metastatic lesions and primary tumours are limited by the small cohorts analysed and the limitations of the current assays to analyse differences. Analysis of protein expression has been much more fruitful and there are many examples of differentially expressed proteins. Protein expression within tumours and metastatic lesions can be probed using the technique of immunohistochemistry (IHC). Numerous studies have been performed using this technique and metastatic lesions differentially express a number of proteins.

There is evidence that EGFR is down-regulated in lymph node (LN) and liver metastases (69.4% vs. 80.6%)[224]. In a combined cohort of colon/bladder/breast and lung cancers and their respective metastatic lymph nodes, *PDCD4, DHFR and HOXD10* were differentially expressed[225]. *FASL* is a ligand that enables tumours to induce apoptosis of lymphocytes and has been found to be up-regulated in liver metastases in a cohort of 21 patients. There is also evidence that Thymidylate Synthase, which predicts clinical response to 5-FU, has lower expression in liver metastases compared to the primary tumour (48% vs. 70%), and this may be responsible for the relative chemoresistance of liver metastases[226]. However, the technique is not without its limitations. IHC is semi-quantitative, it only allows analysis of single protein epitopes/targets and is prone to high levels of variability between groups and individuals.

Another approach of identifying the biological changes that are present in metastatic

lesions involves the use of whole transcriptional profiling. The total mRNA composition of the tumour can be analysed simultaneously and mRNA gene counts may be obtained through a variety of techniques, including differential display RT-PCR, microarrays and RNA sequencing. This approach has been used in a number of studies to document differences in gene expression profiles (GEP) between tumours and their metastatic lesions.

The majority of CRC metastasis GEP studies have focused on liver metastases and have primarily used expression microarrays. In a cohort of 16 CRC-liver metastasis samples, 44 genes were identified as differentially expressed, with 36 being decreased and 8 being increased in liver metastasis tumours. Down-regulated genes included those of proteolysis (MMP1, MMP2) and signal transduction (WNT5A, HIF1). The most up-regulated was CXCL12[227]. In another liver metastasis cohort of 12 patients, 80 genes could be found to be differentially expressed. Among these genes, MMP1/MMP3/COL12A1/CXCL3 were up-regulated in the primary tumour with SOX15/NAT5/SPP1/SERPINA1 up-regulated in liver metastases[228]. A number of other smaller cohorts have been performed looking at liver metastases[229 230], however, the largest cohort consisted of 28 liver metastases and 48 primary tumours from unmatched patients. This identified 778 genes differentially expressed with 341 up-regulated and 437 down-regulated in liver metastases. Gene ontology analysis demonstrated that liver metastases had enrichment for genes associated with immune response and tissue remodelling and down-regulation of genes associated with proliferation and oxidative phosphorylation[231]. There is some evidence that different pathways to metastasis may exist and liver lesions that develop synchronously (present at the time of surgery) and metachronously (within a year post surgery) have very different GEP, despite arising from the same tumour[232]. It is notable that all the aforementioned GEP cohorts are small and more often than not, the studies

stress that liver metastases often resemble their primary tumours more than other metastasis samples, a feature I have termed "paired metastasis clusters" and hope to address in this chapter[233,234,235].

There have been no studies looking at GEP at the other common CRC metastasis sites, namely the lymph nodes (LN), lung or bone. Lymph node metastatic tumour deposits (LNM) are particularly interesting as the presence of lymph node invasion is a histological feature of clinical importance, defining colorectal cancer stage and an important tumour feature determining patient survival.

Patients with lymph node metastases (stage III CRC), have a five-year overall survival of 53%, whereas those without (stage II disease) have overall survival of 87%. The lymph nodes are one of the first sites of metastasis, the analysis of which I hoped has the potential to identify the earliest mutations or genes or gene signatures that are crucial for the process. This may inform the development of targeted therapy to prevent early metastatic cells leaving the tumour. There is also evidence that the lymphatic system is not simply a passive conduit for tumour spread, but plays a key role in stem cell survival, immune modulation and tumour recruitment. Tumour cells actively secrete lymphangiogenic factors such as VEGF-C, VEGF-D, VEGF-A and HGF, and branching of new lymphatic channels into the primary CRC occurs prior to the development of LNM[236]. In melanoma patients, tumour cells remain dormant for long periods of time within the protective microenvironment of the lymphatic endothelium after removal of the primary tumour[237]. Finally, it is within the LN that tumour cells are brought in close proximity with the innate and adaptive immune systems that play key roles in immune-control of tumours[238]. There have been few studies analysing CRC LNM and only one study had identified differential protein expression, that of increased cell cycle protein cyclin D1 and this was based on a can-

didate gene approach[239]. There has been no whole transcriptome profiling of CRC LNM to date.

Whole transcriptome profiling of LNM has been performed in the field of breast cancer research and five cohorts have been published, although the largest cohort consisting of only 15 patients. The results from these cohorts demonstrate that there are genes that are differentially expressed in the LNM, with down-regulation of cell-matrix interaction proteins with an increase in chemokine ligands[240 241 242]. However, the authors noted that there were often large similarities between the LNM and the primary tumours of breast cancer, and LNM are very heterogeneous[243]. Thus when clustering was performed of the gene expression profile of 15 breast cancer patients, the breast tumours clustered closely with their LNM in all but 1 patient ("paired metastasis clusters")[242].

There are limitations with the metastasis studies to date. The number of patients in these datasets are small with less than twenty patients and with multiple-testing being performed, it becomes increasingly challenging to distinguish and validate real molecular changes.

The hypothesis for this chapter proposes that the analysis of CRC LNM has the potential to yield further biological understanding of the process of metastasis. My aims are to provide a descriptive analysis of the early molecular events and gene signalling changes that are crucial for the development of metastasis. Secondly, I aim to provide pilot evidence of the feasibility of profiling immune cell infiltrates proximal to the LNM to detect the presence or absence of an anti-tumour response. Finally, I hope to identify potential biomarkers of poor prognosis within LNM and perform a comparison to matched primary tumours.

5.2 Chapter methods

5.2.0.1 RNA extraction

RNA extraction from cell lines RNA was extracted from cell lines using the *RNeasy Mini kit* (*Qiagen*) following manufacturer's protocols. For histological specimens, samples were deparaffinised and RNA was extracted using the *High Pure FFPET RNA Isolation Kit* (*Roche*).

5.2.0.2 RNA quantification

RNA concentration and quality were analysed using the UV-visible spectrophotometer (*Nanodrop 2000*, ThermoScientific). Where more accurate RNA concentration and quality were required, the *Tape station 4200, Agilent* was utilised as per manufacturer's instructions. RNA quality was assessed by analysing both RNA Integrity Score (RIN score) and the percentage of RNA fragments >200 bps (DV200).

5.2.0.3 RNA sequencing

3' mRNA sequencing was performed using the *Quantseq 3' mRNA-seq kit* (*Lexogen*). This approach was taken when there was low quality, quantity, FFPE specimens or a large number of samples. This kit uses total RNA as input with no poly(A) enrichment and starts with an oligodT primer to initiate first strand synthesis. RNA was removed and second strand synthesis was initiated by random primer binding. A magnetic bead-based purification step was used to allow size selection and for nucleic acid purification. Linker sequences for illumina platforms were attached during the library amplification step. Next generation sequencing was performed on the *Illumina Hiseq 4000 platform* with 75 paired end reads.

5.3 Statistical and Bioinformatics methods

5.3.1 Next generation sequencing and RNA sequencing

Next generation sequencing reads were obtained in fastq format. Quality control and trimming of sequencing reads was performed using QC reports from the automated analysis pipeline from the bioinformatics team at the Genomics Centre (*WTCHG*), using FASTQC. Sequencing reads were demultiplexed using the standard Illumina sequencing software. However, where demultiplexing was not possible (i.e. because multiplexing was performed on the forward as well as reverse primer), I performed this using Cutadapt version 1.9.1[87]. Sequence alignment of reads was performed by the bioinformatics team at the Genomics Centre (*WTCHG*), using HISAT2 version 2.0.4. For RNA sequencing, to perform feature counting, I used HTSeq version 0.6.1 with alignment to a GTF file with features obtained from Ensembl for human assembly version GRCh37. I performed differential comparison by using DESeq version 1.10.1[88]. The package performs normalisation of the counts based on size of sequencing and allows differential expression to be determined from count data through use of the negative binomial distribution and a shrinkage estimator for the distribution's variance.

5.3.2 Hierarchical clustering

Hierarchical clustering was performed using the R package, hclust. This calculation works by assigning each object (i.e. sample GEP) to its own cluster. The hierarchical clustering algorithm uses to *Complete Linkage Method* or *Ward's minimum variance method* to proceed iteratively to join two of the most similar clusters, continuing until there is just a single cluster. The distances between each cluster were calculated by using the Lance-Williams dissimilarity update formula. This formula calculates the differences between new clusters and existing points based on the differences that

were present prior to the formation of the new cluster.

5.3.3 Principal component analysis

Principal component analysis (PCA) was performed using the R package, prcomp. PCA is a method that allows the extraction of important variables from a high dimensional data set, in this case, sample gene expression profile (GEP). The aim is to represent this high dimensional dataset as a two dimensional PCA plot. This technique enables a calculation of the first principal component (1st PCA) and plot it against the second principal component (2nd PCA). The 1st PCA is the linear combination of the original genes that captures the maximum variance in the dataset and the 2nd PCA is the linear combination of the original predictors that captures the 2nd most common variance. As each component is identified in an unsupervised fashion, the technique is an unsupervised approach.

5.4 Results

The Oxford Stage III CRC-Life expectancy, adverse Events and Relapse (OSLER) sample set consists of patients with sporadic stage III CRC undergoing routine NHS clinical care who had given consent for biobanking of their tumour resection samples from 2013-15. This research was approved by the National Research Ethics Service-Committee South Central Oxford C (11/SC/0236). This sample set is representative of CRC patients in a standard clinical care setting (Table 1).

There were 69 patients in this sample set with a mean age of 69.2 years (sd 10.4) and 52.2% of patients were male. Patients had been followed up for a mean duration of 27 months (sd 6.7 months) (Table 5.1).

The mean age of 69 years in this cohort is representative of patients diagnosed with colorectal cancer. Colorectal cancers presents later than rectal cancers with a mean age of diagnosis is 68 years in men and 72 in women.

I had to consider if I should exclude patients with right sided disease as these cohort of patients have developed a tumour from a very different molecular pathway from those with left sided tumours. However, as the molecular pathways that promote CRC LN metastasis remain unknown and not known to be different between molecular subtypes, I elected to include these tumours, but ensured that paired specimens were available for all patients.

Ribonucleic acid (RNA) was extracted from the primary CRC, lymph node metastasis (LNM) from each patient (n=69), and an area of normal lymph node stroma adjacent to but separate from the LNM (n=50) (Fig 5.1). 3' RNA sequencing was performed on a total of 188 samples.

Figure 5.1: Figure demonstrating regions dissected from lymph nodes

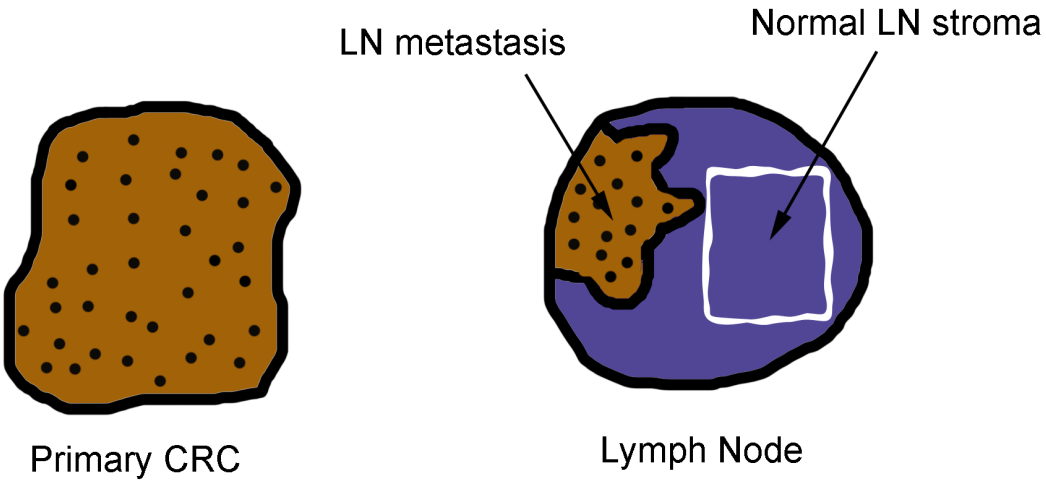


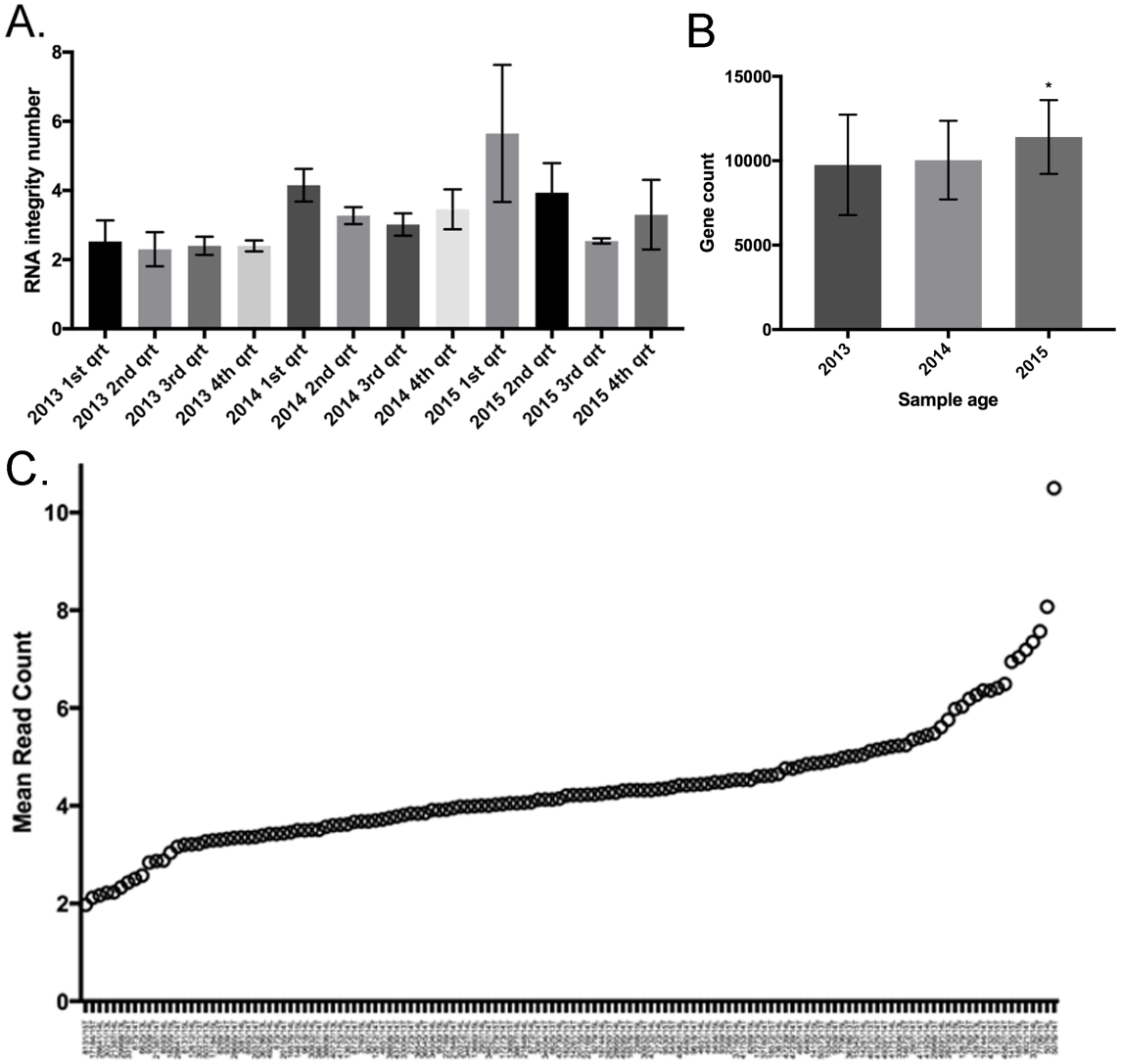
Table 5.1: Baseline demographics and outcomes of OSLER sample set

Oxford sample set (n=69)		
Demographics (%)		
	Age <50	1 (1.4)
	Age 50-59	12 (17.4)
	Age 60-69	25 (36.2)
	Age 70-79	18 (26.1)
	Age 80-89	13 (18.8)
Sex (%)	M	36 (52.2)
	F	33 (47.8)
Tumour Location (%)	Caecum	17 (24.6)
	Right sided	23 (33.3)
	Hepatic flexture	5 (7.2)
	Splenic flexure	1 (1.4)
	Sigmoid	15 (21.7)
	rectosigmoid	8 (11.6)
Nodes sampled (%)	<12	4 (5.8)
	13-19	18 (26.1)
	20-29	36 (52.2)
	30-49	6 (8.7)
	>40 nodes	5 (7.2)
Tumour differentiation (%)	Poor	18 (26.1)
	Moderate	49 (71.0)
	Well	2 (2.9)
Venous invasion (%)	V0	28 (40.6)
	V1	41 (59.4)
Lymphatic invasion (%)	L0	28 (40.6)
	L1	41 (59.4)
Tumour budding (%)	High	39 (56.5)
	Low	15 (21.7)
	Mucinous	9 (13.0)
	Not assessed	6 (8.7)
Microsatellite status (%)	Microsatellite unstable	11 (15.9)
	Microsatellite stable	45 (65.2)
	Results not available	13 (18.8)
Outcome (%)	Alive	59 (85.5)
	Deceased	10 (14.5)
	Relapsed	21 (30.4)
	Non-relapsed	48 (69.6)

5.4.1 Quality analyses of samples

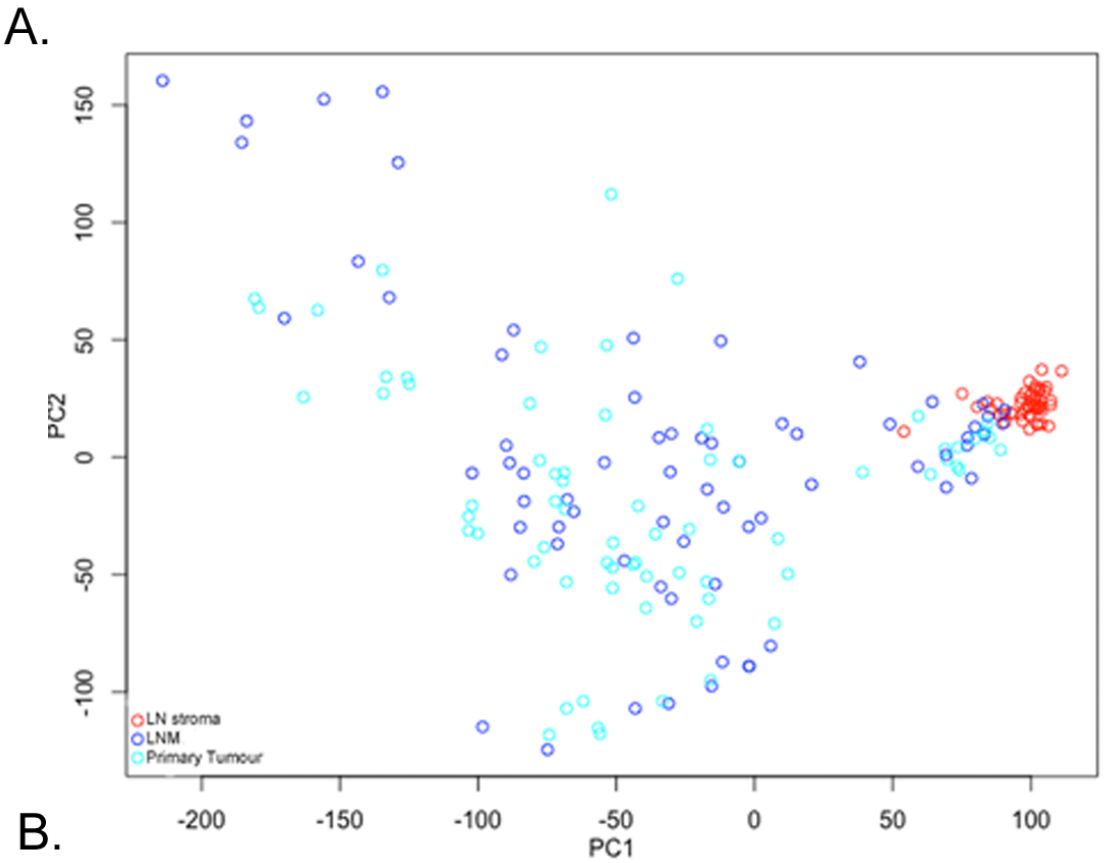
3' RNA sequencing is a technique for RNA sequencing from FFPE specimen and particularly robust as it only requires the analysis of a short stretch of the transcript at the 3' end of the mRNA. The quality of RNA extracted for the FFPE samples from this sample set was sufficient for 3' RNA sequencing with an average RNA integrity score (RIN) of 3.37 (sd=1.88) and a DV200 (proportion of reads less than 200bps) of 24.0%. Although the RIN score was lower for samples stored for more than 3 years (RIN 2.38 vs. 3.54, $p<0.0005$), this was still at a level acceptable for 3' RNA sequencing (Fig 5.2). Following library preparation, both the read count and total number of genes detected were good. The mean read depth was 4.34 M (sd 1.25M). The mean gene count was 10,320 (sd 2,447). There was no association between either the read count or gene count and location of sample. There was no association between the read depth and the age of specimens. However, the mean gene count of specimens from 2015 (11,406) was significantly higher than the mean gene count of specimens from 2013 and 2014 (9,752 and 10,040 respectively) (one way anova, $p=0.01$), reflecting RNA degradation and their lower RIN scores.

Figure 5.2: A. RNA integrity of FFPE specimen and sample age demonstrating a decline in RNA quality with older specimens. B. Lower genes mapped in specimens in 2013/2014 vs 2015. C. Mean read count for specimens



Principal components analysis (PCA) was performed on the 188 GEPs to determine the reliability of the dataset and determine if there were any global changes in GEP between the samples. There was good separation between the LN stroma (LNS), and the samples from the tumour and LNMs (Fig 5.3a). There was a degree of overlap between the tumour samples and the LNM, though PCA identified there were still GEP differences (22% variance from the PCA1, 6% for PCA2) (Fig 5.3b).

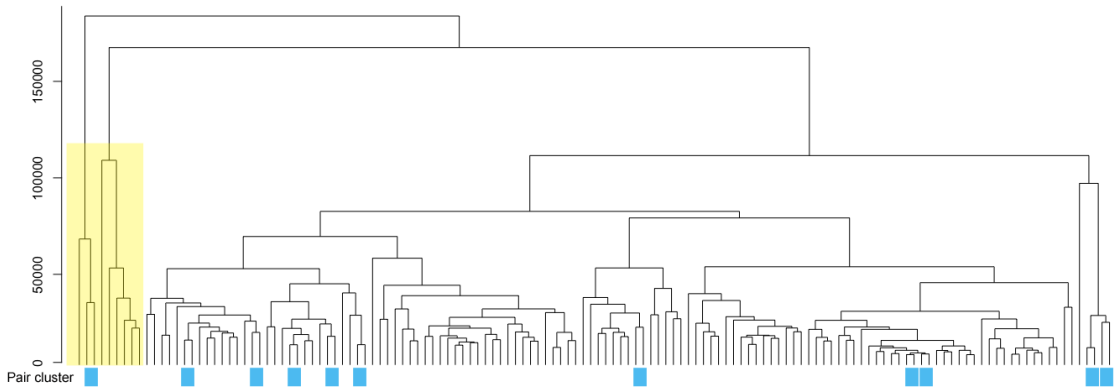
Figure 5.3: A. PCA of all specimens. Lymph node normal stroma (red) cluster very closely, with LNM and primary CRC demonstrating more heterogenous GEP (dark blue and light blue respectively). B. PCA- Primary tumour and LNM. LNM and primary CRC specimens cluster separately suggesting the presence of differentially expressed genes



5.4.2 Clustering of samples

Cluster analysis or clustering, is an approach by which samples that are more similar are grouped together than to those in other groups (clusters) and is a form of exploratory data mining. It is of particular use where data has multiple dimensions and a technique widely used with RNA sequencing. Non-supervised hierarchical clustering using the *Complete Linkage Method* was performed to compare the primary tumours to the LNM. This clustering analysis revealed that 9 patient samples did not cluster with the rest of the cohort. This was likely due to technical reasons as these specimens were performed on a separate pilot sequencing run. These patients were removed from all further analyses (Fig 5.4 yellow box).

Figure 5.4: Non-supervised hierarchical clustering. Each line on the bottom row of the dendrogram represents a sample and they form links based on similarities. Samples in yellow cluster separately from the rest of the cohort (separate pilot sequencing run). Some tumour-LNM paired samples clustered more closely with each other (blue), than any other specimen- "Paired metastasis clustering"



For a small number of patients (n=11), the GEP primary tumour clustered more closely to its paired LNM than any other specimen, i.e. demonstrating high GEP similarities. These were termed *Paired metastasis clustering* (PMC) (Fig 5.4). PMC is a feature that has been observed in a number of other metastasis datasets[233 234 235],

though their significance has not been remarked upon. The majority of my patients (n=46) had large differences in GEP between the primary tumour and LNM and these were termed *Distant metastasis clustering* (DMC). There was also a group, which showed intermediate differences, so termed *Intermediate metastasis clustering* (IMC, defined as a proximity matrix distance of <50,000).

To gain a greater understanding of the significance of GEP clustering distance, I analysed the histological characteristics of these tumours. There was a strong association with both tumour location (p=0.036) and microsatellite status (p=0.001), PMC/IMC were principally found in patients with right-sided tumours with microsatellite instability. Conversely DMC were principally patients with microsatellite stable tumours found on the left side of the colon (Table 5.2). This might arise through two main mechanisms. It is possible that MSI tumours are less capable of generating large gene expression changes than CIN tumours (i.e. a tumour biology effect). Alternatively, this phenomenon may arise from the effect of the microenvironment on the tumour, with the hypermutator phenotype of MSI tumours leading to immune-editing and immune activation or as a result of infiltration of tumour lymphocytes within the tumour.

There was a large difference in risk of disease recurrence or death with DMC, though this did not quite reach statistical significance (14.3% vs. 37.0%, p=0.06) (Table 5.2).

5.4.3 Survivors vs. Non-Survivors

Gene expression profiles of the primary tumours from patients that survived vs. those that died from their cancers were then compared. This supervised analysis identified

Table 5.2: Baseline disease demographics of patients which demonstrate large/small GEP differences between their LNM and primary CRC. DMC patients are associated with the presence of left sided tumours and a trend towards MSS.

	PMS n=11 (%)	IMC n=10(%)	PMC/IMC n=21 (%)	DMC n=46 (%)	PMC/IMC vs. DMS p value
Stage					
pT1/pT2	1 (9.1)	2 (20.0)	3 (14.3)	2 (4.3)	NS
pT3	4 (36.4)	3 (30.0)	7 (33.3)	24 (52.2)	NS
pT4	6 (54.5)	5 (50.0)	11 (52.4)	20 (43.5)	NS
Location					
Right	8 (72.7)	6 (60.0)	14 (66.7)	18 (39.1)	
Left	3 (27.3)	4 (40.0)	7 (33.3)	28 (60.9)	p=0.04
Status					
Rec/death	3 (27.3)	1 (10.0)	3 (14.3)	17 (37.0)	
No-event	8 (72.7)	9 (90.0)	18 (85.7)	29 (63.0)	p=0.06
MSS/MSI					
MSI+	4 (36.4)	4 (40.0)	8 (38.1)	3 (6.5)	p<0.01
MSS	6 (54.5)	6 (60.0)	12 (57.1)	31 (67.4)	NS
Unknown	1 (9.1)	0	1 (4.8)	12 (26.1)	

273 genes that were differentially expressed at p<0.005. Functional annotation using DAVID identified that these genes were involved in *Cell Adhesion*, *Cell Differentiation*, *Cell-Cell Adhesion*, *Regulation of Apoptotic Process* and *Positive Regulation of Migration*. GSEA identified that *KRAS signalling* was significantly enriched in non-survivors compared to survivors (NES 1.38, p<0.005, but with negative enrichment for *MYC signalling* (NES -1.63, p=0.002).

This analysis was repeated, including data on tumour recurrence as well as mortality. 146 genes were differentially expressed p<0.005. Functional annotation identified that these genes were primarily involved in *Protein Phosphorylation*, *Regulation of Apoptosis* and *G1/S Transition of Mitotic Cell Cycle*. GSEA analysis confirmed the findings from the survival analysis. Patients who died or had disease recurrence, had enrichment for *KRAS signalling* (NES 1.23, p=0.044) but negative enrichment for *MYC signalling* (NES -1.65, p=0.004) in their primary tumours.

In summary, the GEP identified that in the OSLER sample set, the degree of enrichment of KRAS signalling appeared to be an important feature of the primary tumour that is associated with both poorer overall survival and higher disease recurrence post surgical resection.

5.4.4 MSI vs. MSS

Another method of stratifying tumours is between those with MSI and with MSS. Around 15% of CRC patients have a tumour that is MSI+, that develops through defective function of the DNA mismatch repair system. These patients have better prognosis than those with MSS tumours. The efficacy of 5FU-based chemotherapy is unclear in MSI+ tumours. A recent systematic review demonstrated clear benefit for 5FU treatment on disease survival and overall survival for MSS patients (HR of 0.62 and 0.65 respectively), whilst for MSI+ patients, the effects were marginal and not significant (HR 0.84 and 0.66)[244]. This classification of MSS vs. MSI+ tumours is likely to become even more important in future, principally with the advent of immunotherapy. The latest phase 2 clinical trial using PD-1 receptor blockade demonstrated that a tumour's MSI status is the most important tumour feature that predicts response to Pembrolizumab. Response rates were 40% for MSI+ tumours and 0% for MSS tumours[245].

MSI/MSS status from the primary tissue was available for 55 patients in this sample set. 294 genes were differentially expressed at $p<0.005$. 217 genes were down-regulated in MSI tumours and functional annotation using DAVID identified that these genes were involved in *Regulation of Cell Proliferation*, *Regulation of Apoptosis* and *Cell Adhesion*. 77 genes were up-regulated and many were involved in *Regulation of Apoptosis* and *Immune Response* (Table 5.3). These GEPs would be in keeping

Table 5.3: Differentially expressed Immune Response genes between MSI+ vs MSS tumours (Reactome gene set)

Immune geneset p=5.68x10-5	log2FC	
JAK2	1.50	Janus kinase 2
HLA-DQA1	1.38	DQA1 major histocompatibility complex
MT2A	1.29	metallothionein 2A
OAS2	1.49	2' 5' oligoadenylate synthetase 2,
GBP4	1.81	guanylate binding protein 4
RIPK2	1.45	receptor interacting serine threonine kinase 2
CALM2	0.66	calmodulin 2 (phosphorylase kinase, delta)
PSME1	0.80	proteasome activator subunit 1
PSMC6	1.05	proteasome 26S subunit, ATPase, 6
CD74	1.09	CD74 molecule, major histocompatibility complex

with the known tumour biology features of MSI vs. MSS tumours. MSS tumours are more aggressive, and this is reflected by the enrichment of cell proliferation gene sets. MSI tumours are also characterised by profuse lymphocytic infiltration and this was reflected in the enrichment of genes involved in "*immune response*" [246].

GSEA was then performed to gain a more comprehensive understanding of signalling networks. MSI+ tumours showed enrichment for the hallmark gene set *Inflammatory Response*, and principally immune signalling pathways *IL6-JAK-STAT3*, *Interferon Gamma Response* and *Interferon Alpha Response*.

The *ESTIMATE* gene set [247] is a collection of genes that were derived from 10 tumour types from expression data from the TCGA. These are genes that are principally expressed in either stromal or immune cells, and the expression of which allows an inference of the fraction of stromal or immune cells within a tumour sample. Use of this gene set confirmed that MSI+ve tumours had more immune infiltrate (NES 1.69, $p<0.001$). Use of the MsigDB immunological gene sets, enabled much deeper analysis of the MSI+ tumours. MSI+ tumours had enrichment of gene sets of stimulated NK cells (GSE22886, NES 2.62), effector T cells (GSE24634, NES 2.39) and

stimulated macrophages (GSE7348, NES 1.65). These cells are principally involved in mediating an anti-tumour response. It was also possible to detect enrichment of a T regulatory cell gene set (GSE 14415, NES 2.33).

In summary, this analysis substantiates features observed in MSI+ tumours, those of enhanced immunogenicity with increased intraepithelial lymphocytes[248]. However, using this dataset, it is also possible to identify other cells of the innate immune response that are co-activated in the primary tissue of MSI tumours (macrophages, NK cells, T regs). The presence of a signature of Tregs is notable as they are a subset of CD4 T cells that maintain tolerance and suppress the anti-tumour response. Their presence is illustrative of the immune's system's dichotomous role in CRC tumorigenesis, particularly in MSI tumours.

5.4.5 Differential expression of primary tumour and LNM

An analysis was performed of genes differentially expressed between the LNM and the paired primary tumour to identify those which facilitated LN metastasis. To minimise effects of cellular admixture/contamination, genes that were principally associated with the LN stroma were removed. This was achieved through a comparison of the GEP of the 50 LN stroma samples (containing no tumour metastatic deposits) with the primary tumours to generate a cancer LN stroma gene set. This CRC LN stroma gene set consisted of 464 genes that were expressed log2FC>2 with a p.value of <1x10-5 (appendix A.4 on page 210). Gene ontology analysis identified these as being involved in *Immune response*, *Leukocyte/Lymphocyte Activation* or *Immune Defense Response*. Restricting the gene set further would make subsequent analysis less sensitive. Genes within this LN stroma signature were removed and not used in any subsequent analyses to identify differentially expressed genes within the LNM vs. primary tumours.

It is unlikely that the removal of these genes added a significant amount of bias to subsequent analyses. These genes make up a relatively small number of gene counts. This approach meant I could not exclude the possibility that expression of these 464 genes in tumours could be over-expressed to supra-normal levels in CRC LNMs (as I had excluded them from analyses). On balance, this "filtering" of a small number of genes felt like a necessary step to eliminate the effects of the immune admixture, a well known issue with bulk RNA-sequencing projects.

Taking a cut off of $p<0.0001$, 218 genes were differentially expressed between the LNM and primary tumour. 194 genes were expressed at higher levels in the LNM (Table 5.4), whereas 24 genes were higher in the primary tumour (Table 5.5). The most significantly enriched gene in the LNM was CXCR4 (log2FC 1.69, p.adj=8.4x10-11) (Table 3). Other genes that were up-regulated in the LNM, were genes involved in regulating apoptosis (*CDK5R1, CARD8, TP53INP1*), cell proliferation (*PTPRC, CXCL12, PTPN6*) and cell motion (*CXCR4, ARHGDIB, DCLK1, ETS1*). There was down-regulation of other genes involved in the regulation of cell proliferation (*TBX2, GREM1, NTN1, VIP*), metalloproteases (*MMP11, MMP2, MMP1*), cytoskeletal organisation (*DES, JPH2, CNN1*) and tumour suppressor genes (*FOXF1, WNT5A, RGMA*) (Table 3). *FOXF1* is the gene that had been previously identified as a mediator of metastasis from my GECKma screen in chapter 4.

The most significantly differentially expressed gene, CXCR4, is a chemokine receptor and has already been identified as a metastasis driver for non-small lung cancer[249], pancreatic ductal adenocarcinoma[250], breast cancer[251] and colon cancer[252]. In colon cancer, high expression of CXCR4 in liver metastases is associated with poor survival and risk of recurrence[253]. Additionally, there is evidence from animal models that CXCR4 expression in colorectal cell lines increases cellular migration and metastasis

Table 5.4: Top 20 genes up-regulated in LNM

Gene	Log2FC	pvalue	Function
CXCR4	1.69	1.38E-14	Chemotaxis, Cell migration, proliferation, MAPK
IGHM	2.02	1.30E-14	Immune Response
C7	2.38	9.57E-13	Cell death
SPN	1.77	3.23E-12	Chemotaxis, Negative regulation of cell adhesion
PTPRC	1.71	5.22E-12	MAPKKK cascade, Cell proliferation,
IL7R	1.79	4.06E-11	Positive regulation of T cell differentiation
RASSF5	1.53	2.81E-10	Tumour suppressor, Ras association domain family
TTN	1.67	4.11E-08	Actin cytoskeleton organization
CD28	2.09	9.32E-08	Positive regulation of T cell activation
TRBC2	1.57	9.33E-08	T cell receptor
TTC39C	1.60	1.71E-07	Cell cycle regulator
PTGDS	1.94	2.74E-07	Prostanoid metabolic process
SELPLG	1.67	3.37E-07	Cell migration
CD40LG	1.93	1.62E-06	Cell proliferation
TBC1D10C	1.59	2.84E-06	Regulation of Ras protein signal transduction
SUSD3	1.65	6.67E-06	Cell migration, Cell Proliferation
SFRP1	1.62	7.56E-06	Negative regulator of apoptosis
NGFR	1.61	1.34E-05	Negative regulation of apoptosis
ALPL	1.75	2.96E-05	Response to antibiotic
GPSM3	1.65	3.65E-05	G-Protein Signaling modulator

rates[254].

To understand signalling pathways that were differentially expressed in the LNM, gene set enrichment analysis was performed. In keeping with previously described drivers of metastasis, I saw an increase in the *JAK-STAT* (NES 2.07, p<0.001), *KRAS signalling* (NES 1.66, p<0.001), *TNF-NFKB signalling* (NES 2.06, p<0.001) and *PI3K-AKT-MTOR signalling* (NES 1.36, p=0.046) pathways. Interestingly, the LNM had a strong negative enrichment for *Oxidative phosphorylation* (NES 1.38, p=0.007). This corroborates findings that had been also observed in a CRC liver metastases GEP paper[231]. Oxidative phosphorylation is a metabolic process primarily used in normal cells to generate ATP for energy and switching from oxidative phosphorylation to a glycolytic metabolic pathway has been proposed as a important

Table 5.5: Top 20 genes down-regulated in LNM

Gene	Log2FC	pvalue	Function
VIP	-2.53	1.20E-09	Regulation of cell proliferation
CNN1	-1.51	8.97E-09	Actin cytoskeleton organization
MMP1	-2.19	9.26E-09	Collagen Catabolic process
DES	-1.94	1.50E-08	Cytoskeleton organization
MYH11	-1.47	7.85E-08	Cellular component morphogenesis
ACTG2	-1.70	8.34E-08	Cellular motility
MAB21L2	-1.92	1.04E-07	Regulation of cell proliferation
GREM1	-1.41	1.57E-07	Regulation of response to external stimulus
HSPB7	-2.06	2.02E-07	Response to abiotic stimulus
NTN1	-1.75	1.67E-06	Positive regulation of cell differentiation
LMOD1	-1.41	1.76E-06	Cytoskeleton organization
JPH2	-1.69	4.66E-06	Junctional Complex formation
RBPMS2	-1.70	1.37E-05	RNA binding protein
RGMA	-1.42	1.83E-05	Tissue morphogenesis
FOXF1	-1.49	1.94E-05	Positive regulation of gene expression
TBX2	-1.56	4.24E-05	Regulation of cell proliferation
GREM2	-1.62	1.22E-04	Cell surface receptor signal transduction
VASN	-1.53	1.41E-04	TGFb cascade
ABCG5	-1.60	1.60E-04	Homeostatic process

mechanism required for metastasis[255], a process also known as the *Warburg effect* (Figure 5.5, Table 5.6).

There are different molecular subtypes of CRC with the most recent meta-analyses identifying 5 main subsets[20]. Each of the CRC subtypes have their own driver genes, driven by different oncogenic signalling networks and their GEP are distinct from one another. A limitation of differential expression analysis is that it does not analyse gene set changes for each patient, but takes an average tumour vs. average LNM approach. This would lead to an underestimation of gene set enrichments. To address this problem, I analysed relative gene expression (rGE) alterations for each gene set for each individual patient.

rGE alterations for each patient were determined by dividing the normalised gene

Figure 5.5: A. Differentially expressed genes between primary tumour and LNM. B. Gene set enrichment between primary tumour and LNM. KRAS signalling and AKT-MTORC signalling genes were highly enriched in LNM compared to primary CRC

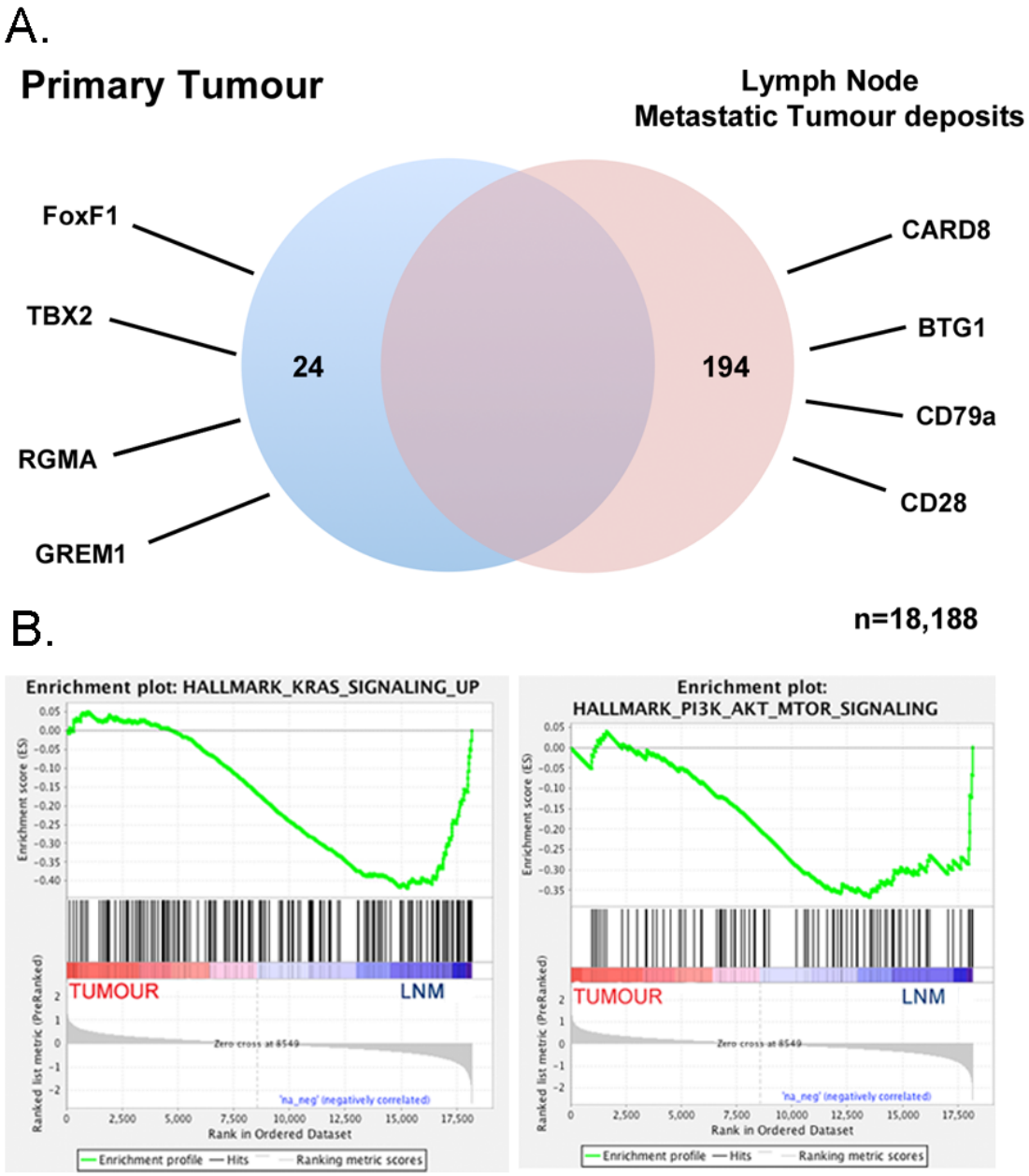


Table 5.6: GSEA analysis of genes differentially expressed between LNM and primary CRC. Full MSigDB gene set results in appendix A.1 on page 205

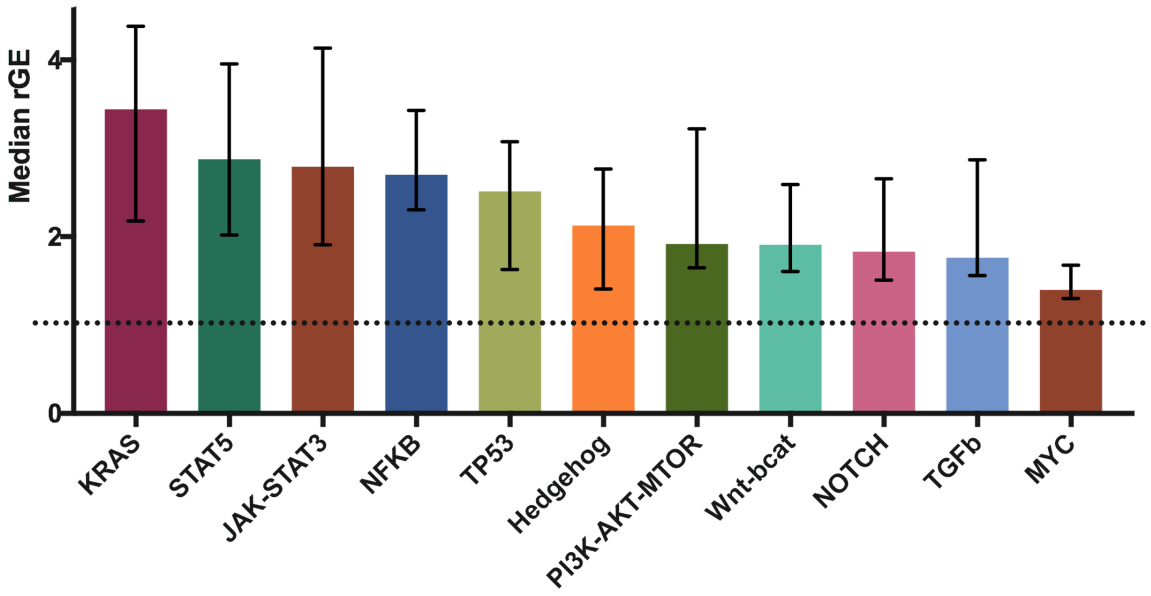
MSigDB	NES	p-value	leading edge genes
KRAS Signalling	1.66	<0.01	IL7R, CXCR4, LCP1
Inflammatory response	2.00	<0.01	GP1BA, IL7R, CCR7
TNF signalling	2.06	<0.01	IL7R, CXCL10, SERPINE1
Jak-stat signalling	2.06	<0.01	CXCL13, LTB, PIK3R5
IL2-stat signalling	2.11	<0.01	LTB, SELL, TNFRSF8
PI3K-Akt-Mtor signalling	1.36	0.04	PRKCB, CXCR4, CAMK4
p53 pathway	0.80	0.89	n/a
Hypoxia	1.20	0.13	CXCR4, CP, ETS1
Oxidative phosphorylation	-1.38	<0.01	NDUFB5, SLC25A12, SLC25A4
Epithelial Mesenchymal transition	-1.33	0.02	MMP1, GREM1, MMP3
Myc targets	-0.46	1.00	PPRC1, PRMT3, TBRG4
Notch Signalling	-1.04	0.38	WNT5A, FZD1, LFNG
Wnt-β catenin	-0.93	0.59	FZD1, NKD1, AXIN1

count data from the LNM by the normalised gene count data from the primary tumour. Taking this approach, it was possible to derive how much enrichment there was for each of the oncogenic hallmarks found in MsigDB. It was possible identify that within the LNM there was up-regulation of more oncogenic signalling pathways, not limited to those previously identified. The highest rGE change was noted for KRAS signalling (3.44x). However, I could identify an up-regulation of other known oncogenic signalling networks including PI3K-AKT-MTOR (1.92x), NOTCH (1.83x) and Hedgehog pathways (2.13x) in the LNM(Figure 5.6).

Links between these pathways and metastasis have been identified primarily through in-vitro studies, but this dataset confirms their importance to human CRC metastasis. Hedgehog signalling has been demonstrated to increase tumour invasiveness[256], Notch through modulating Epithelial Mesenchymal Transition[257] and mTOR through modulation of cell motility by the interaction with S6K and the actin cytoskeleton[62]. Interestingly, whilst there was a significant increase in the median rGE for MYC signalling in LNM, this change was much lower than for other oncogenic signalling

pathways. MYC has pro-metastatic effects in-vitro[58] and it is likely that in stage III CRC patients, this pathway has already been highly activated in the primary tumour to drive cellular proliferation, and further changes in expression are relatively less important in driving LNM.

Figure 5.6: rGE changes of oncogenic signalling networks in LNM



5.4.6 Unsupervised hierarchical clustering

I then looked at the GEP derived from the LNM in more detail. In the previous results section, it was demonstrated that the LNM had a large number of genes that were differentially expressed from the primary tumour. I sought to identify if there were shared biological features or subsets within my LNM samples.

I elected to use clustering analysis to group LNM specimens that were more similar. Unsupervised hierarchical clustering analysis using *Ward's minimum variance method* was performed on LNM GEP. Using this approach, it is possible to identify two clusters that I have termed LN Metastasis Subset 1 (LNMS1) and LN Metastasis

Table 5.7: Associated tumour and patient features with LNMS clustering

	LNMS1, n=15 (%)	LNMS2, n=44 (%)	
Median age at diagnosis	76	69.5	
STAGE			
pT2	0 (0.0)	2 (4.5)	NS
pT3	5 (33.3)	23 (52.3)	NS
pT4	10 (66.7)	19 (43.2)	NS
LOCATION			
Right	11 (73.3)	30 (68.2)	NS
Left	4 (26.7)	14 (31.8)	NS
Size of LNM/mm (sd)	6.6 (3.4)	5.5 (4.0)	
MISMATCH REPAIR STATUS			
MSS	12 (80.0)	25 (56.8)	
MSI+	0 (0.0)	9 (20.5)	p=0.08
Unknown	3	10	

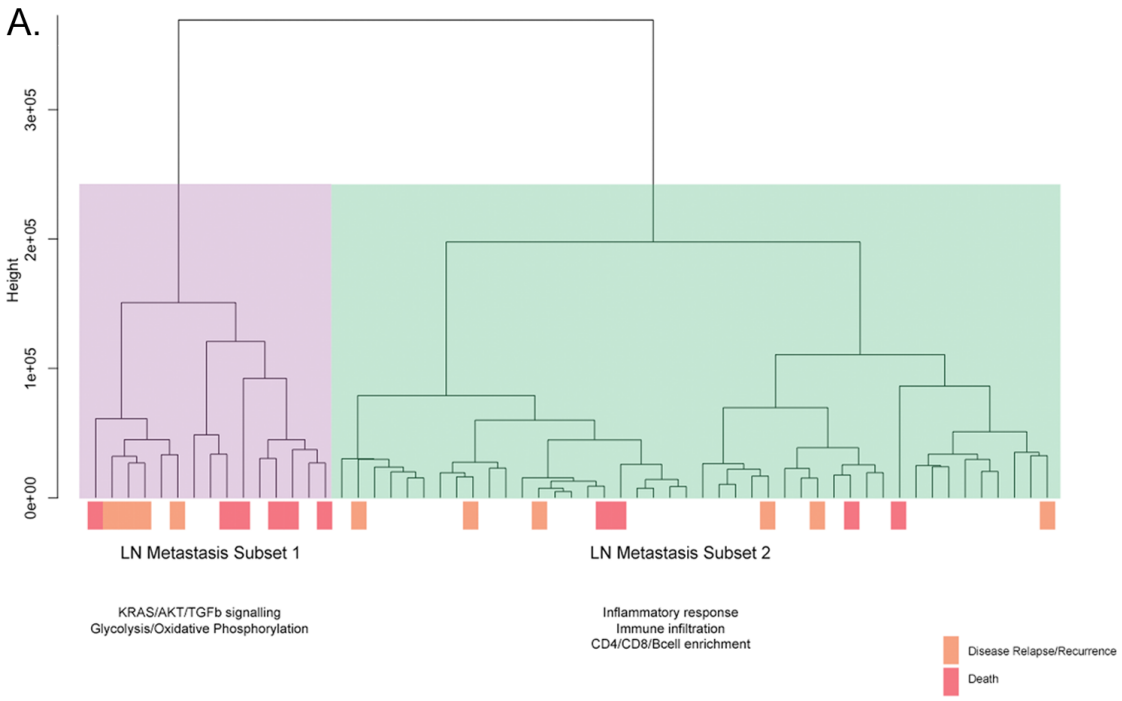
Subset 2 (LNMS2) (Figure 5.7a). When outcome data was overlaid on these clusters, it was possible to observe that patients with LNMS1 had high mortality (red box) and high disease recurrence (orange box). In contrast, patients with LNMS2 had much better prognosis (Figure 5.7b) (Disease free survival 1111 days vs. 614 days, Hazard Ratio 4.02, log rank p<0.001). An analysis was performed of LNMS1 and LNMS2 and primary tumour histological features. There were trends towards higher age at diagnosis and a more aggressive tumour for patients with LNMS1. There was no association with size of the LNMs or with MSI status. (Table 5.7).

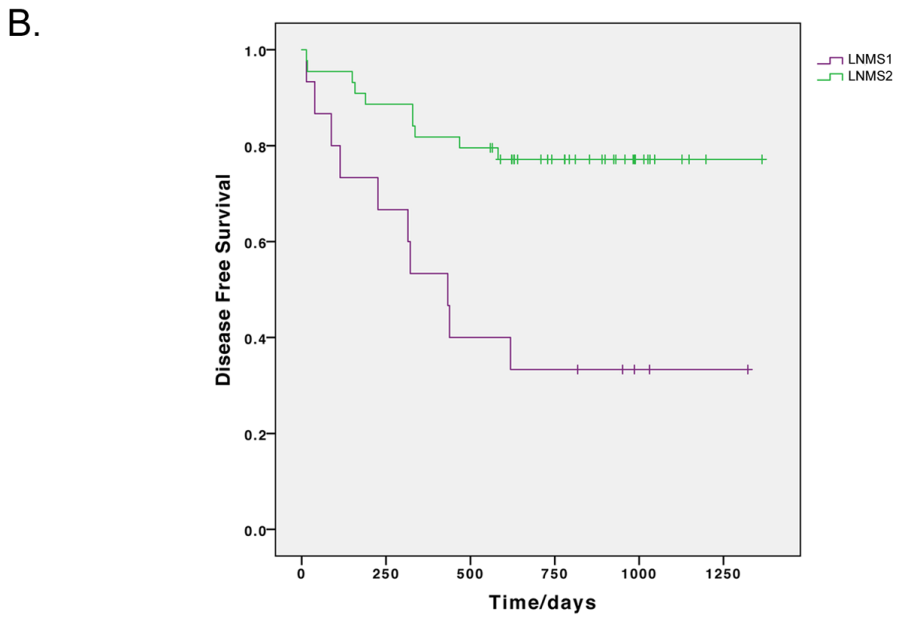
5.4.7 Differences in GEP between LNMS1 vs LNMS2

Differential expression and gene set enrichment were performed to determine key genes or gene sets that defined LNMS1 and LNMS2 signatures. At a cut off of p<0.0001, 250 genes were differentially expressed between LNMS1 and LNMS2. The most significant of which was XIST (Log2FC -3.60, p=4.91x10-12), a tumour repressor that when silenced drives expression of oncogenic X-linked genes[258 259].

However, a possible reason for the identification of XIST, was that there were

Figure 5.7: A. Clustering of LNM, identifying LNMS1 and LNMS2. B. Survival of patients with LNMS1 vs. LNMS2





relatively more male patients within the LNMS1 subtype. Of the 15 patients in LNMS1, 3 were female whereas of the 44 patients in LNMS2, 25 were female. XIST is a gene on the X-chromsome which performs X- inactivation and males do not express the gene.

GSEA identified that LNMS1 patients had enrichment of several oncogenic signalling networks including, *TGFb signalling* (NES 1.49, p=0.016), *KRAS signalling* (1.36, p=0.036), *AKT signalling* (NES 1.20, p<0.001). There was evidence of increased metabolic activity signature (*Oxidative Phosphorylation*, NES 1.49, p<0.001, *Glycolysis*, NES 1.29, p<0.001). Of notable significance, was the decrease of the *Inflammation Signature* (NES -1.33, p=0.026).

The *ESTIMATE* gene set was utilised to infer degree of immune infiltration[247] and this demonstrated markedly less immune infiltration in LNMS1 LNM (NES -2.08, p<0.001). In order to understand how the proportions of immune cells were perturbed, I used the MSigDB C7 immunological signatures. MSigDB C7 is a collection of signatures that are manually curated by the MSigDB consortium that represent cell states and perturbations within the immune systems. These signatures used published human and mouse immunology studies. LNMS2 LNM demonstrated significant enrichment for gene sets demonstrating activation of the adaptive immune system. There was an increase in CD4 T-cell (GSE22886, NES 1.64, p<0.001), CD8 T-cell (GSE22886, NES 1.86, p<0.001) and B cell signatures (GSE29618, 1.67, p<0.001).

In summary, these findings would be consistent with LNMS2 classified patients having a high degree of CD4/CD8 B-cell immune infiltration in their LNM and this in turn was associated good prognosis. Conversely LNMS1 classified patients who had worse prognosis and was associated with activation of oncogenic signalling networks including KRAS signalling. There was a trend for the LNMS classification to be as-

Table 5.8: Founder Gene sets for GSEA

	Positive enrichment	Negative Enrichment
CMS1 vs. others	Cytotoxic T cells/Immune infiltration	Cancer stem cells
CMS2 vs. others	Crypt Base/Myc targets	Stromal Infiltration/Immune
CMS3 vs. others	Starch Sucrose Metabolism	Integrin b3 activation
CMS4 vs. others	Complement Activation	Wnt Activation/Cell cycle

sociated with MSI+ status, though this did not reach statistical significance (p=0.08) and suggests that the LNMS classification is likely to be independent of MSI status.

5.4.8 Association with LNMS with CMS

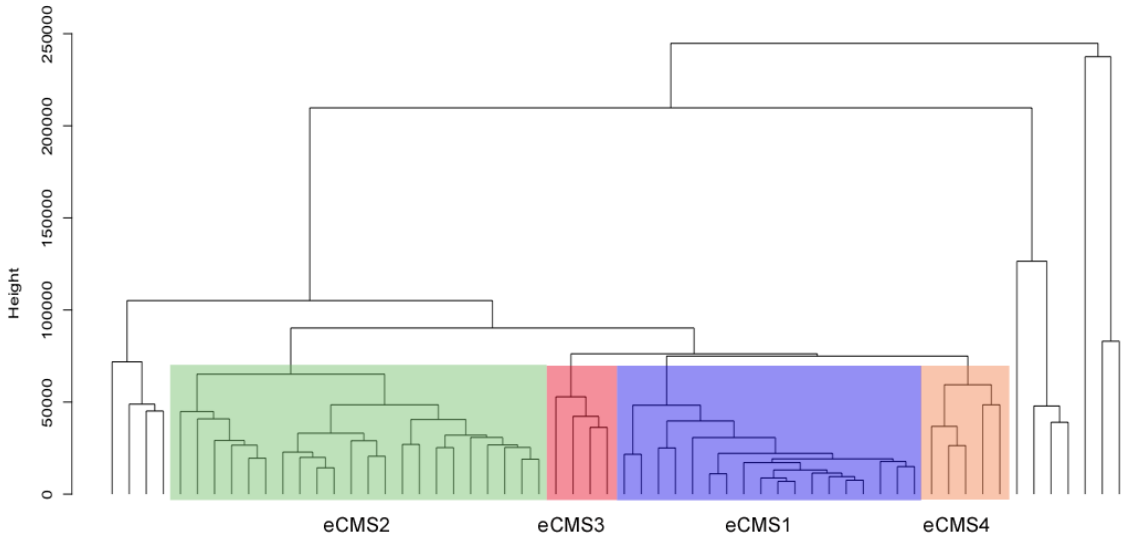
Consensus molecular subtypes (CMS) of CRC were developed through use of GEP of 4151 patients and used expression data in combination with clinical information. Using this technique, it is generally accepted that there are 5 subtypes of CRC, CMS1-5[20]. There was no association of LNMS and tumour histological characteristics, however, I sought to understand if there was an association between LNMS subsets and CMS.

There are no accepted techniques for classifying tumours into the CMS subtypes without access to the original dataset. In order to extrapolate which of my patients belonged to a particular CMS group, I relied on the fact that different gene signatures preferentially group to each CMS subtypes (Table 5.6). For example, CMS1 subtypes of CRC is characterised by the presence of an immune gene set (immune infiltration), with negative enrichment for a cancer stem cell gene set. I performed hierarchical clustering of my primary tumour samples and then on each of the resulting clusters, GSEA for the founder gene sets that are associated with particular CMS groups. Using this approach, it was possible to assign my clusters into extrapolated CMS groups (eCMS) (Table 5.8, 5.9, Figure 5.8).

Table 5.9: Signatures expected to be enriched with eCMS and observed enrichment

	Gene Set	Expected	NES
eCMS1	Immune	+	2.27
	Cancer_stem_cells	-	-1.64
eCMS2	Crypt_Base Signature	+	1.48
	MYC targets	+	2.72
	Immune	-	-2.26
	Stroma	-	-1.55
eCMS3	Starch_Sucrose_metabolism	+	0.79
	Integrin_beta3_Activation	-	-1.14
eCMS4	Stromal	+	1.09
	Wnt_activation	-	-0.73
	Cell_Cycle	-	-1.53

Figure 5.8: Development of eCMS clusters from primary samples from OSLER sample set



The OSLER sample set demonstrated marginally different proportions of the CMS groups with more CMS1 (MSI immune tumours) and with less CMS4 (Mesenchymal). A possible reason was that the OSLER sample set consisted of relatively elderly patients with a high proportion of right sided tumours. eCMS classifications allowed me to perform a comparison between LNMS and eCMS categories. (Table 5.10). Sample sizes are small, but no clear associations can be identified, possibly suggesting

Table 5.10: Associations between eCMS and LNMS1 subsets

	LNMS1	LNMS2
eCMS1	3	15
eCMS2	5	16
eCMS3	2	2
eCMS4	0	5
eCMS-unclassified	5	6

metastasis subtypes arise irrespective of CMS categories.

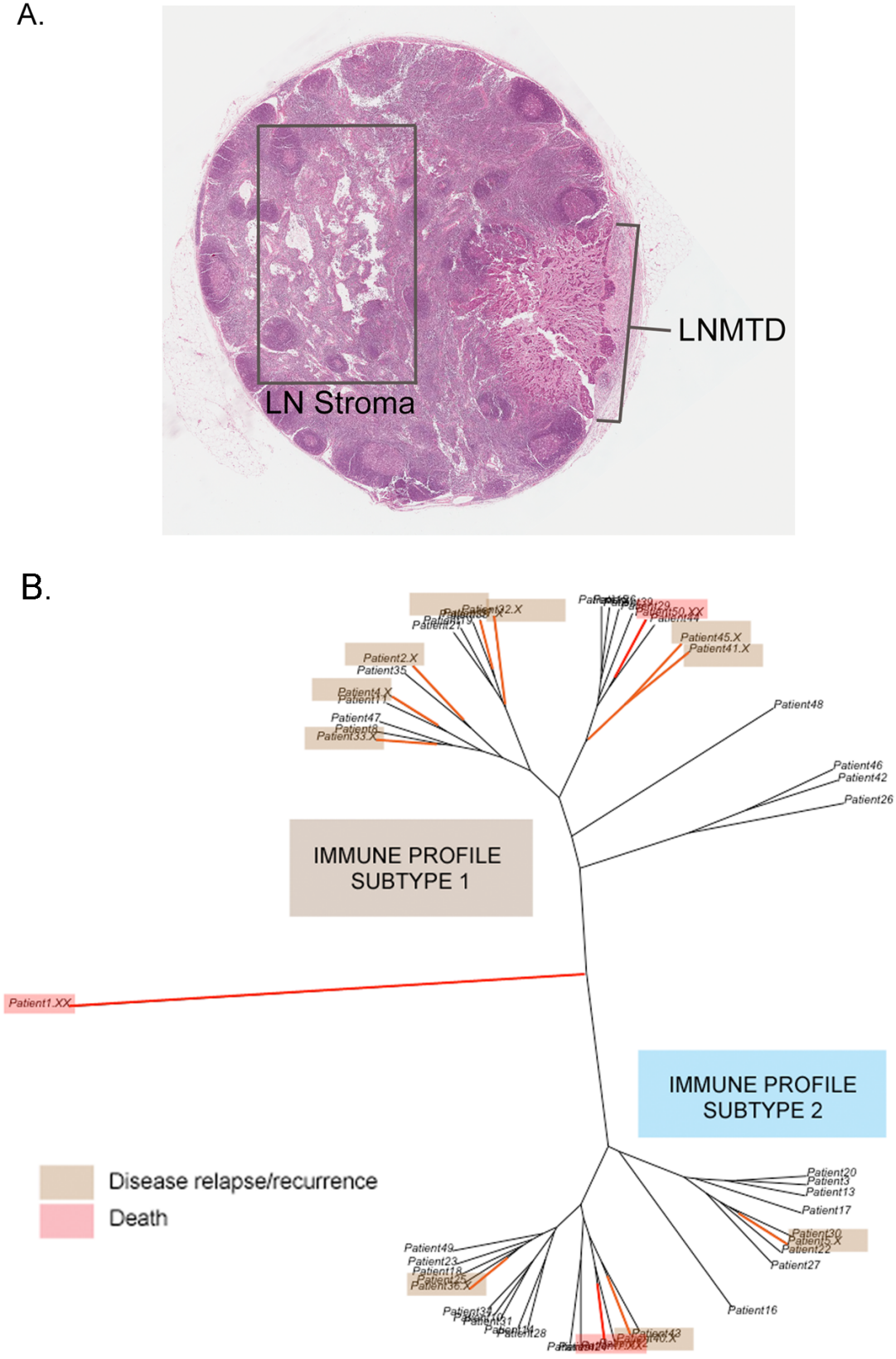
5.4.9 Lymphoid Ecological Niche Immune profiling

The pro-inflammatory and pro-immune response signature within the good prognosis cluster of patients was striking, but not entirely unexpected. Macroscopic analysis of the histological sections demonstrates an abundance of immune cells infiltrating and densely clustered within the LNM tumour colonies. To more accurately determine patient's immune profiles without having an admixture of tumour cells, I used the GEP from regions dissected from the LN stroma directly adjacent to the LNM, but with no evidence of tumour metastasis (Figure 5.9a). I hypothesised that by understanding GEP changes within this lymphoid ecological niche (LEN) proximal to the LNM, it might be able to obtain insights into the presence or absence of an anti-tumour response.

Clear dissection margins were obtained from 50 patients and GEP were obtained by 3' RNA sequencing. To infer if there were different biological groups, unsupervised clustering was performed using the *Complete Linkage* Agglomerative method. One patient did not cluster with the others. However, two immunological profile signatures could be observed and these were termed IPS1 and IPS2 (Figure 5.9b, 5.10a).

There was a marked difference in prognosis, with an IPS1 signature associated

Figure 5.9: A. LN regions. B. Phylogenetic tree illustrating IPS clustering.



with a poorer prognosis than those with IPS2, however, this did not reach statistical significance (Disease recurrence/death of 33.3% vs. 16.0%, Chi-squared test p=0.19)(Figure5.10b). To understand how these signatures differ, an analysis was performed between patients with the two different types of immune signatures. There were 118 genes differentially expressed at p<0.005.

Gene set enrichment revealed that IPS2 patients had increased expression of signatures associated with stimulated CD8 cells (GSE17301, NES 1.34, p=0.010), Macrophages (GSE8621, NES 1.45, p<0.001), NK cells (GSE22886, NES 1.44, p=0.003), Th17 cells (GSE 14308, NES 1.29, p=0.020), CD4 cells (GSE2770, NES 1.41, p=0.006) and dendritic cells (GSE 14000, NES 1.25, p=0.035). These are cells of the innate and adaptive immune system that promote an anti-tumour response[260]. IPS1 patients conversely, had enrichment of cell signatures associated with suppressed anti-tumour activity. There was enrichment for a CD8 T-cell tolerance signature (GSE14699 NES 1.47, p<0.001), evidence of more Treg cells (GSE 14308, NES 1.24, p=0.044) and also activation of the Treg cells present in the tumour (GSE 14415, NES 1.42, p=0.004).

There was no significant association between the identified IPS signatures and the presence of MSI, LNMS or tumour stage; however, patients with an IPS1 signature tended to be older (median age 69.5 vs. 65 years) (Table 5.11).

In summary, unsupervised hierarchical clustering of patient samples identified two subtypes of immune profiles within the neighbouring stroma of LNM in patients with stage III CRC. Patients with IPS1 (poor prognosis) have signatures in keeping with suppressed anti-tumour immunity with Treg activation signatures. Patients with IPS2 (good prognosis) have signatures consistent with an anti-tumour response. Together, these changes in immune profiles could be responsible for the observed trend towards

Figure 5.10: A. Dendrograms illustrating IPS clustering. B. Effect of IPS cluster and disease free survival

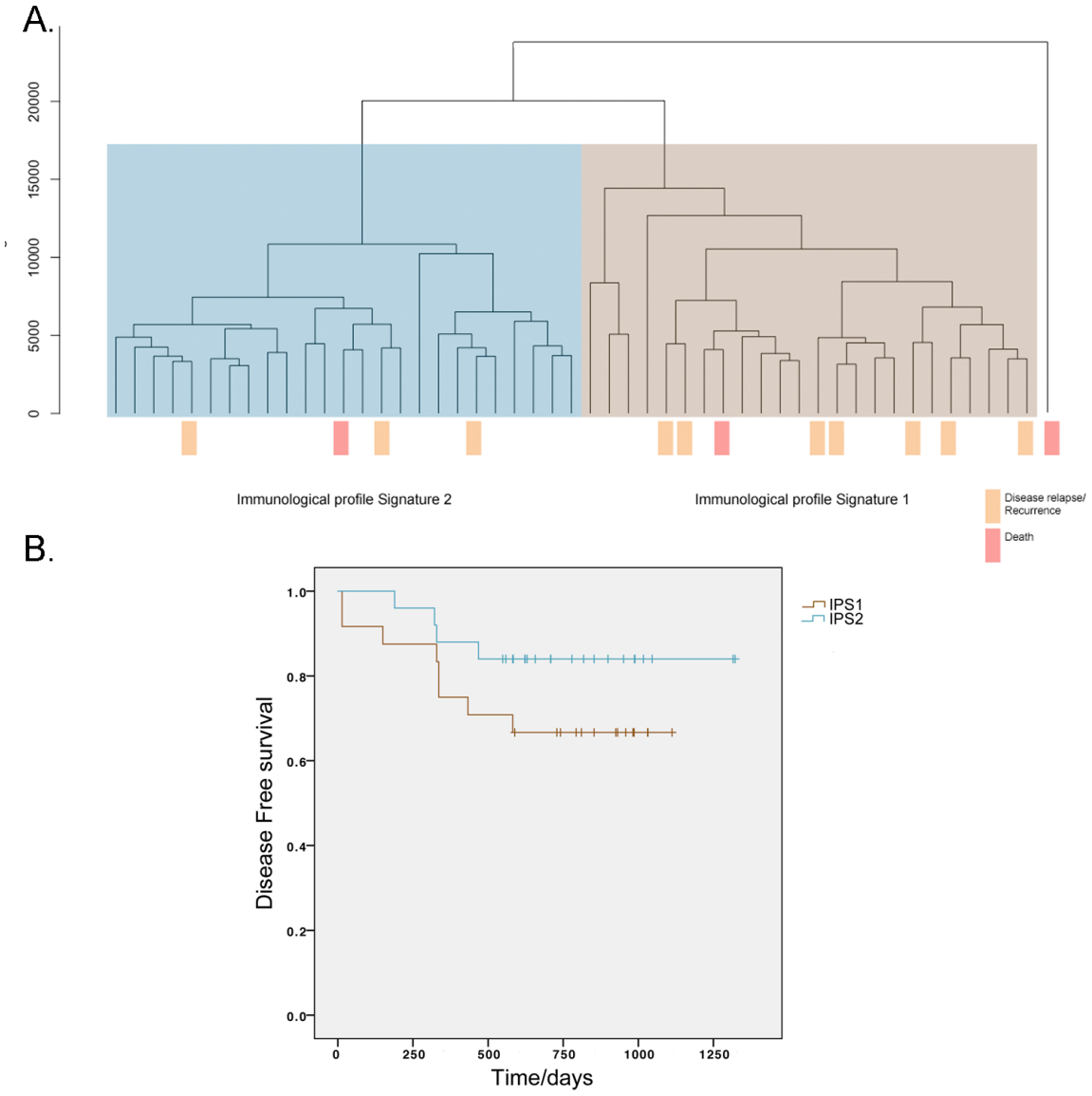


Table 5.11: Associations with tumour features and IPS1 and IPS2 signatures

	IPS1, n=24 (%)	IPS2, n=25 (%)
Median age at diagnosis	69.5	65
Tumour Stage		
pT1/2	1 (4.2)	3 (12.0)
pT3	14 (58.3)	9 (36.0)
pT4	9 (37.5)	13 (52.0)
SEX		
Male	9 (37.5)	13 (52.0)
Female	15 (62.5)	12 (48.0)
Microsatellite status		
MSS	16 (66.7)	16 (64.0)
MSI+	2 (8.3)	4 (16.0)
unknown	6 (25.0)	5 (20.0)
Metastasis subtype		
LNMS1	3 (12.5)	4 (16.0)
LNMS2	18 (75)	14 (56.0)
Unclassified	3 (12.5)	7 (28.0)

poorer survival.

5.5 Discussion

Metastasis is a multistep process involving both lymphoid and haematogenous spread. It is of utility to dissect and gain a greater understanding of the biological changes at each step to improve and inform new therapies for patients with CRC. In this chapter, I have put together the largest cancer cohort of paired tumour-metastatic patients for whole transcriptome analysis. A cohort of this size allows the power to identify differential expression of LNM, identification of putative immune profiles and allows deeper insight into the metastatic process.

During tumorigenesis, clones arise sporadically, each with different gene expression profiles[21]. This process arises either through the development of new mutations, alternations in chromosome structure with resulting SCNAs, chromatin changes or

methylation changes. There is a good understanding of this phenomenon, for example, loss of mismatch repair proteins leads to a high mutation rate, which is in turn associated with CMS1 subtype tumours. What is not clear is how the process of tumour evolution affects metastasis.

What I observed is that LNM have marked changes in expression of genes and signalling networks compared to the primary tumour. LNM are generally more aggressive than the primary tumour and KRAS signalling is preferentially activated. The degree of KRAS activation is in turn associated with tumour recurrence and patient survival. However, activation of KRAS signalling does not exist in isolation, and I demonstrated that within metastatic lesions other networks such as NFKB, Hedgehog and PI3K-AKT-mTOR, and NOTCH are concurrently activated. This activation of multiple oncogenic signalling pathways in metastatic lesions may explain why chemotherapy is much less efficient/inefficient at treating metastatic lesions.

My analyses suggest that there are two subsets of LNM (LNMS1 vs. LNMS2). One subset, LNMS1, is characterised by up-regulation of KRAS/AKT/TGFb signalling and has features of high tumour metabolic activity (oxidative phosphorylation and glycolysis), and these signatures identify patients of poor prognosis. The second subset, LNMS2 is characterised by immune infiltration and identifies patients with good prognosis.

I observed that the ability of a tumour to change GEP may be limited by MSI status. Thus, patients with loss of mismatch repair proteins (MSI) have metastatic lesions that have a GEP profile very similar to the primary tumour, with less changes in oncogenic signalling pathways, and this is in turn associated with good prognosis.

Finally, this data suggests that metastasis is a phenomenon not solely defined by features of the tumour. Within my dataset, I have identified gene expression signatures associated with inflammation and immune cell infiltrates. The presence of these anti-tumour response signatures, with innate and adaptive immune cell infiltration, is a principal feature that identifies LNMS2 patients. This supports the hypothesis that immune cells being trafficked to lymph nodes containing LNM and promote an anti-tumour response. Whilst it is not possible to draw inferences about causality from this dataset, the gene signatures identified here could be useful in predicting patient prognosis.

To gain a greater understanding of the role of the immune system, I analysed the LN stroma adjacent to the LNM, an environment I term the Lymphoid Ecological Niche (LEN). Two types of immune profiles could be delineated and these are largely independent of tumour features. Immune Profile Subtype 1 (IPS1) is characterised by the presence of T regulatory cells that are activated and drive tolerance of CD8 T cells. Immune Profile subtype 2 (IPS2) demonstrates activated innate and adaptive anti-tumour immune responses with evidence of stimulated T/B cells as well as NK, macrophages and dendritic cells.

This dichotomous immune profile may underpin an important clinical feature observed in cancer patients, that of *reactive lymph nodes*. Reactive lymph nodes are clinical signs, consisting of enlarged lymph nodes that do not to contain LNM but have evidence of a profuse immune response. They are usually identified along the lymphatic drainage pathways of large tumours[261] and in breast cancer, the presence of these reactive lymph nodes denotes favourable prognosis[262]. It is possible that LEN immune profile subtypes is a new technique of determining lymph node reactivity and the degree of immune response.

Mean age at diagnosis was different between the two immune profile subtypes. Elderly patients are known to exhibit decreased immune functionality, so-termed immune senescence[263] and this is a trend observed within this dataset. It is possible that the presence of immune senescence may underpin the relative poor prognosis of elderly patients diagnosed with CRC[25].

There are limitations to the work presented in this chapter. I have only presented an analysis of the gene changes that occur in lymph node metastatic lesions. Genes involved in mediating metastasis to lung and liver are likely to be involve genes above and beyond those observed in my work. Secondly, this work presents work defining new associations and further work will be performed to validate and confirm the signatures observed.

In summary, this dataset has helped define the molecular features involved in the first step of metastasis, that of lymph node metastasis. I have achieved this through a greater understanding of the GEP that occur within LNM, and also by providing evidence that it is possible to gain insight into the patient's immune profile by analysing the LEN. It is likely that immune profiles could also be derived through the use of blood markers[264], however, my approach has the benefit of understanding the patient's immunology directly adjacent to metastatic lesions. In future, I believe that a combined immune-tumour analysis would be required for a thorough understanding of tumour metastasis biology. More work will be required to identify factors that determine host immune profiles and the relationship of these to tumour biology.

Chapter 6

Discussion and future directions

Colorectal cancer is the third most common cancer in the United Kingdom and a cause of significant morbidity and mortality. The only treatment available for patients diagnosed with CRC that offers the potential for long-term remission from the disease is surgery. Adjuvant and neoadjuvant chemotherapy add only small incremental benefits in addition to surgical management. The definitive clinical trial to assess the success of fluorouracil and leucovorin (5-FU/LV) in stage III CRC was published in 1999. This trial identified that adjuvant chemotherapy conferred a small disease free survival benefit with one recurrence prevented for every twenty patients treated[265]. The additional benefit gained from the combination of oxaliplatin with 5-FU/LV for stage III CRC patients was discovered in 2004, with an incremental disease free survival benefit of 73.3% vs. 67.4%[266]. This three drug combination, also known as FOLFOX, works non specifically to inhibit DNA synthesis, and this has formed the basis of treatment for stage III CRC for the last decade. The benefits of chemotherapy for patients with stage II CRC is not clear or of low efficacy[267], and targeted therapy (such as EGFR monoclonal antibodies) have only demonstrated small incremental benefits for patients with stage IV disease, with evidence of detrimental effect for earlier stage disease.

In the future, bowel cancer screening is likely to form a big part of our national CRC strategy. Faecal occult blood testing began operating across England in 2009 with patients aged 60-69 years invited to screening every 2 years. Earlier this year, flexible sigmoidoscopy was launched across the UK as a one off screening test for all men and women aged 55. In decades to come, with more patients diagnosed with earlier stage CRC, surgery will remain the mainstay of treatment.

In this context, an understanding of metastasis is of importance. At present, we have very limited tools to predict those who are at risk of disease recurrence or metastasis following successful resection of their tumour. Likewise, there is a relatively poor understanding of the tumour features or biological mechanisms that govern the process. A long-stated aspiration of cancer researchers has been to develop a means to assess the tumour biology of the CRC at diagnosis. Targeted therapies to either prevent the seeding of the tumour before surgery or to treat the presence of established micro-metastases outside the colon following surgery may then be possible.

To improve the understanding of metastasis, new and accurate models of CRC metastasis need to be created. Models of metastasis are currently limited, modelling only the earliest stages of metastasis and animal models are expensive and labour intensive. Results in chapter 3, describe the creation of the OTM model that aims to address many of these limitations. This model involves the orthotopic transplantation of human/mouse CRC directly into the submucosa of the descending colon. There are notable advantages to this approach. Firstly, an in-vivo approach means that it is possible to model the more complex features of CRC metastasis, involving the invasion of normal stroma, intravasation into the circulation and extravasation at target organs. Secondly, orthotopic transplantation models result in metastasis of

the tumour to biologically meaningful sites with high efficiency. Thirdly, the model can be tailored and customised to the research question through the use of immune competent/deficient/transgenic mice and admixtures of cells transplanted into the colon. A final advantage of this approach is that of research resources. Experiments can be performed in any lab with an animal facility, without the need to establish new surgical mouse project licenses. Compared to transgenic models, orthotopic transplantation models use less animal resources and can be completed more rapidly and economically. Altogether, these advantages mean that it becomes feasible to identify the roles of new genes and treatments in a more thorough and timely fashion.

I was able to investigate the role of Serine/Glucocorticoid Regulated Kinase 1 through the use of OTM in chapter 3. *SGK1* is a marker and driver of differentiation during normal homeostasis and expression is down-regulated in CRC. The model enabled me to identify an additional unforeseen function of this intestinal differentiation factor, that of a metastasis suppressor gene. Metastasis rates were significantly reduced when *SGK1* was re-expressed in cancers and this was mediated through inhibition of the pro-metastatic c-MYC signaling cascade. In normal intestinal homeostasis, *SGK1* is expressed in differentiated cells and I propose it functions to prevent intestinal proliferation outside of the intestinal crypts. In cancer, *SGK1* is down-regulated and this gives rise to a more undifferentiated cell that is able to proliferate, migrate and metastasise. Indeed, analysis of human cohort databases demonstrated that the degree to which a primary tumour is able to down-regulate *SGK1* is a robust biomarker of disease recurrence/metastasis and death in patients with CRC.

In chapter 4, I took an approach that would enable me to perform a screen of a large number of genes to determine how many of them had roles in the development

of metastasis. I combined a whole-genome CRISPR/CAS9 screen with the aforementioned OTM model to perform a non-hypothesis driven knockout screen. This approach enabled me to generate a list of metastasis regulating genes and a large number of genes were identified. Perhaps this should not be of great surprise considering that metastasis consists of a number of different steps and numerous genes may be involved in each step. From this gene list, I have validated the role of the pioneer transcription factor, *FOXF1*, as being of foremost importance. In breast cancer, it has been implicated in regulating cell proliferation and in colon cancer cell line HCT116, it regulates cellular migration. I provide evidence that *FOXF1* expression is down-regulated in CRC tumours and reduced further in metastatic lesions and this enables activation of the pro-metastatic mTOR signalling pathway, which is known to drive metastasis.

Chapter 5 reports my experiments aimed at gaining a greater understanding of metastasis through the use of human CRC specimens. A high degree of biological variability is introduced when moving from animal models to human CRC specimens. Human CRC contain intra-tumour heterogeneity, arise from different locations in the colon, from different risk factors, from different immune responses and from patients of different age and gender. To enable whole-transcriptome profiling of a human CRC LN metastasis (LNM) cohort, I utilised 3' RNA sequencing. 3' RNA sequencing is a technology designed to be used with FFPE tissues and this technique allowed the sequencing of a large cohort at relatively low cost. My cohort of 69 patients, the OSLER sample set, is the largest cancer cohort to date for which whole-transcriptome profiling has been performed. My analysis from these LNM specimens demonstrate a more aggressive tumour with over-activation of all known oncogenic signalling networks. Some of the signalling networks up-regulated in human LNMs are the same as those mediated by *FOXF1* and *SGK1*, including c-MYC and MTOR signalling.

KRAS signalling was significantly up-regulated in LNMs where it appeared to partly define LNMS clustering of LNMs and act as a biomarker of poor prognosis. I also identified some tumour-extrinsic features of metastasis. The presence of immune infiltration was associated with good prognosis and it is possible that this technique may in future be useful to help characterise immune profiles of CRC patients.

I have attempted to gain insight into metastasis using many contrasting approaches. Further work would look to identify unifying features of driver genes that promote metastasis in colorectal cancer.

A difficulty is that it is not clear what models, assays, markers or gene networks constitute a "metastasis phenotype". It has often argued that in-vitro assays are unnecessarily reductionist in their approach and that in-vivo studies must be performed with extensive validation using human cohorts. I believe it is a false dichotomy to suggest there might be "metastasis promoters" and "non-metastasis promoter" genes. My work using the genome-wide CRISPR/Cas9 screen has identified many genes, however, there seems to exist a gradient to which a gene may act as a "metastasis promoter". Some genes, like *FOXF1*, may mediate a relatively large effect, but even then, presence of this gene alteration within a cell will not always lead to metastatic deposits. There are probably many more factors that determine metastasis than single gene expression changes.

However, perhaps the strongest inference I can draw from my dataset is of the importance of MTORC signalling in driving metastasis. This has been a consistent finding that I have observed in each of my results chapters, *SGK1* re-expression reduces metastasis and there was a corresponding negative enrichment for MTORC signalling. One of the top hits from the CRISPR screen was of *Deptor*, a negative

regulator of MTORC. *FOXF1* re-expression reduces metastasis and a similar negative enrichment for MTORC signalling was observed. The MTORC inhibitor (sirolimus) was extremely successful in preventing metastasis outgrowth. Finally MTORC signalling was up-regulated in the LNM in chapter 5.

Understanding signalling networks is challenging due to the extensive cross-talk that happens between networks, but MTORC is notable because of the extent to which this happens. Thus MTORC signalling is downstream of the Ras/mapk/erk, Wnt, Akt pathway and there is some data to suggest it interacts with *c-MYC*[268]. It is tempting to hypothesise that MTORC might function as a common node of these signalling pathways and perhaps act to activate metastasis through 4E-BP1 once enough of them become activated. However, much work would be required to prove this experimentally.

There are some limitations to the results outlined in this thesis. The study of metastasis with mouse orthotopic transplantation models is limited by the growth of the primary tumour. Experimental mice do develop distant metastatic lesions, but they often become symptomatic within a few weeks as a result of intestinal obstruction. Whilst this is ideal for understanding tumour biology with regards to metastasis, this does not reflect human metastasis that develops over several years and often post surgical resection. This may limit research into certain genes that require months/years to exert their metastatic effect, for example, DNA polymerase epsilon (POLE) mutations. Secondly, the human cohort is limited in scope. I have analysed the earliest step in the metastasis cascade, that of lymph node metastasis. The genes and networks identified in this thesis are likely to be required for the development of distant metastasis to liver/lung/bone, but is not clear if they would be sufficient or if further genes expression changes would be required to facilitate this

process. Finally, it will also be important to validate the LNMS and IPS signatures obtained from this thesis in a validation cohort and this work will continue following submission of this thesis.

In the years ahead, work will be required to delve more deeply into the drivers of metastasis and I believe four approaches are likely to become more important. Firstly, I support the view that metastasis does not arise out the collective effect of tumour evolution, but arises as chance event in a single clone. Most cells in a human CRC would not be capable of establishing distant lesions. Finding a metastasis-capable clone within a tumour is akin to finding a needle in a haystack. To understand how, when and why a metastasis capable clone arises, single cell sequencing may be a particularly fruitful approach. Secondly, it was beyond the scope of my thesis to challenge or lend evidence to either the "*Soil and Seed*" hypothesis or the "*Filter and Flow*" theories. Different cancers seed to different sites and the underlying reasons are unknown. There is evidence in the literature that tumours may promote the establishment of a pre-metastatic niche, through the secretion of tumour exosomes[50 269] though this mechanism has not yet been defined for CRC. It would be important to gain a greater understanding of this phenomenon. Thirdly, the presence of micro-metastases has yet to be definitively proven. These are tumour deposits that are too small to be detected by conventional methods, but their existence has been used as a justification for disease recurrence, metastasis, relapse and for the failure of chemotherapy. The ability to detect these lesions have been elusive, though it is likely that new targeted sequencing technologies may be able to identify surrogate features of micro-metastases and this could be of great clinical benefit in future. Finally, mutations and gene copy number changes are central hallmarks of cancer. These are the fuel that drive gene expression changes observed in cancers and in metastatic lesions. A key tool to understand the process in more detail would involve an understanding

of both the genetic changes that arise somatically in the tumour and also the genetic background of the host responsible for specifically targeting these gene alterations. In the future, I envisage that these complementary approaches might allow greater insight into a patient's likely outcome from CRC, both from an immunological and a tumour biology perspective and allow us to develop new interventions to improve outcomes from CRC.

6.1 Future experimental plans

These are the next set of experiments planned based on the data in this thesis.

Chapter 3

1. I have demonstrated that MYC co-immunoprecipitates with SGK1 in three cell lines. I believe this adds strength to my assertion that SGK1 participates in a kinase cascade which eventually leads to MYC phosphorylation and degradation. To strengthen this association, I would perform co-immunoprecipitation experiments using C-MYC as bait and SGK1 as target. If the co-immunoprecipitation were observed for both proteins, it would suggest that the protein-protein interactions are strong and more likely to be physiological.

2. I have demonstrated that SGK1 re-expression leads to transcriptome changes which are consistent with MYC signalling downregulation. However, to add more impact to my work, I would like to demonstrate that SGK1 acts primarily through MYC signalling. To achieve this, I would like to re-express SGK1 in cell lines in which there was selective inhibition of MYC signalling. To achieve this, I would use a MYC inhibitor. One MYC inhibitor that I thought would be useful was KJ-Pyr-9. This would help identify if there were other pro-metastatic signalling networks that

were perturbed and equally significant, but also demonstrate the importance of MYC signalling in my model systems.

3. There is likely to be significant cross talk between SGK1 and proteins involved in the MYC signalling cascade. To gain an idea where in the network SGK1 interacts with MYC signalling, I would like to perform selective mutagenesis of different proteins in the MYC signalling cascade. This would enable me to identify the exact target that was sufficient and necessary for SGK1-mediated MYC downregulation. Provisional proteins that I could target were those of the MAKPH pathway, including RAF, MEK, MAPK and ERK.

4. At present, it is unclear how SGK1 was silenced in human CRC, though it is likely that expression of SGK1 might be translationally controlled. To investigate this further, it would be of utility to perform a mutagenesis screen of potential TFBS upstream of SGK1. By ablating these sites and looking at potential effects on SGK1, it would be possible to identify putative repressor sites and important transcription factors. These TFBS could be selectively ablated using CRISPR/Cas9.

5. SGK1 exists as three distinct isoforms. These isoforms are all regulated by post-translational modifications. There is some evidence that different isoforms have different half-lives and cellular localisation[80]. A long-term plan would be to determine the effects of these isoforms and their respective splice variants on CRC cell growth and proliferation. I would ideally also like to determine if these effects are also mediated through downregulation of MYC signalling.

Chapter 4

1. At present in this chapter I was able to demonstrate that there are a number of TFBS that FOXF1 is able to bind to. To add strength to this mechanistic link, it would be necessary to mutate these sites and determine its effect on RPTOR

expression and the likelihood of FOXF1 binding to these sites. Experiments would entail CRISPR-Cas9 mediated mutagenesis and analysis of expression using rt-QPCR and/or protein analysis via Western blotting.

2. Previous studies have demonstrated that there is mislocalisation of FOXF1 in CRC and the degree of which was correlated with lymphatic invasion[214]. However, this experiment was done using a TMA slide in only 50 patients with all stages of CRC. I feel it would be prudent to confirm this analysis, but also to stratify patients by grade of tumour. The current analyses using cancers of all stages is limited as it did not offer any prognostic information beyond that offered by traditional TMN staging. A larger cohort, stratified by grade of tumour would allow me to improve my understanding of the utility of FOXF1 expression as a prognostic biomarker.

3. I have demonstrated the importance of mTOR signalling in CRC metastasis using a tail vein assay in mice treated with Sirolimus. To add more strength to the link between FOXF1 re-expression and inhibition of mTOR signalling, it would be important to perform mTOR signalling in the context of FOXF1 re-expression and knockout cell lines. This would then enable me to identify causality and sufficiency of mTOR inhibition following FOXF1 re-expression.

4. To identify downstream effects of mTOR signalling, I have currently used whole protein quantification. I have not analysed the effect on the phosphorylation sites of these proteins. As mTOR signalling functions predominantly through phosphorylation of different proteins, a phospho-protein western analysis would enable me to identify if FOXF1 might also induce post translational modifications as well as exerting transcriptional control.

Chapter 5

1. This chapter uses high dimensionality data, in this case, the expression of

20,000 genes. This means that there is a high likelihood of false positive results. I would like to establish a validation cohort of another 70 patients and perform RNA sequencing. To do this, I would use the same 3' RNA sequencing technique and use the same clustering algorithm. If I were able to identify the same signatures again, I would be able to say with more certainty that the signatures might be of clinical utility.

2. At present, my immune subtyping signature was underpowered and whilst there was a trend for the different immunological subtypes to be linked with prognosis, it did not reach statistical significance. To address this shortcoming, I would like to expand the number of patients with immune subtyping and determine if this is of utility. This patients would be stage III CRC patients and immune subtyping would be performed on neighbouring normal areas of LN proximal to the primary tumour.

3. I have been successful in identify signatures that characterise certain immunogical subtypes, such as CD8 T cells. To validate the findings of the RNA sequencing at a protein level, it would be useful to perform IHC experiments. I would use stains specific to CD8 T cells, but also consider staining for CD4 T cells and macrophages, cells that were also identified as being differentially present in my transcriptome dataset. IHC experiments would enable me to identify relative differences in cell number, but give me additional anatomical and geographic information within the LNM.

4. This cohort was retrospective analysis of patients who were diagnosed with CRC. The next experiment would entail a prospective collection of tissue in a cohort of patients currently undergoing treatment. Fresh tissue would enable me to address a known criticism of RNA sequencing studies of FFPE specimens, that of degradation of RNA, particularly of long transcripts and the increased likelihood of multiple-mapping of the reads.

5. Finally, I would like to start collection of blood from patients with stage III CRC. If the immunological signatures, particularly CD8 cells can be identified in the

validation cohort of patients, I would be particularly interested to determine if it could be identified from blood. This would mean that at the time of diagnosis, it would be possible to gain an idea of the immune phenotype of the patient and potentially to allow the tailored treatment for drugs like immune checkpoint inhibitors in future.

Appendix A

Appendix

Table A.1: GSEA results for analysis of MSigDB Hallmark gene sets following SGK1 re-expression in CRC cell lines

MSigDB	NES	p-val
HALLMARK_OXIDATIVE_PHOSPHORYLATION	-1.91	5.83E-03
HALLMARK_GLYCOLYSIS	-1.48	1.73E-02
HALLMARK_MTORC1_SIGNALING	-1.65	2.01E-02
HALLMARK_MYC_TARGETS_V1	-1.94	2.72E-02
HALLMARK_REACTIVE_OXYGEN_SPECIES	-1.38	1.22E-01
HALLMARK_CHOLESTEROL_HOMEOSTASIS	-1.28	1.48E-01
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	1.22	1.51E-01
HALLMARK_DNA_REPAIR	-1.26	1.62E-01
HALLMARK_ALLOGRAFT_REJECTION	-1.25	1.63E-01
HALLMARK_PEROXISOME	-1.26	1.84E-01
HALLMARK_HEDGEHOG_SIGNALING	1.25	1.84E-01
HALLMARK_ADIPOGENESIS	-1.13	2.71E-01
HALLMARK_MYC_TARGETS_V2	-1.19	3.02E-01
HALLMARK_ANGIOGENESIS	-1.11	3.03E-01
HALLMARK_FATTY_ACID_METABOLISM	-1.15	3.04E-01

HALLMARK_ANDROGEN_RESPONSE	1.08	3.19E-01
HALLMARK_MITOTIC_SPINDLE	1.10	3.27E-01
HALLMARK_COAGULATION	-1.04	3.71E-01
HALLMARK_APICAL_JUNCTION	1.03	3.97E-01
HALLMARK_KRAS_SIGNALING_DN	1.08	4.04E-01
HALLMARK_HEME_METABOLISM	1.04	4.04E-01
HALLMARK_E2F_TARGETS	-0.91	5.17E-01
HALLMARK_SPERMATOGENESIS	0.93	5.23E-01
HALLMARK_XENOBIOTIC_METABOLISM	-0.96	5.33E-01
HALLMARK_PI3K_AKT_MTOR_SIGNALING	0.90	6.38E-01
HALLMARK_INTERFERON_GAMMA_RESPONSE	-0.74	6.42E-01
HALLMARK_PROTEIN_SECRETION	-0.88	6.45E-01
HALLMARK_ESTROGEN_RESPONSE_EARLY	-0.93	6.46E-01
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	-0.88	6.53E-01
HALLMARK_NOTCH_SIGNALING	0.82	7.09E-01
HALLMARK_HYPOXIA	-0.82	7.37E-01
HALLMARK_MYOGENESIS	0.77	7.68E-01
HALLMARK_UV_RESPONSE_DN	-0.82	7.69E-01
HALLMARK_APICAL_SURFACE	0.80	8.12E-01
HALLMARK_PANCREAS_BETA_CELLS	0.72	8.28E-01
HALLMARK_ESTROGEN_RESPONSE_LATE	-0.80	8.40E-01
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.52	8.45E-01
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.62	8.54E-01
HALLMARK_G2M_CHECKPOINT	0.63	8.56E-01
HALLMARK_KRAS_SIGNALING_UP	0.70	8.57E-01
HALLMARK_P53_PATHWAY	-0.74	8.58E-01
HALLMARK_COMPLEMENT	0.65	8.90E-01
HALLMARK_UV_RESPONSE_UP	-0.61	9.13E-01
HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.51	9.17E-01
HALLMARK_APOPTOSIS	-0.66	9.19E-01
HALLMARK_BILE_ACID_METABOLISM	-0.57	9.58E-01
HALLMARK_INFLAMMATORY_RESPONSE	-0.58	9.72E-01
HALLMARK_IL2_STAT5_SIGNALING	0.48	9.86E-01

HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.40	9.88E-01
HALLMARK_TGF_BETA_SIGNALING	0.35	9.88E-01

Table A.2: GSEA results for analysis of MSigDB Hallmark gene sets using genes identified from GECKmA screen

MSigDB Hallmark Gene Set	NES	p-val
HALLMARK_HEME_METABOLISM	1.45	4.08E-03
HALLMARK_GLYCOLYSIS	1.29	4.47E-02
HALLMARK_APOPTOSIS	1.30	5.34E-02
HALLMARK_KRAS_SIGNALING_UP	1.25	6.98E-02
HALLMARK_PANCREAS_BETA_CELLS	1.38	7.48E-02
HALLMARK_FATTY_ACID_METABOLISM	1.24	9.05E-02
HALLMARK_UV_RESPONSE_DN	1.24	9.91E-02
HALLMARK_ESTROGEN_RESPONSE_EARLY	1.20	1.27E-01
HALLMARK_INFLAMMATORY_RESPONSE	1.19	1.40E-01
HALLMARK_MYC_TARGETS_V1	-1.08	1.67E-01
HALLMARK_INTERFERON_ALPHA	1.19	1.79E-01
HALLMARK_COMPLEMENT	1.15	2.09E-01
HALLMARK_ESTROGEN_RESPONSE_LATE	1.14	2.19E-01
HALLMARK_MYOGENESIS	1.13	2.28E-01
HALLMARK_PEROXISOME	1.16	2.42E-01
HALLMARK_HYPOXIA	1.09	3.01E-01
HALLMARK_TNFA_SIGNALING_VIA_NFKB	1.09	3.02E-01
HALLMARK_ANDROGEN_RESPONSE	1.10	3.14E-01
HALLMARK_IL6_JAK_STAT3_SIGNALING	1.10	3.15E-01
HALLMARK_REACTIVE_OXYGEN_SPECIES	1.10	3.31E-01
HALLMARK_ALLOGRAFT_REJECTION	1.07	3.60E-01
HALLMARK_PI3K_AKT_MTOR_SIGNALING	1.07	3.71E-01
HALLMARK_APICAL_JUNCTION	1.02	4.35E-01
HALLMARK_PROTEIN_SECRETION	1.00	4.58E-01

HALLMARK_COAGULATION	1.02	4.58E-01
HALLMARK_HEDGEHOG_SIGNALING	1.00	4.61E-01
HALLMARK_WNT_BETA_CATENIN_SIGNALING	-0.99	4.73E-01
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.99	4.93E-01
HALLMARK_ADIPOGENESIS	0.95	6.19E-01
HALLMARK_P53_PATHWAY	0.93	6.43E-01
HALLMARK_XENOBIOTIC_METABOLISM	0.94	6.50E-01
HALLMARK_KRAS_SIGNALING_DN	0.92	6.54E-01
HALLMARK_ANGIOGENESIS	0.87	6.58E-01
HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.89	6.71E-01
HALLMARK_OXIDATIVE_PHOSPHORYLATION	0.91	6.88E-01
HALLMARK_MITOTIC_SPINDLE	0.89	7.11E-01
HALLMARK_DNA_REPAIR	0.88	7.19E-01
HALLMARK_BILE_ACID_METABOLISM	0.88	7.20E-01
HALLMARK_IL2_STAT5_SIGNALING	0.86	7.86E-01
HALLMARK_G2M_CHECKPOINT	0.84	8.09E-01
HALLMARK_MTORC1_SIGNALING	0.81	8.50E-01
HALLMARK_TGF_BETA_SIGNALING	0.73	8.65E-01
HALLMARK_MYC_TARGETS_V2	-0.77	8.77E-01
HALLMARK_SPERMATOGENESIS	0.78	8.81E-01
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	0.73	9.11E-01
HALLMARK_UV_RESPONSE_UP	0.72	9.32E-01
HALLMARK_E2F_TARGETS	0.70	9.51E-01
HALLMARK_APICAL_SURFACE	-0.68	9.66E-01

Table A.3: GSEA results for analysis of MSigDB Hallmark gene sets following FOXF1 oe using CRISPR/SAM and Lentivirus

	CRISPR		Lentivirus	
MSigDB	NES	p-val	NES	p-val
HALLMARK_MTORC1_SIGNALING	-1.47	0.00	-1.74	0.00

HALLMARK_MYC_TARGETS_V2	1.74	0.00	-2.79	0.00
HALLMARK_ESTROGEN_RESPONSE	-1.42	0.01	-1.75	0.00
HALLMARK_HEME_METABOLISM	1.41	0.01	1.93	0.00
HALLMARK_PROTEIN_SECRETION	-1.46	0.02	1.82	0.00
HALLMARK_TNFA_SIGNALING	-1.41	0.01	1.51	0.00
HALLMARK_HYPOXIA	-2.01	0.00	1.42	0.00
HALLMARK_GLYCOLYSIS	-1.74	0.00	-1.34	0.01
HALLMARK_MITOTIC_SPINDLE	-1.53	0.00	1.34	0.02
HALLMARK_OXIDATIVE_PHOSPHORYLATION	1.39	0.01	-1.31	0.02
HALLMARK_DNA_REPAIR	1.46	0.01	-1.30	0.03
HALLMARK_APOPTOSIS	1.24	0.07	1.58	0.00
HALLMARK_INTERFERON_GAMMA_RESPONSE	1.22	0.07	2.01	0.00
HALLMARK_INTERFERON_ALPHA_RESPONSE	1.27	0.08	2.17	0.00
HALLMARK_G2M_CHECKPOINT	-1.21	0.09	-1.89	0.00
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	-1.30	0.06	-1.46	0.01
HALLMARK_INFLAMMATORY_RESPONSE	1.20	0.09	1.32	0.03
HALLMARK_IL6_JAK_STAT3_SIGNALING	-1.31	0.08	1.41	0.03
HALLMARK_REACTIVE_OXYGEN_SPECIES	1.41	0.05	1.40	0.05
HALLMARK_P53_PATHWAY	1.32	0.02	1.27	0.06
HALLMARK_IL2_STAT5_SIGNALING	-1.38	0.02	-1.18	0.09
HALLMARK_COMPLEMENT	-1.16	0.15	1.72	0.00
HALLMARK_E2F_TARGETS	1.06	0.30	-2.41	0.00
HALLMARK_MYC_TARGETS_V1	-0.91	0.72	-2.76	0.00
HALLMARK_ADIPOGENESIS	-1.17	0.13	-1.33	0.01
HALLMARK_HEDGEHOG_SIGNALING	1.09	0.33	-1.22	0.16
HALLMARK_ESTROGEN_RESPONSE_EARLY	-1.56	0.00	-1.64	0.00
HALLMARK_COAGULATION	0.89	0.74	1.49	0.00
HALLMARK_TGF_BETA_SIGNALING	-1.17	0.21	1.47	0.02
HALLMARK_NOTCH_SIGNALING	0.91	0.61	-1.55	0.02
HALLMARK_ANDROGEN_RESPONSE	-1.79	0.00	1.38	0.03
HALLMARK_UV_RESPONSE_UP	1.16	0.14	1.28	0.05
HALLMARK_PI3K_AKT_MTOR_SIGNALING	1.32	0.05	1.22	0.12
HALLMARK_UV_RESPONSE_DN	-1.74	0.00	1.16	0.16

HALLMARK_XENOBIOTIC_METABOLISM	-1.12	0.21	-1.13	0.16
HALLMARK_BILE_ACID_METABOLISM	0.85	0.81	1.16	0.18
HALLMARK_SPERMATOGENESIS	-1.04	0.37	1.14	0.22
HALLMARK_CHOLESTEROL_HOMEOSTASIS	-2.33	0.00	-1.12	0.22
HALLMARK_FATTY_ACID_METABOLISM	-1.39	0.02	1.11	0.24
HALLMARK_MYOGENESIS	1.32	0.03	1.06	0.31
HALLMARK_PEROXISOME	-1.21	0.13	1.07	0.33
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	-1.37	0.02	1.05	0.34
HALLMARK_APICAL_SURFACE	-1.18	0.22	-1.06	0.36
HALLMARK_APICAL_JUNCTION	1.02	0.42	-0.99	0.47
HALLMARK_ALLOGRAFT_REJECTION	-0.97	0.53	0.99	0.51
HALLMARK_KRAS_SIGNALING_UP	-1.28	0.04	-0.96	0.60
HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.96	0.53	0.87	0.68
HALLMARK_PANCREAS_BETA_CELLS	-0.74	0.87	-0.87	0.68
HALLMARK_ANGIOGENESIS	-1.84	0.00	0.85	0.68
HALLMARK_KRAS_SIGNALING_DN	0.99	0.49	-0.79	0.97

Table A.4: CRC LN stroma gene set- 464 genes that are were significantly differentially expressed in CRC LN stroma compared to primary CRC

Gene	log2FC	p-value adj
TLR10	3.29	6.68E-26
RNU6-1201P	4.39	7.38E-26
AL591806.1	4.51	6.25E-25
MAL	3.74	1.40E-24
AC104024.2	4.13	3.60E-24
FLT3LG	3.60	8.98E-23
FAM159A	3.61	1.15E-22
RP5-887A10.1	3.99	4.13E-22
CCL21	2.83	4.13E-22

RP11-706D8.3	3.93	1.19E-19
CR2	3.13	1.90E-19
PRKCB	2.35	6.69E-18
TRAC	2.16	1.70E-17
ENAM	3.74	1.99E-17
TRAJ20	3.40	3.75E-17
CTD-3234P18.2	3.92	4.29E-17
SLC12A3	4.16	5.54E-17
SASH3	2.50	5.87E-17
NPAP1P4	3.27	6.22E-17
C11orf21	3.37	1.83E-16
RP11-94L15.2	2.44	2.39E-16
CXCL13	2.09	3.53E-16
LTB	2.13	4.18E-16
TUB	3.37	5.44E-16
CD52	2.32	5.80E-16
ALK	3.21	1.14E-15
TBC1D27	3.11	1.28E-15
FAM65B	2.19	1.82E-15
PKNOX2	3.03	2.49E-15
AL021917.1	2.95	7.91E-15
RP11-861A13.4	3.03	1.27E-14
RP4-800J21.3	3.26	1.84E-14
RNA28S5	2.08	1.94E-14
HLA-DOB	2.64	2.51E-14
RP13-126C7.1	2.96	2.65E-14
CTA-407F11.7	2.66	5.05E-14
RASGRP2	2.29	5.25E-14
CLEC4M	4.41	6.25E-14
P2RX5	2.38	7.44E-14
MS4A1	2.89	7.65E-14
RASAL3	2.06	1.08E-13
CTA-250D10.23	2.49	1.41E-13

STAP1	2.99	1.59E-13
WDFY4	2.36	2.08E-13
CLEC2D	2.86	2.46E-13
CD3E	2.40	2.62E-13
CD19	2.65	2.90E-13
CNR2	3.41	3.05E-13
VPREB3	2.76	3.25E-13
KRT73	3.61	3.59E-13
AL512428.1	2.81	4.15E-13
AL928768.3	3.47	4.58E-13
RP1-92C8.2	3.30	5.02E-13
GFM2	2.80	5.23E-13
COL4A3	2.79	5.57E-13
TNFSF8	3.49	6.41E-13
FGD2	2.53	6.72E-13
PIK3CD	2.01	7.57E-13
SELL	2.69	7.65E-13
MIR142	2.24	8.19E-13
RP3-455J7.4	2.31	8.31E-13
ANKRD55	3.48	9.51E-13
NRBF2P5	3.19	1.41E-12
RP11-414H17.5	3.76	1.49E-12
BANK1	2.56	1.57E-12
CTD-2020K17.1	2.38	1.61E-12
EI24P2	3.54	2.45E-12
FCRLA	2.86	3.03E-12
GVINP1	2.43	3.45E-12
RP11-147L13.2	2.66	4.55E-12
CD3G	2.08	4.68E-12
RP11-160C18.2	3.06	4.99E-12
P2RY8	2.32	5.72E-12
C14orf64	2.68	7.62E-12
RP11-474N24.6	4.05	8.41E-12

RP11-693J15.5	2.73	8.61E-12
RNU6-437P	2.74	1.12E-11
GRAP2	2.44	1.35E-11
IGHD	2.74	1.96E-11
KLHL14	3.85	2.15E-11
PNLIP	2.86	2.24E-11
CCL19	2.24	2.31E-11
ACAP1	2.09	3.04E-11
ADAM28	2.01	3.83E-11
LRMP	2.42	4.11E-11
SNX22	2.52	4.61E-11
TDRD12	2.90	4.62E-11
ZNF879	2.69	4.88E-11
PTPRCAP	2.40	5.01E-11
TESPA1	2.72	5.15E-11
RPS6KB2	2.92	5.67E-11
AC005154.7	2.40	6.96E-11
AC084125.1	2.84	7.14E-11
PLA2G5	2.84	7.60E-11
RNA5SP429	2.95	7.66E-11
AC078899.1	4.26	8.92E-11
USP50	2.58	9.00E-11
POU2AF1	2.12	9.01E-11
S1PR1	2.32	1.09E-10
LIMD2	2.12	1.09E-10
LMO2	2.30	1.18E-10
TMEM156	2.63	1.24E-10
ITK	2.26	1.30E-10
PSMC3IP	2.69	1.45E-10
CCR7	2.53	1.49E-10
ITGAL	2.10	1.51E-10
ANKRD30BL	2.75	1.53E-10
FMO2	2.41	1.66E-10

RNA5SP149	3.88	1.93E-10
BACH2	2.19	1.98E-10
CD3D	2.10	2.09E-10
DOCK8	2.00	2.28E-10
AL109615.1	3.14	2.37E-10
SCIMP	2.41	2.66E-10
LYL1	2.62	2.71E-10
RASGEF1C	3.17	2.90E-10
SETP2	3.21	2.92E-10
CD22	2.63	3.21E-10
FDCSP	2.42	3.38E-10
NUDT22	2.55	3.48E-10
RP11-58K22.4	2.39	4.04E-10
RP11-347P5.1	2.12	4.23E-10
DNASE1L3	2.40	4.25E-10
CD79A	2.33	4.30E-10
TAS2R5	3.32	4.33E-10
RP11-849F2.8	2.10	4.37E-10
ARHGAP9	2.07	4.45E-10
TNFRSF13B	2.47	4.45E-10
RP11-661G16.2	2.92	4.45E-10
LY9	2.43	5.63E-10
LPAL2	3.92	5.87E-10
AL133458.1	2.10	6.14E-10
FCRL2	2.72	6.37E-10
RP11-535A5.1	2.63	6.40E-10
RNU6-1306P	3.54	6.57E-10
RP11-158I9.5	2.75	6.96E-10
FAIM3	2.28	7.35E-10
CTB-133G6.2	2.54	7.37E-10
RCSD1	2.19	7.57E-10
RP5-1139B12.4	2.91	7.98E-10
PARVG	2.20	9.04E-10

LINC00173	2.29	1.04E-09
RN7SL398P	3.55	1.07E-09
LINC00861	2.24	1.10E-09
RP11-375I20.6	2.73	1.20E-09
BIN2	2.24	1.25E-09
GJB6	2.76	1.26E-09
CCDC141	2.68	1.27E-09
AC009960.7	3.56	1.58E-09
BTLA	2.49	1.61E-09
RP11-281P23.1	2.80	1.86E-09
RP11-117D22.2	2.76	1.88E-09
GCGR	3.05	2.01E-09
RP11-428K3.1	2.67	2.13E-09
RP5-1092A3.4	2.31	2.16E-09
FAM196B	3.88	2.24E-09
MEG8	3.37	2.63E-09
CYSLTR1	2.22	2.72E-09
RP11-132N15.3	2.71	2.73E-09
MOBP	2.83	2.77E-09
ADAD2	3.04	3.03E-09
FCRL1	2.46	3.14E-09
CD48	2.45	3.18E-09
SMO	2.97	3.34E-09
RASGRP4	2.72	4.05E-09
FUNDC2P2	3.73	4.10E-09
MMRN1	2.81	4.23E-09
AC008440.10	4.06	4.51E-09
AC073410.1	2.44	4.76E-09
FGF11	2.99	4.96E-09
AICDA	2.84	5.12E-09
RN7SL5P	3.17	5.24E-09
SUGCT	2.28	5.33E-09
AC090627.1	2.14	5.36E-09

DARC	2.28	6.14E-09
GPR171	2.29	6.17E-09
RP11-1277A3.2	2.79	6.96E-09
PCED1B-AS1	2.04	6.98E-09
GLYATL1P3	2.85	6.98E-09
RP11-297C4.2	2.56	7.21E-09
IPCEF1	2.25	7.41E-09
SP140	2.14	8.21E-09
FAM177B	2.26	8.72E-09
RNU6-570P	2.95	8.82E-09
SCARA5	2.65	8.90E-09
RP3-475N16.1	2.44	9.04E-09
PLD4	3.14	9.07E-09
PNPLA7	2.25	9.37E-09
RP11-539I5.1	2.51	1.00E-08
PAX5	2.22	1.01E-08
MDH1B	2.58	1.04E-08
RP11-90E5.1	3.19	1.08E-08
ZNF717	2.49	1.10E-08
ZNF805	2.46	1.12E-08
CTD-2240E14.4	2.74	1.13E-08
ITGB2-AS1	2.17	1.21E-08
SCARNA21	2.30	1.24E-08
LIFR	2.05	1.31E-08
TXK	2.39	1.35E-08
RBM12B-AS1	2.53	1.54E-08
AC004824.2	3.28	1.63E-08
GLRA3	3.70	1.79E-08
AC005086.2	3.21	1.80E-08
ARAP1	2.56	1.92E-08
RNU11	2.03	1.95E-08
RAB39B	2.70	2.25E-08
RP11-474P2.4	2.67	2.34E-08

LILRA4	3.05	2.44E-08
LAT	2.30	2.55E-08
GPR15	3.62	2.69E-08
SLC25A24P1	2.82	2.78E-08
TCL1A	2.54	2.86E-08
AC002480.5	2.30	2.88E-08
RP11-554D14.3	3.32	2.91E-08
FAM129C	2.37	3.40E-08
KEL	2.66	3.66E-08
APOA1-AS	2.43	3.66E-08
RP11-160E2.21	2.60	3.76E-08
RP11-102L12.2	2.64	3.78E-08
WDR81	2.71	3.90E-08
ENPP7P10	3.33	4.29E-08
CCDC22	2.44	4.38E-08
CTD-2328D6.1	2.14	4.53E-08
RP11-265P11.2	2.98	5.27E-08
KIAA0125	2.10	5.48E-08
KRTCAP2	2.13	5.51E-08
HABP4	2.02	5.54E-08
ARHGAP15	2.06	5.58E-08
FCRL3	2.44	5.63E-08
RPL29P14	2.47	5.93E-08
AC007386.3	2.29	6.07E-08
RP11-665A22.1	2.97	6.22E-08
LAX1	2.10	6.22E-08
KLHDC7B	2.28	6.29E-08
SCGB3A1	2.62	6.35E-08
CAMK4	2.01	6.64E-08
RNU1-59P	2.51	6.66E-08
RP11-94H18.1	2.65	7.03E-08
VSIG1	2.43	7.22E-08
AP000439.1	2.84	7.46E-08

AC010468.2	3.07	7.48E-08
AC018755.17	3.77	7.55E-08
U62631.5	2.40	7.97E-08
LINC00528	2.24	8.24E-08
RP11-85G21.3	3.00	8.37E-08
AC005162.4	3.26	8.69E-08
RP11-126O1.4	2.34	8.75E-08
POM121L9P	2.58	8.83E-08
CD79B	2.31	9.37E-08
AC013264.2	2.79	9.37E-08
CTD-3105H18.14	2.23	9.61E-08
CETP	2.45	9.80E-08
IL21R	2.25	1.01E-07
SSTR2	2.68	1.01E-07
CD6	2.07	1.07E-07
TSPAN33	2.19	1.07E-07
RPL21P44	2.21	1.07E-07
RP11-796E2.4	2.28	1.10E-07
RP11-523O18.7	2.18	1.11E-07
TRIM73	2.36	1.15E-07
TLR7	2.23	1.15E-07
RP5-1074L1.1	2.47	1.20E-07
RN7SL337P	2.92	1.22E-07
RP11-1094M14.4	2.81	1.27E-07
KRT16	2.54	1.27E-07
DNASE1L2	2.74	1.30E-07
AL139328.1	2.60	1.31E-07
RPL4P5	2.79	1.31E-07
LRRC4C	2.40	1.33E-07
ZAP70	2.15	1.34E-07
HNRNPA3P3	2.65	1.36E-07
AC092652.1	2.92	1.44E-07
AC006129.2	2.32	1.46E-07

DKFZP779L1853	2.42	1.47E-07
FCER2	2.36	1.48E-07
CD37	2.18	1.51E-07
RN7SL842P	3.26	1.51E-07
MYO1G	2.14	1.56E-07
IGLV6-57	2.09	1.57E-07
CTD-3214H19.6	2.76	1.59E-07
AC008850.3	2.01	1.59E-07
RPL23AP7	3.00	1.59E-07
CNR1	2.25	1.66E-07
RP11-521B24.3	2.72	1.74E-07
AC004791.2	2.45	1.83E-07
SPIB	2.16	1.85E-07
RENBP	2.24	1.89E-07
TPBGL	2.90	1.98E-07
RP11-20B24.7	2.07	2.00E-07
OR1C1	2.98	2.16E-07
KRT1	2.43	2.17E-07
HIST1H2BG	2.18	2.26E-07
COLQ	2.01	2.33E-07
RP11-284N8.3	2.00	2.37E-07
AL137059.1	2.72	2.42E-07
RP11-507J18.2	2.06	2.43E-07
CCL14	3.49	2.48E-07
RP11-830F9.7	3.29	2.51E-07
PNMA3	2.11	2.51E-07
UPP2	3.25	2.51E-07
NCF1C	2.81	2.60E-07
LINC00612	3.59	2.62E-07
RP11-372E1.6	3.16	2.63E-07
REG1B	2.12	2.68E-07
PIPOX	2.21	2.72E-07
RP1-256G22.2	2.57	2.77E-07

TNFRSF13C	3.09	2.89E-07
C12orf39	3.47	2.94E-07
ADCY7	2.06	2.99E-07
MICU3	2.23	3.07E-07
QRICH2	2.36	3.24E-07
HSPE1P13	2.37	3.31E-07
AC106801.1	2.24	3.39E-07
RP11-407N8.4	2.42	3.49E-07
RP11-362K14.5	2.26	3.53E-07
NPM1P25	2.66	3.63E-07
OR5K3	3.59	3.64E-07
RP11-168J18.6	2.95	3.77E-07
RP11-429B14.1	2.17	4.01E-07
AC006116.20	2.38	4.08E-07
AGAP2	2.24	4.14E-07
RGS9	2.85	4.20E-07
RP11-1E22.1	2.47	4.22E-07
C12orf77	3.07	4.63E-07
TSPAN32	2.27	4.65E-07
RP11-72K17.2	3.41	4.75E-07
RIC3	2.27	4.96E-07
PZP	2.29	4.98E-07
RP11-118B18.1	2.54	5.15E-07
RP11-564A8.4	2.97	5.15E-07
RNU1-28P	2.78	5.25E-07
SLC26A1	2.25	5.31E-07
RP11-730A19.5	3.33	5.47E-07
RP11-6B6.3	3.00	5.63E-07
RP11-118F2.2	2.70	6.06E-07
TCHH	2.18	6.08E-07
XXbac-BPG154L12.4	3.08	6.17E-07
SLC26A7	2.68	6.20E-07
ZNF839P1	2.53	6.25E-07

AC097713.5	2.86	6.39E-07
RP11-202G11.2	2.41	6.57E-07
PLCXD3	2.36	6.61E-07
RP4-813F11.4	2.57	6.87E-07
RP11-430C7.5	2.34	7.11E-07
RP11-263C24.1	2.11	7.16E-07
GOLGA6L3	2.92	7.20E-07
RP11-44M6.3	2.23	7.96E-07
RP11-484D2.3	2.23	8.57E-07
CCDC62	2.89	8.66E-07
RP5-1050D4.5	2.88	8.72E-07
RN7SL849P	2.31	8.78E-07
AF131215.3	2.14	8.79E-07
RP11-805I24.1	2.50	8.86E-07
RN7SL377P	2.29	9.41E-07
CTD-3032J10.2	2.06	9.80E-07
OR10A4	2.07	9.82E-07
CLEC4G	2.57	9.89E-07
LYVE1	2.86	1.02E-06
THEMIS	2.07	1.04E-06
LL21NC02-1C16.2	2.90	1.06E-06
RP11-386I8.6	3.03	1.08E-06
RP11-711M9.1	2.30	1.12E-06
VENTXP1	2.79	1.21E-06
OGFR-AS1	2.68	1.26E-06
RP11-732A19.1	2.08	1.26E-06
CHI3L2	2.04	1.29E-06
LINC00877	2.72	1.32E-06
CTD-2510F5.4	2.59	1.38E-06
ESRRA	2.94	1.43E-06
POMC	2.84	1.43E-06
RP11-424C20.2	2.72	1.52E-06
CTD-2165H16.4	2.15	1.54E-06

P2RY10	2.33	1.58E-06
SCARNA17	2.17	1.58E-06
RNU6-1093P	2.46	1.61E-06
RNY4P18	2.70	1.61E-06
ZDHHC4	2.06	1.65E-06
ZNF737	2.50	1.69E-06
RP11-613M10.9	2.34	1.72E-06
LINC00864	2.90	1.74E-06
GP1BA	2.69	1.74E-06
CTD-2193P3.2	2.46	1.92E-06
FLT3	2.47	1.94E-06
GZMM	2.59	1.95E-06
PPATP1	3.12	1.97E-06
IGHV3-53	2.56	1.97E-06
TIFAB	2.48	1.98E-06
IDO2	2.74	2.08E-06
RP11-488C13.1	2.34	2.24E-06
CCDC152	2.07	2.31E-06
GLYCTK-AS1	3.20	2.46E-06
LINC00883	2.21	2.47E-06
CD33	2.08	2.48E-06
TSR3	2.18	2.50E-06
RNA5SP195	2.17	2.66E-06
RP11-876N24.4	2.05	2.68E-06
TTI2	2.29	2.70E-06
TMEM217	2.50	2.74E-06
AC104653.1	2.29	2.87E-06
RP11-330M2.4	2.06	2.99E-06
C1orf228	2.16	3.05E-06
EBF2	2.92	3.06E-06
CD27	2.12	3.15E-06
AC009950.2	2.04	3.15E-06
LIX1	2.25	3.16E-06

RP11-727F15.9	2.49	3.17E-06
OR7G1	2.14	3.20E-06
SLC22A11	2.20	3.34E-06
RP11-495K9.5	2.24	3.40E-06
CNTNAP1	2.09	3.45E-06
CTD-2516F10.2	2.40	3.45E-06
MUC16	2.14	3.49E-06
RNU6-1111P	2.33	3.50E-06
RP4-593C16.3	2.35	3.89E-06
RN7SL274P	3.13	4.02E-06
RP11-570L15.2	2.53	4.03E-06
RP11-152L20.3	2.50	4.03E-06
RPL13AP2	2.82	4.08E-06
AC092295.7	2.11	4.12E-06
ANKHD1-EIF4EBP3	2.44	4.21E-06
APC2	2.59	4.24E-06
RNA5-8SP6	2.33	4.38E-06
SYCE2	2.59	4.51E-06
AC114763.1	2.55	4.54E-06
RN7SKP110	2.96	4.60E-06
PANK4	2.03	4.67E-06
SCARNA20	2.79	4.67E-06
ART4	2.06	4.91E-06
FITM1	2.29	4.93E-06
ZNF671	2.77	4.95E-06
RP5-1125A11.4	2.86	5.03E-06
LINGO3	2.18	5.17E-06
AL353898.3	2.53	5.17E-06
RP11-295M18.6	3.05	5.18E-06
RP11-707M3.3	2.12	5.21E-06
GAPT	2.74	5.23E-06
AC006129.4	2.05	5.27E-06
RP11-423F24.3	3.00	5.27E-06

HIST1H2AK	2.20	5.31E-06
LINC00482	3.19	5.39E-06
TENM1	2.03	5.40E-06
RP11-177N22.3	2.69	5.78E-06
CORO1A	2.08	5.78E-06
EMX2	2.54	6.06E-06
SEPT10P1	2.46	6.10E-06
RP11-325F22.2	2.78	6.88E-06
CLEC4F	3.13	6.90E-06
HIST1H4L	2.23	6.90E-06
CACNA1I	2.68	7.00E-06
IFNG	2.56	7.17E-06
HMGN2P28	2.42	7.17E-06
TMEM108	2.20	7.46E-06
RP11-152P17.2	2.02	7.58E-06
AC116050.1	2.29	7.64E-06
TRAJ36	3.16	7.66E-06
RP11-1057B6.1	2.67	7.73E-06
CTB-167G5.5	2.49	7.86E-06
RP3-521E19.2	2.24	8.63E-06
FAM216B	2.05	8.73E-06
OXCT2P1	2.53	8.79E-06
GALNTL6	3.04	8.97E-06
NCF1B	2.13	9.08E-06
AC092291.2	3.31	9.41E-06
RP4-604G5.1	2.28	9.93E-06

Table A.5: GSEA results for analysis of MSigDB Hallmark gene sets enriched in LNM vs primary CRC

MSigDB	NES	p-value

HALLMARK_ALLOGRAFT_REJECTION	2.77	0.00E+00
HALLMARK_INTERFERON_GAMMA_RESPONSE	2.50	0.00E+00
HALLMARK_IL2_STAT5_SIGNALING	2.11	0.00E+00
HALLMARK_INTERFERON_ALPHA_RESPONSE	2.10	0.00E+00
HALLMARK_COMPLEMENT	2.08	0.00E+00
HALLMARK_IL6_JAK_STAT3_SIGNALING	2.07	0.00E+00
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.06	0.00E+00
HALLMARK_INFLAMMATORY_RESPONSE	2.00	0.00E+00
HALLMARK_KRAS_SIGNALING_UP	1.66	0.00E+00
HALLMARK_PI3K_AKT_MTOR_SIGNALING	1.36	4.62E-02
HALLMARK_ANGIOGENESIS	1.27	1.42E-01
HALLMARK_HYPOXIA	1.20	1.25E-01
HALLMARK_KRAS_SIGNALING_DN	1.16	2.01E-01
HALLMARK_MITOTIC_SPINDLE	1.12	2.43E-01
HALLMARK_ANDROGEN_RESPONSE	1.11	2.64E-01
HALLMARK_UV_RESPONSE_DN	1.09	2.89E-01
HALLMARK_APOPTOSIS	1.09	3.00E-01
HALLMARK_COAGULATION	1.06	3.60E-01
HALLMARK_SPERMATOGENESIS	1.03	4.13E-01
HALLMARK_UV_RESPONSE_UP	1.01	4.47E-01
HALLMARK_ESTROGEN_RESPONSE_LATE	1.00	4.57E-01
HALLMARK_HEME_METABOLISM	0.93	5.98E-01
HALLMARK_APICAL_SURFACE	0.91	6.01E-01
HALLMARK_TGF_BETA_SIGNALING	0.90	6.34E-01
HALLMARK_APICAL_JUNCTION	0.89	7.22E-01
HALLMARK_ESTROGEN_RESPONSE_EARLY	0.83	8.50E-01
HALLMARK_P53_PATHWAY	0.80	8.91E-01
HALLMARK_REACTIVE_OXYGEN_SPECIES	0.79	8.11E-01
HALLMARK_XENOBIOTIC_METABOLISM	0.79	8.80E-01
HALLMARK_G2M_CHECKPOINT	0.77	9.31E-01
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	0.74	9.23E-01
HALLMARK_DNA_REPAIR	0.70	9.76E-01
HALLMARK_PROTEIN_SECRETION	0.69	9.61E-01

HALLMARK_HEDGEHOG_SIGNALING	0.68	9.46E-01
HALLMARK_MTORC1_SIGNALING	0.62	9.98E-01
HALLMARK_E2F_TARGETS	0.54	1.00E+00
HALLMARK_MYOGENESIS	-1.62	0.00E+00
HALLMARK_OXIDATIVE_PHOSPHORYLATION	-1.38	7.46E-03
HALLMARK_EMT	-1.33	2.26E-02
HALLMARK_MYC_TARGETS_V2	-1.11	2.71E-01
HALLMARK_GLYCOLYSIS	-1.11	1.39E-01
HALLMARK_CHOLESTEROL_HOMEOSTASIS	-1.05	3.29E-01
HALLMARK_NOTCH_SIGNALING	-1.04	3.75E-01
HALLMARK_BILE_ACID_METABOLISM	-1.00	4.68E-01
HALLMARK_WNT_BETA_CATENIN_SIGNALING	-0.93	5.93E-01
HALLMARK_FATTY_ACID_METABOLISM	-0.84	9.00E-01
HALLMARK_ADIPOGENESIS	-0.82	9.79E-01
HALLMARK_PANCREAS_BETA_CELLS	-0.72	8.86E-01
HALLMARK_PEROXISOME	-0.71	9.72E-01
HALLMARK_MYC_TARGETS_V1	-0.46	1.00E+00

Table A.6: top 250 differentially expressed genes in the lung metastatic lesions of LS174T-SGK1 vs LS174T-wt.

Gene	baseMean	log2FoldChange	pvalue
PLA2G2A	236.2059782	-9.778063425	3.57E-15
OLFM4	498.1791147	-8.61095435	2.94E-11
CTD-2354A18.1	65.3235788	-8.751079667	1.45E-09
AKR1B1	127.0115564	-6.773981769	2.26E-08
RPL9	107.7783723	-4.667452261	3.52E-08
NUPR1	37.29265543	-7.458190069	7.73E-08
PAGE1	109.1316867	-7.115387232	3.63E-07
EPHB2	72.3888278	-6.290123583	1.07E-06
MUC6	18.7835455	-5.971606942	1.34E-06

CNN3	142.7027098	-6.481181272	1.94E-06
DNAJC15	32.31364907	-4.926909511	1.94E-06
RRP1B	87.68071413	-4.606180883	1.86E-06
BMP4	134.9526007	-4.574268418	2.49E-06
SPINK1	55.15411582	5.892764501	2.57E-06
AP4B1	126.9130927	-6.349215702	3.51E-06
TSPAN1	107.236456	4.112654799	9.52E-06
SNX9	47.26560719	-4.049553465	1.76E-05
MICU2	51.89544652	-4.011649484	2.04E-05
GTSF1	53.54366826	-4.207360995	2.37E-05
ABHD12B	68.05911532	-5.899439313	3.24E-05
ARSJ	83.86623913	-5.200172579	3.60E-05
FUT3	28.08417584	-6.244619409	3.75E-05
IGFL2	379.4625971	-6.701770592	3.84E-05
TPTEP1	75.06751936	-4.188597159	3.78E-05
FASTK	72.80960407	-5.771105036	4.11E-05
CCND2	166.5832645	-5.039658308	4.61E-05
KRT20	49.84954849	4.523285317	4.77E-05
AL589743.1	12.45992449	-6.147033348	5.62E-05
APOBEC3C	51.00909822	-5.06524523	5.95E-05
NT5DC1	49.12846768	-3.711087927	6.09E-05
PI3	60.47730064	5.600293198	6.48E-05
EXOSC4	36.56625383	-5.098426104	6.69E-05
MTND5P30	109.1405382	-8.120216867	9.74E-05
GADD45A	69.74814202	-4.398725006	0.000109742
MYH7	54.64143322	-5.774935804	0.000110588
AC002467.7	24.89938205	-7.308635474	0.000118338
GALNT5	114.0964708	-5.607230249	0.00012306
EGFR	90.84303334	-4.740046769	0.000130731
ELOVL5	81.34643086	-7.415715549	0.000133637
SELK	131.3540448	-3.069120986	0.000136164
PTPRR	24.45057627	-5.735583241	0.000147286
JMY	96.94822846	-5.459403037	0.000163373

MBNL3	138.3180763	-5.968027229	0.000194575
HECTD1	231.2085203	-4.180368855	0.00021357
TRIM33	65.44001414	-3.949870825	0.000229775
MAMDC2-AS1	15.91606871	-5.122395918	0.000245878
STAG3L1	14.78549167	-4.953830418	0.000252391
TRAT1	12.04191714	-6.028652196	0.000287097
AGR2	549.7909257	-3.210439891	0.000323362
DLC1	42.83101618	-5.761279211	0.000323775
MT1F	32.4172607	-5.840269372	0.00032082
ONECUT3	16.83281737	-5.552396549	0.000320474
ILKAP	33.28729044	-4.150524856	0.000337999
NCAPD2P1	52.45588193	-6.961069201	0.000338802
CTA-293F17.1	68.68067321	-5.99819652	0.000350583
MMP7	426.4280898	4.644196391	0.000383592
CTD-2572N17.1	22.06089969	-7.303562485	0.000410644
ST3GAL1	90.50013285	4.007363592	0.000413968
RP11-297P16.4	71.55025057	-5.964948637	0.000434674
MAGEC2	8.274252739	-5.340674466	0.000446453
INO80C	44.05092964	-3.383013592	0.000459816
CTSE	271.8159719	4.064221365	0.000469949
MT1L	10.71354022	-4.992285651	0.000537099
RP9	25.05283068	-4.479423246	0.000528957
TXLNA	56.0337675	3.725348653	0.000538021
VSIG2	21.80226907	4.580270076	0.000514031
CDA	41.47587866	-4.924634663	0.000549724
NBPF1	27.17346753	-3.209455496	0.000586128
RHPN2	24.1853441	4.158107272	0.00058256
IRG1	10.54586111	-6.94319355	0.000598738
CCND1	23.78088945	-3.665798528	0.000624483
CHMP3	48.46333819	4.088218439	0.000622229
KRTAP19-5	17.27293056	-7.095462356	0.000610957
ARPP19	240.2734004	-3.524616692	0.00067865
BPHL	28.34422859	-3.890643359	0.000663383

DSP	381.4699682	-2.844140158	0.000701362
DUSP13	24.20546815	-6.457831285	0.000690016
GPR125	73.71385659	-4.173165542	0.000720578
MGLL	57.84511416	4.452242873	0.000688849
MSX2	26.62536582	-6.36973827	0.000716941
NMUR2	21.69993722	6.720097316	0.000708158
SLC38A9	31.17136693	6.855831111	0.000698278
TM4SF4	76.0820155	4.696084213	0.000712408
KRT8P5	7.720848303	-5.071948368	0.000774938
SLC41A3	121.4146461	-4.49681219	0.000773833
MT1G	36.06943751	-3.819426806	0.000799678
CTD-2307P3.1	14.68760428	-6.287366899	0.000810541
ABHD14A	21.89701669	-3.712185629	0.000823273
CD74	39.49805013	-4.853272857	0.00083356
CA2	21.44132435	5.119747551	0.000866659
CR381653.1	11.57593827	-6.845029526	0.000881521
MRPL16	72.8106356	-4.254426026	0.000883626
FBXL5	82.14161441	-3.67322838	0.000905074
MROH1	27.64125117	3.646762821	0.000896338
BCAT1	45.84800283	-5.590833704	0.000958465
FHL2	140.3907553	-4.586945248	0.000954751
MARVELD2	60.10892997	-3.515833674	0.000963437
NINJ1	31.51474864	-5.4336538	0.000962342
PHLDB2	43.94203393	3.352336511	0.000941284
RAPGEF2	94.70168503	-3.13170209	0.000974536
HNF1B	14.47163411	6.448748053	0.00104264
AP000525.9	26.6934104	-4.773760138	0.001093137
FAM98C	21.14322823	6.583368107	0.001084204
POLG	17.32244936	6.505693674	0.001101569
CTD-2525I3.5	8.980604342	-6.207033453	0.001215293
EREG	361.9222417	-3.524773739	0.001190207
PALLD	23.44444541	-3.086922191	0.001181722
RPS27L	94.18774596	-2.551013207	0.001188614

RPUSD2	20.29734454	6.532110434	0.001218809
ZNF395	20.27249616	-4.314788303	0.001222375
PMP22	52.24064577	-4.373808375	0.001269934
PABPC1L	19.05607007	6.490288911	0.001296286
FAM65A	57.25229849	-5.097704507	0.001343368
KIAA0586	36.66018119	-4.173875151	0.001384538
AGO4	23.59778832	-4.46289337	0.001410991
ERAP2	16.61603915	6.4049149	0.001452611
NF1	138.1102292	-2.433368787	0.001503091
HS6ST3	5.416062461	-6.402142105	0.001563994
NRCAM	32.70413206	-4.258107611	0.001575605
RPS17	21.30146854	-5.200055742	0.001554211
FAM134A	12.32951545	6.263738236	0.00159042
ISYNA1	19.93857947	-4.459021683	0.001636999
CHID1	143.5151938	-3.495355354	0.001681428
RNVU1-14	6.681953704	-6.441135741	0.001685482
MT1H	13.26919313	-5.685709614	0.001740786
CAPRIN2	31.66331758	3.381417411	0.001775452
CTNNB1	368.9145665	-2.766203255	0.00180646
DNMT3A	11.47099067	6.188518467	0.001796453
FAHD2A	20.31432925	5.246835556	0.001904544
SUGP2	33.06526556	-5.238167085	0.001927232
ABHD2	175.9586509	-3.092432145	0.002006065
PTPRS	20.07741536	-4.811048368	0.00203052
RP11-589F5.4	2.888882117	-5.197560784	0.002017606
SLC3A1	16.36397573	6.28915593	0.002038497
WASF3-AS1	18.41783416	6.330310799	0.002034016
GS1-124K5.4	4.267621554	-6.233559341	0.002061784
REG4	608.676499	3.256937513	0.002077397
FAT1	140.8314511	-2.997860733	0.002095039
IKBKE	14.86926895	6.243931008	0.002112927
ACSS1	14.56026111	6.231836392	0.002145059
DUSP4	71.14739236	-4.402924013	0.002133469

PRKCG	17.17459043	6.284307084	0.002169947
AC068014.1	8.927593877	-6.425925871	0.002250552
FLAD1	20.04230024	3.241556851	0.00222656
IPO5	36.65947696	2.649519198	0.002206828
PHLDA1	157.4257367	-4.019633515	0.002265735
PRRG4	12.42391594	6.153807147	0.002249267
TAGLN	15.20818138	6.222247034	0.002306289
AMZ2P1	19.03935207	-5.718240987	0.002446776
CADPS	23.25011673	-4.415647114	0.002489436
DHX40	16.48998994	6.227510268	0.002457608
MTURN	21.06292199	-3.046898844	0.002487967
RP11-304C12.3	16.78644134	6.236822657	0.002434848
SPRY2	161.1701139	-3.63913086	0.002415712
TRIB3	37.57673517	-3.802683226	0.002488434
UBA6	28.58397796	4.135390915	0.002452353
ABL2	28.27640505	5.124184614	0.003300925
AC131025.8	10.08338245	5.922826767	0.003515095
ADD3	29.64115627	4.081962578	0.003630973
AL354822.1	7.000745373	-6.285896722	0.002747632
ALG11	13.09932013	6.027005954	0.003482181
ANKS1B	31.08297035	-4.106920069	0.003012862
ARL6IP5	325.4793455	-3.408481358	0.003715969
ATP1B1	371.1304876	-2.590912833	0.00341792
ATR	72.94639504	-4.168797857	0.002643888
BROX	151.1451791	3.246546667	0.003696722
C2orf78	15.80538946	6.153244335	0.002911838
C6orf222	18.37292646	5.082597596	0.003296924
C7orf65	9.175145926	-5.953482206	0.003238448
CAMSAP3	10.86703576	-5.836793547	0.002626848
CASP10	31.95829798	-3.813228233	0.003210162
CEP72	30.93370537	-5.695307139	0.003408651
CHPF	18.60778249	-4.176009295	0.00354651
CMTM8	15.8567643	-4.992487066	0.003149692

CNP	37.90757829	4.024193035	0.002782383
COMMD9	34.87690005	3.131753976	0.002883551
COX7A2	148.8559597	-2.267843631	0.003299326
CTD-2314B22.3	30.35079675	-4.768382524	0.003101723
CYCSP12	6.212776237	-6.210422295	0.003049265
DDAH1	44.43265409	2.728520729	0.002550448
DDAH2	8.185993544	5.816967749	0.003709022
DHPS	26.01631304	4.372993076	0.003571786
EIF4E	107.5692747	-2.536095837	0.002805731
FAM65B	14.21453073	6.041809073	0.003590483
GALNT14	30.86985707	3.920950586	0.002766856
GAN	25.90299725	4.658162344	0.002696901
GTF2A2	177.1324917	-2.350526383	0.00254336
HAL	15.91788742	6.133303898	0.003095302
HMGN2P21	2.403714146	-5.823426531	0.003493187
HSD3B7	126.5298233	-5.530993657	0.002637737
HULC	10.01367483	-5.759546564	0.002908852
ITGA6	149.8633562	-2.381951397	0.003272306
LEPREL1	32.02745394	-4.234156884	0.003600051
LINC00963	15.90298362	4.953354316	0.00312769
LMBR1	85.00936465	-4.399796407	0.00278692
MAP6D1	8.624792711	-5.892863985	0.003543725
MED15P1	24.79478749	-5.647304153	0.002870749
MORC3	12.02944997	6.027576874	0.00319992
MRPS10	146.9064864	-3.266567136	0.003643835
MT1M	12.15126187	-5.529360626	0.002739925
NPHS1	19.94342154	4.843661869	0.003330344
NTRK2	11.88896703	6.020177914	0.003224756
PEX12	11.94645484	6.034733994	0.003114286
PKD2L1	11.57611997	5.978515011	0.003535926
PNO1	8.291750878	5.834931047	0.00365928
PNPLA6	19.46071345	-5.232882298	0.002888255
PRX	16.75602835	6.148245629	0.003105807

RANBP6	25.83682194	-3.518008487	0.00332551
RGMB	47.98193904	-4.309321302	0.003287508
RP11-356M20.1	14.83445869	6.080596724	0.003356186
RP11-506K6.4	6.015783953	-6.18968349	0.003138578
RP11-651P23.2	6.336448986	-6.226452055	0.002973687
RP11-798K3.3	5.321006608	-5.266170545	0.00362081
RP11-983P16.4	19.12802261	5.126251184	0.003137677
RP4-680D5.4	6.354312854	-5.749473893	0.003432188
RRAD	6.212776237	-6.210422295	0.003049265
RUNX1	42.21274756	-4.37938179	0.002799159
SAMD12	10.99593159	4.603445162	0.002606295
SAMD5	36.54018068	-4.735557198	0.002903191
SETMAR	9.831510244	-5.351557195	0.003303656
SIAE	114.0234402	-4.067750055	0.003396346
SLC30A4	15.22265743	6.157228578	0.00278279
SLC35F6	88.51509656	3.205684083	0.002724613
SPATA18	20.60108229	-5.642840652	0.003425552
SPTBN1	667.0472829	-2.890026719	0.003691815
TXNDC15	34.29906079	-5.66838282	0.003204398
USP31	110.1957884	-4.864083583	0.00329488
VPS25	110.3187282	-3.665619485	0.003626375
YARS2	9.749727795	5.967155792	0.002895354
ZEB1	11.86190158	-3.849028361	0.003524982
ZNF83	27.49505587	3.59674684	0.003399877
RAB17	12.63677917	5.989411333	0.003732722
NCOA7	60.52354289	-3.749586712	0.003802486
ARSF	13.37424925	-4.577997965	0.003998424
AVIL	14.0305773	5.995608676	0.004014229
BNC1	13.69986637	5.984762192	0.00404927
EFCAB2	13.91474118	5.993910282	0.004007656
IFT172	12.11398662	4.657615497	0.003833407
LETMD1	62.9188435	-3.720824792	0.003945005
LPCAT4	26.8478332	4.434661252	0.00403272

PIEZO1	31.32642361	-4.383431544	0.003916308
RAC2	5.566343718	-5.629228742	0.004026346
RBMS2	8.810177147	5.818043868	0.003992498
RP11-18B16.1	13.77283969	5.987525169	0.004047765
TMEM106A	13.55938343	5.996648406	0.003890995
TUBGCP6	15.14391795	6.016601334	0.004040336
TWSG1	32.56462021	2.989968265	0.004031937
VN2R1P	8.214856152	-5.827133731	0.003980284
ZSCAN31	73.60189639	-5.153901027	0.004079307
NME3	6.190982667	-3.749197989	0.004100674

Bibliography

[1] Michal Sheffer, Manny D. Bacolod, Or Zuk, Sarah F. Giardina, Hanna Pincas, Francis Barany, Philip B. Paty, William L. Gerald, Daniel A. Notterman, and Eytan Domany. Association of survival and disease progression with chromosomal instability: A genomic exploration of colorectal cancer. *Proceedings of the National Academy of Sciences*, 106(17):7131–7136, April 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0902232106.

[2] ONS. Overview of the UK population - Office for National Statistics. `www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populatio` 2016.

[3] Rachael Harker. NHS funding and expenditure. 2012.

[4] Melissa M. Center, Ahmedin Jemal, and Elizabeth Ward. International Trends in Colorectal Cancer Incidence Rates. *Cancer Epidemiology and Prevention Biomarkers*, 18(6):1688–1694, June 2009. ISSN 1055-9965, 1538-7755. doi: 10.1158/1055-9965.EPI-09-0090.

[5] CRUK. Bowel cancer statistics. `www.cancerresearchuk.org/health-professional/cance` 2015-05-14T14:57:01+01:00.

[6] Driss Ait Ouakrim, Cécile Pizot, Magali Boniol, Matteo Malvezzi, Mathieu Boniol, Eva Negri, Maria Bota, Mark A. Jenkins, Harry Bleiberg, and Philippe

Autier. Trends in colorectal cancer mortality in Europe: Retrospective analysis of the WHO mortality database. *BMJ*, 351:h4970, October 2015. ISSN 1756-1833. doi: 10.1136/bmj.h4970.

[7] Melina Arnold, Mónica S. Sierra, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, pages gutjnl–2015–310912, January 2016. ISSN , 1468-3288. doi: 10.1136/gutjnl-2015-310912.

[8] Axel Grothey and Daniel J. Sargent. FOLFOX for stage II colon cancer? A commentary on the recent FDA approval of oxaliplatin for adjuvant therapy of stage III colon cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 23(15):3311–3313, May 2005. ISSN 0732-183X. doi: 10.1200/JCO.2005.11.691.

[9] Wendy S Atkin, Rob Edwards, Ines Kralj-Hans, Kate Wooldrage, Andrew R Hart, John MA Northover, D Max Parkin, Jane Wardle, Stephen W Duffy, and Jack Cuzick. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: A multicentre randomised controlled trial. *The Lancet*, 375(9726):1624–1633, May 2010. ISSN 0140-6736. doi: 10.1016/S0140-6736(10)60551-X.

[10] Reid A. Phelps, Stephanie Chidester, Somaye Dehghanizadeh, Jason Phelps, Imelda T. Sandoval, Kunal Rai, Talmage Broadbent, Sharmistha Sarkar, Randall W. Burt, and David A. Jones. A Two-Step Model for Colon Adenoma Initiation and Progression Caused by APC Loss. *Cell*, 137(4):623–634, May 2009. ISSN 0092-8674. doi: 10.1016/j.cell.2009.02.037.

[11] T. Takayama, M. Ohi, T. Hayashi, K. Miyanishi, A. Nobuoka, T. Nakajima, T. Satoh, R. Takimoto, J. Kato, S. Sakamaki, and Y. Niitsu. Analysis of K-ras,

APC, and beta-catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis. *Gastroenterology*, 121(3):599–611, September 2001. ISSN 0016-5085.

[12] Ildiko Mesteri, Günther Bayer, Jochen Meyer, David Capper, Sebastian F. Schoppmann, Andreas von Deimling, and Peter Birner. Improved molecular classification of serrated lesions of the colon by immunohistochemical detection of BRAF V600E. *Modern Pathology*, 27(1):135–144, January 2014. ISSN 0893-3952. doi: 10.1038/modpathol.2013.126.

[13] Shi Yang, Francis A. Farraye, Charline Mack, Oksana Posnik, and Michael J. O'Brien. BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: Relationship to histology and CpG island methylation status. *The American Journal of Surgical Pathology*, 28(11):1452–1459, November 2004. ISSN 0147-5185.

[14] Peter Zauber, Stephen Marotta, and Marlene Sabbath-Solitare. KRAS gene mutations are more common in colorectal villous adenomas and in situ carcinomas than in carcinomas. *International Journal of Molecular Epidemiology and Genetics*, 4(1):1–10, March 2013. ISSN 1948-1756.

[15] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, March 2011. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2011.02.013.

[16] The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407):330–337, July 2012. ISSN 0028-0836. doi: 10.1038/nature11252.

[17] Daniel L Worthley and Barbara A Leggett. Colorectal Cancer: Molecular Fea-

tures and Clinical Opportunities. *The Clinical Biochemist Reviews*, 31(2):31–38, May 2010. ISSN 0159-8090.

[18] Xia Li, Xiaoping Yao, Yibaina Wang, Fulan Hu, Fan Wang, Liying Jiang, Yupeng Liu, Da Wang, Guizhi Sun, and Yashuang Zhao. MLH1 Promoter Methylation Frequency in Colorectal Cancer Patients and Related Clinicopathological and Molecular Features. *PLOS ONE*, 8(3):e59064, March 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0059064.

[19] C. Richard Boland and Ajay Goel. Microsatellite Instability in Colorectal Cancer. *Gastroenterology*, 138(6):2073–2087.e3, June 2010. ISSN 0016-5085. doi: 10.1053/j.gastro.2009.12.064.

[20] Justin Guinney, Rodrigo Dienstmann, Xin Wang, Aurélien de Reyniès, Andreas Schlicker, Charlotte Soneson, Laetitia Marisa, Paul Roepman, Gift Nyamundanda, Paolo Angelino, Brian M. Bot, Jeffrey S. Morris, Iris M. Simon, Sarah Gerster, Evelyn Fessler, Felipe De Sousa E. Melo, Edoardo Missiaglia, Hena Ramay, David Barras, Krisztian Homicsko, Dipen Maru, Ganiraju C. Manyam, Bradley Broom, Valerie Boige, Beatriz Perez-Villamil, Ted Laderas, Ramon Salazar, Joe W. Gray, Douglas Hanahan, Josep Tabernero, Rene Bernards, Stephen H. Friend, Pierre Laurent-Puig, Jan Paul Medema, Anguraj Sadanandam, Lodewyk Wessels, Mauro Delorenzi, Scott Kopetz, Louis Vermeulen, and Sabine Tejpar. The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, advance online publication, October 2015. ISSN 1078-8956. doi: 10.1038/nm.3967.

[21] Marco Gerlinger, Andrew J. Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, Ignacio Varela, Benjamin Phillimore, Sharmin Begum, Neil Q. McDonald, Adam Butler, David Jones, Keiran Raine, Calli Latimer, Clau-

dio R. Santos, Mahrokh Nohadani, Aron C. Eklund, Bradley Spencer-Dene, Graham Clark, Lisa Pickering, Gordon Stamp, Martin Gore, Zoltan Szallasi, Julian Downward, P. Andrew Futreal, and Charles Swanton. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*, 366(10):883–892, March 2012. ISSN 0028-4793. doi: 10.1056/NEJMoa1113205.

[22] Jelena Urosevic, Xabier Garcia-Albéniz, Evarist Planet, Sebastián Real, María Virtudes Céspedes, Marc Guiu, Esther Fernandez, Anna Bellmunt, Sylwia Gawrzak, Milica Pavlovic, Ramon Mangues, Ignacio Dolado, Francisco M. Barriga, Cristina Nadal, Nancy Kemeny, Eduard Batlle, Angel R. Nebreda, and Roger R. Gomis. Colon cancer cells colonize the lung from established liver metastases through p38 MAPK signalling and PTHLH. *Nature Cell Biology*, 16(7):685–694, July 2014. ISSN 1465-7392. doi: 10.1038/ncb2977.

[23] Peter Gassmann and Joerg Haier. The tumor cell-host organ interface in the early onset of metastatic organ colonisation. *Clinical & Experimental Metastasis*, 25(2):171–181, 2008. ISSN 0262-0898. doi: 10.1007/s10585-007-9130-6.

[24] Vassiliki L. Tsikitis, David W. Larson, Marianne Huebner, Christine M. Lohse, and Patricia A. Thompson. Predictors of recurrence free survival for patients with stage II and III colon cancer. *BMC Cancer*, 14:336, 2014. ISSN 1471-2407. doi: 10.1186/1471-2407-14-336.

[25] Matias Riihimäki, Akseli Hemminki, Jan Sundquist, and Kari Hemminki. Patterns of metastasis in colon and rectal cancer. *Scientific Reports*, 6:29765, July 2016. ISSN 2045-2322. doi: 10.1038/srep29765.

[26] Alicia S. Chung, John Lee, and Napoleone Ferrara. Targeting the tumour

vasculature: Insights from physiological angiogenesis. *Nature Reviews Cancer*, 10(7):505–514, July 2010. ISSN 1474-175X. doi: 10.1038/nrc2868.

[27] Ugur Eskiocak, Sang Bum Kim, Peter Ly, Andres I. Roig, Sebastian Biglione, Kakajan Komurov, Crystal Cornelius, Woodring E. Wright, Michael A. White, and Jerry W. Shay. Functional Parsing of Driver Mutations in the Colorectal Cancer Genome Reveals Numerous Suppressors of Anchorage-Independent Growth. *Cancer Research*, 71(13):4359–4365, July 2011. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-11-0794.

[28] Peter Ly, Ugur Eskiocak, Chelsea R. Parker, Kenneth J. Harris, Woodring E. Wright, and Jerry W. Shay. RNAi screening of the human colorectal cancer genome identifies multifunctional tumor suppressors regulating epithelial cell invasion. *Cell Research*, 22(11):1605–1608, November 2012. ISSN 1001-0602. doi: 10.1038/cr.2012.140.

[29] Viktor H. Koelzer and Alessandro Lugli. The Tumor Border Configuration of Colorectal Cancer as a Histomorphological Prognostic Indicator. *Frontiers in Oncology*, 4, February 2014. ISSN 2234-943X. doi: 10.3389/fonc.2014.00029.

[30] Raghu Kalluri and Robert A. Weinberg. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation*, 119(6):1420–1428, June 2009. ISSN 0021-9738. doi: 10.1172/JCI39104.

[31] Zhi-Yu Song, Zu-Hua Gao, and Xian-Jun Qu. A review of CXCR4/CXCL12 axis in colorectal cancer. *Biomedicine & Aging Pathology*, 4(3):285–290, July 2014. ISSN 2210-5220. doi: 10.1016/j.biomag.2014.06.001.

[32] Anan H. Said, Jean-Pierre Raufman, and Guofeng Xie. The Role of Matrix Metalloproteinases in Colorectal Cancer. *Cancers*, 6(1):366–375, February 2014. ISSN 2072-6694. doi: 10.3390/cancers6010366.

[33] Hiroyuki Yamamoto, Fumio Itoh, Yuji Hinod, and Kohzoh Imai. Suppression of matrilysin inhibits colon cancer cell invasion in vitro. *International Journal of Cancer*, 61(2):218–222, April 1995. ISSN 1097-0215. doi: 10.1002/ijc.2910610213.

[34] U. Bork, N. N. Rahbari, S. Schölch, C. Reissfelder, C. Kahlert, M. W. Büchler, J. Weitz, and M. Koch. Circulating tumour cells and outcome in non-metastatic colorectal cancer: A prospective study. *British Journal of Cancer*, 112(8):1306–1313, April 2015. ISSN 0007-0920. doi: 10.1038/bjc.2015.88.

[35] Fanny Grillet, Elsa Bayet, Olivia Villeronce, Luke Zappia, Ebba Louise Lagerqvist, Sebastian Lunke, Emmanuelle Charafe-Jauffret, Kym Pham, Christina Molck, Nathalie Rolland, Jean François Bourgaux, Michel Prudhomme, Claire Philippe, Sophie Bravo, Jean Christophe Boyer, Lucile Canterel-Thouennon, Graham Roy Taylor, Arthur Hsu, Jean Marc Pascussi, Frédéric Hollande, and Julie Pannequin. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. *Gut*, pages gutjnl–2016–311447, July 2016. ISSN , 1468-3288.

[36] T. Lapidot, C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M. A. Caligiuri, and J. E. Dick. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464):645–648, February 1994. ISSN 0028-0836. doi: 10.1038/367645a0.

[37] Radoslaw Zagozdzon and Jakub Golab. Cancer stem cells in haematological malignancies. *Contemporary Oncology*, 19(1A):A1–A6, 2015. ISSN 1428-2526. doi: 10.5114/wo.2014.47127.

[38] Catherine A. O'Brien, Aaron Pollett, Steven Gallinger, and John E. Dick. A human colon cancer cell capable of initiating tumour growth in immunodefi-

cient mice. *Nature*, 445(7123):106–110, January 2007. ISSN 1476-4687. doi: 10.1038/nature05372.

[39] Anne T. Collins, Paul A. Berry, Catherine Hyde, Michael J. Stower, and Norman J. Maitland. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Research*, 65(23):10946–10951, December 2005. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-05-2018.

[40] Chenwei Li, David G. Heidt, Piero Dalerba, Charles F. Burant, Lanjing Zhang, Volkan Adsay, Max Wicha, Michael F. Clarke, and Diane M. Simeone. Identification of Pancreatic Cancer Stem Cells. *Cancer Research*, 67(3):1030–1037, February 2007. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-06-2030.

[41] Muhammad Al-Hajj, Max S. Wicha, Adalberto Benito-Hernandez, Sean J. Morrison, and Michael F. Clarke. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7):3983–3988, April 2003. ISSN 0027-8424. doi: 10.1073/pnas.0530291100.

[42] David Horst, Lydia Kriegl, Jutta Engel, Thomas Kirchner, and Andreas Jung. Prognostic Significance of the Cancer Stem Cell Markers CD133, CD44, and CD166 in Colorectal Cancer. *Cancer Investigation*, 27(8):844–850, January 2009. ISSN 0735-7907. doi: 10.1080/07357900902744502.

[43] Eric C. Anderson, Crystal Hessman, Trevor G. Levin, Marcus M. Monroe, and Melissa H. Wong. The Role of Colorectal Cancer Stem Cells in Metastatic Disease and Therapeutic Response. *Cancers*, 3(1):319–339, January 2011. doi: 10.3390/cancers3010319.

[44] Laurens G. Van der Flier, Jacob Sabates-Bellver, Irma Oving, Andrea Haege-

barth, Mariagrazia De Palo, Marcello Anti, Marielle E. Van Gijn, Saskia Suijkerbuijk, Marc Van de Wetering, Giancarlo Marra, and Hans Clevers. The Intestinal Wnt/TCF Signature. *Gastroenterology*, 132(2):628–632, February 2007. ISSN 0016-5085. doi: 10.1053/j.gastro.2006.08.039.

[45] Nick Barker, Johan H. van Es, Jeroen Kuipers, Pekka Kujala, Maaike van den Born, Miranda Cozijnsen, Andrea Haegebarth, Jeroen Korving, Harry Begthel, Peter J. Peters, and Hans Clevers. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*, 449(7165):1003–1007, October 2007. ISSN 0028-0836. doi: 10.1038/nature06196.

[46] Toshiro Sato, Robert G. Vries, Hugo J. Snippert, Marc van de Wetering, Nick Barker, Daniel E. Stange, Johan H. van Es, Arie Abo, Pekka Kujala, Peter J. Peters, and Hans Clevers. Single Lgr5 stem cells build cryptvillus structures in vitro without a mesenchymal niche. *Nature*, 459(7244):262–265, May 2009. ISSN 0028-0836. doi: 10.1038/nature07935.

[47] Yangyan Jiang, Wenlu Li, Xin He, Hongbo Zhang, Fangzhen Jiang, and Zhigang Chen. Lgr5 expression is a valuable prognostic factor for colorectal cancer: Evidence from a meta-analysis. *BMC Cancer*, 15, December 2015. ISSN 1471-2407. doi: 10.1186/s12885-015-1985-3.

[48] Mariko Shimokawa, Yuki Ohta, Shingo Nishikori, Mami Matano, Ai Takano, Masayuki Fujii, Shoichi Date, Shinya Sugimoto, Takanori Kanai, and Toshiro Sato. Visualization and targeting of LGR5(+) human colon cancer stem cells. *Nature*, 545(7653):187–192, May 2017. ISSN 1476-4687. doi: 10.1038/nature22081.

[49] Felipe de Sousa E. Melo, Antonina V. Kurtova, Jonathan M. Harnoss, Noelyn Kljavin, Joerg D. Hoeck, Jeffrey Hung, Jeffrey Eastham Anderson, Elaine E.

Storm, Zora Modrusan, Hartmut Koeppen, Gerrit J. P. Dijkgraaf, Robert Piskol, and Frederic J. de Sauvage. A distinct role for Lgr5(+) stem cells in primary and metastatic colon cancer. *Nature*, 543(7647):676–680, March 2017. ISSN 1476-4687. doi: 10.1038/nature21713.

[50] Bruno Costa-Silva, Nicole M. Aiello, Allyson J. Ocean, Swarnima Singh, Haiying Zhang, Basant Kumar Thakur, Annette Becker, Ayuko Hoshino, Milica Tešić Mark, Henrik Molina, Jenny Xiang, Tuo Zhang, Till-Martin Theilen, Guillermo García-Santos, Caitlin Williams, Yonathan Ararso, Yujie Huang, Gonçalo Rodrigues, Tang-Long Shen, Knut Jørgen Labori, Inger Marie Bowitz Lothe, Elin H. Kure, Jonathan Hernandez, Alexandre Doussot, Saya H. Ebbesen, Paul M. Grandgenett, Michael A. Hollingsworth, Maneesh Jain, Kavita Mallya, Surinder K. Batra, William R. Jarnagin, Robert E. Schwartz, Irina Matei, Héctor Peinado, Ben Z. Stanger, Jacqueline Bromberg, and David Lyden. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nature Cell Biology*, 17(6):816–826, June 2015. ISSN 1465-7392. doi: 10.1038/ncb3169.

[51] Joseph S. Palumbo, Kathryn E. Talmage, Jessica V. Massari, Christine M. La Jeunesse, Matthew J. Flick, Keith W. Kombrinck, Markéta Jirousková, and Jay L. Degen. Platelets and fibrin(ogen) increase metastatic potential by impeding natural killer cell-mediated elimination of tumor cells. *Blood*, 105(1): 178–185, January 2005. ISSN 0006-4971. doi: 10.1182/blood-2004-06-2272.

[52] S. Karpatkin and E. Pearlstein. Role of platelets in tumor cell metastases. *Annals of Internal Medicine*, 95(5):636–641, November 1981. ISSN 0003-4819.

[53] A. M. Khatib, M. Kontogiannea, L. Fallavollita, B. Jamison, S. Meterissian, and P. Brodt. Rapid induction of cytokine and E-selectin expression in the

liver in response to metastatic tumor cells. *Cancer Research*, 59(6):1356–1361, March 1999. ISSN 0008-5472.

[54] John Condeelis and Jeffrey W. Pollard. Macrophages: Obligate partners for tumor cell migration, invasion, and metastasis. *Cell*, 124(2):263–266, January 2006. ISSN 0092-8674. doi: 10.1016/j.cell.2006.01.007.

[55] K. Kawada, H. Hosogi, M. Sonoshita, H. Sakashita, T. Manabe, Y. Shimahara, Y. Sakai, A. Takabayashi, M. Oshima, and M. M. Taketo. Chemokine receptor CXCR3 promotes colon cancer metastasis to lymph nodes. *Oncogene*, 26(32): 4679–4688, February 2007. ISSN 0950-9232. doi: 10.1038/sj.onc.1210267.

[56] C. F. Rochlitz, R. Herrmann, and E. de Kant. Overexpression and amplification of c-myc during progression of human colorectal cancer. *Oncology*, 53(6):448–454, 1996 Nov-Dec. ISSN 0030-2414.

[57] K. Sato, M. Miyahara, T. Saito, and M. Kobayashi. C-myc mRNA overexpression is associated with lymph node metastasis in colorectal cancer. *European Journal of Cancer*, 30(8):1113–1117, January 1994. ISSN 0959-8049. doi: 10.1016/0959-8049(94)90468-5.

[58] Anita Wolfer and Sridhar Ramaswamy. MYC and Metastasis. *Cancer research*, 71(6):2034–2037, March 2011. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-10-3776.

[59] Marla Lipsyc and Rona Yaeger. Impact of somatic mutations on patterns of metastasis in colorectal cancer. *Journal of Gastrointestinal Oncology*, 6(6):645–649, May 2015. ISSN 2219-679X.

[60] Li Tai Fang, Li Tai Fang, Sharon Lee, Sharon Lee, Helen Choi, Helen Choi, Hong Kwan Kim, Hong Kwan Kim, Gregory Jew, Gregory Jew, Hio Chung

Kang, Hio Chung Kang, Lin Chen, Lin Chen, David Jablons, David Jablons, Il-Jin Kim, and Il-Jin Kim. Comprehensive genomic analyses of a metastatic colon cancer to the lung by whole exome sequencing and gene expression analysis. *International Journal of Oncology*, 44(1):211–221, January 2014. ISSN 1019-6439.

[61] Sun Young Lee, Farhan Haq, Deokhoon Kim, Cui Jun, Hui-Jong Jo, Sung-Min Ahn, and Won-Suk Lee. Comparative Genomic Analysis of Primary and Synchronous Metastatic Colorectal Cancers. *PLOS ONE*, 9(3):e90459, March 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0090459.

[62] Hongyu Zhou and Shile Huang. Role of mTOR signaling in tumor cell motility, invasion and metastasis. *Current Protein & Peptide Science*, 12(1):30–42, February 2011. ISSN 1875-5550.

[63] Ryungsa Kim, Manabu Emi, and Kazuaki Tanabe. Cancer immunoediting from immune surveillance to immune escape. *Immunology*, 121(1):1–14, May 2007. ISSN 0019-2805. doi: 10.1111/j.1365-2567.2007.02587.x.

[64] V. Shankaran, H. Ikeda, A. T. Bruce, J. M. White, P. E. Swanson, L. J. Old, and R. D. Schreiber. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature*, 410(6832):1107–1111, April 2001. ISSN 0028-0836. doi: 10.1038/35074122.

[65] Inge C. van Gool, Tjalling Bosse, and David N. Church. POLE proofreading mutation, immune response and prognosis in endometrial cancer. *Oncoimmunology*, 5(3):e1072675, March 2016. ISSN 2162-4011. doi: 10.1080/2162402X.2015.1072675.

[66] M. W. Heijstek, O. Kranenburg, and I. H. M. Borel Rinkes. Mouse Models of

Colorectal Cancer and Liver Metastases. *Digestive Surgery*, 22(1-2):16–25, May 2005. ISSN 1421-9883. doi: 10.1159/000085342.

[67] Herlen Alencar, Ray King, Martin Funovics, Charles Stout, Ralph Weissleder, and Umar Mahmood. A novel mouse model for segmental orthotopic colon cancer. *International Journal of Cancer*, 117(3):335–339, November 2005. ISSN 0020-7136. doi: 10.1002/ijc.21185.

[68] Ashwani Rajput, Ivan Dominguez San Martin, Rebecca Rose, Alexander Beko, Charles Levea, Elizabeth Sharratt, Richard Mazurchuk, Robert M. Hoffman, Michael G. Brattain, and Jing Wang. Characterization of HCT116 human colon cancer cells in an orthotopic model. *The Journal of Surgical Research*, 147(2):276–281, June 2008. ISSN 0022-4804. doi: 10.1016/j.jss.2007.04.021.

[69] Hong-Jin Chen, Bo-Lin Yang, Yu-Gen Chen, Qiu Lin, Shu-Peng Zhang, and Yun-Fei Gu. A GFP-labeled human colon cancer metastasis model featuring surgical orthotopic implantation. *Asian Pacific journal of cancer prevention: APJCP*, 13(9):4263–4266, 2012. ISSN 1513-7368.

[70] Ehud Zigmond, Zamir Halpern, Eran Elinav, Eli Brazowski, Steffen Jung, and Chen Varol. Utilization of murine colonoscopy for orthotopic implantation of colorectal cancer. *PloS One*, 6(12):e28858, 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0028858.

[71] Kjersti Flatmark, Gunhild M. Maelandsmo, Marita Martinsen, Heidi Rasmussen, and Øystein Fodstad. Twelve colorectal cancer cell lines exhibit highly variable growth and metastatic capacities in an orthotopic model in nude mice. *European Journal of Cancer (Oxford, England: 1990)*, 40(10):1593–1598, July 2004. ISSN 0959-8049. doi: 10.1016/j.ejca.2004.02.023.

[72] María Virtudes Céspedes, Carolina Espina, Miguel Angel García-Cabezas,

Manuel Trias, Alicia Boluda, María Teresa Gómez del Pulgar, Francesc Josep Sancho, Manuel Nistal, Juan Carlos Lacal, and Ramon Mangues. Orthotopic Microinjection of Human Colon Cancer Cells in Nude Mice Induces Tumor Foci in All Clinically Relevant Metastatic Sites. *The American Journal of Pathology*, 170(3):1077–1085, March 2007. ISSN 0002-9440. doi: 10.2353/ajpath.2007.060773.

[73] Kenneth E. Hung, Marco A. Maricevich, Larissa Georgeon Richard, Wei Y. Chen, Michael P. Richardson, Alexandra Kunin, Roderick T. Bronson, Umar Mahmood, and Raju Kucherlapati. Development of a mouse model for sporadic and metastatic colon tumors and its use in assessing drug treatment. *Proceedings of the National Academy of Sciences of the United States of America*, 107(4): 1565–1570, January 2010. ISSN 0027-8424. doi: 10.1073/pnas.0908682107.

[74] Yuan Zhu, James A Richardson, Luis F Parada, and Jonathan M Graff. Smad3 Mutant Mice Develop Metastatic Colorectal Cancer. *Cell*, 94(6):703–714, September 1998. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)81730-4.

[75] Patty Trobridge, Sue Knoblaugh, M. Kay Washington, Nina M. Munoz, Karen D. Tsuchiya, Andres Rojas, Xiaoling Song, Cornelia M. Ulrich, Takehiko Sasazuki, Senji Shirasawa, and William M. Grady. TGF-β receptor inactivation and mutant Kras induce intestinal neoplasms in mice via a β-catenin independent pathway. *Gastroenterology*, 136(5):1680–8.e7, May 2009. ISSN 0016-5085. doi: 10.1053/j.gastro.2009.01.066.

[76] Madeleine Young and Karen R. Reed. Organoids as a Model for Colorectal Cancer. *Current Colorectal Cancer Reports*, 12(5):281–287, October 2016. ISSN 1556-3790, 1556-3804. doi: 10.1007/s11888-016-0335-4.

[77] Toshiro Sato, Daniel E. Stange, Marc Ferrante, Robert G. J. Vries, Johan H.

Van Es, Stieneke Van den Brink, Winan J. Van Houdt, Apollo Pronk, Joost Van Gorp, Peter D. Siersema, and Hans Clevers. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology*, 141(5):1762–1772, November 2011. ISSN 1528-0012. doi: 10.1053/j.gastro.2011.07.050.

[78] Mami Matano, Shoichi Date, Mariko Shimokawa, Ai Takano, Masayuki Fujii, Yuki Ohta, Toshiaki Watanabe, Takanori Kanai, and Toshiro Sato. Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nature Medicine*, 21(3):256–262, March 2015. ISSN 1078-8956. doi: 10.1038/nm.3802.

[79] Jatin Roper, Tuomas Tammela, Naniye Malli Cetinbas, Adam Akkad, Ali Roghanian, Steffen Rickelt, Mohammad Almeqdadi, Katherine Wu, Matthias A. Oberli, Francisco Sánchez-Rivera, Yoona K. Park, Xu Liang, George Eng, Martin S. Taylor, Roxana Azimi, Dmitriy Kedrin, Rachit Neupane, Semir Beyaz, Ewa T. Sicinska, Yvelisse Suarez, James Yoo, Lillian Chen, Lawrence Zukerberg, Pekka Katajisto, Vikram Deshpande, Adam J. Bass, Philip N. Tsichlis, Jacqueline Lees, Robert Langer, Richard O. Hynes, Jianzhu Chen, Arjun Bhutkar, Tyler Jacks, and Ömer H. Yilmaz. In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nature Biotechnology*, advance online publication, May 2017. ISSN 1087-0156. doi: 10.1038/nbt.3836.

[80] Maressa A. Bruhn, Richard B. Pearson, Ross D. Hannan, and Karen E. Sheppard. Second AKT: The rise of SGK in cancer signalling. *Growth Factors (Chur, Switzerland)*, 28(6):394–408, December 2010. ISSN 1029-2292. doi: 10.3109/08977194.2010.518616.

[81] Stefania Segditsas, Oliver Sieber, Maesha Deheragoda, Phil East, Andrew

Rowan, Rosemary Jeffery, Emma Nye, Susan Clark, Bradley Spencer-Dene, Gordon Stamp, Richard Poulsom, Nirosha Suraweera, Andrew Silver, Mohammad Ilyas, and Ian Tomlinson. Putative direct and indirect Wnt targets identified through consistent gene expression changes in APC-mutant intestinal adenomas from humans and mice. *Human Molecular Genetics*, 17(24):3864–3875, December 2008. ISSN 1460-2083. doi: 10.1093/hmg/ddn286.

[82] Francesca Lessi, Andrew Beggs, Mariagrazia de Palo, Marcello Anti, Raffaele Macarone Palmieri, Simona Francesconi, Vito Gomes, Generoso Bevilacqua, Ian Tomlinson, and Stefania Segditsas. Down-Regulation of Serum/Glucocorticoid Regulated Kinase 1 in Colorectal Tumours Is Largely Independent of Promoter Hypermethylation. *PLOS ONE*, 5(11):e13840, November 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0013840.

[83] Kan Wang, Shuchen Gu, Omaima Nasir, Michael Föller, Teresa F. Ackermann, Karin Klingel, Reinhard Kandolf, Dietmar Kuhl, Christos Stournaras, and Florian Lang. SGK1-dependent Intestinal Tumor Growth in APC-deficient Mice. *Cellular Physiology and Biochemistry*, 25(2-3):271–278, 2010. ISSN 1015-8987, 1421-9778. doi: 10.1159/000276561.

[84] Sidi Chen, Neville E. Sanjana, Kaijie Zheng, Ophir Shalem, Kyungheon Lee, Xi Shi, David A. Scott, Jun Song, Jen Q. Pan, Ralph Weissleder, Hakho Lee, Feng Zhang, and Phillip A. Sharp. Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell*, 160(6):1246–1260, March 2015. ISSN 0092-8674. doi: 10.1016/j.cell.2015.02.038.

[85] Neville E. Sanjana, Ophir Shalem, and Feng Zhang. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*, 11(8):783–784, August 2014. ISSN 1548-7091. doi: 10.1038/nmeth.3047.

[86] Ophir Shalem, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl, Benjamin L. Ebert, David E. Root, John G. Doench, and Feng Zhang. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*, 343(6166):84–87, January 2014. ISSN 1095-9203. doi: 10.1126/science.1247005.

[87] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12, May 2011. ISSN 2226-6089. doi: 10.14806/ej.17.1.200.

[88] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.

[89] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, October 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0506580102.

[90] Arthur Liberzon, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P. Mesirov. Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)*, 27(12):1739–1740, June 2011. ISSN 1367-4811. doi: 10.1093/bioinformatics/btr260.

[91] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57, 2009. ISSN 1750-2799. doi: 10.1038/nprot.2008.211.

[92] Wei Li, Han Xu, Tengfei Xiao, Le Cong, Michael I. Love, Feng Zhang, Rafael A.

Irizarry, Jun S. Liu, Myles Brown, and X. Shirley Liu. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*, 15:554, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0554-4.

[93] Martin I. Krzywinski, Jacqueline E. Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, June 2009. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.092759.109.

[94] Human Protein Atlas. Tissue expression of SGK1 - Summary - The Human Protein Atlas. `http://www.proteinatlas.org/ENSG00000118515-SGK1/tissue`, 2017.

[95] M. K. Webster, L. Goya, Y. Ge, A. C. Maiyar, and G. L. Firestone. Characterization of sgk, a novel member of the serine/threonine protein kinase gene family which is transcriptionally induced by glucocorticoids and serum. *Molecular and Cellular Biology*, 13(4):2031–2040, April 1993. ISSN 0270-7306.

[96] T Kobayashi, M Deak, N Morrice, and P Cohen. Characterization of the structure and regulation of two novel isoforms of serum- and glucocorticoid-induced protein kinase. *Biochemical Journal*, 344(Pt 1):189–197, November 1999. ISSN 0264-6021.

[97] G. L. Firestone, J. R. Giampaolo, and B. A. O'Keeffe. Stimulus-Dependent Regulation of Serum and Glucocorticoid Inducible Protein Kinase (SGK) Transcription, Subcellular Localization and Enzymatic Activity. *Cellular Physiology and Biochemistry*, 13(1):1–12, February 2003. ISSN 1421-9778. doi: 10.1159/000070244.

[98] Maria Francisca Arteaga, Diego Alvarez de la Rosa, Jose A. Alvarez, and Ce-

cilia M. Canessa. Multiple translational isoforms give functional specificity to serum- and glucocorticoid-induced kinase 1. *Molecular Biology of the Cell*, 18 (6):2072–2080, June 2007. ISSN 1059-1524. doi: 10.1091/mbc.E06-10-0968.

[99] Maude Tessier and James R. Woodgett. Serum and glucocorticoid-regulated protein kinases: Variations on a theme. *Journal of Cellular Biochemistry*, 98 (6):1391–1407, August 2006. ISSN 0730-2312. doi: 10.1002/jcb.20894.

[100] Juan M. García-Martínez and Dario R. Alessi. mTOR complex 2 (mTORC2) controls hydrophobic motif phosphorylation and activation of serum- and glucocorticoid-induced protein kinase 1 (SGK1). *The Biochemical Journal*, 416 (3):375–385, December 2008. ISSN 1470-8728. doi: 10.1042/BJ20081668.

[101] Florian Lang and Ekaterina Shumilina. Regulation of ion channels by the serum- and glucocorticoid-inducible kinase SGK1. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 27(1):3–12, January 2013. ISSN 1530-6860. doi: 10.1096/fj.12-218230.

[102] Christophe Debonneville, Sandra Y. Flores, Elena Kamynina, Pamela J. Plant, Caroline Tauxe, Marc A. Thomas, Carole Münster, Ahmed Chraïbi, J.Howard Pratt, Jean-Daniel Horisberger, David Pearce, Johannes Loffing, and Olivier Staub. Phosphorylation of Nedd4-2 by Sgk1 regulates epithelial Na+ channel cell surface expression. *The EMBO Journal*, 20(24):7052–7059, December 2001. ISSN 0261-4189. doi: 10.1093/emboj/20.24.7052.

[103] Rosario Amato, Lucia D'Antona, Giovanni Porciatti, Valter Agosti, Miranda Menniti, Cinzia Rinaldo, Nicola Costa, Emanuele Bellacchio, Stefano Mattarocci, Giorgio Fuiano, Silvia Soddu, Marco Giorgio Paggi, Florian Lang, and Nicola Perrotti. Sgk1 activates MDM2-dependent p53 degradation and affects

cell proliferation, survival, and differentiation. *Journal of Molecular Medicine*, 87(12):1221–1239, 2009. ISSN 0946-2716. doi: 10.1007/s00109-009-0525-5.

[104] Minho Won, Kyeong Ah Park, Hee Sun Byun, Young-Rae Kim, Byung Lyul Choi, Jang Hee Hong, Jongsun Park, Jeong Ho Seok, Young-Ho Lee, Chung-Hyun Cho, In Sang Song, Yong Kyung Kim, Han-Ming Shen, and Gang Min Hur. Protein kinase SGK1 enhances MEK/ERK complex formation through the phosphorylation of ERK2: Implication for the positive regulatory role of SGK1 on the ERK function during liver regeneration. *Journal of Hepatology*, 51(1):67–76, July 2009. ISSN 1600-0641. doi: 10.1016/j.jhep.2009.02.027.

[105] Feng Hong, Michelle D. Larrea, Cheryl Doughty, David J. Kwiatkowski, Rachel Squillace, and Joyce M. Slingerland. mTOR-Raptor Binds and Activates SGK1 to Regulate p27 Phosphorylation. *Molecular Cell*, 30(6):701–711, June 2008. ISSN 1097-2765. doi: 10.1016/j.molcel.2008.04.027.

[106] Shunsuke Mori, Shigeyuki Nada, Hironobu Kimura, Shoji Tajima, Yusuke Takahashi, Ayaka Kitamura, Chitose Oneyama, and Masato Okada. The mTOR Pathway Controls Cell Proliferation by Regulating the FoxO3a Transcription Factor via SGK1 Kinase. *PLOS ONE*, 9(2):e88891, February 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0088891.

[107] sgk1. FireBrowse. `http://firebrowse.org/viewGene.html?gene=sgk1`, 2016.

[108] Pau Castel, Haley Ellis, Ruzica Bago, Eneda Toska, Pedram Razavi, F. Javier Carmona, Srinivasaraghavan Kannan, Chandra S. Verma, Maura Dickler, Sarat Chandarlapaty, Edi Brogi, Dario R. Alessi, José Baselga, and Maurizio Scaltriti. PDK1-SGK1 Signaling Sustains AKT-Independent mTORC1 Activation and Confers Resistance to PI3Kα Inhibition. *Cancer Cell*, 30(2):229–242, August 2016. ISSN 1535-6108, 1878-3686. doi: 10.1016/j.ccell.2016.06.004.

[109] U.-M. Fagerli, K. Ullrich, T. Stühmer, T. Holien, K. Köchert, R. U. Holt, O. Bruland, M. Chatterjee, H. Nogai, G. Lenz, J. D. Shaughnessy, S. Mathas, A. Sundan, R. C. Bargou, B. Dörken, M. Børset, and M. Janz. Serum/glucocorticoid-regulated kinase 1 (SGK1) is a prominent target gene of the transcriptional response to cytokines in multiple myeloma and supports the growth of myeloma cells. *Oncogene*, 30(28):3198–3206, July 2011. ISSN 0950-9232. doi: 10.1038/onc.2011.79.

[110] Eeva M. Sommer, Hannah Dry, Darren Cross, Sylvie Guichard, Barry R. Davies, and Dario R. Alessi. Elevated SGK1 predicts resistance of breast cancer cells to Akt inhibitors. *The Biochemical Journal*, 452(3):499–508, June 2013. ISSN 1470-8728. doi: 10.1042/BJ20130342.

[111] Bao-Hong Zhang, Eric D. Tang, Tianqing Zhu, Michael E. Greenberg, Anne B. Vojtek, and Kun-Liang Guan. Serum- and Glucocorticoid-inducible Kinase SGK Phosphorylates and Negatively Regulates B-Raf. *Journal of Biological Chemistry*, 276(34):31620–31626, August 2001. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M102808200.

[112] Fiona J. McDonald. A new SGK1 knockout mouse. *American Journal of Physiology - Renal Physiology*, 294(6):F1296–F1297, June 2008. ISSN 1931-857X, 1522-1466. doi: 10.1152/ajprenal.90245.2008.

[113] Yu Xiaobo, Lin Qiang, Qin Xiong, Ruan Zheng, Zhou Jianhua, Lin Zhifeng, Su Yijiang, and Jian Zheng. Serum and glucocorticoid kinase 1 promoted the growth and migration of non-small cell lung cancer cells. *Gene*, 576(1 Pt 2): 339–346, January 2016. ISSN 1879-0038. doi: 10.1016/j.gene.2015.10.072.

[114] Eva-Maria Schmidt, Shuchen Gu, Vasileia Anagnostopoulou, Konstantinos Alevizopoulos, Michael Föller, Florian Lang, and Christos Stournaras. Serum- and

glucocorticoid-dependent kinase-1-induced cell migration is dependent on vinculin and regulated by the membrane androgen receptor. *FEBS Journal*, 279(7): 1231–1242, April 2012. ISSN 1742-4658. doi: 10.1111/j.1742-4658.2012.08515.x.

[115] Anita Sveen, Trude H. Agesen, Arild Nesbakken, Torleiv O. Rognum, Ragnhild A. Lothe, and Rolf I. Skotheim. Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival. *Genome Medicine*, 3(5):32, May 2011. ISSN 1756-994X. doi: 10.1186/gm248.

[116] S. W. Byers, C. L. Sommers, B. Hoxter, A. M. Mercurio, and A. Tozeren. Role of E-cadherin in the response of tumor cell aggregates to lymphatic, venous and arterial flow: Measurement of cell-cell adhesion strength. *Journal of Cell Science*, 108 (Pt 5):2053–2064, May 1995. ISSN 0021-9533.

[117] T. C. He, A. B. Sparks, C. Rago, H. Hermeking, L. Zawel, L. T. da Costa, P. J. Morin, B. Vogelstein, and K. W. Kinzler. Identification of c-MYC as a target of the APC pathway. *Science (New York, N.Y.)*, 281(5382):1509–1512, September 1998. ISSN 0036-8075.

[118] Chia-Hsin Chan, Szu-Wei Lee, Chien-Feng Li, Jing Wang, Wei-Lei Yang, Ching-Yuan Wu, Juan Wu, Keiichi I. Nakayama, Hong-Yo Kang, Hsuan-Ying Huang, Mien-Chie Hung, Pier Paolo Pandolfi, and Hui-Kuan Lin. Deciphering the transcriptional complex critical for RhoA gene expression and cancer metastasis. *Nature Cell Biology*, 12(5):457–467, May 2010. ISSN 1476-4679. doi: 10.1038/ncb2047.

[119] Kyu Sang Lee, Yoonjin Kwak, Kyung Han Nam, Duck-Woo Kim, Sung-Bum Kang, Gheeyoung Choe, Woo Ho Kim, and Hye Seung Lee. C-MYC Copy-Number Gain Is an Independent Prognostic Factor in Patients with Colorectal

Cancer. *PLOS ONE*, 10(10):e0139727, October 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0139727.

[120] Kerstin Bellmann, Julie Martel, Dominic J. P. Poirier, Mireille M. Labrie, and Jacques Landry. Downregulation of the PI3K/Akt survival pathway in cells with deregulated expression of c-Myc. *Apoptosis: An International Journal on Programmed Cell Death*, 11(8):1311–1319, August 2006. ISSN 1360-8185. doi: 10.1007/s10495-006-8205-1.

[121] Zhongwei Cao, Hua Fan-Minogue, David I. Bellovin, Aleksey Yevtodiyenko, Julia Arzeno, Qiwei Yang, Sanjiv Sam Gambhir, and Dean W. Felsher. MYC Phosphorylation, Activation, and Tumorigenic Potential in Hepatocellular Carcinoma Are Regulated by HMG-CoA Reductase. *Cancer Research*, 71(6):2286–2297, March 2011. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-10-3367.

[122] Markus Welcker, Amir Orian, Jianping Jin, Jonathan A. Grim, J. Wade Harper, Robert N. Eisenman, and Bruce E. Clurman. The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(24):9085–9090, June 2004. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0402770101.

[123] Bruno Amati. Myc degradation: Dancing with ubiquitin ligases. *Proceedings of the National Academy of Sciences of the United States of America*, 101(24): 8843–8844, June 2004. ISSN 0027-8424. doi: 10.1073/pnas.0403046101.

[124] Sharon J. Miller, Bishnu P. Joshi, Ying Feng, Adam Gaustad, Eric R. Fearon, and Thomas D. Wang. In Vivo Fluorescence-Based Endoscopic Detection of

Colon Dysplasia in the Mouse Using a Novel Peptide Probe. *PLoS ONE*, 6(3), March 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0017384.

[125] Alnawaz Rehemtulla, Lauren D. Stegman, Shaun J. Cardozo, Sheila Gupta, Daniel E. Hall, Christopher H. Contag, and Brian D. Ross. Rapid and Quantitative Assessment of Cancer Treatment Response Using In Vivo Bioluminescence Imaging. *Neoplasia*, 2(6):491–495, January 2000. ISSN 1476-5586. doi: 10.1038/sj.neo.7900121.

[126] Jeanne Tie, Lara Lipton, Jayesh Desai, Peter Gibbs, Robert N. Jorissen, Michael Christie, Katharine J. Drummond, Benjamin N. J. Thomson, Valery Usatoff, Peter M. Evans, Adrian W. Pick, Simon Knight, Peter W. G. Carne, Roger Berry, Adrian Polglase, Paul McMurrick, Qi Zhao, Dana Busam, Robert L. Strausberg, Enric Domingo, Ian P. M. Tomlinson, Rachel Midgley, David Kerr, and Oliver M. Sieber. KRAS Mutation Is Associated with Lung Metastasis in Patients with Curatively Resected Colorectal Cancer. *Clinical Cancer Research*, 17(5):1122–1130, March 2011. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-10-1720.

[127] Kristoffer Derwinger, Karl Kodeda, Elinor Bexe-Lindskog, and Helena Taflin. Tumour differentiation grade is associated with TNM staging and the risk of node metastasis in colorectal cancer. *Acta Oncologica*, 49(1):57–62, January 2010. ISSN 0284-186X. doi: 10.3109/02841860903334411.

[128] Florian Lang, Christos Stournaras, and Ioana Alesutan. Regulation of transport across cell membranes by the serum- and glucocorticoid-inducible kinase SGK1. *Molecular Membrane Biology*, 31(1):29–36, February 2014. ISSN 1464-5203. doi: 10.3109/09687688.2013.874598.

[129] Florian Lang, Nathalie Strutz-Seebohm, Guiscard Seebohm, and Undine E.

Lang. Significance of SGK1 in the regulation of neuronal function. *The Journal of Physiology*, 588(Pt 18):3349–3354, September 2010. ISSN 1469-7793. doi: 10.1113/jphysiol.2010.190926.

[130] Emily B. Heikamp, Chirag H. Patel, Sam Collins, Adam Waickman, Min-Hee Oh, Im-Hong Sun, Peter Illei, Archna Sharma, Aniko Naray-Fejes-Toth, Geza Fejes-Toth, Jyoti Misra-Sen, Maureen R. Horton, and Jonathan D. Powell. The AGC kinase SGK1 regulates TH1 and TH2 differentiation downstream of the mTORC2 complex. *Nature Immunology*, 15(5):457–464, May 2014. ISSN 1529-2916. doi: 10.1038/ni.2867.

[131] Jian-An Bai, Gui-Fang Xu, Li-Jun Yan, Wei-Wen Zeng, Qian-Qian Ji, Jin-Dao Wu, and Qi-Yun Tang. SGK1 inhibits cellular apoptosis and promotes proliferation via the MEK/ERK/p53 pathway in colitis. *World Journal of Gastroenterology*, 21(20):6180–6193, May 2015. ISSN 2219-2840. doi: 10.3748/wjg.v21.i20.6180.

[132] Peter M. Wilson, Anthony El-Khoueiry, Syma Iqbal, William Fazzone, Melissa J. LaBonte, Susan Groshen, Dongyun Yang, Kathy D. Danenberg, Sarah Cole, Margaret Kornacki, Robert D. Ladner, and Heinz-Josef Lenz. A phase I/II trial of vorinostat in combination with 5-fluorouracil in patients with metastatic colorectal cancer who previously failed 5-FU-based chemotherapy. *Cancer Chemotherapy and Pharmacology*, 65(5):979–988, April 2010. ISSN 1432-0843. doi: 10.1007/s00280-009-1236-x.

[133] Seung Ah Choi, Seung-Ki Kim, Ji Yeoun Lee, Kyu-Chang Wang, Chanhee Lee, and Ji Hoon Phi. LIN28B is highly expressed in atypical teratoid/rhabdoid tumor (AT/RT) and suppressed through the restoration of SMARCB1. *Cancer Cell International*, 16, April 2016. ISSN 1475-2867. doi: 10.1186/s12935-016-0307-4.

[134] Maren K. Fuentes, Shraddha S. Nigavekar, Thiruvengadam Arumugam, Craig D. Logsdon, Ann Marie Schmidt, Juliet C. Park, and Emina H. Huang. RAGE Activation by S100P in Colon Cancer Stimulates Growth, Migration, and Cell Signaling Pathways. *Diseases of the Colon & Rectum*, 50(8):1230–1240, June 2007. ISSN 0012-3706, 1530-0358. doi: 10.1007/s10350-006-0850-5.

[135] Yan Ning, Philipp C. Manegold, Young Kwon Hong, Wu Zhang, Alexandra Pohl, Georg Lurje, Thomas Winder, Dongyun Yang, Melissa J. LaBonte, Peter M. Wilson, Robert D. Ladner, and Heinz-Josef Lenz. Interleukin-8 is associated with proliferation, migration, angiogenesis and chemosensitivity in vitro and in vivo in colon cancer cell line models. *International Journal of Cancer*, 128(9):2038–2049, May 2011. ISSN 1097-0215. doi: 10.1002/ijc.25562.

[136] Eiji Suzuki, Tetsuya Ota, Kazunori Tsukuda, Atsushi Okita, Kinya Matsuoka, Masakazu Murakami, Hiroyoshi Doihara, and Nobuyoshi Shimizu. Nm23-H1 reduces in vitro cell migration and the liver metastatic potential of colon cancer cells by regulating myosin light chain phosphorylation. *International Journal of Cancer*, 108(2):207–211, January 2004. ISSN 1097-0215. doi: 10.1002/ijc.11546.

[137] N. Nagy, Y. Bronckart, I. Camby, H. Legendre, H. Lahm, H. Kaltner, Y. Hadari, P. Van Ham, P. Yeaton, J.-C. Pector, Y. Zick, I. Salmon, A. Danguy, R. Kiss, and H.-J. Gabius. Galectin-8 expression decreases in cancer compared with normal and dysplastic human colon tissue and acts significantly on human colon cancer cell migration as a suppressor. *Gut*, 50(3):392–401, March 2002. ISSN 0017-5749.

[138] Wies van Roosmalen, Sylvia E. Le Dévédec, Ofra Golani, Marcel Smid, Irina Pulyakhina, Annemieke M. Timmermans, Maxime P. Look, Di Zi, Chantal Pont, Marjo de Graauw, Suha Naffar-Abu-Amara, Catherine Kirsanova, Gabriella Rustici, Peter A. C. 't Hoen, John W. M. Martens, John A. Foekens,

Benjamin Geiger, and Bob van de Water. Tumor cell migration screen identifies SRPK1 as breast cancer metastasis determinant. *The Journal of Clinical Investigation*, 125(4):1648–1664, April 2015. ISSN 1558-8238. doi: 10.1172/JCI74440.

[139] Gromoslaw A. Smolen, Jianmin Zhang, Matthew J. Zubrowski, Elena J. Edelman, Biao Luo, Min Yu, Lydia W. Ng, Cally M. Scherber, Benjamin J. Schott, Sridhar Ramaswamy, Daniel Irimia, David E. Root, and Daniel A. Haber. A genome-wide RNAi screen identifies multiple RSK-dependent regulators of cell migration. *Genes & Development*, 24(23):2654–2665, December 2010. ISSN 1549-5477. doi: 10.1101/gad.1989110.

[140] Kiranmai Gumireddy, Fangxian Sun, Andres J. Klein-Szanto, Jonathan M. Gibbins, Phyllis A. Gimotty, Aleister J. Saunders, Peter G. Schultz, and Qihong Huang. In vivo selection for metastasis promoting genes in the mouse. *Proceedings of the National Academy of Sciences*, 104(16):6696–6701, April 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0701145104.

[141] Kiranmai Gumireddy, Anping Li, Phyllis A. Gimotty, Andres J. Klein-Szanto, Louise C. Showe, Dionyssios Katsaros, George Coukos, Lin Zhang, and Qihong Huang. KLF17 is a negative regulator of epithelial–mesenchymal transition and metastasis in breast cancer. *Nature Cell Biology*, 11(11):1297–1304, November 2009. ISSN 1465-7392. doi: 10.1038/ncb1974.

[142] S. A. Rosenberg, P. Spiess, and R. Lafreniere. A new approach to the adoptive immunotherapy of cancer with tumor-infiltrating lymphocytes. *Science (New York, N.Y.)*, 233(4770):1318–1321, September 1986. ISSN 0036-8075.

[143] Zlatko Trajanoski, Mirjana Efremova, Victoria Klepsch, Pornpimol Charoentong, Francesca Finotello, Dietmar Rieder, Hubert Hackl, Natasch Hermann-

Kleiter, Gottfried Baier, and Anne Krogsdam. Targeting the PD-1/PD-L1 pathway potentiates immunoediting to counterbalance neutral evolution in a mouse model of colorectal cancer. *bioRxiv*, page 099747, January 2017. doi: 10.1101/099747.

[144] Renlong Zou, Xinping Zhong, Chunyu Wang, Hongmiao Sun, Shengli Wang, Lin Lin, Shiying Sun, Changci Tong, Hao Luo, Peng Gao, Yanshu Li, Tingting Zhou, Da Li, Liu Cao, and Yue Zhao. MDC1 Enhances Estrogen Receptor-mediated Transactivation and Contributes to Breast Cancer Suppression. *International Journal of Biological Sciences*, 11(9):992–1005, 2015. ISSN 1449-2288. doi: 10.7150/ijbs.10918.

[145] Ivan P. Gorlov, Peter Meyer, Triantafillos Liloglou, Jonathan Myles, Melanie Barbara Boettger, Adrian Cassidy, Luc Girard, John D. Minna, Reiner Fischer, Stephen Duffy, Margaret R. Spitz, Karl Haeussinger, Stefan Kammerer, Charles Cantor, Rainer Dierkesmann, John K. Field, and Christopher I. Amos. Seizure 6-like (SEZ6L) gene and risk for lung cancer. *Cancer Research*, 67(17): 8406–8411, September 2007. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-06-4784.

[146] M. Nishioka, T. Kohno, M. Takahashi, T. Niki, T. Yamada, S. Sone, and J. Yokota. Identification of a 428-kb homozygously deleted region disrupting the SEZ6L gene at 22q12.1 in a lung cancer cell line. *Oncogene*, 19(54): 6251–6260, December 2000. ISSN 0950-9232. doi: 10.1038/sj.onc.1204031.

[147] Bjørn Naume, Xi Zhao, Marit Synnestvedt, Elin Borgen, Hege Giercksky Russnes, Ole Christian Lingjaerde, Maria Strømberg, Gro Wiedswang, Gunnar Kvalheim, Rolf Kåresen, Jahn M. Nesland, Anne-Lise Børresen-Dale, and Therese Sørlie. Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer.

Molecular Oncology, 1(2):160–171, September 2007. ISSN 1878-0261. doi: 10.1016/j.molonc.2007.03.004.

[148] Er-Yong Lai, Zhen-Guo Chen, Xuan Zhou, Xiao-Rong Fan, Hua Wang, Ping-Lin Lai, Yong-Chun Su, Bai-Yu Zhang, Xiao-Chun Bai, and Yun-Feng Li. DEPTOR expression negatively correlates with mTORC1 activity and tumor progression in colorectal cancer. *Asian Pacific journal of cancer prevention: APJCP*, 15 (11):4589–4594, 2014. ISSN 1513-7368.

[149] Minoru Terashima, Akihiko Ishimura, Masakazu Yoshida, Yutaka Suzuki, Sumio Sugano, and Takeshi Suzuki. The tumor suppressor Rb and its related Rbl2 genes are regulated by Utx histone demethylase. *Biochemical and Biophysical Research Communications*, 399(2):238–244, August 2010. ISSN 1090-2104. doi: 10.1016/j.bbrc.2010.07.061.

[150] Hideki Yamada, Kiyoshi Yanagisawa, Shogo Tokumaru, Ayumu Taguchi, Yuji Nimura, Hirotaka Osada, Masato Nagino, and Takashi Takahashi. Detailed characterization of a homozygously deleted region corresponding to a candidate tumor suppressor locus at 21q11-21 in human lung cancer. *Genes, Chromosomes & Cancer*, 47(9):810–818, September 2008. ISSN 1098-2264. doi: 10.1002/gcc.20582.

[151] Lele Wu, Yuzhi Wang, Yan Liu, Shiyi Yu, Hao Xie, Xingjuan Shi, Sheng Qin, Fei Ma, Tuan Zea Tan, Jean Paul Thiery, and Liming Chen. A central role for TRPS1 in the control of cell cycle and cancer development. *Oncotarget*, 5(17): 7677–7690, July 2014. ISSN 1949-2553.

[152] Teruaki Nomura, Md Matiullah Khan, Sunil C. Kaul, Hai-Dong Dong, Renu Wadhwa, Clemencia Colmenares, Isao Kohno, and Shunsuke Ishii. Ski is a component of the histone deacetylase complex required for transcriptional re-

pression by Mad and thyroid hormone receptor. *Genes & Development*, 13(4): 412–423, February 1999. ISSN 0890-9369.

[153] Kuen-Tyng Lin, Yi-Wei Wang, Chiung-Tong Chen, Chun-Ming Ho, Wen-Hui Su, and Yuh-Shan Jou. HDAC inhibitors augmented cell migration and metastasis through induction of PKCs leading to identification of low toxicity modalities for combination cancer therapy. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 18(17):4691–4701, September 2012. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-12-0633.

[154] Sonia B. Jakowlew. Transforming growth factor-beta in cancer and metastasis. *Cancer Metastasis Reviews*, 25(3):435–457, September 2006. ISSN 0167-7659. doi: 10.1007/s10555-006-9006-2.

[155] Michael Pickup, Sergey Novitskiy, and Harold L. Moses. The roles of TGFβ in the tumour microenvironment. *Nature Reviews Cancer*, 13(11):788–799, November 2013. ISSN 1474-175X. doi: 10.1038/nrc3603.

[156] Liying Geng, Anathbandhu Chaudhuri, Geoffrey Talmon, James L. Wisecarver, and Jing Wang. TGF-Beta Suppresses VEGFA-Mediated Angiogenesis in Colon Cancer Metastasis. *PLOS ONE*, 8(3):e59918, March 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0059918.

[157] S. A. Lamprecht, B. Schwartz, and A. Glicksman. Transforming growth factor-beta in intestinal epithelial differentiation and neoplasia (review). *Anticancer Research*, 9(6):1877–1881, 1989 Nov-Dec. ISSN 0250-7005.

[158] Chuan-Wei Jang, Chun-Hau Chen, Chun-Chieh Chen, Jia-yun Chen, Yi-Hsien Su, and Ruey-Hwa Chen. TGF-β induces apoptosis through Smad-mediated expression of DAP-kinase. *Nature Cell Biology*, 4(1):51–58, January 2002. ISSN 1465-7392. doi: 10.1038/ncb731.

[159] Hua-chuan Zheng, Jing Li, Dao-fu Shen, Xue-feng Yang, Shuang Zhao, Ya-zhou Wu, Yasuo Takano, Hong-zhi Sun, Rong-jian Su, Jun-sheng Luo, and Wen-feng Gou. BTG1 expression correlates with pathogenesis, aggressive behaviors and prognosis of gastric cancer: A potential target for gene therapy. *Oncotarget*, 6 (23):19685–19705, June 2015. ISSN 1949-2553.

[160] Kenji Iwai, Ken-ichi Hirata, Tatsuro Ishida, Shigeto Takeuchi, Tetsuaki Hirase, Yoshiyuki Rikitake, Yoko Kojima, Nobutaka Inoue, Seinosuke Kawashima, and Mitsuhiro Yokoyama. An anti-proliferative gene BTG1 regulates angiogenesis in vitro. *Biochemical and Biophysical Research Communications*, 316(3):628–635, April 2004. ISSN 0006-291X. doi: 10.1016/j.bbrc.2004.02.095.

[161] Yi-Wei Li, Jia-Xing Wang, Xin Yin, Shuang-Jian Qiu, Han Wu, Rui Liao, Yong Yi, Yong-Sheng Xiao, Jian Zhou, Bo-Heng Zhang, and Jia Fan. Decreased Expression of GATA2 Promoted Proliferation, Migration and Invasion of HepG2 In Vitro and Correlated with Poor Prognosis of Hepatocellular Carcinoma. *PLOS ONE*, 9(1):e87505, January 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0087505.

[162] Ke Zhang, Chun Ye, Qin Zhou, Rong Zheng, Xiaoyan Lv, Ye Chen, Zhongguo Hu, Hong Guo, Zheng Zhang, Yidong Wang, Ruizhi Tan, and Yuhang Liu. PKD1 inhibits cancer cells migration and invasion via Wnt signaling pathway in vitro. *Cell Biochemistry and Function*, 25(6):767–774, November 2007. ISSN 1099-0844. doi: 10.1002/cbf.1417.

[163] Debarshi Banerjee, Sonia L. Hernandez, Alejandro Garcia, Thaned Kangsamaksin, Emily Sbiroli, John Andrews, Lynn Ann Forrester, Na Wei, Angela Kadenhe-Chiweshe, Carrie J. Shawber, Jan K. Kitajewski, Jessica J. Kandel, and Darrell J. Yamashiro. Notch suppresses angiogenesis and progres-

sion of hepatic metastases. *Cancer research*, 75(8):1592–1602, April 2015. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-14-1493.

[164] Yan-jun Zhang, Lichun Wei, Mei Liu, Jie Li, Yi-qiong Zheng, Ying Gao, and Xi-ru Li. BTG2 inhibits the proliferation, invasion, and apoptosis of MDA-MB-231 triple-negative breast cancer cells. *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, 34(3): 1605–1613, June 2013. ISSN 1423-0380. doi: 10.1007/s13277-013-0691-5.

[165] Hideaki Sato, Paula Murphy, Shay Giles, John Bannigan, Hajime Takayasu, and Prem Puri. Visualizing expression patterns of Shh and Foxf1 genes in the foregut and lung buds by optical projection tomography. *Pediatric Surgery International*, 24(1):3–11, January 2008. ISSN 0179-0358. doi: 10.1007/s00383-007-2036-1.

[166] Vladimir V. Kalinichenko, Yan Zhou, Brian Shin, Donna Beer Stolz, Simon C. Watkins, Jeffrey A. Whitsett, and Robert H. Costa. Wild-type levels of the mouse Forkhead Box f1 gene are essential for lung repair. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 282(6):L1253–1265, June 2002. ISSN 1040-0605. doi: 10.1152/ajplung.00463.2001.

[167] Xiaomeng Ren, Vladimir Ustiyan, Arun Pradhan, Yuqi Cai, Jamie A. Havrilak, Craig S. Bolte, John M. Shannon, Tanya V. Kalin, and Vladimir V. Kalinichenko. FOXF1 transcription factor is required for formation of embryonic vasculature by regulating VEGF signaling in endothelial cells. *Circulation Research*, 115(8):709–720, September 2014. ISSN 1524-4571. doi: 10.1161/CIRCRESAHA.115.304382.

[168] Blair B. Madison, Lindsay B. McKenna, Diane Dolson, Douglas J. Epstein, and Klaus H. Kaestner. FoxF1 and FoxL1 Link Hedgehog Signaling and the

Control of Epithelial Proliferation in the Developing Stomach and Intestine. *The Journal of Biological Chemistry*, 284(9):5936–5944, February 2009. ISSN 0021-9258. doi: 10.1074/jbc.M808103200.

[169] Pang-Kuo Lo, Ji Shin Lee, Xiaohui Liang, Liangfeng Han, Tsuyoshi Mori, Mary Jo Fackler, Helen Sadik, Pedram Argani, Tej K. Pandita, and Saraswati Sukumar. Epigenetic inactivation of the potential tumor suppressor gene FOXF1 in breast cancer. *Cancer Research*, 70(14):6047–6058, July 2010. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-10-1576.

[170] M. Tamura, Y. Sasaki, R. Koyama, K. Takeda, M. Idogawa, and T. Tokino. Forkhead transcription factor FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness. *Oncogene*, 33(40):4837–4846, October 2014. ISSN 0950-9232. doi: 10.1038/onc.2013.427.

[171] Dmitriy Malin, Il-Man Kim, Evan Boetticher, Tanya V. Kalin, Sneha Ramakrishna, Lucille Meliton, Vladimir Ustiyan, Xiangdong Zhu, and Vladimir V. Kalinichenko. Forkhead Box F1 Is Essential for Migration of Mesenchymal Cells and Directly Induces Integrin-Beta3 Expression. *Molecular and Cellular Biology*, 27(7):2486–2498, January 2007. ISSN 0270-7306, 1098-5549. doi: 10.1128/MCB.01736-06.

[172] D. Milewski, A. Pradhan, X. Wang, Y. Cai, T. Le, B. Turpin, V. V. Kalinichenko, and T. V. Kalin. FoxF1 and FoxF2 transcription factors synergistically promote rhabdomyosarcoma carcinogenesis by repressing transcription of p21Cip1 CDK inhibitor. *Oncogene*, July 2016. ISSN 0950-9232. doi: 10.1038/onc.2016.254.

[173] Pang-Kuo Lo, Ji Shin Lee, and Saraswati Sukumar. The p53-p21WAF1 checkpoint pathway plays a protective role in preventing DNA rereplication induced

by abrogation of FOXF1 function. *Cellular Signalling*, 24(1):316–324, January 2012. ISSN 1873-3913. doi: 10.1016/j.cellsig.2011.09.017.

[174] Logan Fulford, David Milewski, Vladimir Ustiyan, Navin Ravishankar, Yuqi Cai, Tien Le, Sreeharsha Masineni, Susan Kasper, Bruce Aronow, Vladimir V. Kalinichenko, and Tanya V. Kalin. The transcription factor FOXF1 promotes prostate cancer by stimulating the mitogen-activated protein kinase ERK5. *Science Signaling*, 9(427):ra48, May 2016. ISSN 1937-9145. doi: 10.1126/scisignal.aad5582.

[175] Gisela Nilsson and Marie Kannius-Janson. Forkhead Box F1 promotes breast cancer cell migration by upregulating lysyl oxidase and suppressing Smad2/3 signaling. *BMC Cancer*, 16:142, 2016. ISSN 1471-2407. doi: 10.1186/s12885-016-2196-2.

[176] Katherine Minter-Dykhouse, Irene Ward, Michael S. Y. Huen, Junjie Chen, and Zhenkun Lou. Distinct versus overlapping functions of MDC1 and 53BP1 in DNA damage response and tumorigenesis. *The Journal of Cell Biology*, 181(5): 727–735, June 2008. ISSN 0021-9525, 1540-8140. doi: 10.1083/jcb.200801083.

[177] J. Bartkova, Z. Hořej&sbreve, í, M. Sehested, J. M. Nesland, E. Rajpert-De Meyts, N. E. Skakkebæk, M. Stucki, S. Jackson, J. Lukas, and J. Bartek. DNA damage response mediators MDC1 and 53BP1: Constitutive activation and aberrant loss in breast and lung cancer, but not in testicular germ cell tumours. *Oncogene*, 26(53):7414–7422, June 2007. ISSN 0950-9232. doi: 10.1038/sj.onc.1210553.

[178] Renlong Zou, Xinping Zhong, Chunyu Wang, Hongmiao Sun, Shengli Wang, Lin Lin, Shiying Sun, Changci Tong, Hao Luo, Peng Gao, Yanshu Li, Tingting Zhou, Da Li, Liu Cao, and Yue Zhao. MDC1 Enhances Estrogen Receptor-mediated

Transactivation and Contributes to Breast Cancer Suppression. *International Journal of Biological Sciences*, 11(9):992–1005, July 2015. ISSN 1449-2288. doi: 10.7150/ijbs.10918.

[179] Chengfu Yuan, Youquan Bu, Changdong Wang, Faping Yi, Zhengmei Yang, Xiuning Huang, Li Cheng, Geli Liu, Yong Wang, and Fangzhou Song. NFBD1/MDC1 is a protein of oncogenic potential in human cervical cancer. *Molecular and Cellular Biochemistry*, 359(1-2):333–346, August 2011. ISSN 0300-8177, 1573-4919. doi: 10.1007/s11010-011-1027-7.

[180] Francesco P. Fiorentino, Catherine E. Symonds, Marcella Macaluso, and Antonio Giordano. Senescence and p130/Rbl2: A new beginning to the end. *Cell Research*, 19(9):1044–1051, August 2009. ISSN 1001-0602. doi: 10.1038/cr.2009.96.

[181] C. M. Howard, P. P. Claudio, G. L. Gallia, J. Gordon, G. G. Giordano, W. W. Hauck, K. Khalili, and A. Giordano. Retinoblastoma-related protein pRb2/p130 and suppression of tumor growth in vivo. *Journal of the National Cancer Institute*, 90(19):1451–1460, October 1998. ISSN 0027-8874.

[182] Giulia De Falco, Eleonora Leucci, Dido Lenze, Pier Paolo Piccaluga, Pier Paolo Claudio, Anna Onnis, Giovanna Cerino, Joshua Nyagol, Walter Mwanda, Cristiana Bellan, Michael Hummel, Stefano Pileri, Piero Tosi, Harald Stein, Antonio Giordano, and Lorenzo Leoncini. Gene-expression analysis identifies novel RBL2/p130 target genes in endemic Burkitt lymphoma cell lines and primary tumors. *Blood*, 110(4):1301–1307, August 2007. ISSN 0006-4971. doi: 10.1182/blood-2006-12-064865.

[183] Farman Ullah, Taimoor Khan, Nawab Ali, Faraz Arshad Malik, Mahmood Akhtar Kayani, Syed Tahir Abbas Shah, and Muhammad Saeed. Pro-

moter Methylation Status Modulate the Expression of Tumor Suppressor (RbL2/p130) Gene in Breast Cancer. *PLOS ONE*, 10(8):e0134687, August 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0134687.

[184] Eva K. Brinkman, Tao Chen, Mario Amendola, and Bas van Steensel. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, 42(22):e168, December 2014. ISSN 1362-4962. doi: 10.1093/nar/gku936.

[185] Mitsuru Nakanishi, Toshinori Ozaki, Hideki Yamamoto, Takayuki Hanamoto, Hironobu Kikuchi, Kazushige Furuya, Masahiro Asaka, Domenico Delia, and Akira Nakagawara. NFBD1/MDC1 Associates with p53 and Regulates Its Function at the Crossroad between Cell Survival and Death in Response to DNA Damage. *Journal of Biological Chemistry*, 282(31):22993–23004, March 2007. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M611412200.

[186] Silvana Konermann, Mark D. Brigham, Alexandro E. Trevino, Julia Joung, Omar O. Abudayyeh, Clea Barcena, Patrick D. Hsu, Naomi Habib, Jonathan S. Gootenberg, Hiroshi Nishimasu, Osamu Nureki, and Feng Zhang. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517 (7536):583–588, January 2015. ISSN 1476-4687. doi: 10.1038/nature14136.

[187] Helena Pópulo, José Manuel Lopes, and Paula Soares. The mTOR Signalling Pathway in Human Cancer. *International Journal of Molecular Sciences*, 13(2): 1886–1918, February 2012. ISSN 1422-0067. doi: 10.3390/ijms13021886.

[188] Yan-Jie Zhang, Qiang Dai, Dan-Feng Sun, Hua Xiong, Xiao-Qing Tian, Feng-Hou Gao, Mang-Hua Xu, Guo-Qiang Chen, Ze-Guang Han, and Jing-Yuan Fang. mTOR signaling pathway is a target for the treatment of colorectal

cancer. *Annals of Surgical Oncology*, 16(9):2617–2628, September 2009. ISSN 1534-4681. doi: 10.1245/s10434-009-0555-9.

[189] Xiaoyu Qin, Bin Jiang, and Yanjie Zhang. 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle*, 15(6):781–786, March 2016. ISSN 1538-4101. doi: 10.1080/15384101.2016.1151581.

[190] An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature*, 489(7414):57–74, September 2012. ISSN 0028-0836. doi: 10.1038/nature11247.

[191] Marinescu. The MAPPER database: A multi-genome catalog of putative transcription factor binding sites, 2004.

[192] Takeshi Mimura, Natalie Walker, Linda Badri, Amir Lagstein, Anish Wadhwa, and Vibha N. Lama. Foxf1 Transcription Factor In Acute Lung Allograft Rejection. In *A33. NEW FRONTIERS IN LUNG TRANSPLANTATION AND REJECTION*, American Thoracic Society International Conference Abstracts, pages A1289–A1289. American Thoracic Society, May 2013.

[193] Do-Hyung Kim, D. D. Sarbassov, Siraj M. Ali, Jessie E. King, Robert R. Latek, Hediye Erdjument-Bromage, Paul Tempst, and David M. Sabatini. mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell*, 110(2):163–175, July 2002. ISSN 0092-8674.

[194] Kimberly Coffman, Bing Yang, Jie Lu, Ashley L. Tetlow, Emelia Pellicio, Shan Lu, Da-Chuan Guo, Chun Tang, Meng-Qiu Dong, and Fuyuhiko Tamanoi. Characterization of the Raptor/4E-BP1 interaction by chemical cross-linking coupled with mass spectrometry analysis. *The Journal of Biological Chemistry*, 289(8):4723–4734, February 2014. ISSN 1083-351X. doi: 10.1074/jbc.M113.482067.

[195] Justin Lamb, Emily D. Crawford, David Peck, Joshua W. Modell, Irene C. Blat, Matthew J. Wrobel, Jim Lerner, Jean-Philippe Brunet, Aravind Subramanian, Kenneth N. Ross, Michael Reich, Haley Hieronymus, Guo Wei, Scott A. Armstrong, Stephen J. Haggarty, Paul A. Clemons, Ru Wei, Steven A. Carr, Eric S. Lander, and Todd R. Golub. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science (New York, N.Y.)*, 313(5795):1929–1935, September 2006. ISSN 1095-9203. doi: 10.1126/science.1132939.

[196] Srinivas Koduru, Raj Kumar, Sowmyalakshmi Srinivasan, Mark B Evers, and Chendil Damodaran. Notch-1 inhibition by Withaferin-A: A therapeutic target against colon carcinogenesis. *Molecular cancer therapeutics*, 9(1):202–210, January 2010. ISSN 1535-7163. doi: 10.1158/1535-7163.MCT-09-0771.

[197] Francois Ghiringhelli, Boris Guiu, Bruno Chauffert, and Sylvain Ladoire. Sirolimus, bevacizumab, 5-Fluorouracil and irinotecan for advanced colorectal cancer: A pilot study. *World Journal of Gastroenterology : WJG*, 15(34): 4278–4283, September 2009. ISSN 1007-9327. doi: 10.3748/wjg.15.4278.

[198] Nanami Itoh, Shuho Semba, Masafumi Ito, Hiroaki Takeda, Sumio Kawata, and Mitsunori Yamakawa. Phosphorylation of Akt/PKB is required for suppression of cancer cell apoptosis and tumor progression in human colorectal carcinoma. *Cancer*, 94(12):3127–3134, June 2002. ISSN 1097-0142. doi: 10.1002/cncr.10591.

[199] Anjali K Gupta, George J Cerniglia, Rosemarie Mick, Mona S Ahmed, Vincent J Bakanauskas, Ruth J Muschel, and W. Gillies McKenna. Radiation sensitization of human cancer cells in vivo by inhibiting the activity of PI3K using LY294002. *International Journal of Radiation Oncology*Biology*Physics*, 56(3):846–853, July 2003. ISSN 0360-3016. doi: 10.1016/S0360-3016(03)00214-1.

[200] Markus Guba, Philipp von Breitenbuch, Markus Steinbauer, Gudrun Koehl, Stefanie Flegel, Matthias Hornung, Christiane J. Bruns, Carl Zuelke, Stefan Farkas, Matthias Anthuber, Karl-Walter Jauch, and Edward K. Geissler. Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: Involvement of vascular endothelial growth factor. *Nature Medicine*, 8(2):128–135, February 2002. ISSN 1078-8956. doi: 10.1038/nm0202-128.

[201] Daniel J. Boffa, Fulung Luan, Dolca Thomas, Hua Yang, Vijay K. Sharma, Milagros Lagman, and Manikkam Suthanthiran. Rapamycin Inhibits the Growth and Metastatic Progression of Non-Small Cell Lung Cancer. *Clinical Cancer Research*, 10(1):293–300, January 2004. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-0629-3.

[202] Fu L. Luan, Ruchuang Ding, Vijay K. Sharma, W. James Chon, Milagros Lagman, and Manikkam Suthanthiran. Rapamycin is an effective inhibitor of human renal cancer metastasis. *Kidney International*, 63(3):917–926, March 2003. ISSN 0085-2538. doi: 10.1046/j.1523-1755.2003.00805.x.

[203] Zheng Wang, Jian Zhou, Jia Fan, Chang-Jun Tan, Shuang-Jian Qiu, Yao Yu, Xiao-Wu Huang, and Zhao-You Tang. Sirolimus inhibits the growth and metastatic progression of hepatocellular carcinoma. *Journal of Cancer Research and Clinical Oncology*, 135(5):715–722, May 2009. ISSN 0171-5216, 1432-1335. doi: 10.1007/s00432-008-0506-z.

[204] Arkaitz Carracedo, Li Ma, Julie Teruya-Feldstein, Federico Rojo, Leonardo Salmena, Andrea Alimonti, Ainara Egia, Atsuo T. Sasaki, George Thomas, Sara C. Kozma, Antonella Papa, Caterina Nardella, Lewis C. Cantley, Jose Baselga, and Pier Paolo Pandolfi. Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer.

The Journal of Clinical Investigation, 118(9):3065–3074, September 2008. ISSN 0021-9738. doi: 10.1172/JCI34739.

[205] Kathryn E. O'Reilly, Fredi Rojo, Qing-Bai She, David Solit, Gordon B. Mills, Debra Smith, Heidi Lane, Francesco Hofmann, Daniel J. Hicklin, Dale L. Ludwig, Jose Baselga, and Neal Rosen. mTOR inhibition induces upstream receptor tyrosine kinase signaling and activates Akt. *Cancer Research*, 66(3):1500–1508, February 2006. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-05-2925.

[206] Dos D. Sarbassov, Siraj M. Ali, Shomit Sengupta, Joon-Ho Sheen, Peggy P. Hsu, Alex F. Bagley, Andrew L. Markhard, and David M. Sabatini. Prolonged rapamycin treatment inhibits mTORC2 assembly and Akt/PKB. *Molecular Cell*, 22(2):159–168, April 2006. ISSN 1097-2765. doi: 10.1016/j.molcel.2006.03.029.

[207] John J. Arcaroli, Kevin S. Quackenbush, Rebecca W. Powell, Todd M. Pitts, Anna Spreafico, Marileila Varella-Garcia, Lynne Bemis, Aik Choon Tan, Jaclyn M. Reinemann, Basel M. Touban, Arvind Dasari, S. Gail Eckhardt, and Wells A. Messersmith. Common PIK3CA Mutants and a Novel 3′ UTR Mutation Are Associated with Increased Sensitivity to Saracatinib. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18(9), May 2012. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-11-3167.

[208] Robert L. Johnson and James C. Fleet. Animal Models of Colorectal Cancer. *Cancer metastasis reviews*, 32(0):39–61, June 2013. ISSN 0167-7659. doi: 10.1007/s10555-012-9404-6.

[209] Eric W.-F. Lam, Jan J. Brosens, Ana R. Gomes, and Chuay-Yeng Koo. Forkhead box proteins: Tuning forks for transcriptional harmony. *Nature Reviews Cancer*, 13(7):482–495, July 2013. ISSN 1474-175X. doi: 10.1038/nrc3539.

[210] Ji-Hye Paik, Ramya Kollipara, Gerald Chu, Hongkai Ji, Yonghong Xiao, Zhihu

Ding, Lili Miao, Zuzana Tothova, James W. Horner, Daniel R. Carrasco, Shan Jiang, D. Gary Gilliland, Lynda Chin, Wing H. Wong, Diego H. Castrillon, and Ronald A. DePinho. FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. *Cell*, 128(2):309–323, January 2007. ISSN 0092-8674. doi: 10.1016/j.cell.2006.12.029.

[211] Xue-Yuan Dong, Ceshi Chen, Xiaodong Sun, Peng Guo, Robert L. Vessella, Ruo-Xiang Wang, Leland W. K. Chung, Wei Zhou, and Jin-Tang Dong. FOXO1A is a candidate for the 13q14 tumor suppressor gene inhibiting androgen receptor signaling in prostate cancer. *Cancer Research*, 66(14):6998–7006, July 2006. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-06-0411.

[212] Il-Man Kim, Timothy Ackerson, Sneha Ramakrishna, Maria Tretiakova, I.-Ching Wang, Tanya V. Kalin, Michael L. Major, Galina A. Gusarova, Helena M. Yoder, Robert H. Costa, and Vladimir V. Kalinichenko. The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. *Cancer Research*, 66(4):2153–2161, February 2006. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-05-3003.

[213] Yuichi Yoshida, I.-Ching Wang, Helena M. Yoder, Nicholas O. Davidson, and Robert H. Costa. The forkhead box M1 transcription factor contributes to the development and growth of mouse colorectal cancer. *Gastroenterology*, 132(4): 1420–1431, April 2007. ISSN 0016-5085. doi: 10.1053/j.gastro.2007.01.036.

[214] Pang-Kuo Lo, Ji Shin Lee, Hexin Chen, David Reisman, Franklin G. Berger, and Saraswati Sukumar. Cytoplasmic mislocalization of overexpressed FOXF1 is associated with the malignancy and metastasis of colorectal adenocarcinomas. *Experimental and Molecular Pathology*, 94(1):262–269, February 2013. ISSN 1096-0945. doi: 10.1016/j.yexmp.2012.10.014.

[215] Ramon Salazar, Paul Roepman, Gabriel Capella, Victor Moreno, Iris Simon, Christa Dreezen, Adriana Lopez-Doriga, Cristina Santos, Corrie Marijnen, Johan Westerga, Sjoerd Bruin, David Kerr, Peter Kuppen, Cornelis van de Velde, Hans Morreau, Loes Van Velthuysen, Annuska M. Glas, Laura J. Van't Veer, and Rob Tollenaar. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 29(1):17–24, January 2011. ISSN 1527-7755. doi: 10.1200/JCO.2010.30.1077.

[216] Steven Eschrich, Ivana Yang, Greg Bloom, Ka Yin Kwong, David Boulware, Alan Cantor, Domenico Coppola, Mogens Kruhøffer, Lauri Aaltonen, Torben F. Orntoft, John Quackenbush, and Timothy J. Yeatman. Molecular Staging for Survival Prediction of Colorectal Cancer Patients. *Journal of Clinical Oncology*, 23(15):3526–3535, May 2005. ISSN 0732-183X. doi: 10.1200/JCO.2005.00.695.

[217] Meihua Li, Yu-Min Lin, Suguru Hasegawa, Takashi Shimokawa, Kohei Murata, Masao Kameyama, Osamu Ishikawa, Toyomasa Katagiri, Tatsuhiko Tsunoda, Yusuke Nakamura, and Yoichi Furukawa. Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. *International Journal of Oncology*, 24(2):305–312, February 2004. ISSN 1019-6439.

[218] Hyuk-Chan Kwon, Sung-Hyun Kim, Mee-Sook Roh, Jae-Seok Kim, Hyung-Sik Lee, Hong-Jo Choi, Jin-Sook Jeong, Hyo-Jin Kim, and Tae-Ho Hwang. Gene Expression Profiling in Lymph Node-Positive and Lymph Node-Negative Colorectal Cancer. *Diseases of the Colon & Rectum*, 47(2):141–152, February 2004. ISSN 0012-3706, 1530-0358. doi: 10.1007/s10350-003-0032-7.

[219] Efsevia Vakiani, Manickam Janakiraman, Ronglai Shen, Rileen Sinha, Zhaoshi Zeng, Jinru Shia, Andrea Cercek, Nancy Kemeny, Michael D'Angelica, Agnes Viale, Adriana Heguy, Philip Paty, Timothy A. Chan, Leonard B. Saltz, Martin

Weiser, and David B. Solit. Comparative Genomic Analysis of Primary Versus Metastatic Colorectal Carcinomas. *Journal of Clinical Oncology*, 30(24):2956–2962, August 2012. ISSN 0732-183X. doi: 10.1200/JCO.2011.38.2994.

[220] Hideaki Tanami, Hitoshi Tsuda, Satoshi Okabe, Takehisa Iwai, Kenichi Sugihara, Issei Imoto, and Johji Inazawa. Involvement of cyclin D3 in liver metastasis of colorectal cancer, revealed by genome-wide copy-number analysis. *Laboratory Investigation*, 85(9):1118–1129, June 2005. ISSN 0023-6837. doi: 10.1038/labinvest.3700312.

[221] Iain Beehuat Tan, Simeen Malik, Kalpana Ramnarayanan, John R. McPherson, Dan Liang Ho, Yuka Suzuki, Sarah Boonhsui Ng, Su Yan, Kiat Hon Lim, Dennis Koh, Chew Min Hoe, Chung Yip Chan, Rachel Ten, Brian Kp Goh, Alexander Yf Chung, Joanna Tan, Cheryl Xueli Chan, Su Ting Tay, Lezhava Alexander, Niranjan Nagarajan, Axel M. Hillmer, Choon Leong Tang, Clarinda Chua, Bin Tean Teh, Steve Rozen, and Patrick Tan. High-depth sequencing of over 750 genes supports linear progression of primary tumors and metastases in most patients with liver-limited metastatic colorectal cancer. *Genome Biology*, 16:32, February 2015. ISSN 1474-760X. doi: 10.1186/s13059-015-0589-1.

[222] Jeng-Kai Jiang, Yann-Jang Chen, Chi-Hung Lin, I.-Ting Yu, and Jen-Kou Lin. Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases–an analysis by comparative genomic hybridization. *Genes, Chromosomes & Cancer*, 43(1):25–36, May 2005. ISSN 1045-2257. doi: 10.1002/gcc.20167.

[223] Chen Mao, Xin-Yin Wu, Zu-Yao Yang, Diane Erin Threapleton, Jin-Qiu Yuan, Yuan-Yuan Yu, and Jin-Ling Tang. Concordant analysis of KRAS, BRAF, PIK3CA mutations, and PTEN expression between primary colorectal cancer

and matched metastases. *Scientific Reports*, 5:8065, February 2015. ISSN 2045-2322. doi: 10.1038/srep08065.

[224] F. Molinari, V. Martin, P. Saletti, S. De Dosso, A. Spitale, A. Camponovo, A. Bordoni, S. Crippa, L. Mazzucchelli, and M. Frattini. Differing deregulation of EGFR and downstream proteins in primary colorectal cancer and related metastatic sites may be clinically relevant. *British Journal of Cancer*, 100(7): 1087–1094, March 2009. ISSN 0007-0920. doi: 10.1038/sj.bjc.6604848.

[225] Raffaele Baffa, Matteo Fassan, Stefano Volinia, Brian O'Hara, Chang-Gong Liu, Juan P Palazzo, Marina Gardiman, Massimo Rugge, Leonard G Gomella, Carlo M Croce, and Anne Rosenberg. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. *The Journal of Pathology*, 219 (2):214–221, October 2009. ISSN 1096-9896. doi: 10.1002/path.2586.

[226] Carlo Aschele, Domizia Debernardis, Gianni Tunesi, Frank Maley, and Alberto Sobrero. Thymidylate Synthase Protein Expression in Primary Colorectal Cancer Compared with the Corresponding Distant Metastases and Relationship with the Clinical Response to 5-Fluorouracil. *Clinical Cancer Research*, 6(12): 4797–4802, December 2000. ISSN 1078-0432, 1557-3265.

[227] Dong Hyuk Ki, Hei-Cheul Jeung, Chan Hee Park, Seung Hee Kang, Gui Youn Lee, Won Suk Lee, Nam Kyu Kim, Hyun Chul Chung, and Sun Young Rha. Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *International Journal of Cancer*, 121(9):2005–2012, November 2007. ISSN 1097-0215. doi: 10.1002/ijc.22975.

[228] K. H. Koh, H. Rhee, H. J. Kang, E. Yang, K. T. You, H. Lee, B. S. Min, N. K. Kim, S. W. Nam, and H. Kim. Differential Gene Expression Profiles of

Metastases in Paired Primary and Metastatic Colorectal Carcinomas. *Oncology*, 75(1-2):92–101, September 2008. ISSN 1423-0232. doi: 10.1159/000155211.

[229] Rempei Yanagawa, Yoichi Furukawa, Tatsuhiko Tsunoda, Osamu Kitahara, Masao Kameyama, Kohei Murata, Osamu Ishikawa, and Yusuke Nakamura. Genome-Wide Screening of Genes Showing Altered Expression in Liver Metastases of Human Colorectal Cancers by cDNA Microarray. *Neoplasia (New York, N.Y.)*, 3(5):395–401, September 2001. ISSN 1522-8002.

[230] Antonello D'Arrigo, Claudio Belluco, Alessandro Ambrosi, Maura Digito, Giovanni Esposito, Antonella Bertola, Michele Fabris, Valentina Nofrate, Enzo Mammano, Alberta Leon, Donato Nitti, and Mario Lise. Metastatic transcriptional pattern revealed by gene expression profiling in primary colorectal carcinoma. *International Journal of Cancer*, 115(2):256–262, June 2005. ISSN 0020-7136. doi: 10.1002/ijc.20883.

[231] H.-M. Lin, A. Chatterjee, Y.-H. Lin, A. Anjomshoaa, R. Fukuzawa, J. L. McCall, and A. E. Reeve. Genome wide expression profiling identifies genes associated with colorectal liver metastasis. *Oncology Reports*, 17(6):1541–1549, June 2007. ISSN 1021-335X.

[232] M A Pantaleo, A Astolfi, M Nannini, P Paterini, G Piazzi, G Ercolani, G Brandi, G Martinelli, A Pession, A D Pinna, and G Biasco. Gene expression profiling of liver metastases from colorectal cancer as potential basis for treatment choice. *British Journal of Cancer*, 99(10):1729–1734, November 2008. ISSN 0007-0920. doi: 10.1038/sj.bjc.6604681.

[233] Astrid Koehler, Frauke Bataille, Cornelia Schmid, Petra Ruemmele, Annette Waldeck, Hagen Blaszyk, Arndt Hartmann, Ferdinand Hofstaedter, and Wolfgang Dietmaier. Gene expression profiling of colorectal cancer and metas-

tases divides tumours according to their clinicopathological stage. *The Journal of Pathology*, 204(1):65–74, September 2004. ISSN 1096-9896. doi: 10.1002/path.1606.

[234] Stephanie Roessler, Guoling Lin, Marshonna Forgues, Anuradha Budhu, Shelley Hoover, R. Mark Simpson, Xiaolin Wu, Ping He, Lun-Xiu Qin, Zhao-You Tang, Qing-Hai Ye, and Xin Wei Wang. Integrative Genomic and Transcriptomic Characterization of Matched Primary and Metastatic Liver and Colorectal Carcinoma. *International Journal of Biological Sciences*, 11(1):88–98, January 2015. ISSN 1449-2288. doi: 10.7150/ijbs.10583.

[235] Ja-Rang Lee, Chae Hwa Kwon, Yuri Choi, Hye Ji Park, Hyun Sung Kim, Hong-Jae Jo, Nahmgun Oh, and Do Youn Park. Transcriptome analysis of paired primary colorectal carcinoma and liver metastases reveals fusion transcripts and similar gene expression profiles in primary carcinoma and liver metastases. *BMC Cancer*, 16:539, 2016. ISSN 1471-2407. doi: 10.1186/s12885-016-2596-3.

[236] Sinem Karaman and Michael Detmar. Mechanisms of lymphatic metastasis. *The Journal of Clinical Investigation*, 124(3):922–928, March 2014. ISSN 1558-8238. doi: 10.1172/JCI71606.

[237] F. Meier, S. Will, U. Ellwanger, B. Schlagenhauff, B. Schittek, G. Rassner, and C. Garbe. Metastatic pathways and time courses in the orderly progression of cutaneous melanoma. *The British Journal of Dermatology*, 147(1):62–70, July 2002. ISSN 0007-0963.

[238] Thomas F. Gajewski, Hans Schreiber, and Yang-Xin Fu. Innate and adaptive immune cells in the tumor microenvironment. *Nature Immunology*, 14(10): 1014–1022, October 2013. ISSN 1529-2908. doi: 10.1038/ni.2703.

[239] Judith A. McKay, Joy J. Douglas, Val G. Ross, Stephanie Curran, Fareeda Y.

Ahmed, Joseph F. Loane, Graeme I. Murray, and Howard L. McLeod. Expression of Cell Cycle Control Proteins in Primary Colorectal Tumors Does Not Always Predict Expression in Lymph Node Metastases. *Clinical Cancer Research*, 6(3):1113–1118, March 2000. ISSN 1078-0432, 1557-3265.

[240] Yumei Feng, Baocun Sun, Xiaoqing Li, Liang Zhang, Yun Niu, Chunhua Xiao, Liansheng Ning, Zhiyi Fang, Yuli Wang, Lina Zhang, Jing Cheng, Wei Zhang, and Xishan Hao. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. *Breast Cancer Research and Treatment*, 103(3):319–329, July 2007. ISSN 0167-6806, 1573-7217. doi: 10.1007/s10549-006-9385-7.

[241] Mika Suzuki and David Tarin. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: Clinical implications. *Molecular Oncology*, 1(2):172–180, September 2007. ISSN 1574-7891. doi: 10.1016/j.molonc.2007.03.005.

[242] B. Weigelt, L. F. A. Wessels, A. J. Bosma, A. M. Glas, D. S. A. Nuyten, Y. D. He, H. Dai, J. L. Peterse, and L. J. van't Veer. No common denominator for breast cancer lymph node metastasis. *British Journal of Cancer*, 93(8):924–932, September 2005. ISSN 0007-0920. doi: 10.1038/sj.bjc.6602794.

[243] Xishan Hao, Baocun Sun, Limei Hu, Harri Lähdesmäki, Valerie Dunmire, Yumei Feng, Shi-Wu Zhang, Huamin Wang, Chunlei Wu, Hua Wang, Gregory N. Fuller, W. Fraser Symmans, Ilya Shmulevich, and Wei Zhang. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *Cancer*, 100(6):1110–1122, March 2004. ISSN 1097-0142. doi: 10.1002/cncr.20095.

[244] Elizabeth M. Webber, Tia L. Kauffman, Elizabeth O'Connor, and Katrina AB Goddard. Systematic review of the predictive effect of MSI status in colorectal cancer patients undergoing 5FU-based chemotherapy. *BMC Cancer*, 15:156, 2015. ISSN 1471-2407. doi: 10.1186/s12885-015-1093-4.

[245] Dung T. Le, Jennifer N. Uram, Hao Wang, Bjarne R. Bartlett, Holly Kemberling, Aleksandra D. Eyring, Andrew D. Skora, Brandon S. Luber, Nilofer S. Azad, Dan Laheru, Barbara Biedrzycki, Ross C. Donehower, Atif Zaheer, George A. Fisher, Todd S. Crocenzi, James J. Lee, Steven M. Duffy, Richard M. Goldberg, Albert de la Chapelle, Minori Koshiji, Feriyl Bhaijee, Thomas Huebner, Ralph H. Hruban, Laura D. Wood, Nathan Cuka, Drew M. Pardoll, Nickolas Papadopoulos, Kenneth W. Kinzler, Shibin Zhou, Toby C. Cornish, Janis M. Taube, Robert A. Anders, James R. Eshleman, Bert Vogelstein, and Luis A. Jr. Diaz. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Medicine*, 372(26):2509–2520, June 2015. ISSN 0028-4793. doi: 10.1056/NEJMoa1500596.

[246] David Tougeron, Emilie Fauquembergue, Alexandre Rouquette, Florence Le Pessot, Richard Sesboüé, Michèle Laurent, Pascaline Berthet, Jacques Mauillon, Frédéric Di Fiore, Jean-Christophe Sabourin, Pierre Michel, Mario Tosi, Thierry Frébourg, and Jean-Baptiste Latouche. Tumor-infiltrating lymphocytes in colorectal cancers with microsatellite instability are correlated with the number and spectrum of frameshift mutations. *Modern Pathology*, 22(9):1186–1195, June 2009. ISSN 0893-3952. doi: 10.1038/modpathol.2009.80.

[247] Kosuke Yoshihara, Maria Shahmoradgoli, Emmanuel Martínez, Rahulsimham Vegesna, Hoon Kim, Wandaliz Torres-Garcia, Victor Treviño, Hui Shen, Peter W. Laird, Douglas A. Levine, Scott L. Carter, Gad Getz, Katherine Stemke-Hale, Gordon B. Mills, and Roel G. W. Verhaak. Inferring tumour purity and

stromal and immune cell admixture from expression data. *Nature Communications*, 4:2612, October 2013. ISSN 2041-1723. doi: 10.1038/ncomms3612.

[248] Frank A. Sinicrope and Daniel J. Sargent. Clinical implications of microsatellite instability in sporadic colon cancers. *Current Opinion in Oncology*, 21(4):369–373, July 2009. ISSN 1531-703X. doi: 10.1097/CCO.0b013e32832c94bd.

[249] Sebastiano Cavallaro. CXCR4/CXCL12 in Non-Small-Cell Lung Cancer Metastasis to the Brain. *International Journal of Molecular Sciences*, 14(1):1713–1727, January 2013. ISSN 1422-0067. doi: 10.3390/ijms14011713.

[250] JINGHUI GUO, WENHUI LOU, YUAN JI, and SHUNCAI ZHANG. Effect of CCR7, CXCR4 and VEGF-C on the lymph node metastasis of human pancreatic ductal adenocarcinoma. *Oncology Letters*, 5(5):1572–1578, May 2013. ISSN 1792-1074. doi: 10.3892/ol.2013.1261.

[251] Fangfang Liu, Ronggang Lang, Jia Wei, Yu Fan, Lifang Cui, Feng Gu, Xiaojing Guo, Gordon A. Pringle, Xinmin Zhang, and Li Fu. Increased expression of SDF-1/CXCR4 is associated with lymph node metastasis of invasive micropapillary carcinoma of the breast. *Histopathology*, 54(6):741–750, May 2009. ISSN 1365-2559. doi: 10.1111/j.1365-2559.2009.03289.x.

[252] Joseph Kim, Hiroya Takeuchi, Stella T. Lam, Roderick R. Turner, He-Jing Wang, Christine Kuo, Leland Foshag, Anton J. Bilchik, and Dave S. B. Hoon. Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 23(12):2744–2753, April 2005. ISSN 0732-183X. doi: 10.1200/JCO.2005.07.078.

[253] Joseph Kim, Takuji Mori, Steven L. Chen, Farin F. Amersi, Steve R. Martinez, Christine Kuo, Roderick R. Turner, Xing Ye, Anton J. Bilchik, Donald L. Mor-

ton, and Dave S. B. Hoon. Chemokine Receptor CXCR4 Expression in Patients With Melanoma and Colorectal Cancer Liver Metastases and the Association With Disease Outcome. *Annals of Surgery*, 244(1):113–120, July 2006. ISSN 0003-4932. doi: 10.1097/01.sla.0000217690.65909.9c.

[254] Teppei Murakami, Kenji Kawada, Masayoshi Iwamoto, Masatoshi Akagami, Koya Hida, Yuki Nakanishi, Keitaro Kanda, Mayumi Kawada, Hiroshi Seno, Makoto Mark Taketo, and Yoshiharu Sakai. The role of CXCR3 and CXCR4 in colorectal cancer metastasis. *International Journal of Cancer*, 132(2):276–287, January 2013. ISSN 1097-0215. doi: 10.1002/ijc.27670.

[255] Valéry L. Payen, Paolo E. Porporato, Bjorn Baselet, and Pierre Sonveaux. Metabolic changes associated with tumor metastasis, part 1: Tumor pH, glycolysis and the pentose phosphate pathway. *Cellular and molecular life sciences: CMLS*, 73(7):1333–1348, April 2016. ISSN 1420-9071. doi: 10.1007/s00018-015-2098-5.

[256] Ann Hanna and Lalita A. Shevde. Hedgehog signaling: Modulation of cancer properies and tumor mircroenvironment. *Molecular Cancer*, 15:24, 2016. ISSN 1476-4598. doi: 10.1186/s12943-016-0509-3.

[257] Yi-Yang Hu, Min-Hua Zheng, Rui Zhang, Ying-Min Liang, and Hua Han. Notch signaling pathway and cancer metastasis. *Advances in Experimental Medicine and Biology*, 727:186–198, 2012. ISSN 0065-2598.

[258] Eda Yildirim, James E. Kirby, Diane E. Brown, Francois E. Mercier, Ruslan I. Sadreyev, David T. Scadden, and Jeannie T. Lee. Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice. *Cell*, 152(4), February 2013. ISSN 0092-8674. doi: 10.1016/j.cell.2013.01.034.

[259] Sarah M. Weakley, Hao Wang, Qizhi Yao, and Changyi Chen. Expression

and Function of a Large Non-coding RNA Gene XIST in Human Cancer. *World Journal of Surgery*, 35(8):1751–1756, August 2011. ISSN 0364-2313. doi: 10.1007/s00268-010-0951-0.

[260] Hye-Jung Kim and Harvey Cantor. CD4 T-cell Subsets and Tumor Immunity: The Helpful and the Not-so-Helpful. *Cancer Immunology Research*, 2(2):91–98, February 2014. ISSN 2326-6066, 2326-6074. doi: 10.1158/2326-6066.CIR-13-0216.

[261] B. Lores, J. M. García-Estevez, and C. Arias. Lymph nodes and human tumors (review). *International Journal of Molecular Medicine*, 1(4):729–733, April 1998. ISSN 1107-3756.

[262] J W Berg, A G Huvos, L M Axtell, and G F Robbins. A new sign of favorable prognosis in mammary cancer: Hyperplastic reactive lymph nodes in the apex of the axilla. *Annals of Surgery*, 177(1):8–12, January 1973. ISSN 0003-4932.

[263] Anthony D Foster, Amogh Sivarapatna, and Ronald E Gress. The aging immune system and its relationship with cancer. *Aging health*, 7(5):707–718, October 2011. ISSN 1745-509X. doi: 10.2217/ahe.11.56.

[264] Marissa Fessenden. The cell menagerie: Human immune profiling. *Nature*, 525 (7569):409–411, September 2015. ISSN 0028-0836. doi: 10.1038/525409a.

[265] N. Wolmark, H. Rockette, E. Mamounas, J. Jones, S. Wieand, D. L. Wickerham, H. D. Bear, J. N. Atkins, N. V. Dimitrov, A. G. Glass, E. R. Fisher, and B. Fisher. Clinical trial to assess the relative efficacy of fluorouracil and leucovorin, fluorouracil and levamisole, and fluorouracil, leucovorin, and levamisole in patients with Dukes' B and C carcinoma of the colon: Results from National Surgical Adjuvant Breast and Bowel Project C-04. *Journal of Clinical On-*

cology: Official Journal of the American Society of Clinical Oncology, 17(11): 3553–3559, November 1999. ISSN 0732-183X. doi: 10.1200/jco.1999.17.11.3553.

[266] Thierry André, Corrado Boni, Lamia Mounedji-Boudiaf, Matilde Navarro, Josep Tabernero, Tamas Hickish, Clare Topham, Marta Zaninelli, Philip Clingan, John Bridgewater, Isabelle Tabah-Fisch, Aimery de Gramont, and Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) Investigators. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *The New England Journal of Medicine*, 350(23):2343–2351, June 2004. ISSN 1533-4406. doi: 10.1056/NEJMoa032709.

[267] Diogo Assed Bastos, Suilane Coelho Ribeiro, Daniela de Freitas, and Paulo M. Hoff. Combination therapy in high-risk stage II or stage III colon cancer: Current practice and future prospects. *Therapeutic Advances in Medical Oncology*, 2(4):261–272, July 2010. ISSN 1758-8340. doi: 10.1177/1758834010367905.

[268] Michael Pourdehnad, Morgan L. Truitt, Imran N. Siddiqi, Gregory S. Ducker, Kevan M. Shokat, and Davide Ruggero. Myc and mTOR converge on a common node in protein synthesis control that confers synthetic lethality in Myc-driven cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29):11988–11993, July 2013. ISSN 1091-6490. doi: 10.1073/pnas.1310230110.

[269] Jaclyn Sceneay, Mark J. Smyth, and Andreas Möller. The pre-metastatic niche: Finding common ground. *Cancer and Metastasis Reviews*, 32(3-4):449–464, December 2013. ISSN 0167-7659, 1573-7233. doi: 10.1007/s10555-013-9420-1.