

# SMARCB1 Maintains Lineage Fidelity in Clear Cell Renal Cell Carcinoma

*Ludovic Julian Wesolowski*

The University of Cambridge

Darwin College

September 2021

*This thesis is submitted for the degree of Doctor of Philosophy*

# Preface

## *Thesis declaration*

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. Finally, it does not exceed the prescribed word limit for the School of Clinical Medicine Degree Committee.

## *Animal work*

All mouse work, including subcutaneous injection, calliper measurements, dissection and tumour processing were performed by Veronica Caraffini (V.C), Jianfeng Ge (J.G) and Alyson Speed (A.S) with support from the Anne McLaren Laboratory of Regenerative Medicine animal facility and the Department of Pathology. All analysis and plotting were the result of my own work.

## *Bioinformatic support*

The pre-processing (trimming, aligning and peak calling) of FASTQ files for ATAC-seq experiments were performed by Shoko Hirose (S.H). Analysis and plotting of scRNA-seq data displayed in the Introduction of this thesis was also performed by S.H. The future work discussed at the end of this thesis was the result of a collaboration with Debora Sesia (D.S) from the Ciriello group at the Swiss Institute of Bioinformatics. Specifically, D.S performed the permutation based statistical analysis to identify changes in the dependency profiles of different categories of cell lines.

## Abstract

Lineage-specific transcription factors have emerged as a promising class of essential genes in cancer. The best examples of leveraging this phenomenon in the clinic is targeting the androgen receptor and the oestrogen receptor in prostate and breast cancer respectively. Despite the success of these therapies, the mechanisms that maintain lineage fidelity in advanced cancer clones, and whether lineage factor pathways could be exploited in other cancer types remain poorly understood. Using clear cell renal cell carcinoma (ccRCC) as a model, I characterise mechanisms that underlie lineage factor dependence in cancer. Using CRISPR/Cas9 loss-of-function screening coupled with *in vitro* and *in vivo* validation I show that the loss of SMARCB1, a member of the SWI/SNF chromatin remodelling complex, confers an advantage to ccRCC cells upon inhibition of the essential renal lineage factor PAX8. *SMARCB1* knockout (KO) leads to large-scale loss of a kidney-specific enhancer program, conversion to a cellular state resembling that of rhabdoid tumours, and the re-activation of proliferative pathways. Using a second CRISPR/Cas9 screen, I show that these proliferative pathways are underpinned by the acquisition of new transcriptional dependencies. These dependencies represent rare essential genes across different lineage-specific and oncogenic pathways, a principle validated in a large-scale CRISPR/Cas9 screening data set comprising hundreds of cancer cell lines. In summary, dependence on tissue-specific lineage factors in cancer can be modulated via epigenetic remodelling.

## Acknowledgements

Firstly, I would like to thank Sakari Vanharanta and Saroor Patel for excellent scientific mentoring and technical support. Due to their efforts, I developed hugely as a researcher through a challenging four years. I would also like to thank all the additional current and past members of the lab for their support including, Jianfeng Ge, Veronica Caraffini, Anna Dyas, Leti Castillon, Alyson Speed, Shoko Hirose, Paulo Rodrigues, Saiful Effendi Bin Syafruddin, Muhammad Nazhif Zaini, Emma Richardson, Erika Vojtasova and Anabel Zelceski. A special thank you to Paulo for technical support and interesting scientific discussions, and to Jianfeng for experimental support especially during holidays/absences.

The work contained within this thesis would not have been possible without the support of the core facilities teams at the Hutchison MRC Research Centre. Thank you to the porters, media kitchen, glass wash, finance, human resources, building administration and health and safety teams for supporting this work. In addition, the excellent flow cytometry facility at the Cambridge Institute for Medical Research and the Genomic core at the Cancer Research UK Cancer Institute enabled the execution of complex/important experiments which were pivotal for this project.

# Contents

<b>1</b>	<b>The Introduction</b>	<b>15</b>
1.1	An introduction to cancer	16
1.1.1	Sustaining proliferative signalling and evasion of growth suppression	16
1.1.2	Resistance to cell death	17
1.1.3	Enabling replicative immortality	18
1.1.4	Angiogenesis	19
1.1.5	Invasion and metastasis	20
1.1.6	Reprogramming Energy/ Metabolism	22
1.1.7	Evading Immune Destruction	23
1.1.8	Lineage plasticity in cancer progression and therapy resistance	24
1.1.9	Summary of the cancer hallmarks	26
1.2	Introduction to kidney cancer	27
1.2.1	Clear cell renal cell carcinoma	29
1.2.2	Genetics of ccRCC	31
1.2.3	Metabolic re-wiring in ccRCC	36
1.2.4	ccRCC Therapy	40
1.3	Transcription factors in cancer	46
1.3.1	Regulation of transcription	47
1.3.2	Transcriptional deregulation in cancer	50
1.3.3	Lineage addiction	51
1.4	Identifying lineage specific transcription factor dependencies in kidney cancer	53
1.5	Problem statement	57
<b>2</b>	<b>Materials and methods</b>	<b>59</b>
2.1	Cell lines and cell culture	60
2.2	Drug treatment	60
2.3	Plasmids	60
2.4	Lentiviral transduction	62
2.5	Immunoblotting	62
2.6	Reverse transcription and quantitative polymerase chain reaction (qPCR)	63
2.7	In vitro proliferation assays	63
2.8	Fluorescent activated cell sorting and analysis	64
2.9	RNA-seq	64
2.10	ATAC-seq	65
2.11	ATAC-seq analysis	66

2.12	scRNAseq analysis .....	67
2.13	Animal studies.....	68
2.14	Pooled CRISPR-Cas9 screening.....	68
2.15	The encyclopedia of DNA elements (ENCODE) analysis .....	69
2.16	Cancer cell line encyclopedia (CCLE).....	70
2.17	DepMap analysis.....	70
2.18	Pan-cancer Lineage dependency (LD) analysis .....	71
2.19	Statistical analysis .....	72
<b>3</b>	<b>Identifying modulators of PAX8 dependency in ccRCC .....</b>	<b>73</b>
3.1	Introduction .....	74
3.2	Main .....	75
3.3	Summary .....	86
<b>4</b>	<b>SMARCB1 inactivation promotes large scale chromatin remodelling and cellular instability.....</b>	<b>88</b>
4.1	Introduction .....	89
4.2	Main .....	90
4.3	Summary .....	100
<b>5</b>	<b>Reprogramming to resist.....</b>	<b>102</b>
5.1	Introduction .....	103
5.2	Main .....	103
5.3	Summary .....	116
<b>6</b>	<b>The Discussion .....</b>	<b>118</b>
6.1	Lineage plasticity as a resistance mechanism.....	119
6.2	Acquired transcriptional dependencies replace PAX8 signalling.....	124
6.3	Heterogeneity and lineage plasticity .....	126
6.4	Targeting PAX8 in ccRCC .....	129
6.5	Future work.....	131
6.5.1	Supporting experiments and open questions.....	131
6.5.2	Generalisability of the study - an experimental approach .....	133
6.5.3	Generalisability of the study - an informatic approach .....	134
<b>7</b>	<b>Closing Remarks.....</b>	<b>148</b>
<b>8</b>	<b>References.....</b>	<b>150</b>

## List of figures

Figure 1.....	21
Figure 2.....	28
Figure 3.....	30
Figure 4.....	32
Figure 5.....	38
Figure 6.....	48
Figure 7.....	52
Figure 8.....	54
Figure 9.....	55
Figure 10.....	56
Figure 11.....	75
Figure 12.....	76
Figure 13.....	77
Figure 14.....	78
Figure 15.....	79
Figure 16.....	80
Figure 17.....	81
Figure 18.....	82
Figure 19.....	83
Figure 20.....	84
Figure 21.....	85
Figure 22.....	86
Figure 23.....	89
Figure 24.....	91
Figure 25.....	92
Figure 26.....	93
Figure 27.....	94
Figure 28.....	95
Figure 29.....	97
Figure 30.....	98
Figure 31.....	99
Figure 32.....	100
Figure 33.....	103
Figure 34.....	105
Figure 35.....	106
Figure 36.....	107
Figure 37.....	108
Figure 38.....	109
Figure 39.....	110
Figure 40.....	111
Figure 41.....	111
Figure 42.....	112
Figure 43.....	113
Figure 44.....	114

Figure 45 .....	115
Figure 46 .....	117
Figure 47 .....	121
Figure 48 .....	126
Figure 49 .....	135
Figure 50 .....	136
Figure 51 .....	137
Figure 52 .....	139
Figure 53 .....	142
Figure 54 .....	145
Figure 55 .....	146

## List of tables

Table 1.....	61
Table 2.....	61
Table 3.....	61
Table 4.....	138
Table 5.....	141

## List of abbreviations

<b>4E-BP1</b>	4E-binding protein 1	<b>MM</b>	Multiple myeloma
<b>ACTL</b>	Actin-like	<b>MOI</b>	Multiplicity if infection
<b>Acyl-CoA</b>	Acyl-Coenzyme A	<b>mRNA</b>	Messenger RNA
<b>AML</b>	Acute myeloid leukaemia	<b>MRT</b>	Malignant rhabdoid tumours
<b>AP-1</b>	Activator protein 1	<b>mSigDB</b>	Molecular signatures database
<b>APML</b>	Promyelocytic leukaemia	<b>mTOR</b>	Mammalian target of rapamycin
<b>APML</b>	Acute promyelocytic leukaemia	<b>mTORC1</b>	mTOR Complex 1
<b>ARID1A</b>	AT-Rich Interaction Domain 1A	<b>MYEOV</b>	Myeloma overexpressed gene
<b>ARNT</b>	Aryl Hydrocarbon Receptor Nuclear Translocator	<b>NB</b>	Neuroblastoma
<b>ASS1</b>	Argininosuccinate synthase 1	<b>ncBAF</b>	non-canonical BAF
<b>AT/RT</b>	Atypical teratoid rhabdoid tumours	<b>NE</b>	Neuroendocrine
<b>ATAC-seq</b>	Assay for Transposase-Accessible Chromatin using sequencing	<b>NGS</b>	Next generation sequencing
<b>ATCC</b>	American Type Culture Collection	<b>NHL</b>	Non-Hodgkin lymphoma
<b>ATP</b>	Adenosine triphosphate	<b>NHR</b>	Nuclear hormone receptor
<b>ATRA</b>	All-trans retinoic acid	<b>NKX2.1</b>	NK2 homeobox 1
<b>BAF</b>	BRG1/BRM associated factor	<b>NOXA</b>	Phorbol-12-myristate-13-acetate-induced protein 1
<b>B-ALL</b>	B cell acute lymphoblastic leukaemia	<b>NR2F1</b>	Nuclear receptor subfamily 2 group F member 1
<b>BAP1</b>	BRCA1 Associated Protein 1	<b>NR3C1</b>	Nuclear Receptor Subfamily 3 Group C Member 1
<b>BC</b>	Bladder carcinoma	<b>NSCLC</b>	Non-small cell lung cancer
<b>BCL</b>	B-cell lymphoma	<b>NTC</b>	Non-targeting control
<b>BHLHE40</b>	Basic Helix-Loop-Helix E40	<b>OA</b>	Ovarian adenocarcinoma
<b>BHLHE41</b>	Basic helix-loop-helix e41	<b>ODD</b>	Oxygen dependant domain

BORIS	Brother of the Regulator of Imprinted Sites	P/S	Penicillin-streptomycin
BRC	Breast carcinoma	P8 <sub>1</sub>	PAX8 shRNA-1
BRD	Bromodomain containing	P8 <sub>1</sub> Ctrl(A)	shPAX8-1, sgNTC, acutely Dox treated
BrD	Bromodomain	P8 <sub>1</sub> Ctrl(LT)	shPAX8-1, sgNTC, 2-3 month Dox treated
BRDC	Breast ductal carcinoma	P8 <sub>1</sub> Ctrl(MT)	shPAX8-1, sgNTC, 1-2 month Dox treated
CCA	Cholangiocarcinoma	P8 <sub>1</sub> S1 <sub>1</sub> (A)	shPAX8-1, sgSMARCB1-1, acutely Dox treated
CCLE	Cancer cell line encyclopaedia	P8 <sub>1</sub> S1 <sub>1</sub> (A)	shPAX8-2, sgSMARCB1-1, acutely Dox treated
CCND1	Cyclin D1	P8 <sub>1</sub> S1 <sub>1</sub> (LT)	shPAX8-1, sgSMARCB1-1, 2-3 month Dox treated
ccRCC	Clear cell renal cell carcinoma	P8 <sub>1</sub> S1 <sub>1</sub> (MT)	shPAX8-1, sgSMARCB1-1, 1-2 month Dox treated
cCREs	Candidate cis-regulatory regions	P8 <sub>1</sub> S1 <sub>2</sub> (A)	shPAX8-1, sgSMARCB1-2, acutely Dox treated
CDH1	Cadherin 1	P8 <sub>1</sub> S1 <sub>2</sub> (A)	shPAX8-2, sgSMARCB1-2, acutely Dox treated
CDH6	Cadherin 6	P8 <sub>1</sub> S1 <sub>2</sub> (LT)	shPAX8-1, sgSMARCB1-2, 2-3 month Dox treated
CDK	cyclin dependant kinase	P8 <sub>1</sub> S1 <sub>2</sub> (MT)	shPAX8-1, sgSMARCB1-2, 1-2 month Dox treated
CDKN2A	Cyclin-dependent kinase inhibitor 2A	P8 <sub>2</sub>	PAX8 shRNA-2
cDNA	Complementary DNA	padj	Adjusted <i>P value</i>
CHD1	Conserved chromatin remodelling and assembly factor	PAS	Per-Arnt-Sim
chrCC	Chromophobe renal cell carcinoma	PAX	Paired Box
CNBD1	Cyclic Nucleotide Binding Domain Containing 1	PBAF	Polybromo-associated BAF complexes
CNV	Copy number variation	PBRM1	Polybromo 1
CPT1A	Carnitine palmitoyl transferase	PCC	Pearson's correlation coefficient
CRISPR	Clustered regularly interspaced short palindromic repeats	PCR	Polymerase chain reaction
CRPC	Castration resistant prostate cancer	PCT	Proximal convoluted tubule
CRPC-NE	Castration resistant prostate cancer with neuroendocrine features	PD	Programmed cell death
CTCF	CCCTC-binding factor	PDGF	Platelet-derived growth factor

CTLA-4	Cytotoxic T-lymphocyte-associated-antigen	PDGFB	Platelet-derived growth factor subunit B
Ctrl.Ctrl(A)	shRenilla, sgNTC, acutely Dox treated	PDK	Phosphorylation-dependent kinase
CXCL14	C-X-C Motif Chemokine Ligand 14	PHD	Prolyl hydroxylase domain
CXCR4	C-X-C Motif Chemokine Receptor 4	PHOX2A	Paired mesoderm homeobox protein 2A
D-2HG	D-2-hydroxyglutarate	PI3K	Phosphatidylinositol 3-kinase
DA	Differentially accessible	PIC	Pre-initiation complex
DE	Differentially expressed	PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
DepMap	Dependency map	PIP2	Phosphatidylinositol (4, 5)-biphosphate
DKK1	Dickkopf WNT Signalling Pathway Inhibitor 1	PIP3	Phosphatidylinositol (3, 4, 5)-triphosphate
DNA	Deoxyribonucleic Acid	PML	Promyelocytic leukaemia protein
DNase	Deoxyribonuclease	PML	Promyelocytic leukaemia protein
DNMT	DNA methyltransferase	POU2AF1	POU domain class 2-associating factor 1
Dox	Doxycycline	POU3F2	POU Class 3 Homeobox 2
DPF1	Double PHD Fingers 1	PRAD	Prostate adenocarcinoma
EAC	Endometrial adenocarcinoma	PRC2	Polycomb repressor complex 2
EAC	Endometrial adenocarcinoma	pRCC	Papillary renal cell carcinoma
EGFR	Epidermal growth factor receptor	PRDM1	PR/SET Domain 1
eIF4E	Eukaryotic Translation Initiation Factor 4E	PTEN	Phosphatase and tensin
EMT	Epithelial to mesenchymal transition	PUMA	p53 upregulated modulator of apoptosis
ENCODE	The Encyclopaedia of DNA Elements	qPCR	Quantitative polymerase chain reaction
EPAS1	Endothelial PAS domain-containing protein 1	Q-Q plot	Quantile-quantile plots
EPC	Exocrine pancreatic cancer	RARA	Retinoic acid receptor alpha
ER	Estrogen receptor	RB	Retinoblastoma protein
eRNA	Enhancer RNA	RCC	Renal cell carcinoma

esBAF	Embryonic stem cell-specific SWI/SNF complex	RHEB	Ras homolog enriched in brain
ESCC	Oesophageal squamous cell carcinoma	RMS	Rhabdomyosarcoma
ESCs	Embryonic stem cells	RNA	Ribonucleic acid
ETS1	ETS Proto-Oncogene 1	RNA-seq	RNA sequencing
EWS	Ewing sarcoma	ROS	Reactive oxygen species
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit	RTK	Receptor tyrosine kinase
FAS	Fas Cell Surface Death Receptor)	RUNX1	Runt-related transcription factor 1
FBP1	Fructose-1,6-bisphosphatase 1	S <sub>1</sub>	SMARCB1 sgRNA-1
FBS	Foetal bovine serum	S <sub>2</sub>	SMARCB1 sgRNA-2
FDR	False discovery rate	S6K	S6 kinase
FKBP5	FKBP Prolyl Isomerase 5	SCD1	Stearoyl-CoA desaturase
FLI1	Friend leukaemia integration 1 transcription factor	SCLC	Small cell lung cancer
GA	Gastric adenocarcinoma	SCNAs	Somatic copy number alteration
GATA3	GATA Binding Protein 3	scRNA-seq	Single-cell RNA-seq
GDP	Guanosine diphosphate	SDF1	Stromal cell-derived factor 1
GLUT1	Glucose transporter 1	SDF1	Stromal cell-derived factor 1
GSEA	Gene set enrichment analysis	SERDs	Selective estrogen receptor degraders
GTE <sub>x</sub>	Genotype-Tissue Expression	SERMs	Selective estrogen receptor modulators
GTP	Guanosine triphosphate	SETD2	SET domain containing 2
GWAS	Genome-wide association studies	SFRP1	Secreted frizzled-related protein 1 precursor
H3K27 <sub>ac</sub>	Acetylation of lysine 27 on histone H3	SFRP1	Secreted frizzled-related protein 1 precursor
HA	Higher accessibility	sgRNA	Single guide RNA
HATs	Histone acetyltransferases	shRNA	short hairpin RNA
HCC	Hepatocellular carcinoma	SLC16A7	Solute Carrier Family 16 Member 7

HDAC1	Histone Deacetylase 1	SMARC	Matrix Associated, Actin Dependent Regulator Of Chromatin
HDACs	Histone deacetylases	SNPs	Single nucleotide polymorphisms
HIF	Hypoxia inducible factor	SOX	SRY-related HMG-box
HNF	Hepatocyte nuclear factor	SQSTM1	Sequestosome 1
HRE	HIF responsive element	SS18L1	Synovial sarcoma translocation gene on chromosome 18-like 1
HSCs	Haemopoietic stem cells	STC2	Stanniocalcin 2
HSP90	Heat shock protein 90	SWI/SNF	Switch/sucrose non-fermentable
ICI	Immune checkpoint inhibitors	SYP	Synaptophysin
IDO	Indoleamine 2,3-dioxygenase	TADs	Topologically associated domains
IDO-1	Indoleamine 2,3-Dioxygenase 1	T-ALL	T cell acute lymphoblastic leukaemia
IFN	Interferon	TBX	T-box
IL	Interleukin	TCA	Tricarboxylic acid
IPS	Induced pluripotent stem cell	TCF3	Homozygous transcription factor 3 gene
IRF2	Interferon regulatory factor 2	TCGA	The Cancer Genome Atlas
IRF4	Interferon regulatory factor 4	TEAD	TEA Domain Transcription Factor
JMJ	Jumonji	TET	Ten eleven translocation
KDM5C	Lysine demethylase 5C	TF	Transcription factor
KLF6	Krüppel-like factor 6	TK	Tyrosine kinase
KLK2	Krüppel-like factor 2	TLR	Toll-like receptor
KLK3	Kallikrein Related Peptidase 3	TMPRSS2	Transmembrane protease serine 2
KO	Knockout	TNF	Tumour Necrosis Factor
L2HGDH	L-2-Hydroxyglutarate Dehydrogenase	TPM	Transcripts per million
LA	Lower accessibility	TSC	TSC Complex Subunit
LD	Lineage dependency	TSG	Tumour suppressor gene

LMT	Lung mesothelioma	TSP-1	Thrombospondin 1
lncRNA	Long non-coding RNA	TSS	Transcriptional start site
LOF	Loss-of-function	UATN	Upper aerodigestive tract neoplasm
LOH	Loss of heterozygosity	UQCRH	Ubiquinol-Cytochrome C Reductase Hinge Protein
mAb	Monoclonal antibody	VCAN	Versican
MAPK	Mitogen-activated protein kinase	VEGF-A	Vascular Endothelial Growth Factor A
MECP2	Methyl CpG binding protein 2	VEGFR	Vascular endothelial growth factor receptor
MED12	Mediator Complex Subunit 12	VHL	Von Hippel-Lindau syndrome
MED12	Mediator of RNA polymerase II transcription	WES	Whole-exome sequencing
Mel	Melanoma	WNT	Wingless-related integration site
miR	MicroRNA	ZEB1	Zinc Finger E-Box Binding Homeobox 1
MITF	Melanocyte inducing transcription factor	ZNFX1	Zinc finger NFX1-type containing 1

1

## The Introduction

## 1.1 An introduction to cancer

Cancer is an evolutionary process that describes the conversion of normal cells into a state of uncontrolled proliferation and survival. The substrates for this process are genetic and epigenetic alterations, caused for example by oncogenic viruses, environmental mutagens, and erroneous cellular replication. Broadly speaking tumours can be classified as either benign or malignant, depending on whether the tumour is localised or invasive/spreading<sup>1</sup>. Most cancers can be further classified into epithelial (carcinoma), mesenchymal (sarcoma), hematopoietic or neuroectodermal, according to their tissue of origin<sup>1</sup>. Cancer that bears little to no histological resemblance to a tissue of origin is said to be anaplastic<sup>1</sup>. To develop, cancer must acquire traits to overcome significant roadblocks, including but not limited to regulated mitogenic signalling, mortality, energetics, immune surveillance, and physical space. This is achieved through the manipulation of existing cellular pathways, which are best summarised as the hallmarks of cancer<sup>2,3</sup>.

### 1.1.1 Sustaining proliferative signalling and evasion of growth suppression

Cell growth and division is a tightly controlled process essential for development and, the turnover and homeostasis of adult tissues. It is regulated through the action of growth factors and their cognate cell surface receptors, cell-cell contact signalling, and membrane-permeable messengers (such as hormones)<sup>4</sup>. These pathways can be de-regulated in cancer through hyperactivation, for example, activating mutations in BRAF occur in ~40% of melanoma patients, which triggers the constitutive activation of the growth-promoting mitogen-activated protein kinase (MAPK) pathway<sup>5</sup>. Alternatively, negative feedback mechanisms which function to dampen mitogenic signalling can be lost in cancer, for example, the proto-oncogene rat sarcoma virus (Ras). In its Guanosine-5'-triphosphate (GTP)-bound active form, Ras functions to signal via the MAPK pathway. In doing so it hydrolyses the GTP to

Guanosine-5'-diphosphate (GDP) and becomes inactive, an example of intrinsic negative feedback. Mutations in Ras can interfere with its hydrolytic activity and prevent the reversion to a GDP-bound inactive state, thereby becoming constitutively active<sup>1,6</sup>. Other examples of growth factor signalling that can be co-opted in cancer include the phosphoinositide 3-kinase (PI3K) -AKT, the mammalian target of rapamycin (mTOR), patched-smoothed, and canonical and non-canonical Wnt signalling<sup>1,4</sup>. These pathways do not operate in isolation, instead, there is cross-talk and they can be simultaneously triggered by the same cell surface receptor/environmental cue<sup>1,4</sup>. In addition to sustained proliferative signalling, cancer cells must overcome strong negative regulators of proliferation, which are known as tumour suppressor genes. Tumour suppressor genes tend to be inactivated through loss-of-function (LOF) mutations; a prototypical example is the gene retinoblastoma protein (*RB*)<sup>2</sup>. Alongside tumour protein 53 (TP53), RB is a critical arbiter of proliferative, senescence, or apoptotic programs. Progression through the stages of the cell cycle (growth 1, synthesis, growth 2, and mitosis) is mediated by cyclin dependant kinases and cyclins. In response to mitogenic signalling a cell can progress through ~2/3 of growth 1 before reaching a restriction point (R point). RB incorporates extracellular and intracellular signals to determine whether a cell should progress through the R point and continue the cycle<sup>4,7</sup>. Cancer cells with RB LOF mutations are thus missing a critical gatekeeper of cell-cycle progression<sup>7</sup>.

### 1.1.2 Resistance to cell death

In response to certain stimuli, cells can undergo a process of programmed cell death known as apoptosis. Apoptosis is a necessary process in the development and also maintenance of healthy adult tissue<sup>4</sup>. In response to various internal and external physiologic stressors, the balance of anti-apoptotic proteins (i.e. B-cell lymphoma 2; BCL-2 and BCL-extra-large; BCL-XL) and pro-apoptotic proteins (i.e. Phorbol-12-myristate-13-acetate-induced protein 1; NOXA and p53 upregulated modulator of

apoptosis; PUMA) can shift to instigate a signalling cascade culminating in caspase-mediated proteolysis, cellular disassembly, and eventual phagocytosis<sup>8</sup>. In the process of tumorigenesis, cancer cells are exposed to a range of physiological stresses which can induce apoptosis<sup>8</sup>. For example, the tumour suppressor TP53 can act to sense DNA damage caused by hyperproliferation or defects in the DNA repair machinery<sup>9</sup>. In response to DNA breaks and other chromosomal abnormalities, TP53 upregulates PUMA and NOXA. PUMA and NOXA bind to anti-apoptotic proteins, which leads to the loss of mitochondrial integrity and the release of cytochrome-c. Cytochrome-c is then capable of inducing caspase activity and hence cellular degradation<sup>9</sup>. The most common route to overcome this process is via the loss of TP53 tumour suppressor function, indeed *TP53* is the most commonly mutated tumour suppressor gene in cancer<sup>10</sup>.

### 1.1.3 Enabling replicative immortality

Larger, longer-lived animals have a greater number of cells exposed over a longer period to mutagens and hence in principle should have a higher risk of acquiring oncogenic mutations and cancer. However, in practice, this is not the case<sup>11</sup>. A study of 15 mammalian species, demonstrated an inverse correlation between life span and mutational rate, such that the mutational burden at the end of life was consistent between species<sup>12</sup>. Further, Mice have >1000 times fewer cells and a >30 times shorter life span and yet the incidence rate of cancer in aged mice is approximately 90%, whereas in humans it is closer to 25%<sup>11</sup>. Larger organisms have evolved mechanisms capable of safeguarding against the heightened DNA mutation and cancer risk to reduce the chance of pre-reproductive mortality<sup>11</sup>. One such mechanism is to restrict the number of replicative cycles most somatic cells can undergo before the induction of senescence followed by crisis and apoptosis. This is achieved by limiting the expression of telomerase, an enzyme responsible for lengthening multiple tandem hexanucleotide repeats sequences (telomeres) at the end of chromosomes<sup>11,13</sup>. Telomeres shorten progressively during

sequential cell divisions until they reach a length where they can no longer protect chromosomes from end-to-end fusion events. The formation of dicentric chromosomes results in karyotype scrambling, senescence and apoptosis<sup>4</sup>. In order to form a macroscopic tumour, cancer cells need to acquire unlimited replicative potential, termed immortalisation. The aberrant expression of telomerase has been detected in ~90% of immortalised cell lines, including cancer cells<sup>3</sup>. To date, a diverse array of mechanisms for the reactivation of telomerase have been described, including epigenetic modifications, promoter mutations and the induction or suppression of positive and negative regulators respectively<sup>14</sup>. For example, recurrent mutations in the promoter of the telomerase gene (TERT) have been reported in melanoma patients, and have since been shown to positively regulate telomerase expression in model systems<sup>15-17</sup>. Similar mutations have been detected in other cancer types, for example, mutations in the TERT promoter and 5'UTR have also been described in renal cancer and have been associated with telomere length maintenance<sup>18</sup>.

#### 1.1.4 Angiogenesis

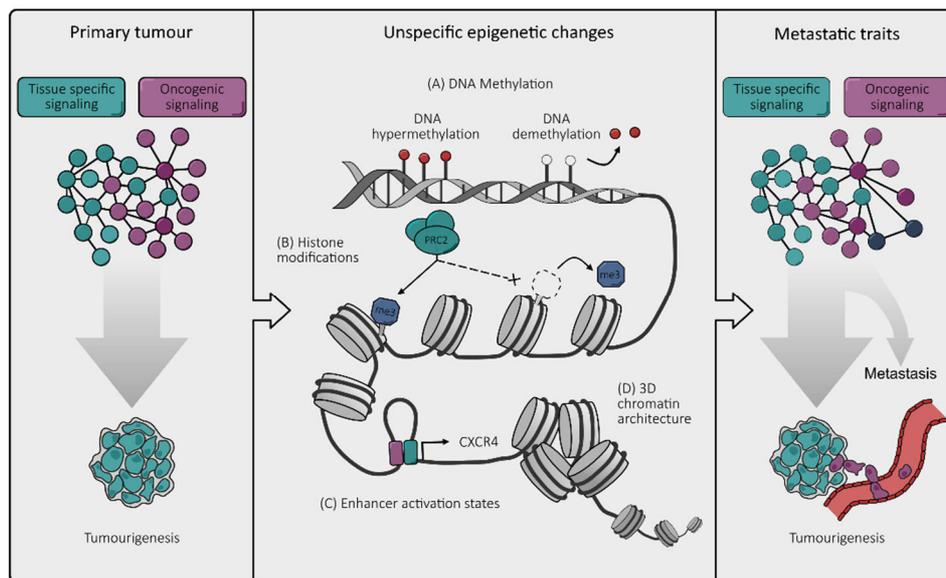
Tumours do not form in isolation, but in a microenvironment populated by vasculature and a host of immune and stromal cells<sup>19</sup>. Similarly, to normal tissue, cancers need vasculature for the delivery of nutrients and oxygen as well as the removal of metabolic waste and carbon dioxide<sup>1,4</sup>. To grow to a size beyond approximately 1-2mm<sup>3</sup> a tumour needs to generate new vasculature to ensure viable growth conditions in an otherwise hypoxic centre<sup>20</sup>. The generation of new blood vessels (vasculogenesis) and the creation of new branches from an already existing blood vessel (angiogenesis) are tightly regulated processes that occurs predominantly during embryogenesis. In adults, the generation of new vasculature is largely limited to the process of wound healing and the female reproductive cycle<sup>3</sup>. In these instances, angiogenesis is triggered transiently via an 'angiogenic switch'. The angiogenic switch is controlled by a balance of pro and anti-angiogenic factors, for example, vascular endothelial growth

factor-A (VEGF-A) and thrombospondin-1 (TSP-1), respectively<sup>20</sup>. In cancer the angiogenic switch is constantly on, supporting the formation of large macroscopic tumours<sup>3</sup>. For example, ~98% of clear cell renal cell carcinomas (ccRCC) have biallelic inactivation of Von Hippel-Lindau syndrome (*VHL*), which leads to the stabilisation of HIF2A protein, which in turn positively regulates the expression of *VEGF-A*<sup>21</sup>. Treatment with the vascular endothelial growth factor receptor (VEGFR) tyrosine kinase inhibitor sunitinib leads to tumour regression and has been approved as a first-line treatment for metastatic ccRCC patients<sup>22</sup>.

### 1.1.5 Invasion and metastasis

The spread of cancer from the primary tumour to secondary sites either in the same organ or a secondary organ is called metastasis. Metastatic spread is the leading cause of cancer-related deaths worldwide, in part because it makes surgical resection very challenging<sup>23</sup>. Multiple routes towards metastasis formation have been described. A simple example, cells disseminate from the primary tumour as groups or individual cells, enter the blood or lymphatic vasculature, extravasate at a distal site, survive in a foreign microenvironment and establish a secondary growth<sup>24</sup>. Despite often being depicted as a neat multi-step cascade, metastasis is a highly inefficient process, for example only 0.02% of melanoma cells injected into the portal vein of mice developed into macro-metastases, and even isolated metastatic cell lines suffer extensive attrition during colonisation<sup>25,26</sup>. A diverse range of cellular mechanisms that enable successful metastasis has been described, including actin cytoskeleton rearrangement, the co-option of immune cells from the microenvironment and extracellular matrix degradation<sup>27-29</sup>. Despite these efforts, the regulation of metastatic traits remains poorly understood. Large-scale bulk and multi-region sequencing efforts have failed to identify driver mutations that are specific to cancer metastases<sup>23</sup>. Nevertheless, specific transcriptional signatures correlate with metastasis in several cancer types and cancer cell lines with stable, highly metastatic phenotypes, can

be isolated from multiple different human cancers<sup>30</sup>. Emerging evidence suggests that metastatic traits are not acquired through new mutations but rather the expansion or tuning of oncogenic signalling, achieved through non-specific epigenetic modifications<sup>30</sup> (**Figure 1**). For example, in ccRCC loss of Polycomb repressive complex 2 (*PRC2*) histone H3 Lys27 trimethylation marks enables the VHL-HIF-dependant upregulation of chemokine (C-X-C motif) receptor 4 (*CXCR4*), a potent driver of metastasis<sup>31</sup>.



**Figure 1** | Cell-type specific and genetically activated signalling pathways conspire to drive the oncogenic program (left). The output of this collaborative program can be modulated through changes in the epigenome, including (a) DNA methylation, (b) covalent histone modifications, (c) chromatin accessibility and (d) higher order chromatin changes. These changes are largely unspecific, thus the phenotypes that emerge are determined by the active oncogenic program. If a particular trait conveys a growth/survival advantage in a particular context, for example a metastatic bottleneck, it is selected for. In this way the oncogenic program that facilitated primary tumour formation evolves through unspecific epigenetic alterations to enable metastatic spread. Figure taken from Patel et al<sup>23</sup>.

### 1.1.6 Reprogramming Energy/ Metabolism

To facilitate growth, proliferating cancer cells need to generate enough energy to support cell replication, satisfy the demands of anabolic macromolecular biosynthesis and establish redox homeostasis in the face of elevated reactive oxygen species production<sup>1</sup>. This is achieved through metabolic re-wiring, most often downstream of oncogenic signalling<sup>3,32</sup>. The prototypical example is the Warburg effect<sup>32</sup>. Under aerobic conditions, cells metabolise glucose to pyruvate (aerobic glycolysis), which in turn undergoes oxidative phosphorylation to produce the maximal amount of ATP per glucose molecule. In a low oxygen environment, pyruvate produced by glycolysis is instead reduced to lactate (anaerobic glycolysis), which can be used as a substrate for the citric acid cycle<sup>4</sup>. In 1956 Otto Warburg observed that even in the presence of oxygen, cancers upregulate anaerobic glycolysis, prioritising reducing glucose into lactate, which is paradoxically ~18 less efficient in terms of ATP production than complete oxidative phosphorylation<sup>32-34</sup>. Although a universally accepted rationale for the Warburg effect remains elusive, there are multiple possible explanations. In the presence of sufficient glucose, the prioritisation and upregulation of anaerobic glycolysis can produce more molecules of ATP in a given time frame<sup>35</sup>. Further, the extra glycolytic intermediates produced can be diverted into biosynthetic pathways required for the assembly of new cellular organelles<sup>35</sup>.

In addition to glucose metabolism, the increased uptake and utilisation of glutamine is also frequently observed in cancer<sup>36</sup>. Glutamine is imported into the mitochondria, where it is converted into glutamate and an ammonium ion by the enzyme glutaminase<sup>36</sup>. Glutamate can then be further metabolised to  $\alpha$ -ketoglutarate, which can be utilised in the tricarboxylic acid (TCA) cycle for ATP production. Although proliferating and cancer cells do use glutamine for energy production, the majority of  $\alpha$ -ketoglutarate is used for biomass production for cellular replication<sup>36</sup>. The carbon from glutamine can be utilised for the production of amino and fatty acids, and the nitrogen contributes to *de novo* biosynthesis of nucleotides (purines and pyrimidines)<sup>36-39</sup>.

As discussed, changes in metabolic programs downstream of mitogenic and nutrient-sensing signalling pathways help support deregulated cell division in cancer primarily through altered energetics and biomass production<sup>40</sup>. However, metabolic re-wiring is not simply a passive consequence of the oncogenic programme, certain metabolites can in turn influence signalling dysregulation, so-called oncometabolites<sup>40</sup>. Arguably the best characterised example of an oncometabolite is D-2-hydroxyglutarate (D-2HG)<sup>40</sup>. Certain point mutations in the enzyme isocitrate dehydrogenase (*IDH*), most frequently observed in gliomas and leukaemia's, results in neomorphic activity to produce D-2HG<sup>40-42</sup>. D-2HG is structurally similar to  $\alpha$ -ketoglutarate, and hence competitively inhibits enzymes that use  $\alpha$ -ketoglutarate, such as the ten eleven translocation (TET) family of 5-methylcytosine hydroxylases and Jumonji (JMJ) family of histone demethylases<sup>40,43</sup>. High levels of D-2HG affect the function of TET and JMJ family members, resulting in CpG island and histone hypermethylation, which has been associated with mitogenic signalling and disease progression<sup>41,44-47</sup>.

### 1.1.7 Evading Immune Destruction

Over the last two decades, the role of the immune system in the development and maintenance of the cancer phenotype has become a major player in modern-day cancer therapeutics<sup>48,49</sup>. Initially, the role of the immune system in the detection and removal of premalignant cells, termed immunosurveillance, was described and has since gained experimental support<sup>50,51</sup>. Cancers raised in immunodeficient mice were unable to give rise to tumours in their immunocompetent counterparts, however, the reciprocal was true<sup>51,52</sup>. The incidence of tumorigenesis in response to carcinogens was also increased in immunodeficient mice<sup>51,52</sup>. Further, immunodeficient transplant patients have a drastically higher incidence of cancer<sup>53</sup>. However, the vast majority of these cancers could be attributed to oncogenic viruses, suggesting a role for the immune system specifically in suppressing viral-induced tumours<sup>3,53</sup>. Of note, unlike the immunodeficient mice, Immunosuppressed patients have major deficiencies in B

and T cell compartments but are still able to mount an immune response through natural killer cells and other innate immune cells, leaving room for interpretation<sup>3</sup>.

Despite immunosurveillance, cancers are still able to develop. Beyond the removal of early pre-malignant cells, the immune system is involved throughout the multistage process of tumorigenesis and metastasis and is now leveraged as a new therapeutic paradigm<sup>48,49,51</sup>. The prevailing theory for the involvement of the immune system in cancer – immune editing, is characterised by three phases: elimination, equilibrium and escape<sup>51</sup>. Elimination is the removal of pre-malignant cells through the activity of both the innate and adaptive immune systems. Cancer cells with low immunogenicity survive elimination and are maintained in a slow-cycling or quiescent state. This is the longest of the three phases, possibly lasting years before the development of detectable tumours<sup>54</sup>. During equilibrium, cancer cells acquire new somatic alterations from which immune suppressing traits can be selected. The final phase of immune editing, escape, describes the emergence of cancer clones capable of circumventing immune suppression and developing into macroscopic tumours<sup>51</sup>. Multiple escape mechanisms have been described; a downregulation of antigen presentation mediated by cell-intrinsic or extrinsic adaptations, resistance to immune cell-mediated killing, and finally suppression of immune cell activation through the direct expression/secretion of immune-modulatory factors or the recruitment of immunosuppressive inflammatory cells<sup>55–57</sup>. Circumventing immune escape mechanisms and the boosting of immune-mediated killing forms the basis of modern-day immunotherapy, and holds great promise towards the clinical management of cancer<sup>49</sup>.

#### 1.1.8 Lineage plasticity in cancer progression and therapy resistance

Although lineage plasticity is not one of the cancer hallmarks put forward by Hanahan and Weinberg, it has been widely associated with cancer stem cells, metastatic disease, and more recently as a shared pathway towards therapeutic resistance<sup>58,59</sup>. The best characterised contribution of lineage plasticity in

cancer is the epithelial to mesenchymal transition (EMT), which has been reported in a range of cancer lineages. The EMT is a normal physiological transition, required during various stages of embryogenesis and tissue repair, and wound healing<sup>58,60,61</sup>. Conventionally, an EMT describes the downregulation of epithelial markers (e.g. E-cadherin) and the upregulation of mesenchymal genes (e.g. Zinc Finger E-Box Binding Homeobox 1; *ZEB1*, *SNAIL*, and *SLUG*), towards a state of enhanced migratory and metastatic potential<sup>1,58</sup>. Different EMT programs have been described and cells can exist in primed transitional states along the epithelial to mesenchymal spectrum<sup>62-65</sup>. The EMT process is readily reversible (i.e. mesenchymal to epithelial transition - MET), which in theory is required for metastatic colonisation at distal sites<sup>58</sup>. Interestingly, in a *TP53/P TEN* mutant model for breast cancer, it was shown that almost all founding clones could exist in an epithelial or mesenchymal state, supporting a model whereby most tumour cells in a tumour have inherent lineage plasticity<sup>66</sup>. Although the exact contribution of the EMT to metastasis and cancer progression is a subject for debate, it remains a shared characteristic across cancer types that correlates with disease progression<sup>58,67</sup>.

Lineage plasticity can also cause resistance to a range of cancer therapeutics, including chemotherapy, targeted therapies, and immune therapy<sup>59</sup>. *SNAIL* and *SLUG* expression have been functionally linked to chemotherapy resistance in breast and ovarian cancer, through antagonising p53-mediated apoptosis and inducing a stem-like state<sup>68,69</sup>. Similar de-differentiation mechanisms have also been shown to confer resistance to MAPK inhibition in melanoma and a hedgehog antagonist in basal cell carcinoma, through the activation of alternative pathways which can support cancer cell survival<sup>70-73</sup>. In a mouse model for acquired resistance to adoptive cytotoxic T-cell transfer therapy in melanoma, the pro-inflammatory cytokine, tumour necrosis factor (TNF)- $\alpha$ , induced a loss of melanocytic antigens through a reversible de-differentiation mechanisms, thereby circumventing T-cell mediated killing<sup>74,75</sup>. In addition to EMT and de-differentiation, lineage transdifferentiation has also been reported to lead to therapy resistance. The histological transformation from adenocarcinoma to high-grade neuroendocrine triggers therapy resistance in both prostate and breast cancer patients treated with androgen depletion and anti-epidermal growth factor receptor (EGFR) therapy respectively<sup>59</sup>. The key

molecular drivers of this process are *RB1* and *TP53* mutations, which enable transformation through the activity of MYC, AKT-mTOR, and SOX2 signalling<sup>59,76–78</sup>. Notably, tumours presenting with neuroendocrine histology and molecular features (i.e. expression of Chromogranin A; *CHGA* and Synaptophysin; *SYP*), also have an upregulation of stemness related factors (i.e. *SOX2* and *HOXA2*), implying a functional similarity to the aforementioned EMT and de-dedifferentiation based resistance mechanisms<sup>79</sup>.

### 1.1.9 Summary of the cancer hallmarks

The hallmarks of cancer as described by Hanahan and Weinberg neatly summarises common traits required for malignant transformation; the deregulation of mitogenic signalling, resistance to cell death, replicative immortality, recruitment of new vasculature, altered metabolism, immune escape/co-option, and the colonisation of distal sites<sup>3</sup>. In accordance with Darwinian evolution, hallmark traits are acquired through the selection of certain genomic and epigenetic alterations, which provide a growth and survival advantage. The link between DNA mutations and cancer hallmarks has been exhaustively demonstrated and has provided sound logic for the rational design of therapeutic interventions<sup>3</sup>. However, despite extensive bulk and multi-regional DNA sequencing, the association between some cancer traits and discrete genomic alterations has not been made. Instead, it is becoming clear that complex phenotypes in cancer, such as the metastatic cascade, can be controlled by heritable epigenetic alterations, such as aberrant enhancer activation and changes in DNA methylation<sup>23,30</sup>.

Although the above summary focussed largely on cell-intrinsic mechanisms, it is important to consider that cancer is not a cell-autonomous disease, but rather requires a compendium of heterotypic interactions with a host of accessory cell types, collectively referred to as the cancer organ or the tumour microenvironment<sup>3</sup>. For example, the formation of macroscopic lesions requires the

recruitment of endothelial cells for angiogenesis, and certain immune populations have been implicated in metastasis by facilitating cancer cell dissemination and intravasation through the secretion of various bioactive molecules<sup>80,81</sup>.

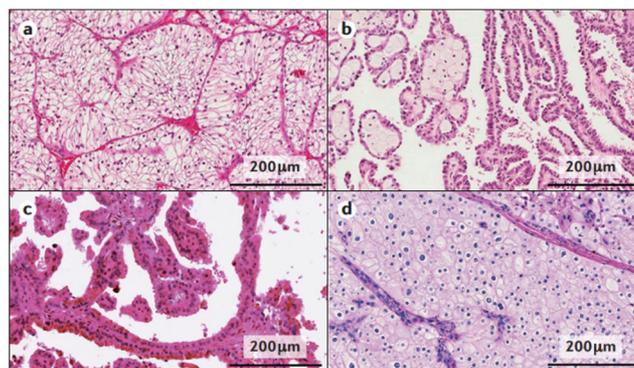
As mentioned, cancer hallmark traits are selected from genetic and epigenetic alterations. Certain enabling characteristics of cancer cells can create more diversity for selection, such as genomic instability. Defects in the DNA repair or DNA damage sensing machinery can lead to an increase in the number of mutations acquired through successive DNA replications, heightening the chance of acquiring an oncogenic alteration<sup>3</sup>. Similarly, mutations in key epigenetic regulators can destabilise the 3D architecture of the genome, allowing for suppression of tumour suppressor genes or the activation of previously dormant oncogenes<sup>82</sup>. Both genetic and epigenetic instability are by nature unspecific, but by creating heterogeneity they enable the selection of advantageous gene programs to support tumour progression and adapt to environmental challenges.

Although lineage plasticity is not one of the original hallmarks, it has an emerging role in acquired resistance across a broad range of therapeutic modalities in several cancer types, in my opinion qualifying it has an emerging hallmark<sup>59</sup>. Alternatively, given the contribution of lineage plasticity to other hallmarks (e.g. metastasis and immune evasion) it could be classified as an enabling characteristic alongside genomic and epigenetic instability. Irrespective, a better understanding of the molecular drivers of lineage infidelity could have far-reaching implications for the design of novel therapeutics and the war on acquired resistance<sup>83</sup>.

## 1.2 Introduction to kidney cancer

Kidney cancer is the 7<sup>th</sup> most common cancer in the UK, approximately 13,000 new cases are reported annually accounting for 4% of all new cancer cases<sup>84</sup>. The incidence rate is higher in men compared to

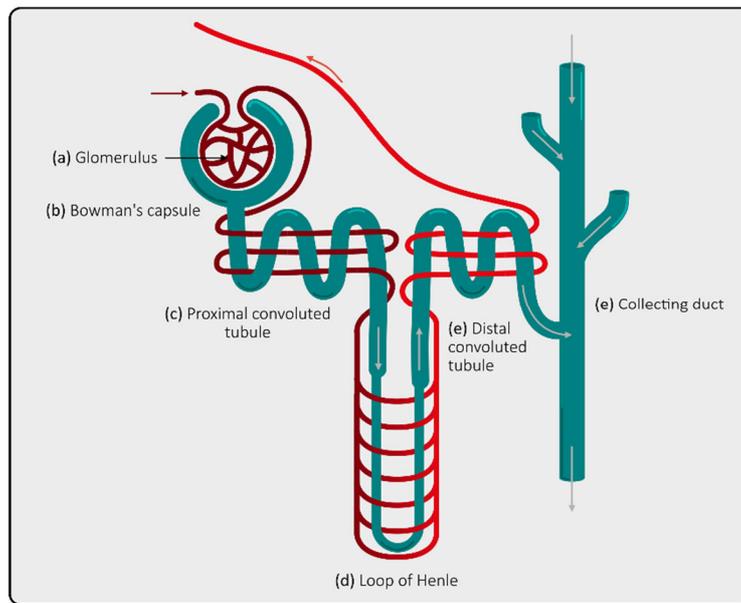
women and more economically developed countries compared to their less developed counterparts<sup>84,85</sup>. Over the last decade, the number of kidney cancer diagnoses has increased by ~1/3, likely due to the advent and implementation of more sensitive screening approaches<sup>84</sup>. In line with this, the incidence rate appears to have plateaued in the last few years<sup>86</sup>. In the UK, there are 4,600 kidney cancer-related deaths, accounting for 3% of all annual cancer deaths<sup>84</sup>. Kidney cancer can be broadly classified by the cell type of origin; renal cell carcinoma (RCC) arises from the kidney epithelium and accounts for ~90% of cases and, transitional cell carcinoma occurs in the renal pelvis and makes up 5-10% of all kidney cancers<sup>87,88</sup>. RCC encompasses >10 histological and molecular subtypes, of which the most common three are, clear cell RCC (ccRCC), papillary RCC (pRCC), and chromophobe RCC (chRCC)(**Figure 2**)<sup>89,90,90</sup>. In children, kidney cancer accounts for ~7% of all cancer diagnoses. Unlike in adults, RCC accounts for a small minority (~1%), instead, the majority of diagnosed paediatric kidney cancers are Wilm's tumours (~90%), clear cell sarcoma (~2-3%), and malignant rhabdoid tumours (~1-2%)<sup>91</sup>.



**Figure 2** | Three main histological subtypes of RCC. **(a)** ~75% of RCC are ccRCC. **(b-c)** ~15% of RCC are pRCC, which histologically appear as either **(b)** basophilic or **(c)** eosinophilic. **(d)** ~5% of RCC are chRCC. Figure taken from Hsieh et al<sup>90</sup>.

### 1.2.1 Clear cell renal cell carcinoma

The vast majority (75-85%) of RCC arising in the kidney epithelium is of the clear cell phenotype. The 'clear cell' refers to the accumulation of lipid and glycogen in the cytoplasm of tumour cells, which appear as clear droplets. In line with these observations, ccRCC is often considered a metabolic disease due to widespread metabolic re-programming and the convergence of disease-causing mutations on key metabolic pathways, such as the nutrient-sensing PI3K-mTOR pathway<sup>92,93</sup>. Marker gene based classification (derived from RNAseq of micro dissected kidney structures) and a more recent scRNA-seq study have identified the proximal convoluted tubule (PCT) as the most probable origin of ccRCC (**Figure 3**)<sup>94,95</sup>. The majority of RCC cases are sporadic, however, 2-4% are associated with inherited tumour syndromes<sup>96</sup>. The most frequent example of a heritable disease predisposing to ccRCC is the von Hippel–Lindau syndrome, which is characterised by the germline inactivation of one allele of the *VHL* gene, increasing the likelihood of the biallelic loss of VHL, which is a hallmark initiating event in ccRCC development (discussed below)<sup>97</sup>. Interestingly, high-grade RCC tumours (including ccRCC) can present with rhabdoid and sarcomatoid features. Rhabdoid and sarcomatoid renal carcinoma histologically and transcriptionally resemble a de-differentiated state, are resistant to targeted therapies, and correlate with poor patient outcome<sup>98</sup>.

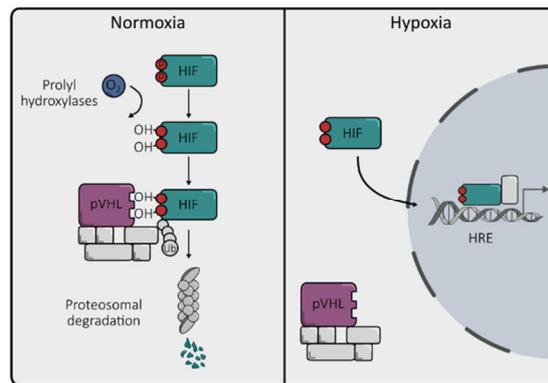


**Figure 3** | A schematic of the nephron. The kidneys, along with the ureter, bladder, prostate (in men) and urethra, form the urinary system, which primarily serves to filter metabolic waste, maintain an ion balance and regulate water levels in the blood. The kidney is at the interface between the blood and the rest of the urinary system. Electrolytes, fatty acids, amino acids, and other metabolic products are either retained and reabsorbed, or they are filtered and excreted by specialised structures called nephrons. Upon reaching the kidney, renal arteries divide into incrementally smaller vessels to form **(a)** the glomerulus, a portal exchange microvascular bed, which is in contact with the Bowman's capsule. Pressure from the blood flow forces water and ions (filtrate) out of the blood into the nephron via the capsule. The filtrate travels from the capsule into **(b)** the proximal convoluted tubule (PCT), **(c)** the loop of Henle, and **(d)** distal convoluted tubule (DCT) before reaching the **(e)** collecting duct. As the filtrate passes along the nephron, nutrients and water are reabsorbed and waste products such as urea and ammonia are retained and excreted into the collecting duct, through a regulated process of active and passive diffusion. Figure adapted from [www.nagwa.com](http://www.nagwa.com)<sup>99</sup>.

## 1.2.2 Genetics of ccRCC

### 1.2.2.1 VHL-HIF

As mentioned previously, ccRCC is characterised by the biallelic inactivation of the *VHL* gene, achieved through point mutations, deletions, and epigenetic silencing<sup>100–102</sup>. *VHL* inactivation is a truncal mutational event in ccRCC, occurring in >90% of patient tumours. VHL forms a functional E3 ubiquitin ligase complex with elongin B/C, cullin 2, and RING-box protein 1 (henceforth known as the VHL complex), which can recognise and polyubiquitinate the oxygen degradation domain (ODD) of hypoxia-inducible factor (HIF) $\alpha$ , targeting it for proteasomal degradation (**Figure 4**). In the presence of oxygen (normoxia), HIF $\alpha$  is hydroxylated on conserved prolines within the ODD domain by prolyl hydroxylase domain (PHD)-containing enzymes, which is necessary and sufficient for recognition by the VHL complex. Thus, the proteasomal degradation of HIF $\alpha$  only occurs in the presence of oxygen. Whereas under hypoxic conditions, or in the event of VHL inactivation, HIF $\alpha$  is no longer prolyl-hydroxylated and so is not recognised by VHL. Stabilised HIF $\alpha$  protein can therefore heterodimerise with aryl hydrocarbon receptor nuclear translocator (ARNT), translocate to the nucleus, bind to HRE DNA elements, and trigger the transcriptional activation of hypoxia-inducible genes (**Figure 4**)<sup>4</sup>.



**Figure 4** | The regulation of HIF by VHL. Under normoxic conditions, two proline residues on HIF are hydroxylated by prolyl hydroxylases, enabling recognition by the VHL complex, followed by ubiquitination and proteasomal degradation. In hypoxic conditions, the proline residues on HIF are not hydroxylated and therefore is not recognised by the VHL complex. Non-degraded HIF can translocate to the nucleus, functionally dimerise, and bind to HIF responsive DNA elements (HRE) to regulate gene expression. Figure adapted from Alberts et al<sup>4</sup>.

There are two isoforms of *HIF $\alpha$* , *HIF1 $\alpha$* , and *HIF2 $\alpha$* , which interestingly seem to have antagonistic roles in tumour development. *HIF1 $\alpha$*  has been shown to reduce the tumorigenic competence of ccRCC models by inhibiting mitochondrial biogenesis and cellular respiration<sup>103,104</sup>. The *HIF1 $\alpha$*  gene resides on chromosome 14 (14q23.2) and may explain, at least in part, the observed frequency of 14q deletions in ccRCC, however, this has not been definitively proven<sup>104,105</sup>. In contrast, there is also evidence that *HIF1 $\alpha$*  can promote tumour development, for example through the re-wiring of glucose metabolism<sup>106,107</sup>. Further, the expression of *HIF1 $\alpha$*  protein is detected in ~70% of ccRCC tumours and the gene is infrequently biallelically inactivated<sup>106</sup>. In contrast, *HIF2 $\alpha$*  has been established as a key oncogenic driver in ccRCC, by promoting tumour angiogenesis, cellular proliferation and survival, metabolic reprogramming, metastasis, and therapy resistance<sup>108</sup>. This is achieved through the regulation of target genes, such as; platelet-derived growth factor (*PDGF*)- $\beta$ , transforming growth factor (*TGF*)- $\alpha$ , *c-Met*, cyclin D1 (*CCND1*), and stromal cell-derived factor 1 (*SDF1*) and its receptor *CXCR4*<sup>109</sup>.

In line with this, inhibition of HIF2 $\alpha$  using the small molecule inhibitor PT2977 (Belzutifan) has been FDA approved as a therapy for the treatment of VHL-associated tumours arising in the kidney<sup>110</sup>. Interestingly, PT2977 has shown efficacy in the treatment of rarer non-renal VHL-associated tumours including central nervous system hemangioblastomas and pancreatic neuroendocrine tumours, highlighting the importance of this signalling axis in tumorigenesis<sup>110</sup>. Currently FDA approval is restricted to VHL-associated tumours in the kidney, pancreas and central nervous system but in the future this may be expanded to include additional organs such as the eye<sup>110</sup>.

#### 1.2.2.2 Copy number and single nucleotide variation

*VHL* loss is not sufficient to trigger a neoplastic transformation, as evidenced by the long latency of tumour initiation in patients harbouring *VHL* germline mutations (30-40 years) and the observation that the loss of VHL in mice models does not lead to tumorigenesis, suggesting that additional genetic alterations are required<sup>97,111,112</sup>. Genome-wide integrated analysis of copy number variations (CNV) and transcriptomics have identified a range of deletions and amplifications, associated with putative oncogenes and tumour suppressors<sup>113,114</sup>. The most frequently observed CNV (~91% of patients), is the deletion of a 43 megabase region of chromosome 3p, which contains the putative tumour suppressor genes *VHL*, Polybromo 1 (*PBRM1*), BRCA1 associated protein-1 (*BAP1*), and SET domain containing 2 (*SETD2*) (discussed below)<sup>115</sup>. The aforementioned genes are typically inactivated by a loss of heterozygosity (LOH) event followed by a somatic mutation in the remaining copy<sup>18,116,117</sup>. The next most frequently deleted region is the loss of 14q (49%), however, the contribution of this CNV to ccRCC progression is less clear. The deletion of one copy of *HIF1 $\alpha$*  and L-2-Hydroxyglutarate Dehydrogenase (*L2HGDH*) is thought to underly the selective advantage of 14q loss<sup>104,118-120</sup>. However, ~25 additional genes which are frequently deleted, and whose low expression correlates with a worse prognosis, are yet to be experimentally validated<sup>105</sup>. Whilst 3p and 14q deletions are associated with the loss of

putative tumour suppressor genes, the amplification of certain genomic regions is associated with the activation of oncogenes<sup>100</sup>. For example, the amplification of a region containing ~60 genes on chromosome 5q35 is observed in ~67% of ccRCC patients<sup>100</sup>. Multiple lines of evidence have linked 5q gains to the increased mRNA expression of Sequestosome 1 (*SQSTM1*), Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (*EZH2*), Stanniocalcin 2 (*STC2*), and Versican (*VCAN*), which have in turn been attributed oncogenic functions in ccRCC through *in vitro* and *in vivo* mechanistic studies<sup>114,121–123</sup>. Multiple comprehensive genome-wide association studies (GWAS) in ccRCC patient tumours have identified susceptibility loci and single nucleotide polymorphisms (SNPs) associated with ccRCC incidence and prognosis<sup>124–126</sup>. For example, the 12p12.1 risk allele was mapped to rs7132434, a functional variant in an enhancer for the gene single nucleotide polymorphisms (SNPs), which was subsequently shown to increase Basic helix-loop-helix e41 (*BHLHE41*) expression through enhanced (Activator protein 1) AP-1 binding and contribute towards oncogenic signalling through the induction of Interleukin 11 (*IL-11*) expression<sup>124,125</sup>. Another notable example includes a SNP located in the 11q13.3 susceptibility locus, situated intergenically flanking myeloma overexpressed gene (*MYEOV*) and *CCND1*<sup>126</sup>. The identified rs7948643 variant impaired the recruitment of HIF2 $\alpha$  to a newly identified *CCND1* enhancer, reducing the risk of developing ccRCC. Interestingly the SNP does not alter the HIF2 $\alpha$  DNA motif and so likely affects the binding of a co-factor capable of recruiting/ stabilising HIF2 $\alpha$  at the *CCND1* enhancer<sup>127</sup>.

### 1.2.2.3 DNA methylation and somatic alterations

DNA hypermethylation correlates with poor patient prognosis in all three major subtypes of RCC. In ccRCC, as mentioned previously, a copy of *VHL* can be lost through aberrant DNA methylation in ~7% of cases<sup>100</sup>. A further 289 other genes are methylated in at least 5% of tumours, the most significant of which include cyclin-dependent kinase inhibitor 2A (*CDKN2A*; p16), Ubiquinol-Cytochrome C Reductase

Hinge Protein (*UQCRH*), and the wingless-related integration site (WNT) pathway regulators secreted frizzled-related protein 1 precursor (*SFRP1*) and Dickkopf WNT Signaling Pathway Inhibitor 1 (*DKK1*)<sup>100,128</sup>. Interestingly, the epigenetic silencing of *SFRP1* and *DKK1* correlates with patient outcome, although the mechanism remains poorly understood<sup>129–131</sup>. Abnormal DNA hypermethylation profiles have been linked to somatic alterations in the non-redundant histone H3 lysine 36 trimethyltransferase, *SETD2*, which may contribute towards tumorigenesis by introducing intra-tumoral heterogeneity, through lineage plasticity and/or genomic instability<sup>100,128</sup>. In addition to *SETD2* (mutated in 12% of cases), significantly mutated genes in ccRCC include *PBRM1* (38%), *BAP1* (10%), *MTOR* (8%), *KDM5C* (5%), *ARID1A* (3.5%), *PTEN* (3%), and *TP53* (3%)<sup>100,128</sup>. Interestingly, *PBRM1* and *BAP1* mutations are largely mutually exclusive, and broadly characterise two major mutational subtypes of ccRCC with different prognoses<sup>132</sup>. The exact contribution of the inactivation of the chromatin modifiers, *SETD2*, *PBRM1*, *BAP1*, and *KDM5C* is yet to be fully elucidated<sup>90</sup>. *MTOR* and *PTEN* mutations lead to the dysregulation of the PI3K-AKT-mTOR pathway, a key metabolic signalling cascade, discussed in more detail below<sup>123</sup>. As mentioned previously in the cancer hallmarks section, *TP53* is a key tumour suppressor gene (TSG), whose loss-of-wildtype-function supports cancer progression. Unlike most TSGs, the majority of *TP53* mutations are missense single residue substitutions. Collectively these mutations are referred to as mutp53, they include alterations that can interrupt canonical target gene activation, exert trans-dominant repression over the wild-type counterpart, or confer neomorphic activity<sup>133</sup>. Without biallelic inactivation, which is common for most TSGs (e.g. *PBRM1*, *SETD2*, etc), mutp53 can exert a range of pro-tumorigenic effects, including metabolic re-wiring to support uncontrolled proliferation<sup>133</sup>.

## Intra-tumoral genetic heterogeneity

Multiregional sequencing efforts have allowed for the construction of tumour phylogenies, whereby ubiquitous mutations form the ‘trunk’ of the tree, and mutations found in some subclones make up the ‘branches’<sup>134</sup>. In ccRCC, the inactivation of *VHL* is a truncal event. By contrast, oncogenic drivers such as *PBRM1*, *BAP1*, and *SETD2* mutations are examples of ‘branched’ or subclonal mutations<sup>117,135</sup>. In a recent study, 1,206 regions from 101 primary ccRCC tumours were profiled using a bespoke panel consisting of 110 putative ccRCC driver genes. This analysis revealed seven deterministic evolutionary subtypes of ccRCC characterised by specific combinations of driver genes, which correlated with different intrinsic tumour properties and clinical outcomes. Within these subtypes the parallel evolution of *BAP1*, *PTEN*, and *SETD2* mutations, and SCNAs (e.g. 14q loss) was observed, emphasising their importance in ccRCC development<sup>135</sup>. Interestingly, the comparison of matched metastatic and primary tumours revealed that the cognate metastases harboured fewer subclonal mutations and were enriched for SCNAs including 9p21.3 and 14q loss and 8q24.1 gain<sup>136,137</sup>. Taken together, these studies have provided an insight into the genomic architecture of primary and metastatic ccRCC cancers, demonstrating the complex and heterogeneous nature of the disease.

### 1.2.3 Metabolic re-wiring in ccRCC

#### 1.2.3.1 Glucose metabolism

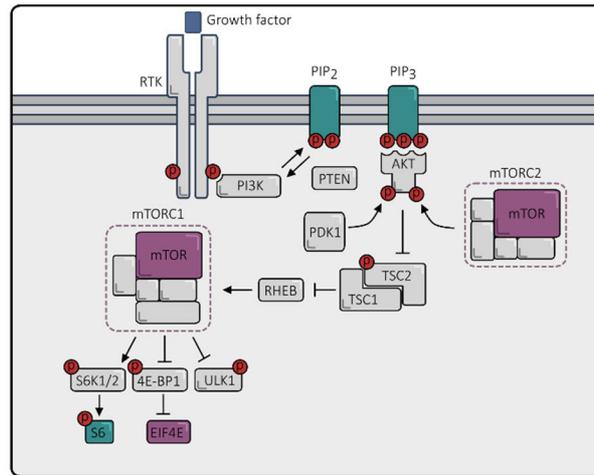
The Warburg effect is an archetypal example of metabolic re-wiring, commonly observed in a range of malignancies. In ccRCC, the increased expression of the glucose transporter 1 (*GLUT1*), as well as glycolytic metabolites and enzymes, suggests an upregulation of glucose metabolism for lactate fermentation – a hallmark of the Warburg effect<sup>138</sup>. Further, the expression of fructose-1,6-

bisphosphatase 1 (*FBP1*), a negative regulator of glycolysis, is downregulated in ccRCC tissue and its exogenous re-introduction inhibits tumour growth in a xenograft model<sup>139</sup>. The inactivation of *VHL* and de-regulation of HIF $\alpha$  expression is thought to be a critical driver of the glycolytic switch in ccRCC<sup>138</sup>. HIF1 $\alpha$  promotes the expression of GLUT1 for enhanced glucose import, as well as the glycolytic enzymes hexokinase 1 and 2 and glyceraldehyde 3-phosphate dehydrogenase<sup>140–142</sup>. Further HIF1 $\alpha$  upregulates the enzyme lactate dehydrogenase (LDH), increasing the conversion of pyruvate to lactate, thus shifting glucose metabolism away from the TCA cycle<sup>143,144</sup>. Finally, the microRNA miR-210 is highly expressed in ccRCC in a HIF1 $\alpha$  dependent manner and has been demonstrated to negatively affect mitochondrial respiration<sup>145–147</sup>.

#### 1.2.3.2 PI3K-AKT-mTOR pathway

Two of the most significantly mutated genes in ccRCC are *MTOR* and *PTEN*, which are positive and negative regulators of the PI3K-AKT-mTOR pathway respectively. In addition, the positive regulator PIK3CA and the negative regulator TSC Complex Subunit 1/2 (*TSC1/2*) are also frequently mutated in ccRCC<sup>100</sup>. In sum, the PI3K-AKT-mTOR pathway is hyperactivated in the majority of ccRCC patients and is mutated in approximately 28% of cases<sup>22,100,148,149</sup>. A simplified overview of the signalling pathway is summarised in **Figure 5**. In response to growth factor binding, receptor tyrosine kinases (RTK) dimerization, and transphosphorylation, PI3K is recruited and activated (through phosphorylation) at the cell surface. Activated PI3K catalyses the conversion of phosphatidylinositol (4, 5)-biphosphate (PIP2) to phosphatidylinositol (3, 4, 5)-triphosphate (PIP3), which in turn triggers the phosphorylation and activation of Akt via phosphorylation-dependent kinase (PDK)-1. Activated AKT can activate or inhibit a plethora of downstream targets involved in multiple cancer hallmarks. An important target of Akt in ccRCC are *TSC1/2*. AKT phosphorylates *TSC1/2* thereby inactivating it and preventing the hydrolysis of GTP- Ras homolog enriched in brain (RHEB). Constitutively GTP bound RHEB facilitates the

assembly and activation of the mTOR Complex 1 (mTORC1), which triggers protein synthesis and cell growth via S6 kinase (S6K), 4E-binding protein 1 (4E-BP1) and Eukaryotic Translation Initiation Factor 4E (eIF4E)<sup>150</sup>.



**Figure 5 |** A simplified schematic overview of the PI3K-AKT-mTOR signalling pathway. Figure adapted from Weinberg<sup>1</sup>.

### 1.2.3.3 Lipid and fatty acid metabolism

Alterations to lipid metabolism in ccRCC were first reported in 1987, since then great progress has been made towards deciphering the molecular underpinnings<sup>138,151</sup>. Generally speaking, lipids are utilised either as an energy source via  $\beta$ -oxidation in the mitochondria, storage, or as a building block for the plasma membrane<sup>138</sup>. In ccRCC, there is a downregulation of  $\beta$ -oxidation and an upregulation of lipid droplet accumulation, which is at least in part mediated through deregulated HIF signalling<sup>152-154</sup>. For example, HIF directly represses the expression of the rate-limiting enzyme carnitine palmitoyltransferase (*CPT1A*), a protein responsible for importing fatty acyl-CoA into the mitochondria

for  $\beta$ -oxidation<sup>153,155,156</sup>. If fatty acyl-CoA is not imported into the mitochondria it is converted back into a fatty acid by FAS and elongated and further modified into unsaturated fatty acids, triglycerides, and phospholipids (required for the plasma membrane) by stearoyl-CoA desaturase (SCD1)<sup>138</sup>. The mRNA expression of *SCD1* is upregulated in ccRCC and the knock-down (KD) of SCD1 protein or treatment with a small molecule inhibitor reduces viability in cell models<sup>157</sup>. Further, there is an accumulation of long-chain fatty acids in ccRCC, and increased expression of the enzyme Fas Cell Surface Death Receptor (*FAS*) correlates with poor patient outcome<sup>155,156,158</sup>. Taken together there is a downregulation of  $\beta$ -oxidation of fatty acids and an upregulation of lipid storage and production of phospholipids for membrane synthesis. The synthesis of phospholipids is important for the construction of a new cell membrane to support proliferation. The the pro-tumorigenic effect of lipid droplet accumulation is less clear, emerging evidence suggests that it may protect against endoplasmic reticulum stress and/or scavenge reactive oxygen species (ROS)<sup>123,152,159</sup>.

#### 1.2.3.4 Amino acid metabolism

The metabolism of key amino acids is de-regulated in ccRCC, namely tryptophan, glutamine, and arginine. The upregulation or dependence on pathways involving these amino acids has provided potentially interesting opportunities for the design of non-genetic-based therapies<sup>138</sup>.

Tryptophan is an essential amino acid that is the substrate for several metabolic pathways, including the kynurenine pathway<sup>160</sup>. In ccRCC the levels of the kynurenine pathway metabolites kynurenine and quinolinate and the expression of the rate-limiting enzyme indoleamine 2,3-dioxygenase (*IDO*)-1/2 are increased, suggesting heightened tryptophan metabolism through the kynurenine pathway<sup>138,161</sup>. The contribution of tryptophan metabolism to ccRCC pathogenesis is not well understood<sup>138</sup>. Kynurenine and quinolinate have been linked to an immunosuppressive microenvironment in both immune-privileged sites (e.g. the eyes and testis) and multiple malignancies<sup>160,162</sup>. In line with this, an IDO-1

inhibitor sensitised previously refractory ccRCC tumours to treatment with the immunotherapy interferon (IFN) $\alpha$ <sup>161</sup>.

Glutamine is important for protein assimilation, lipid synthesis, and energy production and is the precursor to glutathione (an antioxidant)<sup>36</sup>. Multiple studies have demonstrated an increase in glutamine metabolism in ccRCC<sup>93,155,163</sup>. In ccRCC, glutamine is either used in the reductive carboxylation pathway to produce fatty acids (an energy source) and L-2-hydroxyglutarate (an oncometabolite that influences DNA methylation), or in the oxidized glutathione pathway to scavenge ROS, promoting cell survival<sup>93,120,155,164</sup>.

Argininosuccinate synthase 1 (ASS1) is the rate-limiting enzyme for the production of arginine from citrulline in the production of urea in the urea cycle, an important process for ammonia detoxification in the liver and kidney cortex<sup>138</sup>. Interestingly, in ccRCC ASS1 is frequently lost or downregulated which induces a dependence on extracellular arginine, referred to as arginine auxotrophy<sup>156,165</sup>. The competitive growth advantage of becoming arginine auxotrophic is unclear but does present an interesting therapeutic opportunity<sup>138</sup>. Arginine deprivation in a xenograft mouse model of ccRCC prevented tumorigenesis<sup>165</sup>. Arginine depletion as a potential therapeutic is being pursued in a range of cancers including hepatocellular carcinoma, non-small cell lung cancer, acute myeloid leukaemia, breast carcinoma, non-Hodgkin lymphoma and melanoma, but not currently ccRCC<sup>166</sup>. Although these trials have shown some promise, reactivation of ASS1 may curtail this approach<sup>165</sup>.

#### 1.2.4 ccRCC Therapy

Patients can be diagnosed with RCC after presenting with flank pain, a palpable abdominal mass, gross haematuria, and paraneoplastic syndromes, or incidental detection during a non-invasive radiological scan (i.e. CT scan)<sup>90</sup>. The American Joint Committee on Cancer (AJCC) employs the tumour, node, and

metastasis (TNM) framework to classify patients with renal tumours into different stages (I-IV)<sup>167</sup>. The stage of the disease dictates the treatments a patient receives. For example, patients with a tumour mass of  $\leq 7$ cm (T1), no metastases in a regional lymph node (N0), and no distal metastases (M0) are classified as stage I disease (T1, N0, M0), whereas patients with tumours  $> 7$  cm radiologically localised to the kidney are classified as stage II (T2, N0, M0)<sup>168</sup>. Typically, patients with stage I or II disease receive a simple, partial, or radical nephrectomy depending on the size of the tumour<sup>168</sup>. If early-stage renal cancer is comorbid with a terminal disease, arterial embolisation and external beam radiotherapy (EBRT) can be added as part of a palliative care regime<sup>168</sup>. Approximately 42% of diagnoses are stage I and 6% are stage II and the respective 5-year survival rates are 95% and 88%<sup>90,169</sup>. Stage III is classified as a T1 or T2 tumour in conjunction with a regional lymph node metastasis (T1/2, N1, M0) or a tumour that extends into major veins or perinephric tissues with an intact Gerota's fascia (a fibrous connective tissue that encapsulates the kidneys) with or without a regional lymph node metastasis (T3, N0/1, M0)<sup>168</sup>. Patients with stage III disease are also treated by nephrectomy but in combination with renal vein and/or vena cava resection and/or lymph node dissection<sup>168</sup>. Preoperative embolization and preoperative or postoperative EBRT can also be used in conjunction with nephrectomy<sup>168</sup>. Palliative options for patients with a comorbid terminal disease include tumour embolisation, EBRT and, nephrectomy<sup>168</sup>. The annual percentage of stage III diagnoses is 18% and the 5-year survival rate drops to 59%<sup>90,169</sup>. Approximately 18% of diagnoses are stage IV disease, which includes T1, T2, and T3 tumours, in combination with a regional lymph node metastasis and/or a distal metastasis (T1/2/3, N0/1, M1), or a tumour that invades beyond Gerota's fascia with or without detectable regional lymph node or distal metastases (T4, N0/1, M0/1)<sup>168,169</sup>. Depending on the exact TNM classification, stage IV patients are treated with radical or cytoreductive nephrectomy in combination with a systemic drug regiment<sup>168</sup>. Although the 5-year survival rate for stage IV renal cancer has improved markedly over the last 15 years due to the advent of new therapies, it remains as low as 20%<sup>90</sup>. Broadly speaking the classes of drugs used to treat stage IV disease can be classified as anti-angiogenic treatments, mTOR inhibitors, and immune therapies, each of which will be discussed in the following sections<sup>168</sup>. The above

summary provides an illustrative overview of the classification and treatment of renal tumours but is by no means exhaustive. Within each disease stage, the TNM classification can be defined beyond the aforementioned parameters and available treatments can vary depending on patient-specific variables (such as comorbidities).

#### 1.2.4.1 Anti-angiogenic and mTOR inhibitors

Due to the highly vascularised nature of ccRCC tumours (driven by VHL-HIF-VEGF axis), small molecules targeting the VEGF RTK receptor (e.g. Sunitinib, Sorafenib, Pazopanib, Axitinib, and Cabozantinib) and a monoclonal antibody (Bevacizumab) against the ligand VEGF, have been developed and have shown some efficacy in the treatment of metastatic ccRCC<sup>170</sup>. Sunitinib, pazopanib, and cabozantinib have been approved as first-line single agents whereas, axitinib is a second-line single-agent therapy<sup>90</sup>. Sunitinib and pazopanib both significantly improved overall response rate (ORR), progression-free survival (PFS), and overall survival (OS) compared to the placebo<sup>171,172</sup>. Although both RTK inhibitors affected patient outcomes similarly, in 11 of 14 health-related quality of life scores pazopanib outperformed sunitinib and more patients discontinued pazopanib treatment (24% vs 20%) in part due to the higher incidence of liver toxicity<sup>173</sup>. Cabozantinib is the most recently approved VEGF RTK inhibitor (second line -2016, first-line - 2017), compared to sunitinib in a phase II trial, it increased the PFS from 5.6 months to 8.2 months<sup>174,175</sup>. The high frequency of mutations and subsequent deregulation of the PI3K-AKT-mTOR pathway led to the application of mTOR inhibitors for the treatment of ccRCC. Indeed, Everolimus and Temsirolimus have been approved as monotherapies in the second-line setting and as a first line therapy for patients with poor-risk status<sup>176,177</sup>. The combination therapies, bevacizumab and IFN $\alpha$ , and Lenvatinib with Everolimus have also been approved as first and second-line therapies in certain ccRCC disease contexts<sup>178-180</sup>. The development

of targeted therapies centred on the VHL-HIF-VEGF and PI3K-AKT-mTOR axes have improved the quality of care for metastatic ccRCC patients, however, the impact on the 5-year survival rate is still moderate<sup>90</sup>.

#### 1.2.4.2 Immune checkpoint therapy

Over the last two decades, great progress has been made towards leveraging the adaptive and innate anti-tumour activity of the immune system for therapy<sup>49</sup>. Arguably the most significant development is the discovery of immune checkpoint inhibitors (ICI). The two best described immune checkpoint proteins are programmed cell death (PD)-1 and the cytotoxic T-lymphocyte-associated-antigen (CTLA)-4<sup>49,181</sup>. PD-1 and CTLA-4 are expressed on the surface of T cells, when bound to their respective ligands (PD-1 and PD-L1/2, CTLA-4 and CD80/CD86) on antigen-presenting cells or macrophages, they trigger a negative feedback cascade, disengaging activated T cells and reducing the amplitude of the inflammatory response<sup>181</sup>. Cancer cells can co-opt this physiological regulatory checkpoint to inhibit the anti-tumour immune response in the tumour microenvironment, for example through the ectopic expression of PD-L1 and CD80/86<sup>49,181</sup>. To date, a range of monoclonal antibodies (mAb) targeting PD-1, PD-L1, and CTLA-4 have been developed which disrupt the immune checkpoint, triggering a larger and more sustained anti-tumour immune response<sup>49</sup>. The only FDA-approved ICI monotherapy in ccRCC is the use of nivolumab (anti-PD-1 antibody) as a second-line therapy<sup>181</sup>. In a phase III clinical trial comparing nivolumab to everolimus, nivolumab improved ORR (25% vs 5%) and increased the OS from 19.6 months to 25 months<sup>182</sup>. Currently, there is no phase III data to test the activity of nivolumab as a first-line single-agent therapy<sup>181</sup>. Although the moderate increase in OS with nivolumab versus everolimus may appear disheartening, it may also not represent the full potential of this therapy<sup>181</sup>. The observed difference in the OS was primarily driven by a small number of patients who responded to ICI and survived long term<sup>182</sup>. Currently, a focus of ongoing research is to increase the proportion of

patients who will respond to this potentially curative therapy and uncover biomarkers for affective patient stratification<sup>181</sup>.

#### 1.2.4.3 ICI combination therapy

PD-L1 and CTLA-4 have complementary roles in tempering the immune response. CTLA-4 acts at the point of T cell activation in the lymph nodes, whereas PD-L1 acts at the level of the tumour<sup>183</sup>. Combined blockade of these tumour checkpoints in melanoma and colorectal cancer models exhibited enhanced disease control versus either therapy alone<sup>184,185</sup>. Of note, in a toxicology study in Cynomolgus Macaques, the combination of Ipilimumab and Nivolumab triggered a stronger self-reactive immune response compared to a single agent<sup>184</sup>. A similar increase in toxicity has also been reported in human trials<sup>186,187</sup>. Despite heightened toxicity, the combination of nivolumab and ipilimumab (CTLA-4 mAb) significantly improved patient outcomes in the metastatic ccRCC context and was FDA approved in 2017<sup>188,189</sup>.

A second approach that is being investigated is the combination of ICIs with already FDA-approved VEGF inhibitors<sup>181</sup>. The rationale for this approach is based on preliminary studies correlating VEGF expression to immune evasion, possibly through a reduction in mature antigen-presenting dendritic cells or an expansion of the suppressive myeloid compartment<sup>189,190</sup>. Further, In a phase I clinical study, bevacizumab (VEGF mAb) treatment increased infiltration of tumour-specific T cells and in an analogous immunotherapeutic approach, bevacizumab in combination with INF $\alpha$  (an immune activator) improved PFS compared INF $\alpha$  alone<sup>179,191,192</sup>. Multiple phase III trials have demonstrated the utility of combining immune checkpoint and VEGFR inhibitors. For example, the combination of axitinib plus avelumab was evaluated in 886 patients versus sunitinib in the phase III JAVELIN Renal 101 trial, demonstrating increased median PFS (HR 0.69, 95% CI 0.56–0.84; P<0.001) and ORR (51.4% versus 25.7%)<sup>193</sup>. Interestingly, a phase III trial is combining both combination strategies by comparing cabozantinib plus

nivolumab plus ipilimumab versus nivolumab plus ipilimumab (NCT03937219). It will be interesting to observe how the synergistic/additive efficacy of this drug regiment will stack up against the possible side effects. In summary, the combination of ICIs and VEGFR inhibitors has become the new standard of care in the treatment of metastatic ccRCC, surpassing the precious gold standard treatment with the TKI sunitinib<sup>181</sup>. However, there is clearly scope to expand the curative effects of ICI blockade to a greater subset of patients.

#### 1.2.4.4 Additional therapies

As mentioned previously, the biallelic loss of VHL and the stabilisation of HIF2 $\alpha$  is a key oncogenic driver in ccRCC. Recently a first-in-class HIF2 $\alpha$  antagonist (PT2399) was developed which prevents the dimerization of HIF2 $\alpha$  and its binding partner HIF1 $\beta$ /ARNT. PT2399 binds to a pocket in the Per-Arnt-Sim (PAS)-B domain of HIF2 $\alpha$ , preventing dimerization and subsequent DNA binding at HREs. Of note, PT2399 is a selective HIF2 $\alpha$  inhibitor, treatment with PT2399 led to the reduction of HIF2 $\alpha$  target genes without affecting HIF1 $\alpha$  targets<sup>194</sup>. Treatment with PT2399, and its derivative PT2385, outperformed sunitinib in reducing tumour growth in xenograft assays, without causing a reduction in body weight<sup>194-196</sup>. A more recently published derivative with increased potency and pharmacokinetics (PT2977) received FDA approval in 2021 for the treatment of VHL-associated tumours in the kidney, pancreas, and CNS<sup>110</sup>. Combinations with HIF2 $\alpha$  inhibitors are also being explored. PT2385 in combination with nivolumab is currently undergoing a phase I safety trial in 50 participants and two trials with PT2977 in combination with Lenvatinib and Pembrolizumab or Cabozantinib are currently recruiting<sup>197,198</sup>. HIF2 $\alpha$  inhibition resistant tumours (de novo and acquired) have been reported in both pre-clinical models and an early phase trial<sup>194,199</sup>. The dominant route to resistance appears to be a mutation that disrupts drug binding, but there is preliminary evidence for the contribution of *TP53* mutations<sup>194,199,200</sup>.

As mentioned above, ccRCC tumours exhibit glutamine addiction, which could be targeted therapeutically. To date, four early phase clinical trials have been launched testing the small molecule inhibitor of glutaminase (CB-839) as a single agent or in combination with either everolimus, nivolumab, or cabozantinib (NCT02071862, NCT02071862, NCT02771626, and NCT03428217 respectively).

#### 1.2.4.5 Cautionary statement

It is worth noting that due to the number of trialled small molecules and antibodies, their combinatorial application, and utilisation in different disease settings, the current list of regulatory approved treatments for ccRCC is rapidly evolving. Therefore, it is likely that the landscape of approved drugs presented above could soon be quite different.

### 1.3 Transcription factors in cancer

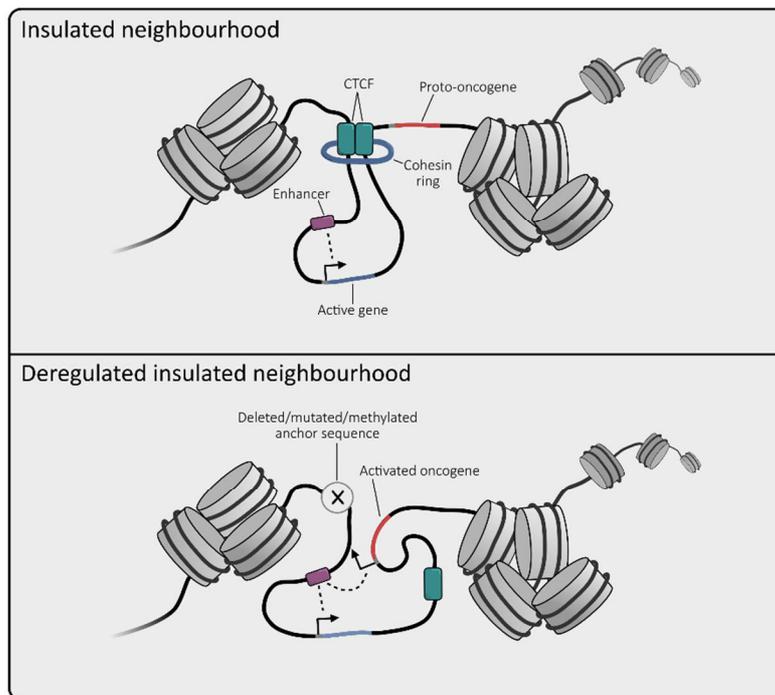
Although great progress has been made towards the curative treatment of metastatic ccRCC, there remains a clear unmet clinical need. A possible source of novel therapeutics, which have already shown some efficacy in ccRCC, are transcription factors (TFs)<sup>201</sup>. Oncogenic alterations in a range of growth-promoting signalling pathways converge on TFs as effectors to drive cancer hallmarks<sup>202</sup>. TFs are proteins with an activation or repression domain and a DNA binding domain capable of recognising specific nucleotide sequences (DNA motifs)<sup>203</sup>. Through the recruitment of co-activators TFs can regulate RNA polymerase activity in a gene-specific manner<sup>204</sup>. The activity of TFs is frequently dysregulated in cancer through amplifications, deletions, chromosomal rearrangements, and somatic alterations which can result in gain or loss of function<sup>201,205</sup>. Further, TFs can act as key nodes canalising the signalling output of a range of hyperactivated cellular pathways and orchestrate complex

transcriptional programs through the direct regulation of their pleiotropic target genes<sup>206,207</sup>. As such, TFs represent targetable bottlenecks with potential implications for multiple cancer hallmarks<sup>207,208</sup>. Compared to designing small molecules targeting kinases, it has proven challenging to disrupt the protein-DNA or protein-protein activity of TFs<sup>202</sup>. In part due to the positively charged convex protein-DNA and the flatter protein-protein interfaces, which contrast to the deep pockets present in enzyme active sites<sup>209,210</sup>. Despite the technical challenges, multiple examples of targeting TFs or other transcriptional trans-factors have shown efficacy in the clinic for a range of cancer types<sup>202</sup>.

### 1.3.1 Regulation of transcription

Humans are assimilated from a plethora of specialised cell types, and with the advent and widespread adoption of single-cell sequencing technologies, new cell types are being continuously described<sup>211</sup>. A specific cell type is defined by the combination of genes expressed, which to a large extent is determined by TFs<sup>212-214</sup>. Approximately 1600 TFs are known or have been predicted based on their sequence<sup>203</sup>. The majority bind to free/accessible DNA, but a distinct class known as pioneer factors, can bind nucleosomal DNA and establish *de novo* gene expression, often through the recruitment of histone acetyltransferases (HATs) (discussed below)<sup>215,216</sup>. The binding sites of TFs are located at gene promoters or in intergenic/intronic regions which have cis-regulatory activity – termed enhancers<sup>217</sup>. Enhancers tend to be occupied by multiple TFs and can be > 1 million base pairs away or even on different chromosomes from their target gene<sup>217,218</sup>. At promoters, TFs recruit transcriptional co-activators (e.g. mediator) and the machinery required for assembling the pre-initiation complex, the prelude to elongation and mRNA synthesis<sup>216</sup>. The initiation and maintenance of mRNA elongation is also closely regulated<sup>204</sup>. For example, cyclin dependant kinase (CDK)7 and CDK9 initiate elongation by phosphorylating serine residues in the C-terminal domain of RNA polymerase II<sup>219</sup>. TFs bound at enhancers similarly recruit DNA polymerase, resulting in the bidirectional transcription of unstable

enhancer RNA (eRNA), which without being translated influences enhancer maintenance and dynamics<sup>220</sup>. The 3D architecture of the genome is arranged so that enhancers become physically juxtaposed to gene promoters through DNA looping, enabling transcriptional regulation<sup>221</sup>. The selection of genes with which discrete enhancers can interact are physically restrained through the binding of CCCTC-binding factor (CTCF) and the formation of a cohesin ring<sup>221</sup>. Looped DNA held in place by a cohesin ring is referred to as an insulated neighbourhood, and is the mechanistic basis of higher-order topologically associated domains (TADs)(Figure 6)<sup>222,223</sup>.



**Figure 6** | Insulated neighbourhoods are DNA loops mediated through the binding of CTCF to anchor points and the formation of a cohesin ring. This limits the promoters a certain enhancer can interact with (upper panel). Deletions, non-synonymous mutations, and aberrant DNA methylation can disrupt CTCF binding and the formation of the cohesin ring. This can expand the number of target genes of a particular enhancer, sometimes triggering the ectopic activation of an oncogene.

Afferent signalling through the actions of TFs can be highly dynamic due to their short protein half-life, in part because their activity is frequently coupled to their destruction<sup>208</sup>. For example, Smad TFs are phosphorylated via TGF $\beta$  signalling and CDK8/9 activity, simultaneously activating them and enabling their recognition by specific ubiquitin ligases for proteasomal degradation<sup>224</sup>. Given the sheer complexity detailed above, the efficient organisation of transcriptional apparatus is paramount for timely gene expression. Emerging evidence suggests that TFs, co-activators, DNA polymerase are concentrated at gene promoters and enhancers by the formation of condensates through liquid-liquid phase separation, enabling the efficient assembly of the pre-initiation complex (PIC)<sup>204</sup>.

The genome in eukaryotic cells consists of nucleic acids and associated proteins (histones), collectively referred to as chromatin<sup>225</sup>. Molecules of 147 base pairs of DNA are wrapped around octamers of histone proteins, forming nucleosomes, which are in turn packaged into chromosomes<sup>225</sup>. Broadly speaking, chromatin can exist in two states, one which enables gene activity (euchromatin) and one which represses gene expression (heterochromatin)<sup>226</sup>. Euchromatin and Heterochromatin states are determined by how closely the DNA is associated with histones, a closer association prevents non-pioneering TFs from being able to recognise and interact with their DNA binding motif<sup>226</sup>. The association of the DNA with histones to a large extent is determined by two classes of proteins, histone modifiers, and ATP-dependant chromatin modellers<sup>226</sup>. A large number of covalent histone tail modifications (e.g. methylation, phosphorylation, acetylation, ect), which are dynamically laid down and removed by chromatin-modifying enzymes, have been described and are the subject of a number of excellent reviews<sup>227,228</sup>. One of the earliest and arguably best characterised examples of histone-modifying enzymes are HATs. HATs catalyse the acetylation of lysine residues in the histone tail, altering the charge of the histone so it repels the negatively charged DNA, loosening the association of the DNA with the histone<sup>229</sup>. Histone deacetylases (HDACs) are capable of removing these acetylation marks (e.g. H3K27ac)<sup>229</sup>. HATs are often referred to as writers because they place epigenetic marks, whereas HDACs are 'erasers' because they remove them<sup>229</sup>. The third class of chromatin modifiers, 'readers', are proteins that are capable of recognising covalent modifications on histone tails. For example,

bromodomain (BrD) containing proteins (i.e. bromodomain containing (BRD)7 and BRD9) can recognise acylated lysine residues. BrD subunits are found in the SWI/SNF ATP-dependant chromatin remodelling complexes, which hydrolyse ATP to physically displace the DNA from the histone, 'opening up' the DNA region and enabling the recruitment of transcriptional machinery<sup>229</sup>. In addition to chromatin modifications, the DNA itself can be modified without altering the nucleotide sequence<sup>229</sup>. The methylation of the fifth position of the cytosine ring in the CpG dinucleotides by DNA methyltransferase enzymes (DNMTs) can alter gene expression in three main ways; methylation at promoters and enhancers preventing the binding of TFs, methylation of CTCF loop anchors to prevent CTCF protein binding thus altering TADs, and methylation marks can be recognised by chromatin-associated proteins (e.g. methyl CpG binding protein 2; MECP2) which can in turn influence gene transcription<sup>229-233</sup>. Collectively these reversible, heritable, and non-genetic determinants of gene expression are referred to as the epigenome<sup>229</sup>. A number of additional well-described epigenetic regulators not discussed above are also important for gene regulation, such as polycomb silencing, long non-coding RNA (lncRNA), X inactivation, imprinting, histone variants, and bivalent chromatin<sup>226,229</sup>.

### 1.3.2 Transcriptional deregulation in cancer

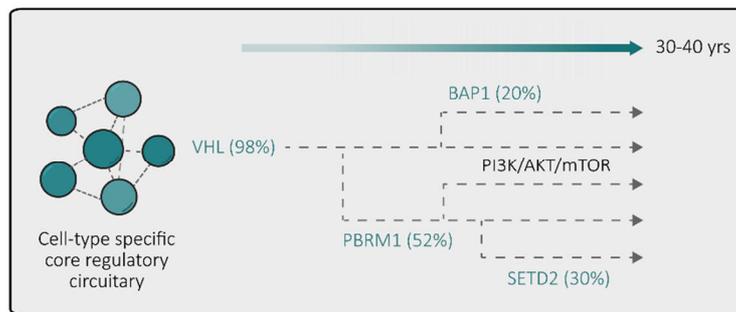
Transcriptional programs are deregulated in cancer through genetic alterations at two levels, trans-acting (e.g. TFs, co-activators, and chromatin regulators) and cis-acting (e.g. enhancers, promoters and insulators)<sup>208</sup>. With regards to cis-acting alterations, amplifications in MYC and point mutations in TP53 are two of the most frequent genetic alterations in cancer<sup>10</sup>. Mutations are also common in chromatin regulators, for example, a member of the switch/sucrose non-fermentable (SWI/SNF) chromatin remodelling complexes are mutated in ~20% of cancers, including the frequent biallelic inactivation of *PBRM1* in ccRCC<sup>234,235</sup>. Translocations can also disrupt transcriptional programs, for example, the reciprocal translocation (15;17) which fuses the retinoic acid receptor alpha (*RARA*) to Promyelocytic

leukemia protein (*PML*), is a driver event in acute promyelocytic leukemia (APML)<sup>208,236</sup>. The fusion *PML/RAR $\alpha$*  oncogene drives tumorigenesis by suppressing the expression of genes essential for granulocytic differentiation, through the recruitment of co-repressors (i.e. HDAC, DNA methyltransferase (DnmT)1 and Dnmt3a). Treatment with all-trans retinoic acid (ATRA) dissociates *PML/RAR $\alpha$*  from the DNA, inducing differentiation and tumour regression in APML<sup>236</sup>. Mutations in trans-acting factors include the aforementioned mutations in the *TET* promoter that induce an aberrant upregulation of the telomerase gene, the translocation of a strong enhancer into the TAD of an oncogene, the focal amplification of an enhancer regulating an oncogene and mutations in anchorage sites which can disrupt an isolated neighbourhood triggering aberrant activation of an oncogene or repression of a TSG<sup>237–242</sup>. Somatic mutation, disruptive translocation, and aberrant methylation of the CTCF-DNA binding region of anchorage sites in insulated neighbourhoods have been described in multiple cancer types<sup>242</sup> (**Figure 6**). In a recent study, microdeletions that eliminated boundary sites in T cell acute lymphoblastic leukemia (T-ALL) lead to oncogene activation when reproduced in non-malignant cells<sup>242</sup>.

### 1.3.3 Lineage addiction

In recent years, targeting lineage-specific TF networks has emerged as a promising non-genetic therapeutic paradigm<sup>243</sup>. As mentioned, cell identity is predominately determined by TFs through the gene regulatory pathways they define. Although approximately half of all TFs are expressed in all cell types, the expression of a small fraction of TFs is restricted to discrete cell lineages, referred to as master TFs<sup>212,213,239</sup>. Master TFs tend to be highly expressed and are necessary and sufficient to establish the transcriptional programs which underpin cell identity<sup>244</sup>. They tend to co-bind with other master TFs to enhancer elements and in particular to large clusters of enhancers (sometimes referred to as stretch or super-enhancers)<sup>208</sup>. They recruit additional factors and act within insulated neighbourhoods,

to drive the expression of tissue-specific genes<sup>208,221,244</sup>. Cancers show a degree of tissue specificity, and the development of different cancer types requires specific chromatin landscapes and combinations of driver mutations<sup>208</sup>. Interestingly, many core transcriptional networks that are important for cell identity are maintained in tumorigenesis and members of these networks have been shown to represent cellular dependencies in a range of cancer models (**Figure 7**)<sup>243,245–249</sup>. The best examples of this include targeting the estrogen receptor (ER) in breast cancer and the androgen receptor (AR) in prostate cancer, neither of which are commonly mutated<sup>250</sup>.



**Figure 7** | Schematic overview of the interaction between the tissue-specific and genetically activated pathways, using ccRCC as an example. Somatic alterations occur in a certain cellular context, which determines their oncogenic potential. The cellular context of a specific cell type is determined by a small number of master TFs. **Abbreviations:-** LD: Lineage dependency, All: All-non-RCC cell lines.

The ER and AR are examples of nuclear hormone receptors (NHRs)<sup>250</sup>. NHRs have a ligand-binding domain and a DNA binding domain. The ligand-binding domain enables their interaction with small molecules – hormones. When bound to their respective hormone, NHRs can positively or negatively regulate gene expression through the recruitment of co-factors<sup>251</sup>. The nature of the ligand-binding domain has made the design of small molecule antagonists far easier than with other TFs<sup>202</sup>.

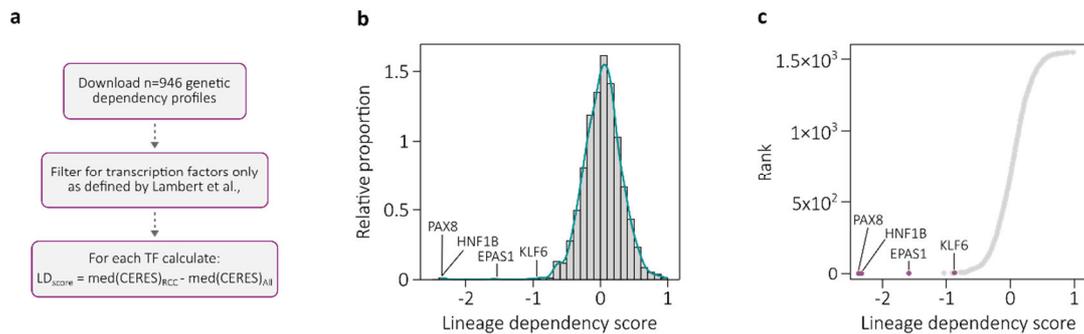
The ER has been established as a key oncogenic driver in ~75% of breast cancers<sup>252</sup>. Two methods of pharmacological inhibition of the ER have shown efficacy, selective oestrogen receptor degraders

(SERDs) and selective oestrogen receptor modulators (SERMs)<sup>252</sup>. SERMs, such as tamoxifen, interact with the ligand-binding domain of the NHR preventing the binding of estrogen. Without bound estrogen, the estrogen receptor does not undergo a conformational change enabling the recruitment of co-activators and instead recruits co-repressors<sup>253</sup>. SERDs, such as fulvestrant, also bind to the ligand binding domain but instead target the estrogen receptor for protein degradation<sup>252</sup>.

The AR regulates gene expression after binding androgen hormones, such as testosterone. Similar to the estrogen receptor, the androgen receptor is important for tissue homeostasis and the sex hormone response in the prostate<sup>254</sup>. The binding of androgen to the ligand binding domain of the AR displaces bound chaperone heat shock protein 90 (HSP90), enabling the NHR to translocate to the nucleus and regulate gene expression<sup>255</sup>. A number of drugs that interfere with androgen binding to the AR have been developed, which have shown efficacy in the clinic<sup>256</sup>.

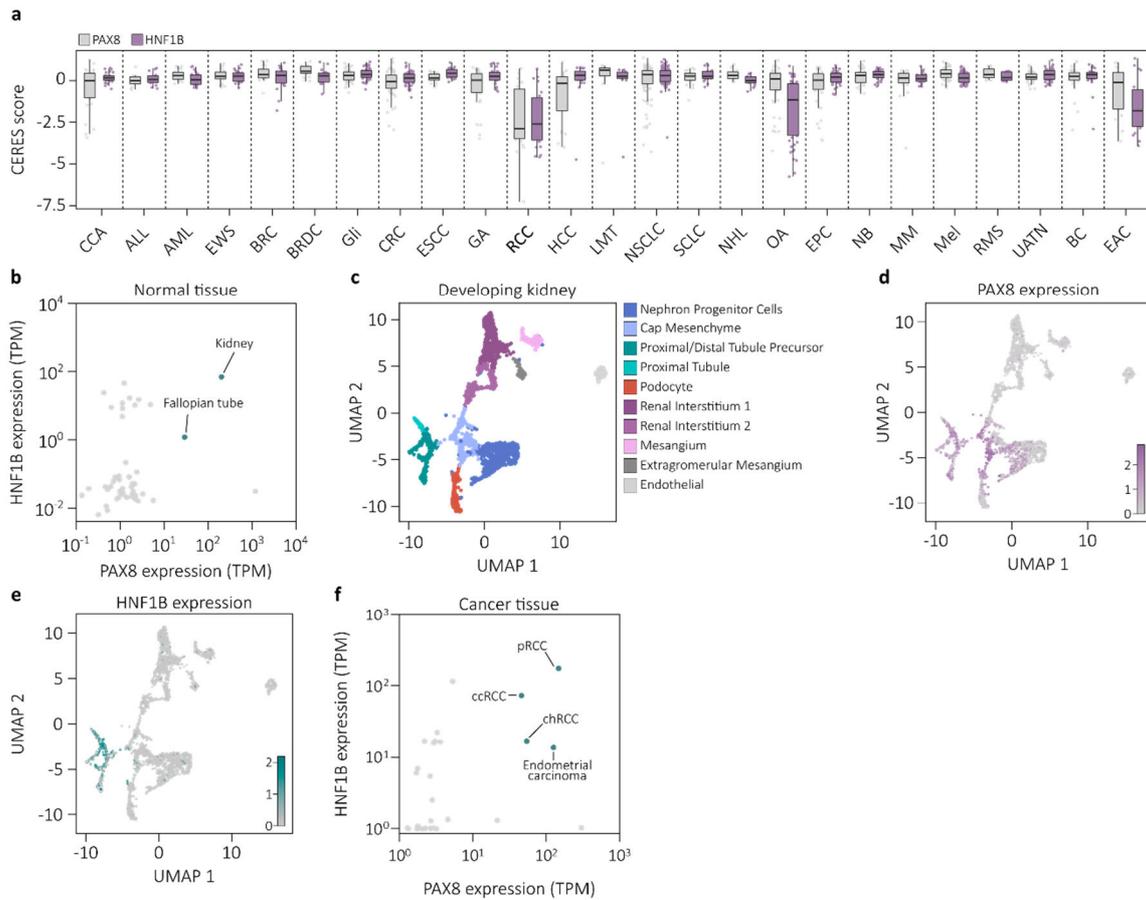
#### 1.4 Identifying lineage specific transcription factor dependencies in kidney cancer

The success of NHR based targeted therapies suggests that lineage factor dependencies could be exploitable in non-hormone receptor-driven cancers, especially in instances where subclonal heterogeneity often leads to rapid resistance to oncogene-targeted therapies, such as in ccRCC. To identify candidate TF lineage dependencies in ccRCC, I leveraged the genetic dependency data from the collaborative DepMap genome-wide CRISPR-Cas9 pooled screening project and applied a similar workflow to Rauscher et al., depicted in **Figure 8a**<sup>257–259</sup>. In short, for each TF (~1600), I compared the median dependency score (CERES) in RCC cell lines to the median CERES score in all other cell lines from all remaining lineages. In this way, a score was assigned to each TF (lineage dependency score – LD score) based on how specifically it was a dependency in RCC cell lines (**Figure 8b-c**). A large negative score represents a strong RCC-specific dependency.



**Figure 8** | Prediction of RCC specific transcriptional dependencies **(a)** Workflow for generation of lineage dependency scores (LD). RCC: all RCC cell lines, All: all non-RCC cell lines. **(b)** Frequency distribution of LD scores for each TF. **(c)** Ranked plot of LD scores for each TF.

This analysis identified two very strong outlying tissue-specific dependencies, PAX8 and HNF1B. Out of the 27 lineages profiled, the co-depletion of paired Box (PAX)8 and hepatocyte nuclear factor(HNF)1B was only seen in one additional lineage, endometrial adenocarcinoma (EAC) **(Figure 9a)**. Although the depletion of PAX8/HNF1B was relatively weaker in EAC compared to RCC. PAX8 was also specifically depleted in ovarian adenocarcinoma and HNF1B was weakly depleted in hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA) **(Figure 9a)**. Notably, the genes Eedothelial PAS domain-containing protein 1 (*EPAS1*; encodes HIF2 $\alpha$ ) and Krüppel-Like factor 6 (*KLF6*) also scored **(Figure 8b-c)**. HIF2 $\alpha$  is a bona fide oncogenic driver, its protein expression is deregulated after the biallelic loss of *VHL*, a mutation that is highly specific to ccRCC<sup>10</sup>. Recently, our group identified *KLF6* as a tissue-specific dependency in ccRCC through a stretch/super-enhancer-based ranking approach <sup>260</sup>. *KLF6* was subsequently shown to support ccRCC pathogenesis through the regulation of lipid metabolism and the induction of PDGFB and mTOR signalling<sup>260</sup>.

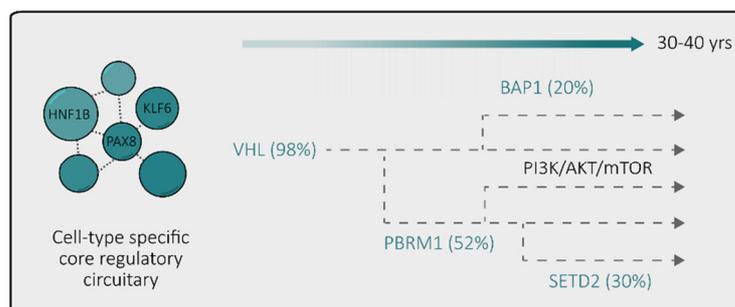


**Figure 9** | PAX8 and HNF1B as lineage specific dependencies. **(a)** box plots of CERES scores for PAX8 (grey) and HNF1B (purple) for 25 cancer lineages with  $\geq 10$  cell lines. **(b)** Median PAX8 and HNF1B RNA expression (TPM) in 54 normal tissues from the GTEx project. **(c)** Unsupervised UMAP analysis of scRNAseq data from fetal human kidney. **(d)** PAX8 and HNF1B mRNA expression in the cell types identified from (c). **(e-f)** PAX8 and HNF1B mRNA expression superimposed on Unsupervised UMAP from (c). **(g)** Median PAX8 and HNF1B RNA expression (TPM) in 34 cancer tissues from the TCGA project. Acknowledgements:- As stated in the Preface, the scRNAseq analysis and plotting for panels c-d was performed by S.H using data from Young et al<sup>94</sup>.

A literature search confirmed that PAX8 and HNF1B both play important roles in renal development. Heterozygous mutations in *HNF1B* are the most common monogenic cause of developmental renal disease, including congenital anomalies of the kidney and urinary tract (CAKUT) a frequent cause of renal failure in children<sup>261,262</sup>. The biallelic loss of *PAX8* in the developing mouse embryo does not affect

renal development, however, the kidneys of *PAX8*<sup>-/+</sup>*PAX2*<sup>-/+</sup> mice compared to *PAX2*<sup>-/+</sup> mice have a significant reduction in the number of nephrons, poor distal convoluted tubule differentiation, and a partial defect in branching morphogenesis <sup>263</sup>.

Further, I confirmed that the co-expression of *PAX8* and *HNF1B* was highly specific in normal and neoplastic kidney tissue. Analysis of *PAX8/HNF1B* mRNA expression in normal tissue from the Genotype-Tissue Expression Project (GTEx project; 53 tissue types) showed the specific co-expression of *PAX8/HNF1B* in the fallopian tubes and kidney medulla and cortex (**Figure 9b**). Looking more closely at the kidney, a re-analysis of a scRNAseq study of the developing kidney, showed that the co-expression of *PAX8* and was enriched in the proximal tubule cluster, which is the proposed origin of ccRCC (**Figure 9c-e**)<sup>94</sup>. Additional functional validation comes from a cellular re-programming study which demonstrated that HNF1β and PAX8 in combination with two additional TFs can convert fibroblasts into renal tubular epithelial cells, which resemble the proximal tubule epithelium<sup>264</sup>. Finally, using mRNA expression data from the Cancer Genome Atlas (TCGA; 34 cancer types), I confirmed that the specific co-expression of *PAX8/HNF1B* was maintained in cancer arising from the kidney (**Figure 9f**). Taken together, PAX8 and HNF1B are important factors in renal development and tissue homeostasis, their expression is maintained in RCC and inhibition of these factors is able to undermine the oncogenic program (**Figure 10**).



**Figure 10** | PAX8, HNF1B and KLF6 are members of the core regulatory circuitry of RCC tumours.

## 1.5 Problem statement

Transcriptional lineage factor dependencies are observed across a range of malignancies, making them an attractive target class for therapy development. A Lineage targeted approach might be especially effective in cancers with prevalent subclonal heterogeneity that often leads to rapid resistance to oncogene-targeted therapies<sup>265</sup>. Multi-region sequencing of ccRCC patient tumours revealed seven evolutionary distinct molecular subtypes with different combinations of driver mutations<sup>135,136,266</sup>. To date, targeted therapies in ccRCC have only been moderately successful in the clinic, in part due to the complex clonal architecture of ccRCC tumours<sup>90,267,268</sup>. The majority of ccRCC cell lines are dependent on the expression of *PAX8* and *HNF1B* and so inhibiting these factors holds great promise as a therapeutic paradigm, to make this heterogeneous disease more therapeutically tractable (**Figure 8a**). However, the role of lineage factors in cancer and the mechanisms that maintain lineage fidelity in advanced cancers remain poorly understood. Further, beyond targeting the NHRs in breast and prostate cancer, it is unclear what the long-term consequences and possible resistance mechanisms to lineage targeted therapy are.

The three aims of my PhD are thus as follows:

**Aim 1:** *to identify regulators of lineage dependence in ccRCC models.*

**Aim 2:** *to characterise the molecular consequence of losing regulators important for safeguarding lineage identity.*

**Aim 3:** *uncover the functional relevance of lineage reprogramming in response to lineage targeted therapies*

To approach this project, I started by targeting a lineage dependency in a ccRCC model and then used a clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-based genetic screen to identify genes capable of modulating the proliferative phenotype. Taking forward the most strongly

enriched hit from the screen, I profiled the transcriptome and chromatin landscape to determine the molecular consequence following the KO of this factor. Finally, I used a second CRISPR/Cas9-based genetic screen to determine the functional outcome of the observed reprogramming event.

2

## Materials and Methods

## 2.1 Cell lines and cell culture

The Human ccRCC cell lines used were 786-M1A and UOK101. 786-M1A cells are the metastatic derivatives of 786-O (obtained from American Type Culture Collection - ATCC), and they have been previously described<sup>31</sup>. Renal cancer cells were cultured in RPMI-1640 medium (Sigma) supplemented with 1% (v/v) penicillin-streptomycin (P/S) and 10% (v/v) fetal bovine serum (FBS). Human embryonic kidney HEK293T cells (obtained from ATCC) were cultured in DMEM medium (ThermoFisher Scientific) supplemented with 1% P/S and 10% FBS. Cell line identity was authenticated by short-tandem repeat profiling. Cell lines were routinely profiled for mycoplasma using the MycoAlert™ Mycoplasma Detection Kit (Lonza, LT07-318).

## 2.2 Drug treatment

Doxycycline (Sigma) was diluted in RPM1 medium to a final concentration of 0.1, 0.3, 0.6, or 1µg/ml (as specified in results) from a stock concentration of 1mg/ml, before adding to the cells. Doxycycline-infused media was replenished every 2-3 days depending on the length of the treatment.

## 2.3 Plasmids

The plasmids used in this study are listed in **Table 1**. Derivatives of gifted plasmids were generated by restriction enzyme cloning. The Broad institute tool (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>) was used to design sgRNAs (**Table 1**) and standard methods were used for cloning into pKLV2. Sequences for shRNA were taken from FellMann et al (**Table 2**) and the restriction enzymes, EcoRI-HF and XhoI, were used for cloning into the miRE vectors<sup>269</sup>.

Original plasmid	Addgene ID	Kind gift from	Derivatives	Ref
pCW-Cas9	50661		LT3-Cas9-Blast	
pKLV2-U6-gRNA(BbsI)-PGKpuro-2A-BFP	67974	Kosuke Yusa	pKLV2-U6-gRNA(BbsI)-PGKhygro-2A-BFP	
LT3-eGFP-miRE-PGK-puro		Johannes Zuber	LT3-eGFP-miRE-PGK-BFP	
LT3-dsRed-miRE-PGK-Venus				
psPAX2	12260			
pMD2.G	12259			

**Table 1** | Plasmids used in this study.

Gene	sgRNA ID	Sequence
non-targeting control	NTC18	GAGTGTCGTCGTTGCTCCTA
SMARCB1	sgRNA_1	GTTCTACATGATCGGCTCCG
SMARCB1	sgRNA_2	GTTCTACATGATCGGCTCCG

**Table 2** | sgRNA constructs

Gene	shRNA ID	Sequence
Renilla	shRenilla	CAGGAATTATAATGCTTATCTA
PAX8	sh1503	ATCCATTATTAACACAACCTCTA
PAX8	sh786	ACCGACTAAGCATTGACTCACA

**Table 3** | shRNA constructs

## 2.4 Lentiviral transduction

The plasmid of interest (**table 3**) and the viral packaging plasmids psPAX2 and pMD2.G were co-transfected into HEK293T cells using Fugene 6 (Promega E269A) according to the manufacturer's protocol. Viral supernatants were harvested 48h following transfection and filtered through a 0.45µM PVDF sterile filter. One day prior to transduction  $2.5 \times 10^5$  cells were seeded on a 6-well plate. Immediately before transduction, the medium was changed to RPMI containing 6-8µg/mL Polybrene (Millipore). Fresh or frozen viral supernatant was added to the cells accordingly. 24h after transduction the media was changed to fresh RPMI. Antibiotic selection media were added two days post-transduction: 4µg/ml puromycin (Invivogen), 800µg/ml hygromycin (Invivogen), or 25µg/ml blasticidin (Invivogen).

## 2.5 Immunoblotting

Cell pellets were washed once with ice-cold PBS and lysed with RIPA lysis buffer (Sigma) containing 1x protease inhibitor cocktail (Sigma): cells were incubated on ice for 30m and vortexed every 10m, followed by centrifugation at 14,000 RPM at 4°C. The protein lysate (supernatant) was collected and quantified using the Pierce™ BCA protein assay kit (Thermo Scientific), before being stored at -80°C. For Immunoblotting, 10µl of protein (15-30µg) was added to 4µl of loading buffer (Invitrogen, NuPAGE) and 1 of β-mercaptoethanol (Sigma) and topped up to a total volume of 20µl with H<sub>2</sub>O. Samples were boiled at 95°C for 5m and then centrifuged for 20s. Following centrifugation, samples were loaded onto a polyacrylamide gel alongside a precision plus protein standards kaleidoscope ladder (BIO-RAD) and run at 80V for 15m followed by 100V for 2h. Proteins were transferred onto a PVDF membrane (Millipore) for approximately 2-3h at 100v. The membrane was subsequently blocked in 5% milk (dissolved in 0.1% PBS-Tween, PBST) for 1h. For immunoblotting, primary antibodies were diluted in

5% milk and incubated overnight with the membrane at 4°C. Membranes were washed three times with PBST before incubation with the secondary anti-mouse (DAKO, cat no. P 0447, 1:10,000) or anti-rabbit antibodies (DAKO, cat no. P 0448, 1:10,000) conjugated to horseradish peroxidase (HRP) for 2h. Membranes were washed three times with PBST and were developed with Luminata™ Classico Western HRP substrate (Millipore) using a film processor. The following primary antibodies were used: PAX8 (Santa Cruz, cat no. Sc-81353, 1:250), HNF1β (Atlas, cat no. HPA002083, 1:5000, β-actin (Abcam, cat no. ab8227, 1:20,000), SMARCB1/SNF5 (Bethyl laboratories, cat no. A301-087A, 1:2500).

## 2.6 Reverse transcription and quantitative polymerase chain reaction (qPCR)

Cells were pelleted at 500g for 3m and stored at -80°C before processing. RNA was extracted using the RNeasy Mini Kit (Qiagen), according to the manufacturer's instructions. Reverse transcription-polymerase chain reaction (PCR) was performed using the High-Capacity complementary DNA (cDNA) Reverse Transcription Kit (Thermo Scientific) to generate cDNA from 500ng of RNA. The StepOnePlus™ Real-Time PCR instrument (Thermo Scientific) was used with TaqMan reagents (Thermo Scientific). Samples were run in triplicate, normalised to the housekeeping gene TATA-box binding protein (TBP), and analysed using the double delta Ct method. Taqman probes used: PAX8 (Hs00247586\_m1) and TBP (Hs00427620\_m1).

## 2.7 In vitro proliferation assays

6x10<sup>3</sup> 786-M1A cells were seeded in triplicate on a 24-well cell culture plate and analysed using the InCuCyte ZOOM™ instrument (Essen Bioscience). Bright-field images were acquired in 9 independent

locations within each well every 2 hours. Confluency was measured by applying a predefined cell-specific mask to each image, which distinguished the cells from the background. For competition assays, dsRED/eGFP was used to gate all cells expressing an shRNA, and BFP/eGFP/mCherry was used to measure the abundance of two co-cultured cell populations. The proportion of the competing cell populations was measured by flow cytometry on an LSR Fortessa (BD Biosciences) and compared to day 0. The following gating approach was used: FSC-A, FSC-W, SSC-A to distinguish single cells from debris, and then dsRed (561nm, 610/20nm), mCherry (532nm, 610/20nm), BFP (405nm, 450/50nm) or GFP (488nm, 515/20nm), venus (488, 530/30) channels for discriminating between cell populations.

## 2.8 Fluorescent activated cell sorting and analysis

Fluorescent activated cell analysis was performed using a BD LSRFortessa flow cytometer. FlowJo software (BD Biosciences) was used to analyse flow cytometry data and generate plots. Fluorescent activated cell sorting was carried out by the Flow Cytometry Core Facility at the Cambridge Institute for Medical Research.

## 2.9 RNA-seq

Cells were seeded 24h before as four replicates per condition in 6 well plates. The medium was aspirated, and the cells were lysed on ice in buffer RLT (RNeasy Plus Mini Kit Qiagen). Total RNA was extracted using the RNeasy Mini Kit (Qiagen), according to the manufacturer's instructions. RNA quality was determined using Agilent RNA Nano 6000 kit (Agilent 5067-1511) and RNA concentration was determined using a NanoDrop 1000 Spectrophotometer. Library preparation was performed using the QuantSeq 3 mRNA-Seq Library Prep Kit FWD for Illumina and PCR Add-on Kit for Illumina (Lexogen),

with 300ng input RNA. The size of the final libraries was determined using the Agilent 4200 TapeStation System using the high sensitivity D1000 reagents (5067-5592). The concentration of the libraries was determined using the Qubit Flex Fluorometer (Thermo Fisher). The libraries were pooled in equimolar concentrations and submitted for deep sequencing on the Illumina HiSeq4000 platform (SE50). The processing of the FASTQ files to a read count table was performed on the BlueBee® Genomics Platform. In short, reads were trimmed using Bbduk (v35.92) from the bbmap suite, aligned with STAR aligner (v2.5.2a) to hg38, and counted using HTSeq-count (v0.6.0). Differentially expressed (DE) genes were determined from the read counts table using DESeq2 (v1.26.0)<sup>270</sup>. The custom rhabdoid signature was generated using common DE genes upon SMARCB1 re-introduction in TTC1240 and G401 cell lines, which satisfied  $\log_2FC < -0.5$  and  $padj < 0.05$ <sup>271</sup>. The top 500 up and downregulated genes with a  $padj < 0.05$  upon SMARCB1 knockout (KO) were used to generate gene signatures for validation with CCLE expression data. The ARID1A KO signature was generated using the top 500 upregulated genes after ARID1A KO in MCF7 luminal breast cancer cells<sup>272</sup>. The NE signature was generated by taking the top 500 varimax PCA loadings along a principle component which simultaneously separated NEPC from CRPC and normal prostate tissue, and SCLC from normal lung and LUAD tissue<sup>273</sup>. Gene set enrichment analysis with custom and mSigDB (v7.2.1) signatures was performed using the R package ClusterProfiler (v3.14.3)<sup>274</sup>. Data wrangling and presentation (MA plots) were achieved using the R packages Tidyverse (v1.3.0) and ggpubr (v0.4.0) respectively.

## 2.10 ATAC-seq

786-M1A cells were treated with 0.6µg/ml doxycycline for 6 days before harvesting at 70% confluency. On the day of harvest, cells were trypsinized, counted, nuclei extracted, and 50,000 cells were used for the ATAC-seq protocol as previously described<sup>275</sup>. ATAC libraries were generated with the Illumina Nextera DNA library preparation kit (FC-121-1030) and purified for amplification with the minElute PCR

purification kit (Qiagen 28004). The libraries were amplified for a total of 8-12 cycles using custom Nextera PCR primers and NEBNext PCR master mix (NEB M0541S). The amplified libraries were purified using Agencourt AMPureXP reagents (A63880), profiled on the Agilent 4200 TapeStation System using the high sensitivity D5000 reagents (5067- 5589), pooled in equimolar concentrations, and submitted for sequencing on the Illumina HiSeq4000 platform (SE50).

## 2.11 ATAC-seq analysis

Adapters and low-quality bases (quality < 20) were trimmed from read ends using cutadapt (version 2.10)<sup>276</sup>. Reads were mapped to hg38 using BWA (version 0.7.17)<sup>277</sup>. Low quality reads (mapping quality < 20) and reads mapping to ENCODE blacklisted regions and regions other than chr1-22, chrX and chrY were removed using deepTools<sup>278,279</sup>. Reads were corrected for Tn5 offset (+ve strand: +4bp, -ve strand: -5bp). Peaks were called using MACS2 (version 2.2.7.1) with the following parameters “-f BAM -bdg -g 2913022398 -nomodel -nolambda -shift -100 -extsize 200”<sup>280</sup>. A consensus peak file for DE analysis was generated by extending peak summits to a fixed 501bp window, ranking called peaks by their qvalue, and iterating down the list, removing any overlapping peaks with a lower qvalue. This produced a consensus peak file containing the coordinates of the most significant peak called at a particular locus. The read count table was generated by extending reads to the modal length of 250bp and counting the number of uniquely mapped reads falling within consensus peaks using RSamtools (v2.2.3). Peaks were filtered for  $-\log_{10}(q) < 20$  and differentially accessible (DA) peaks (FC+/-2 and padj<0.001) were determined using DESeq2<sup>270</sup>. Homer (v4.11) was used for *de novo* and known motif enrichment analysis on +/- 50bp flanking the summits of DA regions, compared to a set of high confidence unchanged regions<sup>281</sup>. The R package ChIPseeker (v1.22.1) was used to determine genomic annotations for peaks. To determine the correlation between gene expression and epigenetic changes,

sequentially larger windows around LA and HA peaks were created with GenomicRanges (v1.38.0), and a hypergeometric based test (pHyper) was used to determine whether genes captured within these windows were significantly enriched for down or up-regulated genes, respectively. ESeq (v1.111) was used to determine genes that fell within these windows (GColoc function) and to create a set of matched controls (Controls function)<sup>282</sup>. For data visualisation, EnhancedVolcano (v.1.4.0) was used for volcano plots and ESeq (v1.111) was used for genomic tracks, heatmaps, and metagene plots<sup>282</sup>.

## 2.12 scRNAseq analysis

The expression count matrix for CD45 negative fetal kidney samples 4834STDY7002876, 4834STDY7002881, 4834STDY7002886 were downloaded from Young et al and were processed as follows<sup>94</sup>. (1) Cells with reads mapping to <1500 distinct mRNA transcripts or expressed >20% mitochondrial genes were removed. (2) Normalisation was performed with SCTransform<sup>283</sup>. (3) Integration anchors (n=3000) were identified using the function FindIntegrationAnchors and batch correction was performed using the Seurat function IntegrateData (normalisation.method="SCT")<sup>284,285</sup>. (4) The first 30 principal components were used to calculate UMAPs and clustering was achieved using the FindNeighbors and FindClusters (resolution=0.3) functions. The clusters were biologically annotated based on the expression of published marker genes<sup>286–288</sup>. (5) Clusters that expressed less than 15 marker genes and the immune marker CD45 were filtered out. (6) UMAP and violin plots were made using the R packages Tidyverse, Rcolorbrewer and ComplexHeatmap.

## 2.13 Animal studies

All animal protocols were approved by the Home Office (UK) and the University of Cambridge Animal Welfare and Ethical Review Body (PFCB122AA). Five to seven-week-old athymic female nude mice (Charles River Laboratories) were injected subcutaneously with 500 000 cells in each flank, using 100 $\mu$ L of 1:1 phosphate buffered saline (PBS)/Matrigel Matrix (BD) solution. The tumour growth was measured by calliper and tumour volume was calculated as follows,  $V = (\text{length} \times \text{width}^2) \times 0.5$ .

## 2.14 Pooled CRISPR-Cas9 screening

Cells were transduced with a lentiviral library at a low multiplicity of infection (MOI; <0.3) to ensure 1000x sgRNA representation. An MOI of <0.3 was used so that >85% of cells had a single sgRNA integration. After 48h following transduction, the cells expressing the integrated library were selected with puromycin or hygromycin for 5 days. For doxycycline naïve cells, the screen was initiated after antibiotic selection by supplementing the medium with 0.6 $\mu$ g/ml doxycycline to induce the expression of Cas9, otherwise, the screen was considered to have started 24h post-transduction. Cells were cultured for 17-21 days after screen initiation and two replicates at various time points were collected for each condition. For time points that required FACS, enough cells to ensure >130x coverage were harvested, otherwise, >500x coverage was maintained. Day 17 of the chromatin regulator screen was the only time point that required sorting. Given that the primary focus of this screen was to look for enriched hits and that the initial coverage was ~1000x, an sgRNA coverage of 130x was tolerated. Genomic DNA was isolated using the QIAamp DNA mini kit (Qiagen 51304) and libraries were created by amplifying the cassette containing the sgRNAs using KAPA HiFi HotStart ReadyMix (Roche) and custom primers adapted from for pKLV2<sup>289</sup>. Libraries were purified using Agencourt AMPure XP

(Beckman-Coulter A63880) beads, profiled using the Agilent 4200 TapeStation System using the high sensitivity D1000 reagents (5067-5592), quantified using the Qubit dsDNA HS assay kit (Thermo), and pooled in equimolar concentrations for sequencing on the Illumina HiSeq4000 platform (SE50). FASTQ files from sequencing were aligned to the sgRNA library and counted, using the command `mageck count` (v0.5.9), which tolerated no mismatches. Raw fold changes from normalised counts were calculated for each sgRNA construct for each gene, compared to Day 0 or the plasmid library. The top performing 3 sgRNAs (depleted or enriched depending on context) for each gene averaged across the experimental screen arm (i.e.  $P8_{1/2}$  for the TF screen) were used to calculate the final beta scores (normalised FC) for each gene. Beta scores were calculated using `mageck mle` (v0.5.9) with the normalisation method set to median and permutations set to 1000.

## 2.15 The encyclopedia of DNA elements (ENCODE) analysis

DNase I hypersensitivity profiles for cell lines (n=97), adult primary cells and tissues (n=125), and embryonic tissues (n=282) were downloaded from SCREEN (<https://screen.encodeproject.org/>), with chromatin accessibility annotations (open/closed) determined for each of the 926,535 candidate cis-regulatory regions (cCREs) identified by the ENCODE project. For clustering, the top 250,000 most variable regions across samples were selected to generate a pairwise Pearson correlation matrix. To identify clusters, samples were first ranked based on the number of additional samples with which they highly correlated (Pearson's correlation coefficient (PCC)>0.6). Starting with the top-ranked sample, all samples which correlated (PCC>0.6) were identified and labelled as cluster 1. These samples were then removed from the matrix, the matrix was re-ranked and cluster 2 was identified by the same means. This process was repeated until the size of the cluster dropped below a cut-off of five samples. This resulted in 21 clusters, incorporating 376/504 samples. Clusters were given a biological annotation based on their sample composition, and clusters that could not be annotated were removed (n=1). Cluster-specific cis-regulatory regions were defined as peaks that were present in 80% of cluster

samples and appeared in no more than two additional clusters. Overlap analysis between cluster peaks sets and +/-25bp flanking the summit of DA regions from this study was performed using the R package ChipPeakAnno, requiring a stringent 100% overlap.

## 2.16 Cancer cell line encyclopedia (CCLE)

Mutational and gene expression data were downloaded from the DepMap portal (<https://depmap.org/portal/download/>). Differential expression analysis with raw RNA-seq counts was performed as above. Biallelic inactivation of the essential gene *VHL* is a truncal initiation event in ccRCC, RCC cell lines were considered ccRCC if they had a 'damaging' or 'other non-conserving' mutation in *VHL*, which also corresponded with resistance to *VHL* KO (CERES > -0.5). To prioritise functionally relevant mutations in *ARID1A* and *SMARCB1*, 'damaging' or 'other non-conserving' with a TCGA or COSMIC hotspot were considered.

## 2.17 DepMap analysis

Dependency data for 990 cell lines were downloaded from the DepMap portal (<https://depmap.org/portal/download/>). Where appropriate, centred CERES scores were calculated using the scale function in R.

## 2.18 Pan-cancer Lineage dependency (LD) analysis

Genome-wide CRISPRcas9 genetic dependency data for 946 cell lines was downloaded from the DepMap project ([www.depmap.org/portal/](http://www.depmap.org/portal/))<sup>257,290</sup>. The list of ~18000 genes was filtered for ~1600 TFs, as defined by Lambert et al<sup>203</sup>. The Lineage of a cell line was defined by their lineage annotation (e.g. skin) and lineage subtype annotation (e.g. melanoma). Of the 81 defined lineages (e.g. skin\_melanoma), 56 were removed because they were underrepresented (<10 cell lines). A lineage dependency score ( $LD_{score}$ ) for each TF in each lineage context was calculated according to:  $LD_{score} = \text{mean}(\text{CERES})_{\text{lineage } x} - \text{mean}(\text{CERES})_{\text{remaining lineages}}$ . The  $LD_{score}$  was a measure of how specific a particular transcriptional dependency was to a certain lineage context; a larger negative score denotes a stronger and more specific dependency. Simultaneously, the Kruskal Wallance statistical test in conjunction with a Benjamini Hochberg correction was implemented to derive a corresponding  $p$ -value for each  $LD_{score}$ . To define putative core regulatory circuitry (CRC) for each of the 25 lineages, three levels of filtering were applied. (1) Based on the distribution of maximum  $LD_{scores}$  for each TF, a cut-off of  $LD_{score} < -1.2$  and a  $p$ -value  $< 0.05$  was implemented. The distribution of maximum  $LD_{scores}$  was plotted by selecting the lowest possible  $LD_{score}$  for each TF (**Figure 49b**). For example, PAX8 and HNF1B had the most negative  $LD_{score}$  in the RCC lineage context, and so these scores were used. (2) Examples of LDs that were strong cellular dependencies were selected by filtering for putative LDs for which the majority (>50%) of cell lines in their respective lineage had a CERES score of  $\leq -0.5$ . (3) Putative LDs which were pan-cancer dependencies but were more strongly depleted in a particular lineage were also removed. This was accomplished by using a plot of the distribution of the median CERES score across all cell lines for each putative LD. Based on the bimodal distribution of the data, a cut-off median CERES  $> -0.2$  was identified. (**Figure 51b**). After filtering, CRC predictions were available for 10/25 lineages. To identify lineage-resistant cell lines within each of the 10 lineages, a distribution of the averaged CERES scores of LDs in each cell line of their respective lineage was plotted (**Figure 52a-d**). Based on the bimodal distribution of the data, a cut-off of average CERES score  $> -0.45$  was used to identify lineage-resistant cell lines.

A permutation-based statistical method was used to identify acquired transcriptional dependencies or lack of dependencies in lineage resistant versus lineage sensitive lines within each lineage. (1) Compute the observed effect size using Cohen's D (lineage resistant vs lineage sensitive) relative to each TF. (2) Create a permuted dataset by randomly dividing cell lines into sensitive and resistant whilst maintaining the original number of observations for each of the two categories. (3) Compute the effect size with the permuted data. (4) Repeat steps 2 and 3 1000 times. (5) For each  $t \in \text{seq}(0, \max(|\text{observed effect size}|), \text{by} = 0.001)$ , estimate the false discovery rate (FDR) associated to the threshold  $t$ :

$$\hat{FDR}(t) = \frac{\text{expected number of false positives}}{\text{number of observed positives}} \quad \hat{FDR}(t) = \frac{E[\#(|X_{ran}| > t)]}{\#(|X_{obs}| > t)}$$

The numerator is the mean (over the 1000 permutations) number of false positives (features with absolute effect size  $> t$ ). The denominator is the number of observed features with absolute effect size  $> t$ . (6) The smallest  $t$  for which  $FDR(t) < 0.1$  was used as a threshold to call significant features.

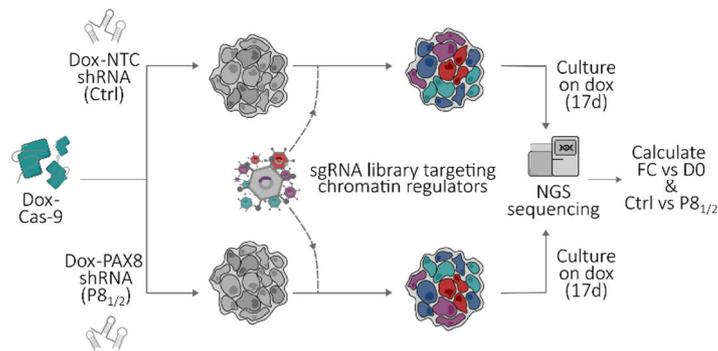
## 2.19 Statistical analysis

Statistical analyses were performed in R. Kruskal-Wallis was used for competitive proliferation assays and comparison of dependency data between subtypes. For Kaplan-Meier curves of tumour free progression, the logrank test was used. The hypergeometric distribution (phyper) test was to measure the significance of ATAC/DNAse I, gene set, and genomic region overlaps. Pearson correlation was used for correlation analysis. For all tests, a p-value of q-value of  $< 0.05$  was considered significant.

# Identifying Modulators of PAX8 Dependency in ccRCC

### 3.1 Introduction

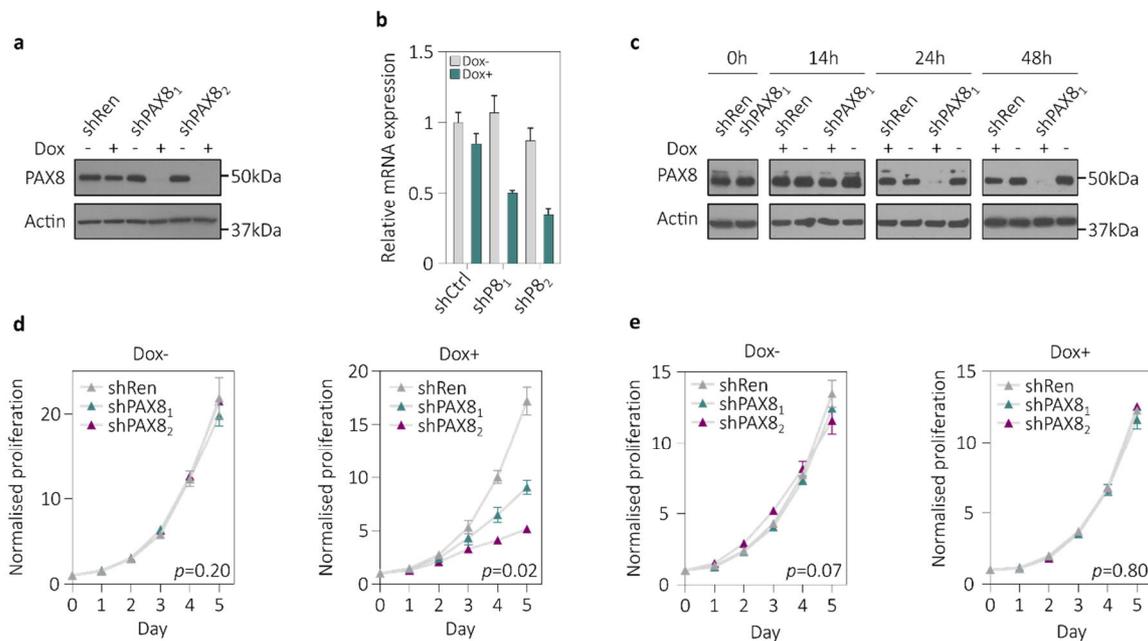
The advent of CRISPR-Cas9 genome editing has had far reaching implications for molecular biological research and medicine. The ability to efficiently edit the genome, with less off-target activity than previously available methods, has expedited mechanistic study and enabled the design of complex experiments, such as pooled loss-of-function genetic screening, henceforth known as CRISPR-cas9 screening<sup>291,292</sup>. To date there are many different screening protocols with various advantages and disadvantages, but each has the same core tenants. A simple example: a cell line expressing Cas9 is transduced at a low MOI with a library of vectors containing a single sgRNA construct targeting a gene, cells are passaged for approximately 12 doubling events and the relative abundance of sgRNAs at the end of the screen is compared to day 0 by next generation sequencing (NGS). Collaborative efforts from the Sanger and Broad institutes have leveraged the power of genome-wide CRISPR/Cas9 screening (~18,000 genes) to create a dependency map of almost a thousand cancer models to date<sup>257</sup>. Whilst the wealth of this data set cannot be overstated, there remains an opportunity for more nuanced mechanistic applications for CRISPR/Cas9 screens, which are difficult to scale. To look for resistance mechanisms to lineage factor therapy, I performed a CRISPR/Cas9 screen in 786-M1A cells to uncover genes capable of modulating the dependence of ccRCC lines on PAX8 (**Figure 11**). The screen had two arms, a control with wild type *PAX8* expression, and an experimental with *PAX8* suppressed. The primary aim of the screen was to identify genes specifically enriched in the experimental arm, and hence compensated for PAX8 loss. Due to the complex nature of the screen a smaller, more focused library was required. An sgRNA library targeting chromatin regulators was selected because they are key developmental mediators and have been shown to modulate the efficiency of cellular re-programming events<sup>293-295</sup>.



**Figure 11** | A schematic of a pooled CRISPR-Cas9-based loss of function screen using an sgRNA library targeting chromatin regulators to identify genes capable of modulating the dependence of ccRCC cell lines on PAX8 inhibition.

### 3.2 Main

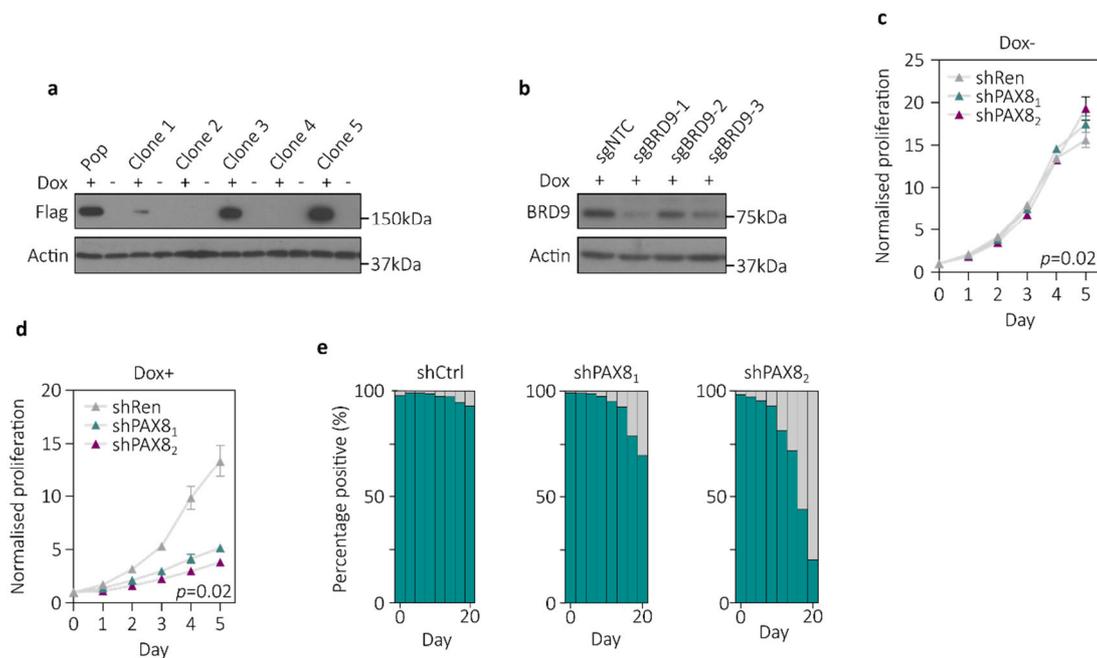
For the experimental arm, I selected two shRNAs that could effectively suppress *PAX8* expression (**Figure 12a-b**). For the control arm, a non-targeting shRNA against renilla was used. I used shRNAs for two reasons: (1) a KD more closely mimics the action of a drug than a CRISPR-cas9 KO, which is relevant when screening for resistance mechanisms for potential therapy, and (2) they suppress protein expression faster and so establish a *PAX8* KD background before library selection has begun (**Figure 12c**). In line with genome-wide screening data, both *PAX8* shRNAs triggered a strong proliferative phenotype in renal cancer cells (786-M1A) but not in HeLa cells which do not express *PAX8* (**Figure 12d-e**)<sup>296</sup>. Confirming no phenotype in HeLa cells suggested that the proliferative effect was due to the KD of *PAX8* protein instead of an off-target effect<sup>297,298</sup>.



**Figure 12** | (a) Western blot of PAX8 expression in 786-M1A cells expressing doxycycline-inducible shRNAs targeting PAX8 (PAX8<sub>1/2</sub>) or renilla control (Ctrl). Cells were treated with 0.6µg/ml doxycycline for three days. (b) Relative PAX8 mRNA expression as determined by qRT-PCR in 786-M1A cells treated with 0.6µg/ml doxycycline for three days. (c) Western blot time course of PAX8 expression in 786-M1A cells expressing a doxycycline-inducible shRNAs. Cells were treated with 0.6µg/ml doxycycline. N to be determined. (d) Confluency-based cellular proliferation assay, with 786-M1A cells expressing inducible shRNAs. ‘Dox +’ cells were pre-treated with 0.6µg/ml doxycycline for 3 days before assay start. Three technical replicates per condition. Error bars are SD. PAX8 shRNA combined for statistical test by Kruskal- Wallis. (e) Confluency-based cellular proliferation assay, with HeLa cells expressing inducible shRNAs. ‘Dox +’ cells were pre-treated with 0.6µg/ml doxycycline for 3 days before assay start. Three technical replicates per condition. Error bars are SD. PAX8 shRNA combined for statistical test by Kruskal- Wallis.

To ensure efficient Cas9 activity, we selected a single cell clone with tight dox-inducible control of Cas9 expression and high genome editing efficiency (Figure 13a-b). Similar to the parental cells, the selected clone, 786-M1A-C6, was sensitive to PAX8 suppression (Figure 13c-d). The shRNAs were fused to a fluorescent reporter, which allowed me to monitor expression and select for cells that maintained PAX8 KD throughout the experiment (Figure 13e)<sup>269</sup>. The strong selective pressure against PAX8 KD produced fast growing ‘escaped cells’ with normal PAX8 expression, and in a screen would result in the non-

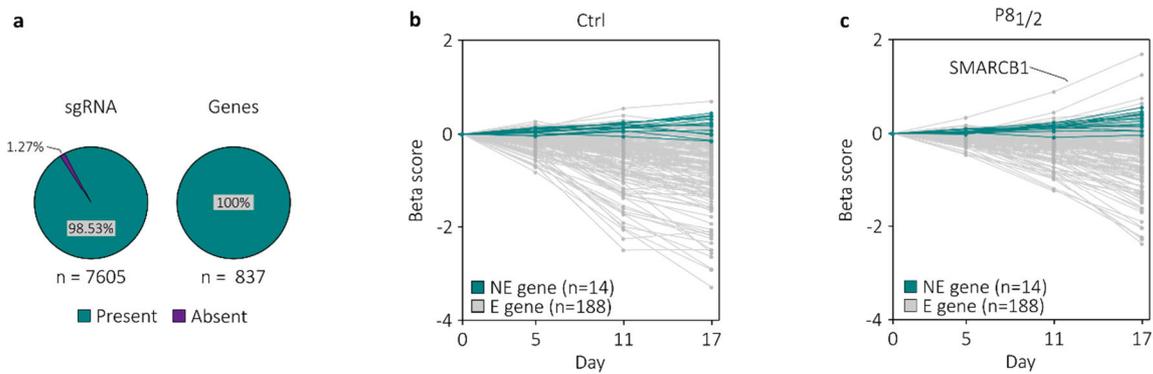
specific amplification of random sgRNAs. Without sorting, this unspecific bias would have made the screen difficult to interpret.



**Figure 13** | **(a)** Western blot analysis of Cas9 expression in 786-M1A doxycycline-inducible Cas9 single-cell clones. Cells were treated with 0.6µg/ml doxycycline for 3 days. **(b)** Western blot analysis of BRD9 expression in 786-M1A clone 6 (786-M1A-C6), expressing sgRNA targeting BRD9 or a non-targeting control. Cells were treated with 0.6µg/ml doxycycline for 5 days. **(c-d)** Confluency-based cellular proliferation assay, with 786-M1A Cas9 clone 6 (786-M1A-C6) cells expressing inducible shRNAs. ‘Dox +’ cells were pre-treated with 0.6µg/ml doxycycline for 3 days before assay start. Three technical replicates per condition. Error bars are SD. PAX8 shRNA combined for statistical test by Kruskal- Wallis. **(e)** Escaper assay with 786-M1A-C6, showing the percentage of cells expressing the fluorophore dsRed as a measure of shRNA expression. Three technical replicates per condition. Error bars are SD.

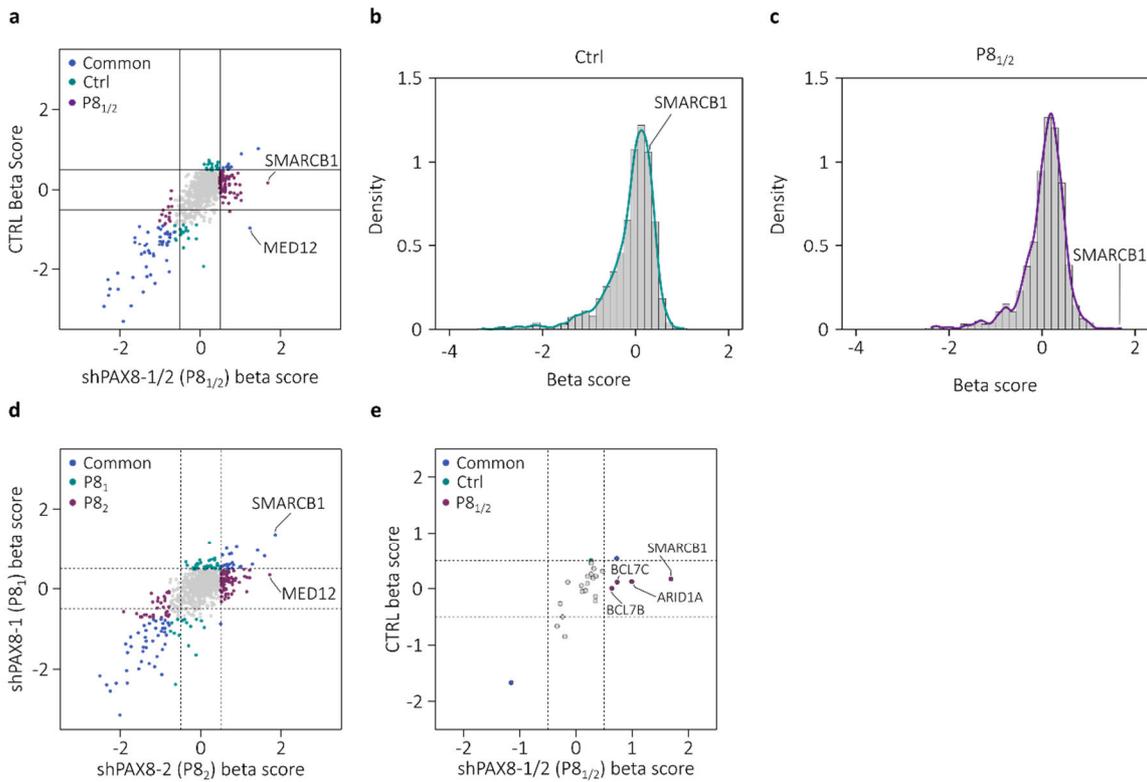
The screen was run for 17 days to allow for sufficient time for sgRNA selection and to ensure enough non-escaped cells for sorting. Time points were also collected on days 0, 5, and 11 for quality assurance. Over 98% of sgRNAs and all 837 genes were represented at the start of the screen (**Figure 14a**). To test whether the screen selection occurred as it should, I measured the change in sgRNA abundance for

genes classified in the DepMap project as ‘essential’ or ‘non-essential’. As expected, there was a negative selection for known essential genes and no selection for non-essential genes (**Figure 14a-c**)<sup>257</sup>.



**Figure 14** | **(a)** Pie chart of percentage sgRNA constructs and genes represented at day 0 of the chromatin regulator CRISPR-Cas9 screen. **(b-c)** Beta scores of essential (n=188) and non-essential genes (n=14) at time points throughout the screen for the control arm (b) and experimental arm (c).

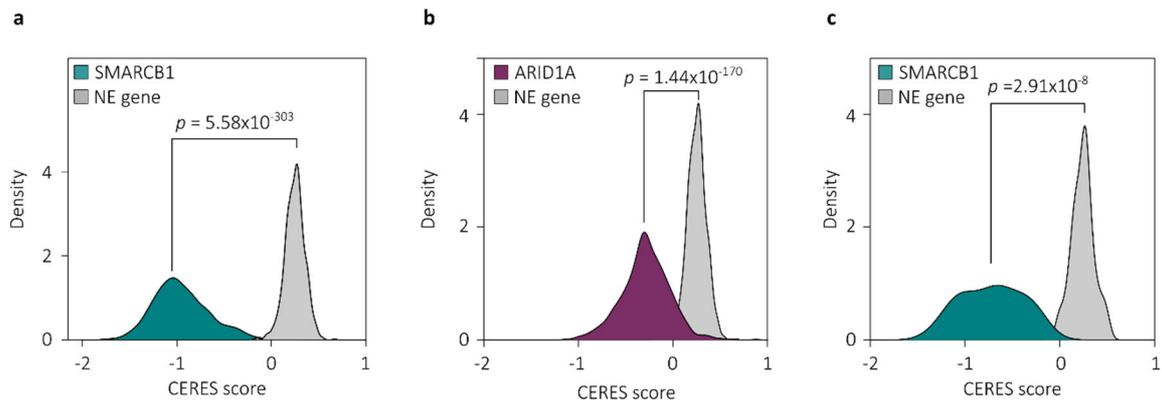
The most enriched PAX8 KD condition (P8<sub>1/2</sub>)-specific hit from the screen was SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin (SMARCB1), a member of the SWI/SNF chromatin remodelling complex, which has a key role in chromatin landscape maintenance in development and tissue homeostasis (**Figure 15a-c**)<sup>299,300</sup>. SMARCB1 sgRNAs were enriched in both PAX8 shRNA conditions, reducing the chance that SMARCB1 was enriched in response to an shRNA off-target effect (**Figure 15d**). Three additional SWI/SNF complex members were also specifically enriched in the PAX8 KD condition, of which AT-Rich interaction Domain 1A (ARID1A) showed the strongest enrichment (**Figure 15e**).



**Figure 15** | (a-e) Changes in sgRNA abundance over time, measured by calculating beta scores using the top 3 three enriched sgRNAs per gene relative to day 0, from two technical replicates. Highlighted points have a beta score <-0.5 or >0.5 and an adjusted p-value < 0.05. (a) The beta scores for the ctrl arm of the screen versus the pooled experimental arm (P8<sub>1/2</sub>). (b-c) Distribution of beta scores at day 17, for the control arm (b) and the pooled experimental arm (c). The bin containing the beta score for SMARCB1 is labelled. (d) The beta scores for each of the experimental arms of the screen. (e) The beta scores for the ctrl arm of the screen versus the pooled experimental arm, filtered for SWI/SNF complex members.

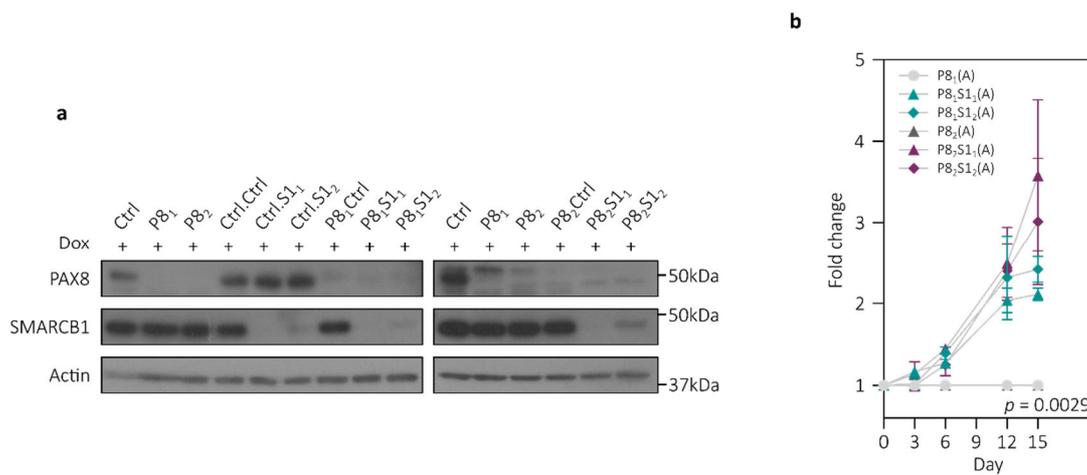
ARID1A and SMARCB1 inactivation have recently been shown to reduce the sensitivity of breast cancer cells to oestrogen receptor inhibition, supporting my findings in a second cancer type<sup>272,301</sup>. Interestingly, unlike ARID1A, and in contrast to its effect in our screen, SMARCB1 is a prototypical pan-cancer essential gene (Figure 16a-b)<sup>257</sup>. SMARCB1 is a strong dependency in RCC cell lines but was not depleted in the control arm of the screen, suggesting 786-M1A-C6 cells were better able to tolerate its loss (Figure 15a and Figure 16c). The second strongest hit from the screen after SMARCB1 was Mediator

Complex Subunit 12 (MED12), a mediator for enhancer-bound transcription factors and the transcription apparatus at gene promoters (**Figure 15a,d**)<sup>302</sup>. However, MED12 sgRNAs were only significantly enriched in one PAX8 shRNA condition (P8<sub>2</sub>) and not the other (P8<sub>1</sub>), and so was not followed up.



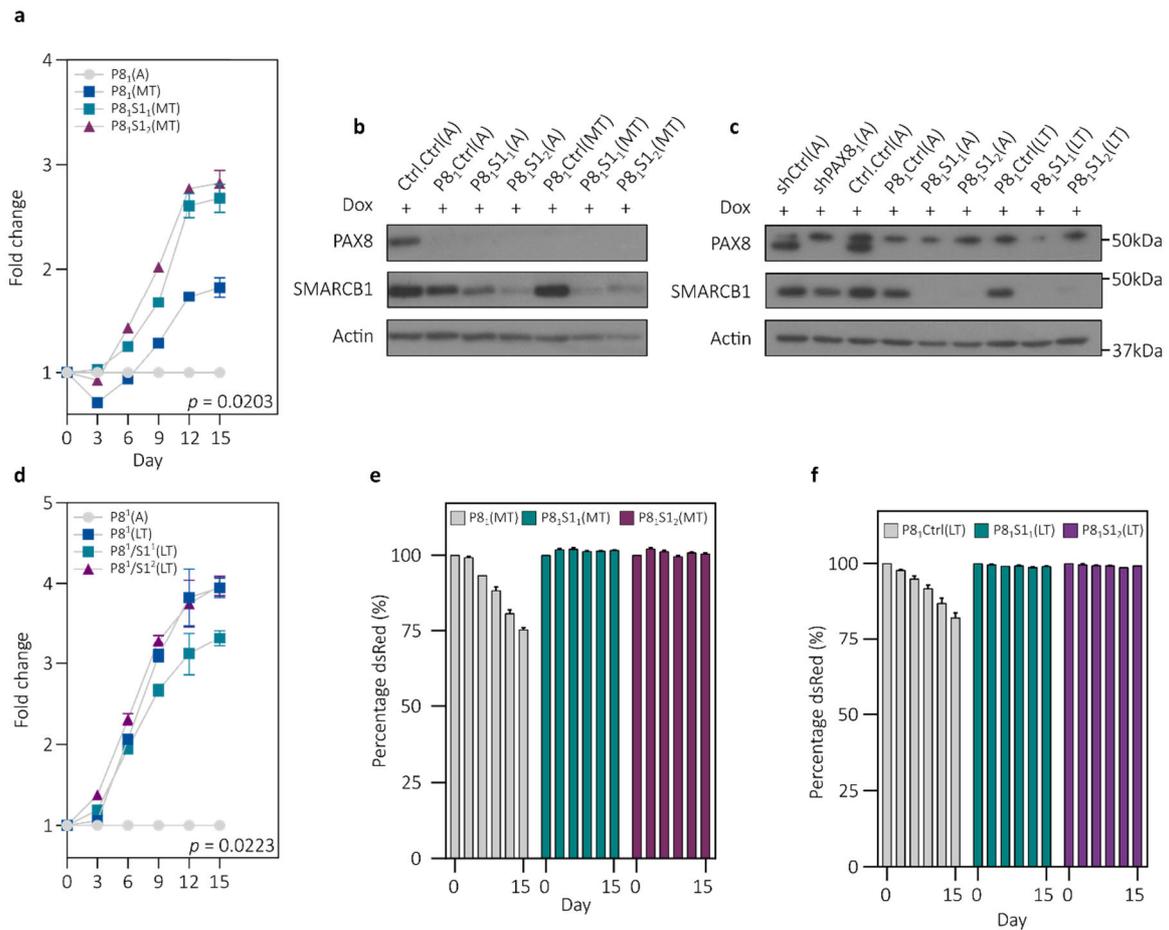
**Figure 16** | **(a)** Genetic dependency data for 946 cell lines from the DepMap project. Distribution of CERES scores for the gene *SMARCB1* and an example non-essential gene (Cyclic Nucleotide Binding Domain Containing 1; *CNBD1*). Statistical testing using Kruskal-Wallis. **(b)** Genetic dependency data from the DepMap project (n=946). Distribution of CERES scores for *ARID1A* and an example non-essential gene (*CNBD1*). Kruskal-Wallis test. **(c)** Genetic dependency data for renal cell carcinoma lines (n=21) from the DepMap project. Distribution of CERES scores for *SMARCB1* and an example non-essential gene (*CNBD1*). Kruskal-Wallis test.

We validated the results from the screen by competitive cellular proliferation assays, using one PAX8 shRNA (P8<sub>1</sub>) and two additional *SMARCB1* sgRNAs (S1<sub>1/2</sub>) (**Figure 17a**). Acute PAX8 KD/*SMARCB1* KO (P8<sub>1/2</sub>S1<sub>1/2</sub>(A)) resulted in a proliferative rescue but the cells appeared to grow less stably and were sensitive to passaging (**Figure 17b**).



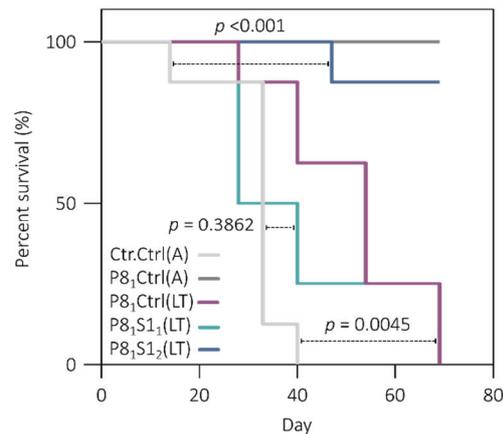
**Figure 17** | (a) Western blot analysis of PAX8 and SMARCB1 expression in 786-M1A-C6 cells expressing combinations of sgRNAs (ctrl or S11/2) and inducible shRNAs (ctrl or P81/2). Cells were treated for 6 days with 0.6µg/ml doxycycline. (b) Competitive proliferation assays, using 786-M1A Cas9 clone 6 (786-M1A-C6) cells expressing combinations of sgRNAs (Ctrl or S1<sub>1/2</sub>) and inducible shRNAs (Ctrl or P8<sub>1/2</sub>), competed against cells with a PAX8 KD (P8<sub>1/2</sub>). Three technical replicates per condition. Doxycycline was added at day 0. Error bars are SD. Statistical testing using Kruskal-Wallis.

Given time (~1 month, mid-term, MT) they began to grow more robustly (**Figure 18a**). Using FACS of the fluorescent marker attached to the PAX8 shRNA, we were able to maintain a PAX8 KD state and validated by western blot that the SMARCB1 KO was maintained (**Figure 18b**). SMARCB1 loss thus provides an immediate proliferative advantage to PAX8 KD cells but with a significant stability trade-off that can be selected against over time. Over a period of ~2-3 months (long-term, LT) the P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells maintained a similar growth phenotype but cells without SMARCB1 KO (P8<sub>1</sub>Ctrl) also adapted to PAX8 suppression (**Figure 18c-d**). Critically, despite a partial proliferative rescue, a selective pressure to regain PAX8 suppression was maintained in P8<sub>1</sub>Ctrl(LT/MT) cells but not their SMARCB1 KO counterparts, as evidenced by the gradual loss of PAX8 shRNA expression in P8<sub>1</sub>Ctrl(LT/MT) cells (**Figure 18e-f**).



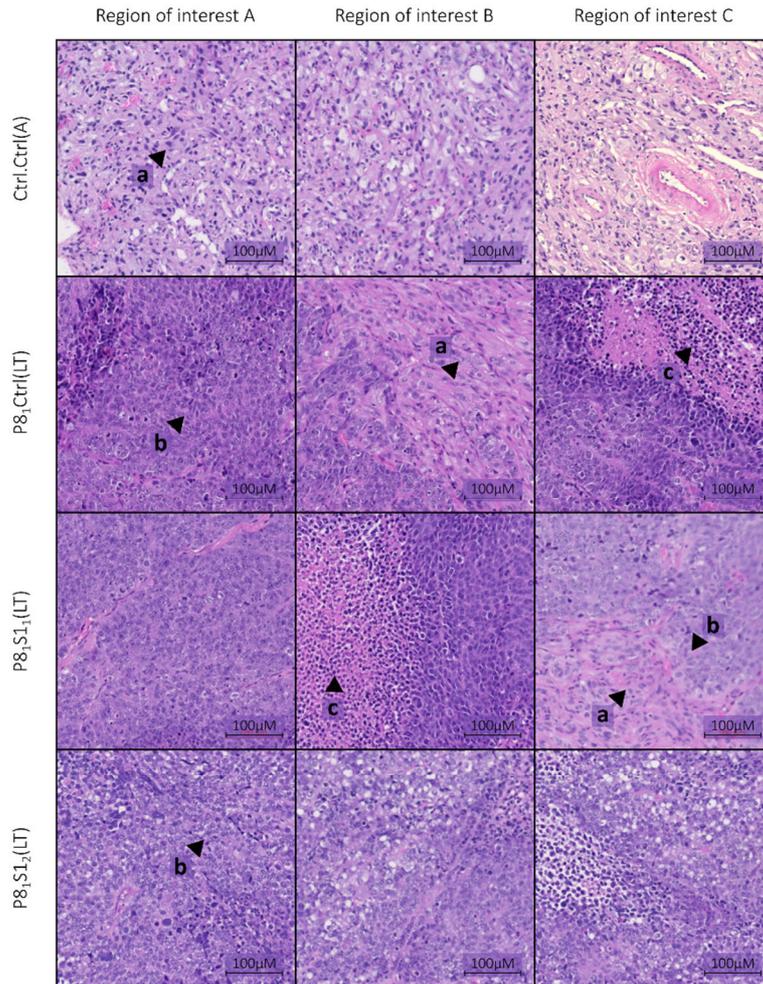
**Figure 18** | **(a)** Competitive proliferation assays, using 786-M1A Cas9 clone 6 (786-M1A-C6) cells expressing combinations of sgRNAs (Ctrl or S<sub>1/2</sub>) and inducible shRNAs (Ctrl or P8<sub>1/2</sub>), competed against cells with a PAX8 KD (P8<sub>1/2</sub>). Three technical replicates per condition. were pre-treated for ~1 month with doxycycline before the start of the assay (midterm, MT). Error bars are SD. Statistical testing using Kruskal-Wallis. **(b)** Western blots of PAX8 and SMARCB1 expression for 786-M1A-C6 cells expressing combinations of sgRNAs (ctrl or S<sub>1/2</sub>) and inducible shRNAs (ctrl or P8<sub>1</sub>), cultured for ~1month (MT) on 0.6µg/ml doxycycline, compared to acutely (A) treated cells. **(c)** Western blots of PAX8 and SMARCB1 expression for 786-M1A-C6 cells expressing combinations of sgRNAs (ctrl or S<sub>1/2</sub>) and inducible shRNAs (ctrl or P8<sub>1</sub>), cultured for ~2-3 months (MT) on 0.6µg/ml doxycycline, compared to acutely (A) treated cells. **(d)** Competitive proliferation assays, using 786-M1A Cas9 clone 6 (786-M1A-C6) cells expressing combinations of sgRNAs (Ctrl or S<sub>1/2</sub>) and inducible shRNAs (Ctrl or P8<sub>1/2</sub>), competed against cells with a PAX8 KD (P8<sub>1/2</sub>). Three technical replicates per condition. were pre-treated for ~2-3 months with doxycycline before the start of the assay (midterm, MT). Error bars are SD. Statistical testing using Kruskal-Wallis. **(e-f)** Escaper assay with 3 technical replicates using 786-M1A-C6 cells, showing the percentage of cells expressing the fluorophore dsRed as a measure of PAX8 shRNA expression for cells pre-cultured on doxycycline for (e) ~1 month (MT) or (f) ~2-3 months, normalised to day 0. Cells were sorted at the beginning of the assay to ensure a starting point of 100% dsRed. Error bars are SD.

Tumour growth *in vivo* was totally abrogated by PAX8 KD (P8<sub>1</sub>(A)), partially rescued for P8<sub>1</sub>Ctrl(LT) cells, and completely rescued for P8<sub>1</sub>S1<sub>1</sub>(LT) cells. P8<sub>1</sub>S1<sub>2</sub>(LT) cells, on the other hand, showed only weak tumorigenicity (**Figure 19**).



**Figure 19** | Kaplan-Meier analysis of tumour-free survival in athymic mice, after subcutaneous injection of the following 786-M1A-C6 cell lines; Ctrl.Ctrl(A), P8<sub>1</sub>Ctrl(A), P8<sub>1</sub>Ctrl(LT), P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT). Acute cells were treated with 0.6µg/ml doxycycline for 3 days prior to injection. Long term cells were maintained on 0.6µg/ml doxycycline. Four mice were injected with 5x10<sup>5</sup> cells in both flank per condition. Logrank test. Acknowledgment:- As stated in the Preface, the mouse work was completed by V.C, J.G and A.S. Analysis and plotting was my own work.

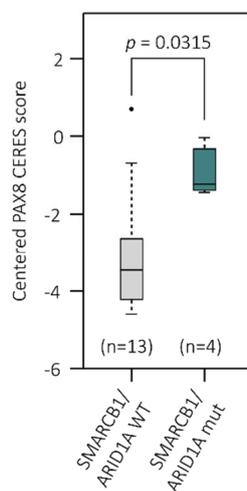
Histological analysis revealed marked changes in the morphology of P8<sub>1</sub>Ctrl(LT) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells compared to the control tumours. Ctrl.Ctrl(A) tumours exhibited the expected high-grade ccRCC phenotype with sarcomatoid dedifferentiation, whereas tumours arising from P8<sub>1</sub>Ctrl(LT) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells showed high-grade undifferentiated histology, with large regions of necrosis and an appearance of neuroendocrine differentiation (**Figure 20**). In accordance with the *in vitro* observations, there were regions of detectable sarcomatoid histology reminiscent of Ctrl.Ctrl(A) tumours in P8<sub>1</sub>Ctrl(A) tumours, providing evidence of PAX8 KD escaped cells *in vivo* (**Figure 20**). Unlike the *in vitro* findings, potential escaped cells were detectable in one of the P8<sub>1</sub>S1<sub>1</sub>(LT) but not P8<sub>1</sub>S1<sub>2</sub>(LT) (**Figure 20**).



**Figure 20** | Haemotoxylin and Eosin (H&E) staining of tumours from Figure 19. Tumours were harvested at day 47 after subcutaneous injection. **(a)** Examples of high-grade ccRCC with sarcomatoid dedifferentiation. **(b)** high-grade undifferentiated histology with neuroendocrine differentiation. **(c)** Regions of necrosis. Acknowledgments:- As stated in the Preface, the mouse work was completed by V.C, J.G and A.S. Analysis and plotting was my own work.

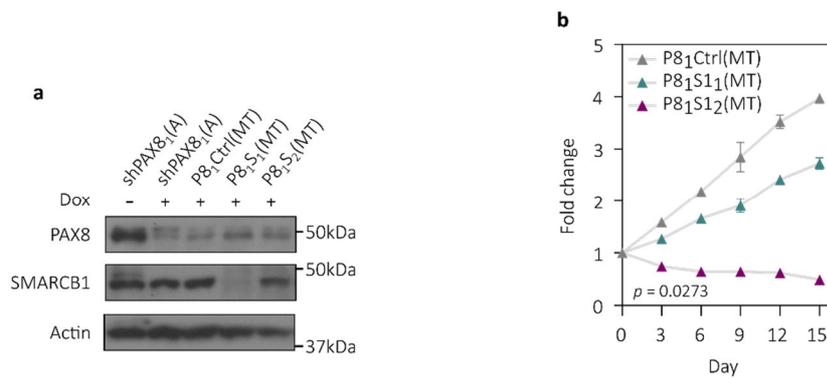
To expand our study to additional ccRCC models, we took a systematic approach and identified *VHL* mutant ccRCC cell models which have a non-synonymous *SMARCB1* (n=3) or *ARID1A* (n=1) mutation from the cell line encyclopaedia (CCLE) and compared their PAX8 inhibition sensitivity to their *SMARCB1/ARID1A* wild type counterparts (n=13) using loss-of-function data from DepMap<sup>257,303</sup>. We chose to use *VHL* mutant renal cancer lines because *VHL* inactivation is a highly specific truncal event

in ccRCC<sup>10</sup>. Damaging non-synonymous *VHL* mutations were determined based on resistance to KO, as *VHL* is a known essential gene. *SMARCB1* and *ARID1A* mutations were prioritised if they were annotated as a TCGA/COSMIC hotspot or predicted as ‘damaging’. In line with the findings from our screen, *SMARCB1/ARID1A* mutant lines showed strong resistance to PAX8 KO compared to their WT counterparts (**Figure 21**). However, there were also two *VHL* mutant cell lines that showed similar resistance to PAX8 KO but did not have a *SMARCB1*, *ARID1A* or any SWI/SNF mutation, indicating that PAX8 inhibition resistance can arise through several mechanisms.



**Figure 21** | PAX8 centred CERES dependency score for *VHL*-mutant RCC lines from the DepMap project, which are either WT (n=13) or mutant for *ARID1A* (n=1) or *SMARCB1* (n=3). Kruskal-Wallis.

In line with this, PAX8 depletion in UOK101 cells, another *VHL* mutant ccRCC cell line, resulted in the quick emergence of a resistant population which maintained the essential status of *SMARCB1* (**Figure 22a-b**). One month’s culture on doxycycline resulted in resistant P8<sub>1</sub>Ctrl cells, partially resistant P8<sub>1</sub>S1<sub>2</sub> cells which had selected against the *SMARCB1* KO, and P8<sub>1</sub>S1<sub>1</sub> cells which had maintained a *SMARCB1* KO but had not adapted to PAX8 KD.



**Figure 22** | (a) Western blot of PAX8 and SMARCB1 expression in UOK101 cells, expressing a combination of sgRNAs (ctrl or  $S_{1/2}$ ) and an inducible shRNA targeting PAX8 ( $P8_1$ ). Acute (A) cells were pre-treated with 0.6 $\mu$ g/ml of doxycycline for three days and midterm (MT) cells were cultured on 0.6 $\mu$ g/ml doxycycline for ~1month. (b) Competitive proliferation assay with UOK101 cells expressing a combination of sgRNAs (ctrl or  $S_{1/2}$ ) and a PAX8 inducible-shRNA ( $P8_1$ ), cultured for ~1 month on doxycycline (MT), competed against cells with a PAX8 KD. To establish a PAX8 KD to compete against, doxycycline was added 3 days before starting the assay. Three technical replicates per condition. Error bars are SD.

### 3.3 Summary

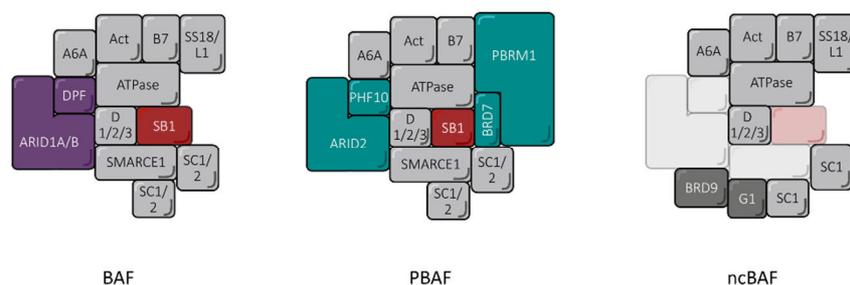
Several factors were associated with lineage factor independence in ccRCC, two of the top 5 enriched hits from the screen were SWI/SNF complex members, which are critical epigenetic developmental mediators. SMARCB1 and ARID1A have also been implicated in resistance to tamoxifen treatment in luminal breast cancer, generalising these findings to another cancer type. *SMARCB1* was the strongest hit from the screen, the KO of which was capable of rescuing PAX8 loss both *in vitro* and *in vivo*, through a process that initially introduced instability but provided a substrate for the rapid selection of pro-proliferative traits. Naturally occurring *SMARCB1* mutations in patient-derived ccRCC lines provide resistance to PAX8 KO in the DepMap data, expanding these observations to additional RCC models. This data set contained two cell lines that overcame PAX8 sensitivity without a detectable SWI/SNF mutation, implicating multiple routes to adaptation. This was validated by the emergence of resistance

in P8<sub>1</sub>Ctrl(LT) cells for both 786-M1A-C6 and UOK101 cells, in a SMARCB1 independent manor. Notably, despite an increase in proliferation, the emergence of shRNA ‘escapers’ was observed for 786-M1A P8<sub>1</sub>Ctrl(LT) cells, suggesting an incomplete de-coupling from PAX8 pro-tumorigenic signalling.

# SMARCB1 Inactivation Promotes Large-Scale Chromatin Remodelling and Cellular Instability

## 4.1 Introduction

SMARCB1 is a subunit of the SWI/SNF chromatin remodelling complexes<sup>299</sup>. The SWI/SNF complexes were originally described in yeast as a multi-subunit protein complex capable of regulating the expression of inducible genes involved in mating-type switching and growth on sucrose medium, through the ATP-dependant mobilisation of nucleosomes, thereby enabling the binding of transcriptional activators to DNA<sup>304–309</sup>. To date, three mSWI/SNF complexes have been described: the BRG1/BRM-associated factor complexes (BAFs), the polybromo-associated BAF complexes (PBAFs), and the non-canonical BAF complexes (ncBAFs; **Figure 23**)<sup>310–313</sup>. SMARCB1 is a core subunit in BAF and PBAF but not the more recently described ncBAF complex<sup>314</sup>.



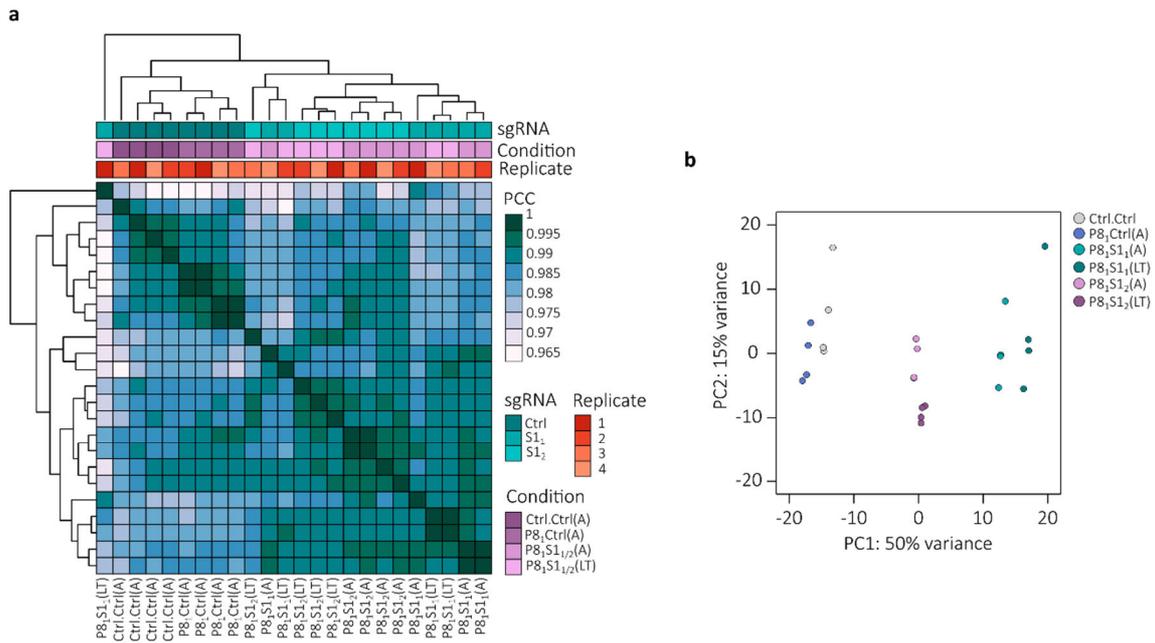
**Figure 23** | Schematic of subunit composition of mammalian BAF, PBAF and ncBAF complexes. Purple, cyan, and dark grey highlighted subunits have complex specificity, whereas light grey subunits are shared by two or more complexes. Red represents SMARCB1. Abbreviations :- A6A:ACTLA6A, Act:B-actin, ATPase:SMARCA2/4, B7:BCL7A/B/C, DPF:DPF1/2/3, D1/2/3:SMARCD1/D2/D3, SB1:SMARCB1, SC1/2:SMARCC1/C2, G1:GLTSCR1/1L. Schematic adapted from Michel et al., 2018.

Interestingly SWI/SNF complex members are the most commonly mutated chromatin regulator in cancer and one of the most frequently mutated tumour suppressors overall, mutated in ~20% of human cancers<sup>234,315</sup>. *SMARCB1* is mutated in >98% of Malignant rhabdoid tumours (MRT) and atypical teratoid rhabdoid tumours (AT/RT), which are aggressive paediatric tumours that occur predominately in the

kidney or soft tissue and central nervous system respectively<sup>316-318</sup>. MRT and AT/RT tumours have very stable genomes and biallelic loss of *SMARCB1* is often the only detectable mutation, suggesting that *SMARCB1* loss alone is sufficient for tumorigenesis in the right context<sup>319-323</sup>. To understand the role of *SMARCB1* in PAX8 inhibition resistance, we performed RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) to measure changes in the transcriptome and chromatin accessibility upon *SMARCB1* loss.

## 4.2 Main

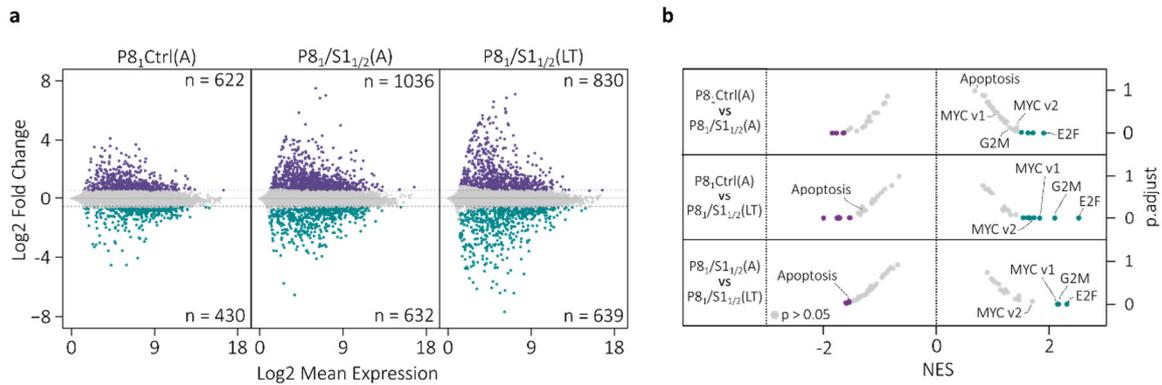
The pre-processing of the RNAseq FASTQ sequencing files passed quality control measures, as determined by FASTQC and the percentage of uniquely mapped reads (~80%). To assess replicate quality, I plotted a pairwise Pearson correlation heatmap with hierarchical clustering and a PCA plot (**Figure 24a-b**). All sample replicates correlated highly, with the exception of P8<sub>1</sub>S1<sub>1</sub>-R4. This sample was removed before running DEseq2 to determine differentially expressed genes. On a per-sample basis, Ctrl.Ctrl(A) and P8<sub>1</sub>Ctrl(A) cells clustered and the *SMARCB1* KO conditions clustered. Within the *SMARCB1* KO cluster, the P8<sub>1</sub>S1<sub>1</sub>(A) and P8<sub>1</sub>S1<sub>2</sub>(A) samples were more closely related to their long-term equivalents, as opposed to clustering by time point. Of note, P8<sub>1</sub>S1<sub>2</sub>(A/LT) cells were more closely related to Ctrl.Ctrl(A)/Ctrl.P8<sub>1</sub>(A) cells than P8<sub>1</sub>S1<sub>1</sub>(A/LT) cells.



**Figure 24** | (a) Heat map of pair-wise Pearson correlation coefficient matrix with hierarchical clustering, for normalised RNAseq samples. (b) PCA plot of RNAseq normalised expression data, using the first two principal components.

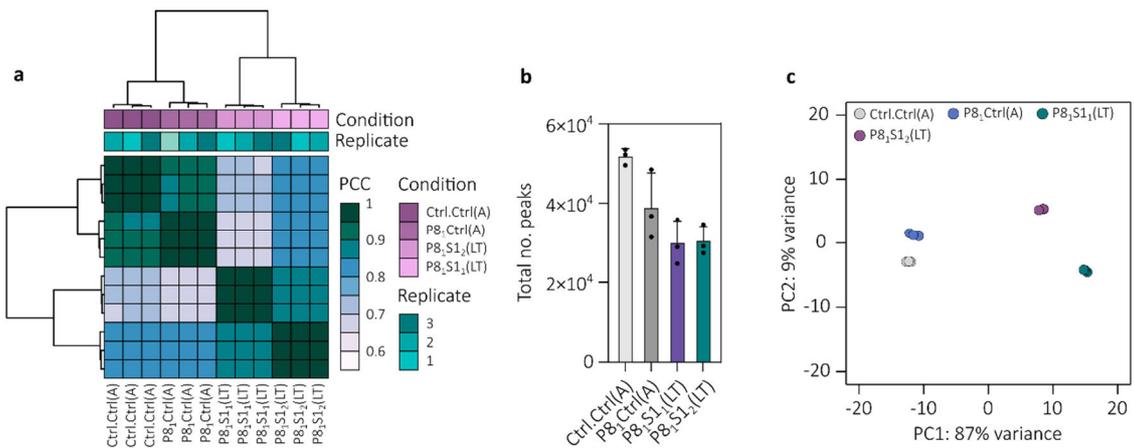
Using DEseq2 I detected differentially expressed (DE) genes across conditions, with more gene expression changes in the P8<sub>1</sub>S1<sub>1/2</sub>(A/LT) conditions compared to P8<sub>1</sub>Ctrl(A) (**Figure 25a**). To draw biological insights from the DE genes I performed gene set enrichment analysis (GSEA), to look for up and downregulated signalling pathways. I opted to use GSEA instead of a hypergeometric-based pathway overlap approach because it does not require defining an arbitrary cut-off for DE genes and is more sensitive to subtle transcriptional changes<sup>324,325</sup>. GSEA draws on a database (MsigDB) of nine major collections of gene sets, hallmarks, and Collection 1-8 (C1-8). To gain an initial overview of signalling changes I used the hallmarks collection, which is a summary of well-defined biological states or processes, assimilated from gene sets in C1-C8 that have coherent expression. In accordance with the proliferative phenotype, we found that acute SMARCB1 loss triggered an increase in proliferative gene signatures (MYC\_V1, MYC\_V2, G2M, E2F), which increased over time (P8<sub>1</sub>S1<sub>1/2</sub>(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT) (**Figure 25b**). The hallmark apoptosis signature was also reduced in the comparison between acute and

long-term SMARCB1 KO, supporting the observation that SMARCB1 simultaneously triggers heightened proliferation and instability, and over time clones that can tolerate SMARCB1 loss are selected for (Figure 25b).



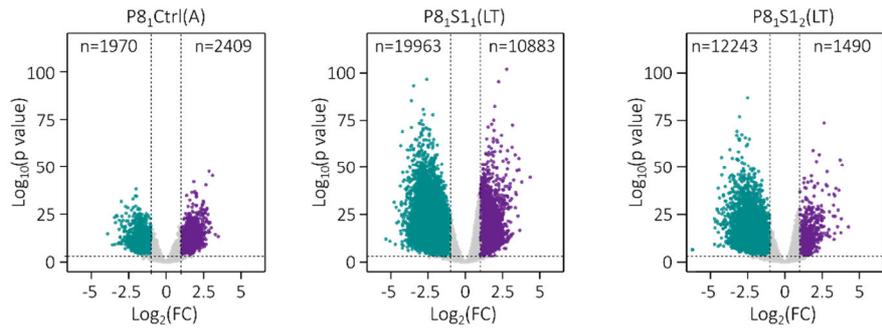
**Figure 25 | (a)** MA plots of RNA-seq differential expression analysis from 786-M1A-C6 cells. Gene expression fold change was calculated relative to shRen/sgNTC (Ctrl.Ctrl ) cells. Highlighted points have a FC >1.5 or <(-1.5) and p.adjust < 0.05. **(b)** Gene set enrichment analysis (GSEA) using the hallmarks collection from mSigDB, for different comparisons as indicated on the left. Highlighted points (purple/cyan) have a p.adjust < 0.05.

For the ATACseq experiment, the pre-processing of FASTQ files passed quality control and all replicates were highly correlated (Figure 26a). I detected ~72,000 high confidence peaks in total across conditions, with control and P8<sub>1</sub>Ctrl(A) having a similar number of peaks and P8<sub>1</sub>S1<sub>1/2</sub>(LT) having substantially less (Figure 26b). The sample clustering of ATACseq samples bore a strong resemblance to the RNAseq experiment; Ctrl.Ctrl(A) and P8<sub>1</sub>Ctrl(A) clustered and the SMARCB1 KO samples clustered, and P8<sub>1</sub>S1<sub>2</sub>(LT) cells correlated more highly with the Ctrl.Ctrl(A)/P8<sub>1</sub>Ctrl(A) cluster than P8<sub>1</sub>S1<sub>1</sub>(LT) (Figure 26a,c).



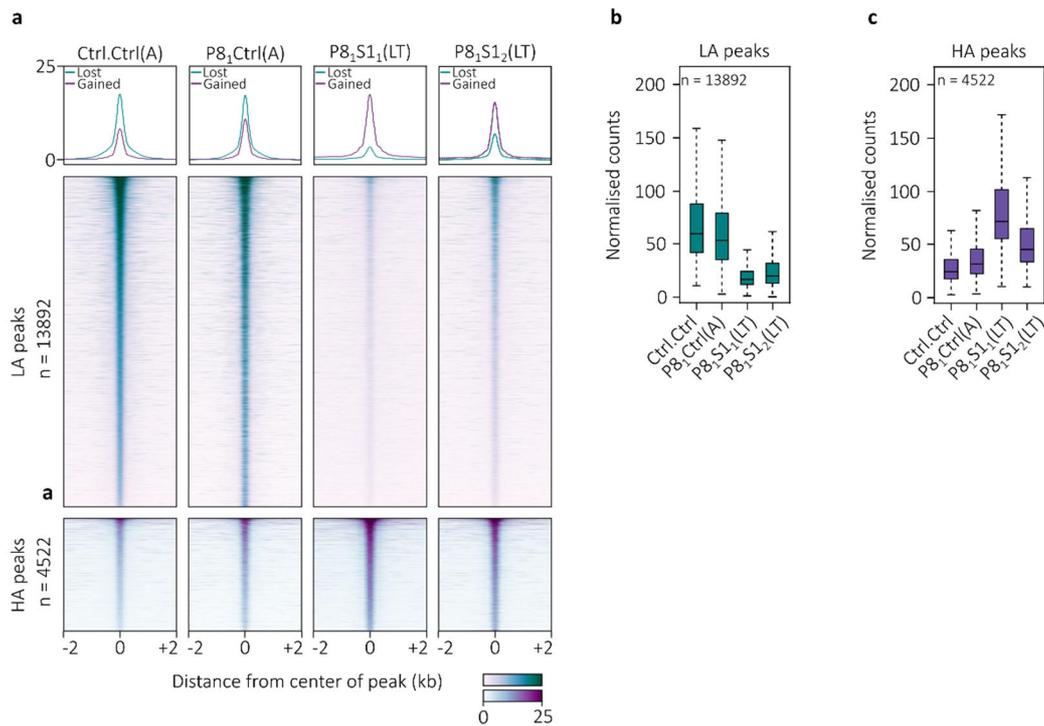
**Figure 26** | **(a)** Heat map of a pair-wise Pearson correlation coefficient matrix with hierarchical clustering, for all ATACseq samples using normalised counts across a consensus list of all peaks merged across conditions. **(b)** Total number of called peaks per condition. Q value < 0.01. Error bars are SD. **(c)** PCA plot of ATACseq normalised count data for a consensus list of all peaks merged across conditions, using the first two principal components.

I performed differential expression analysis using the list of high confidence consensus peaks, to identify differentially accessible (DA) regions between samples. Compared to Ctrl.Ctrl(A) cells the number and magnitude of lower accessibility (LA) regions were far greater for the SMARCB1 KO cells than P8<sub>1</sub>Ctrl(A) cells (**Figure 27a-c**). In contrast, for the same comparison, I detected a similar number of higher accessibility (HA) regions for P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells and approximately four times more for P8<sub>1</sub>S1<sub>1</sub>(LT) (**Figure 27a-c**). In line with the correlative analysis, P8<sub>1</sub>S1<sub>1</sub>(LT) cells had a greater number of both LA and HA regions compared to P8<sub>1</sub>S1<sub>2</sub>(LT) cells.



**Figure 27** | Volcano plots showing differentially accessible ATAC-seq regions for Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A), P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT). Highlighted points satisfy FC >2 or <(-2) and p.adjust < 0.001.

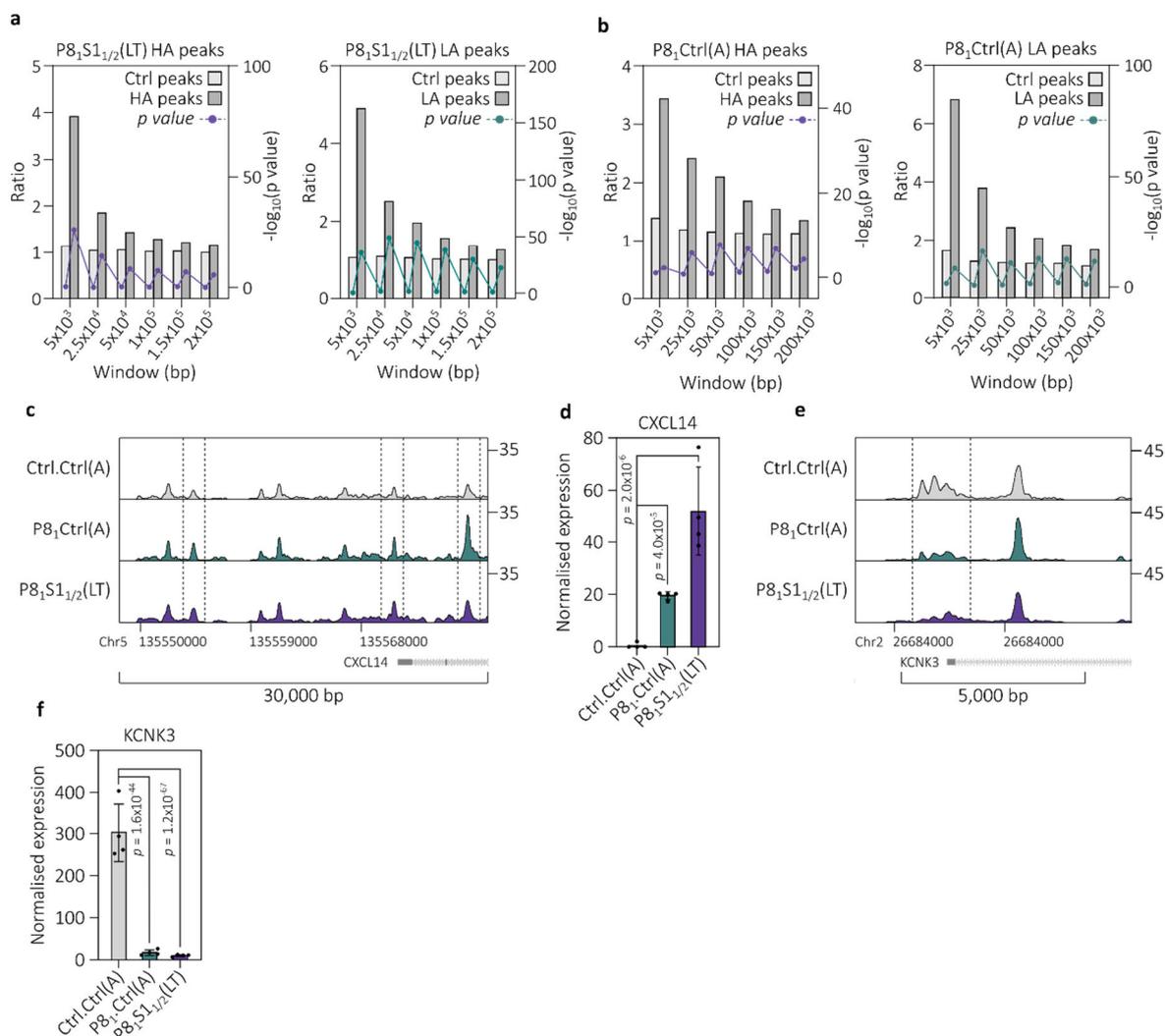
Despite differences, there was a large proportion of common differentially accessible regions in the SMARCB1 KO conditions compared to the control, 18,414 in total, 13892 of which had lower accessibility and 4522 higher accessibility (**Figure 28a-b**).



**Figure 28** | (a) Heatmaps showing normalised ATAC-seq signal  $\pm$  2kb centred on summits of differentially accessible (DA) regions, defined by Ctrl.Ctrl vs P8<sub>1</sub>S1<sub>1/2</sub> (FC >2 or <(-2), p.adjust <0.001). Top panels show the average signal for higher accessible and lower accessible regions. (b) Tukey plot of normalised counts across samples from (a) for high and low accessibility peak sets, defined by Ctrl.Ctrl vs P8<sub>1</sub>S1<sub>1/2</sub> (LT) (FC >2 or <(-2), p.adjust <0.001). For boxplots, the centre line shows the median, the box bounds represent the first and third quartiles and the whiskers extend to the highest and lowest values, no further than 1.5×IQR.

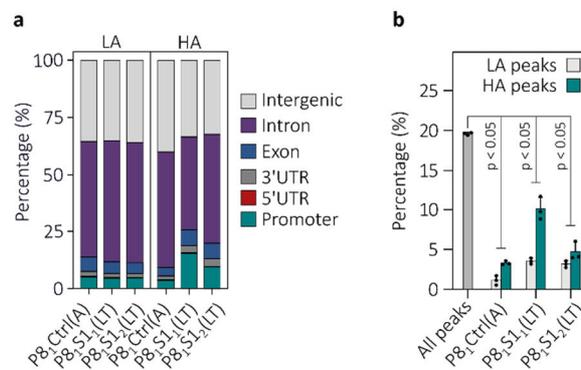
Given that the chromatin landscape informs gene expression, for quality assurance purposes, I measured the correlation of open chromatin and transcriptional changes. Due to the complex 3D architecture of the DNA-chromatin complex, if the DA region is not located at the transcriptional start site (TSS) of a gene (i.e. cis-regulatory region), it is very challenging to determine which gene it may affect<sup>326</sup>. However, on a population level, the probability that a cis-regulatory region interacts with a gene depreciates the further it is from the TSS<sup>327–329</sup>. I leveraged this principle by creating sequentially larger windows around HA and LA regions and tested whether they were significantly enriched for the TSS of up and downregulated genes respectively. As a control, I created matched sets of peaks from the ~72,000 high confidence consensus peak list with similar genomic attributes. As expected, the

changes in chromatin accessibility correlated well with gene expression changes and depreciated the larger the genomic window used (**Figure 29a-b**). Specific illustrative examples include the CXCL14 loci for the correlation of HA regions and upregulated gene expression, and the KCNK3 loci for LA region and downregulated gene expression equivalent (**Figure 29c-f**).



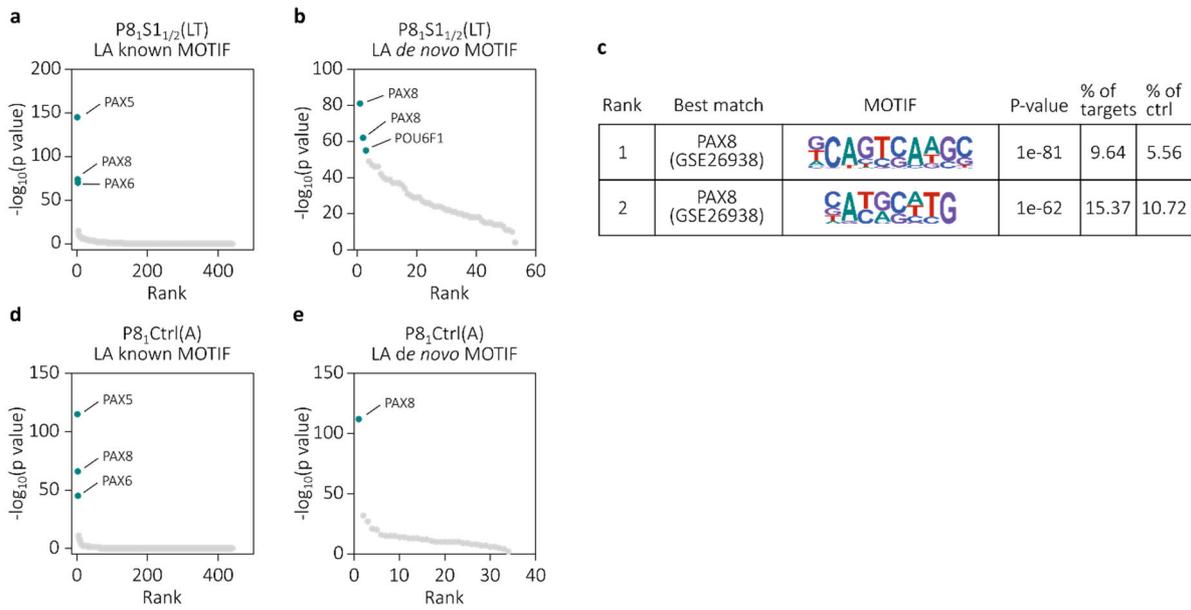
**Figure 29** | (a-b) Correlation of ATAC-seq peak and transcriptional changes, for (a) Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT) and (b) Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A). Left y-axis, the ratio of the number of down/up-regulated genes found within windows created around LA/HA regions compared to the number of expressed genes (universe) also found within the same genomic windows. Left y-axis, p-value, hypergeometric test. Matched Ctrl peaks for lower and higher accessible regions were generated from the consensus list of all peaks merged across conditions. (c) An example region of gained ATAC-seq signal in the CXCL14 locus for P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) compared to Ctrl.Ctrl(A) cells. (d) Normalised CXCL14 mRNA expression as determined by RNAseq for Ctrl.Ctrl(A), P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. P-values calculated by DESeq2. Error bars are SD. (e) An example region of lost ATAC-seq signal in the KCNK3 locus for P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) compared to Ctrl.Ctrl(A) cells. (f) Normalised KCNK3 mRNA expression as determined by RNAseq for Ctrl.Ctrl(A), P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. P-values calculated by DESeq2. Error bars are SD.

To begin to understand the significance of the large-scale changes in the genome observed, I first annotated the differentially accessible regions based on their location in the genome and found that the majority were intronic and intergenic and that there was a statistically significant underrepresentation of promoter annotations (**Figure 30a-b**). Without the accompanying chromatin Immunoprecipitation sequencing data for histone marks such as H3K27ac and H3K4me1 it was not possible to determine whether these regions are enhancers and so for the purpose of this work will be classified as cis-regulatory regions<sup>217</sup>.



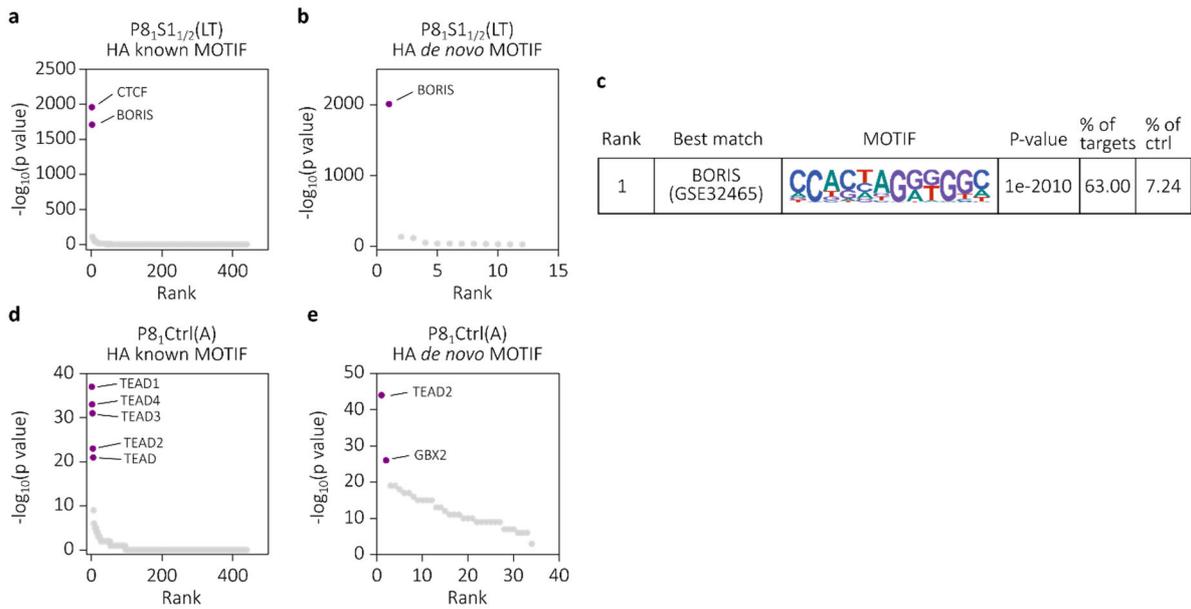
**Figure 30| (a)** Stacked bar plots of genomic annotations for LA and HA regions from comparisons, Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A), P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) (FC >2 or <(-2), p.adjust <0.001). **(b)** Percentage of regions annotated as a promoter in the consensus list of all peaks merged across conditions (dark grey) and lower and higher accessible regions from (Fig. 15). Two-tailed hypergeometric test.

Next, I performed motif analysis for the differentially accessible peak sets. As expected, the PAX motif was highly enriched in the LA set using both known and de novo motif analysis for P8<sub>1</sub>S1<sub>1/2</sub>(LT) and P8<sub>1</sub>Ctrl(A) (**Figure 31a-e**).



**Figure 31** | (a-b) Ranked plots of DNA motif analysis, either *de novo* or using a database of known binding motifs for LA regions from Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). (c) Examples of the top two scoring motifs from *de novo* motif analysis of LA regions from Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). (d-e) Ranked plots of DNA motif analysis, either *de novo* or using a database of known binding motifs for LA regions from Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A).

Unlike the LA regions, the most significantly enriched HA motifs for P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) differed. For P8<sub>1</sub>Ctrl(A) cells the most enriched motif was TEAD, whereas for P8<sub>1</sub>S1<sub>1/2</sub>(LT) it was the CTCF/BORIS motif, which strikingly was detected in 63% of HA peaks versus 7.24% of background peaks. (Figure 32a-e).



**Figure 32 | (a-b)** Ranked plots of DNA motif analysis, either *de novo* or using a database of known binding motifs for HA regions from Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). **(c)** Example of the top scoring motif from *de novo* motif analysis of HA regions from Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). **(d-e)** Ranked plots of DNA motif analysis, either *de novo* or using a database of known binding motifs for HA regions from Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A).

### 4.3 Summary

In the context of PAX8 KD, SMARCB1 KO promotes an unstable cellular state that can result in apoptosis but also provides an opportunity for the selection of proliferative traits. Over time clones that were able to tolerate the loss of SMARCB1 or have adapted, emerge and apoptotic programs are downregulated. Changes in the transcriptional program exerted by the loss of SMARCB1 are associated with large-scale epigenetic reprogramming, characterised predominately by the loss of accessibility at regions enriched for the PAX motif. This is in line with previous reports, demonstrating that SMARCB1 maintains enhancers and that the SWI/SNF complex has a role in regulating lineage-specific enhancer programs<sup>271,330,331</sup>. Interestingly, newly acquired HA regions were strongly enriched for the CTCF/BORIS

motif, which has been linked to the formation of undifferentiated rhabdoid tumours, the maintenance of a naive pluripotent stem cell state, and metastasis<sup>314,332-334</sup>.

5

## Reprogramming to Resist

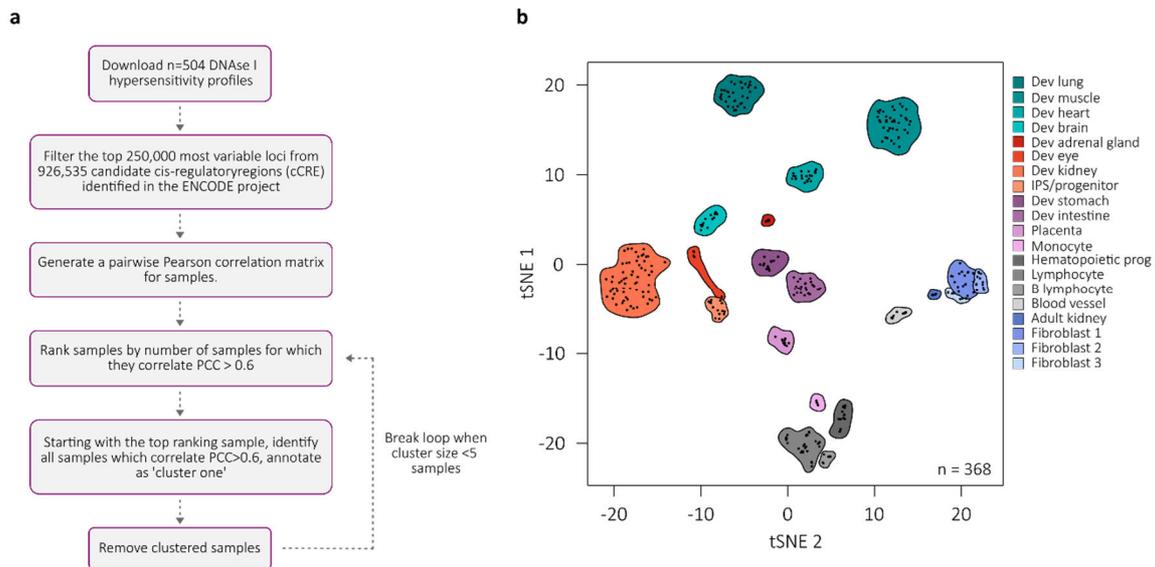
## 5.1 Introduction

Upon SMARCB1 loss, large-scale cis-regulatory reprogramming was observed. Specifically, the loss of regions enriched for the PAX8 motif, a key mediator of renal development, and a gain in accessibility at regions enriched for the CTCF-BORIS motif, which has been associated with the maintenance of a naive pluripotent stem cell state<sup>333,334</sup>. These data suggest SMARCB1 may promote resistance to PAX8 inhibition through a loss of renal differentiation. To test this hypothesis, I initially focused on changes in the cis-regulatory program as opposed to gene changes because they show a stronger lineage specificity than gene expression programs<sup>335</sup>. Using DNase hypersensitivity I profiles from the ENCODE project, spanning a range of cell lines and primary and developmental tissues, I defined sets of key regulatory regions which showed strong tissue specificity. This collection was then used to functionally annotate the DA accessible peak sets from my experiment. Guided by the overlap I ran targeted GSEA analysis using custom signatures and a public collection of cell-type specific gene signatures to validate these findings on a transcriptional level.

## 5.2 Main

First, I downloaded 504 DNase I hypersensitivity open chromatin profiles from the ENCODE project, spanning a range of adult and developmental cell types, and clustered the samples into tissue-specific clusters (**Figure 33a-b**). I implemented a clustering approach summarised in Fig. 20a. In short, pair-wise Pearson correlation coefficients were calculated for each sample, a cut-off of  $PCC > 0.6$  was used to define samples as 'connected', clusters were derived by prioritising the most connected samples and hence the largest clusters. This approach uses an extensively validated method for calculating correlation between samples and attempts to explain the data according to the principle of parsimony.

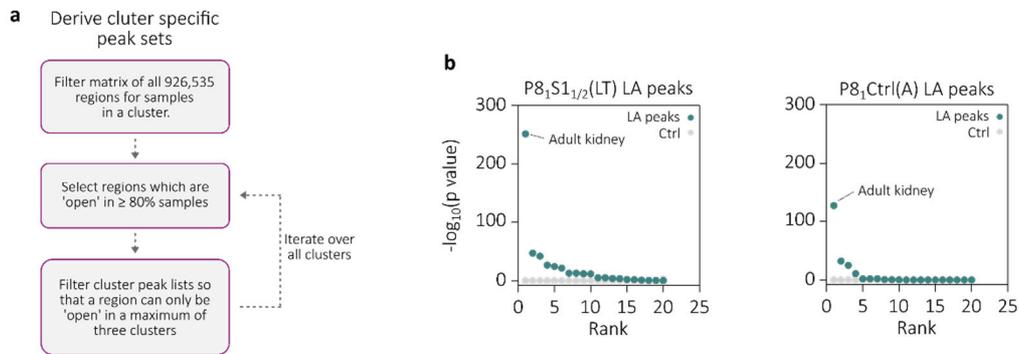
Using a minimum cluster size of five, 368/504 samples clustered into 21 groups. Of the 21 clusters, 20 were easily annotated by the tissue of origin, based on the sample composition (Figure 33b). The cluster which could not be annotated was removed from further analysis.



**Figure 33 | (a)** Schematic of Pearson correlation coefficient-based method used to cluster ENCODE DNase I hypersensitivity profiles. Samples include; cell lines (n=97), adult primary cells and tissues (n=125), and embryonic tissues (n=282). **(b)** tSNE plot based on the top 250,000 most variable regions of 376 DNase I hypersensitivity profiles from the ENCODE project, including cell lines, primary tissue, and embryonic tissue. Clusters were identified by clustering method from (a) and labelled based on sample composition.

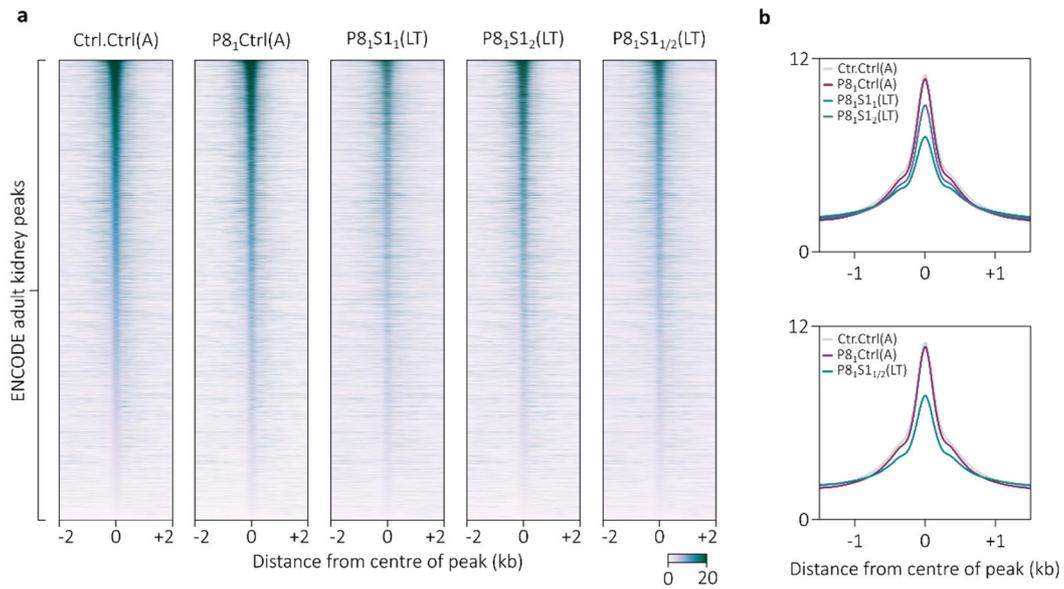
I derived cluster-specific peak sets using the workflow outlined in Figure 21a. The guiding principles used to define the arbitrary cut-offs were: (1) a cluster-specific peak should be commonly 'open' in a cluster (in  $\geq 80\%$  samples), (2) discrete adult tissues can arise from similar evolutionary trajectories, therefore the same peak can appear in multiple clusters ( $\leq 3$  clusters), and (3) a certain complexity (i.e. peak set size) should be maintained to ensure an accurate result from downstream overlap analyses (set sizes ranged from 10157-31419 peaks) (**Figure 34a**). Using the cluster-specific peak sets, I ran an overlap analysis with the differentially accessible ATAC-seq regions. Similar to the motif analysis, the

kidney-specific clusters were most enriched for the lower accessibility peaks sets for both P8<sub>1</sub>Ctrl and P8<sub>1</sub>S1<sub>1/2</sub> (Figure 34b).



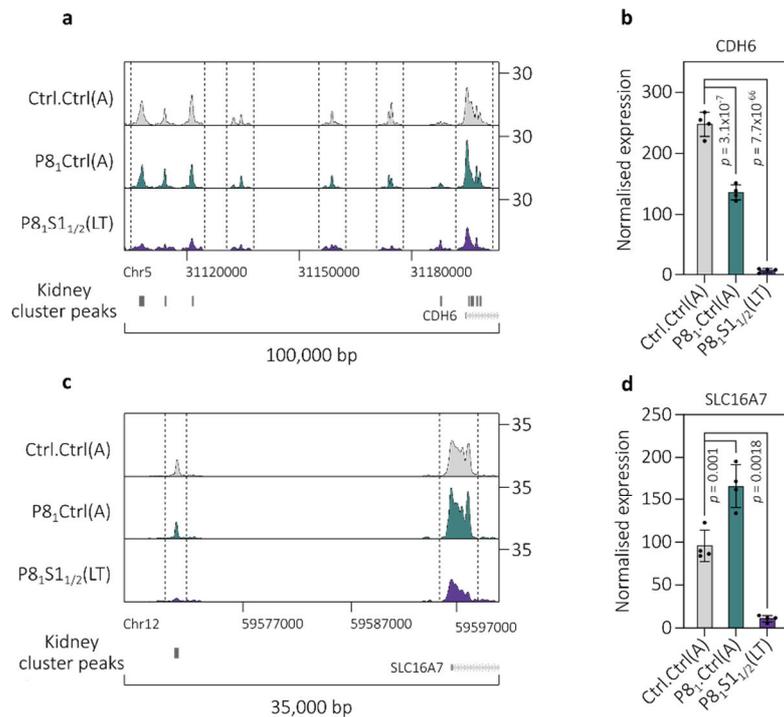
**Figure 34** | (a) Schematic of method used to derive cluster specific peak sets for each cluster. (b) Ranked plots of overlap analysis between cluster-specific peak sets generated from LA regions defined by Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT) and Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A). The same matched controls from Fig 17 were used, see materials and methods for more information.

However, the global loss of signal at these peak sets was substantially greater for P8<sub>1</sub>S1<sub>1/2</sub> compared to P8<sub>1</sub>Ctrl, suggesting that SMARCB1 loss triggers a widespread loss of renal epithelial epigenetic identity (Figure 35a-b). Of note, the renal specific signal was less for P8<sub>1</sub>S1<sub>2</sub> compared to P8<sub>1</sub>S1<sub>1</sub> (Figure 35a-b).



**Figure 35 | (a)** Heatmaps showing normalised ATAC-seq signal +/- 2kb centred on peak summits for the ENCODE adult kidney cluster region set. **(b)** Metagene plots of normalised ATAC-seq signal +/- 2kb centred on peak summits for the ENCODE adult kidney cluster region set.

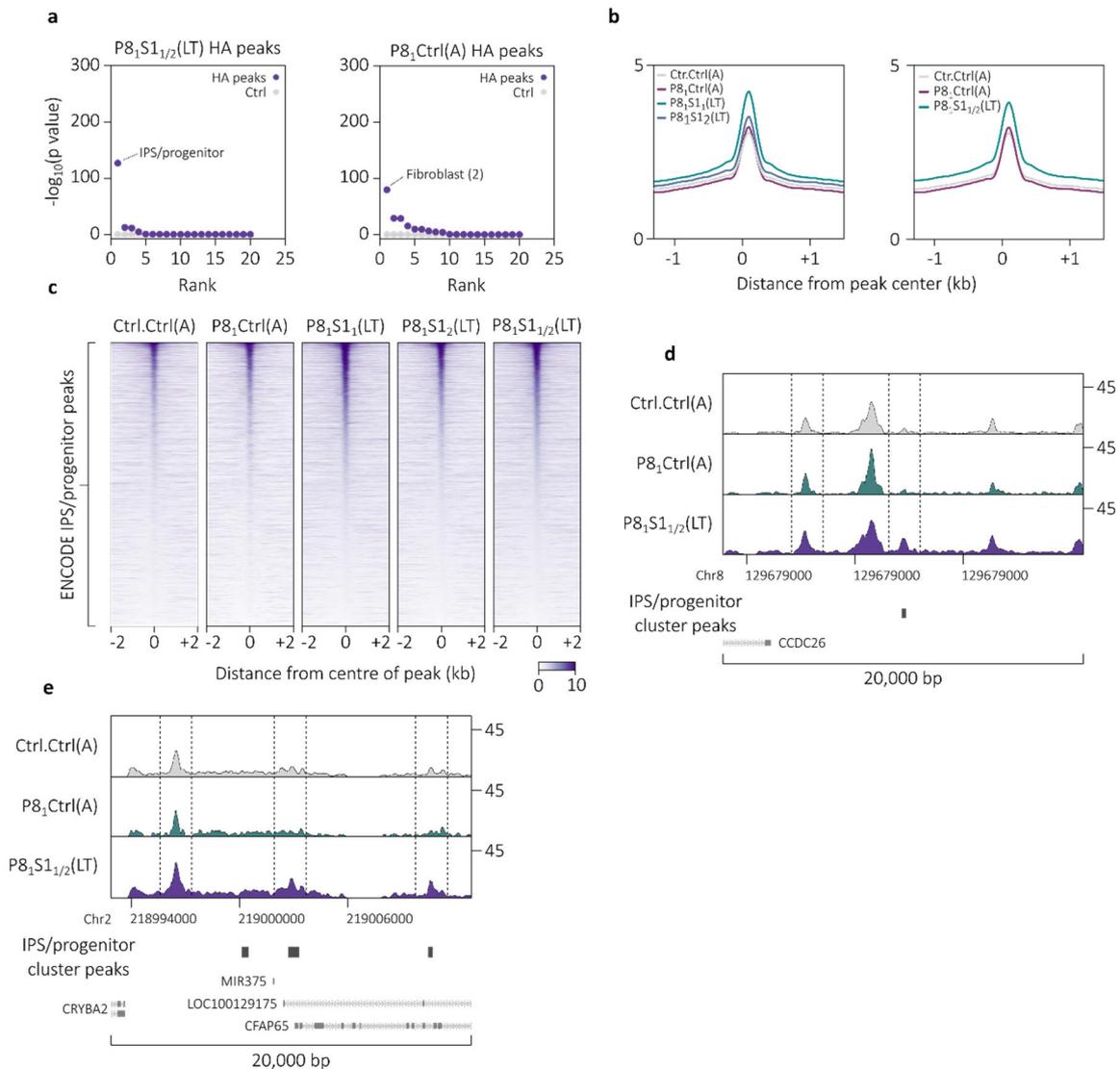
The loss of renal identity was supported by specific genomic loci harbouring known proximal tubule marker genes as defined by single-cell RNA-seq (scRNA-seq) experiments. For example, there was a loss of ATAC-seq signal at kidney cluster-specific peaks in the Cadherin-6 (*CDH6*) and Solute Carrier Family 16 Member 7 (*SLC16A7*) loci, which was concordant with a reduction in mRNA expression (**Figure 36a-d**)<sup>336</sup>.



**Figure 36** | (a) An example region of lost ATAC-seq signal in the CDH6 locus overlapping with kidney cluster specific regions, for P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells compared to Ctrl.Ctrl(A) cells. (b) Normalised CDH6 mRNA expression as determined by RNAseq for Ctrl.Ctrl(A), P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. P-values calculated by DEseq2. Error bars are SD. (c) An example region of lost ATAC-seq signal in the SLC16A7 locus overlapping with kidney cluster specific regions, for P8<sub>1</sub>S1<sub>1/2</sub>(LT) compared to Ctrl.Ctrl(A) cells. (d) Normalised SLC16A7 mRNA expression as determined by RNAseq for Ctrl.Ctrl(A), P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. P-values calculated by DEseq2. Error bars are SD.

The higher accessibility peaks for P8<sub>1</sub>S1<sub>1/2</sub> overlapped most strongly with an induced pluripotent stem cell (IPS)/progenitor cluster, which was not significantly enriched in the P8<sub>1</sub>Ctrl HA regions (**Figure 37a**). Unlike the loss of renal identity signal, the gain of the IPS/progenitor signal was moderate and did not represent the adoption of the whole program but rather specific aspects (**Figure 37c-e**). P8<sub>1</sub>S1<sub>2</sub> showed a greater gain in IPS/progenitor signal than P8<sub>1</sub>S1<sub>1</sub> cells. The strongest overlapping peak set for P8<sub>1</sub>Ctrl HA regions belonged to a fibroblast cluster (**Figure 37a**). Interestingly the ENCODE adult kidney cluster correlates highly with the fibroblast clusters, which suggests that PAX8 may be a critical differentiator

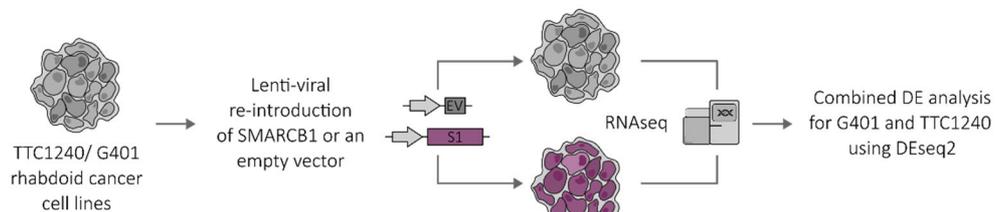
between kidney and fibroblast lineages. In line with this, PAX8 in combination with HNF1B, HNF4a, and EMX2 can reprogram fibroblasts into renal tubular epithelium<sup>264</sup>.



**Figure 37** | (a) Ranked plots of overlap analysis between cluster-specific peak sets generated from HA regions defined by Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT) and Ctrl.Ctrl(A) vs P8<sub>1</sub>Ctrl(A). The same matched controls from Fig 17 were used, see materials and methods for more information. (b) Metagene plots of normalised ATAC-seq signal +/- 2kb centred on peak summits for the ENCODE IPS/progenitor cluster region set. (c) Heatmaps showing normalised ATAC-seq signal +/- 2kb centred on peak summits for the ENCODE IPS/progenitor cluster region set. (d-e) Example regions of gained ATAC-seq signal overlapping with IPS/progenitor cluster peaks, for P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells compared to Ctrl.Ctrl(A) cells.

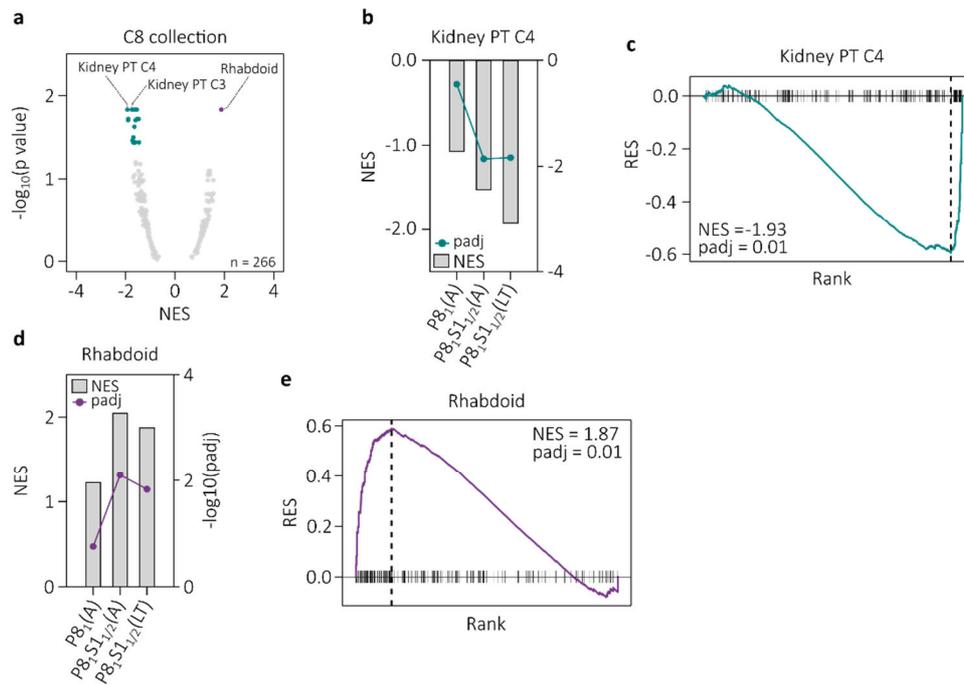
The global loss of the renal epithelial signal in conjunction with the gain of discrete modules from the IPS/progenitor cluster at the chromatin accessibility level supports the notion that SMARCB1 may maintain a lineage-differentiated cellular state.

To test this at the gene expression level, I used the molecular signatures database (mSigDB) cell-type specific signature collection (C8), supplemented with a signature that we derived from SMARCB1 re-introduction experiments in rhabdoid tumour cell lines (**Figure 38**)<sup>271</sup>.



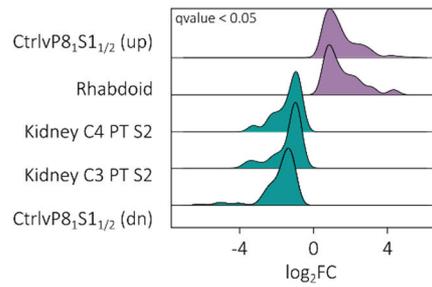
**Figure 38** | Schematic overview of the generation of a SMARCB1 mutant rhabdoid cancer signature.

The two most significantly downregulated signatures in the *SMARCB1* KO lines were from renal proximal tubules, the proposed origin of ccRCC (**Figure 39a**)<sup>90</sup>. The loss of renal transcriptional identity followed a similar pattern to the epigenetic changes: PAX8 KD alone showed a negative enrichment for the proximal epithelial signature C4 but failed to reach significance ( $p < 0.05$ ) and  $P8_1S1_{1/2}(A)$  showed significant downregulation of the signature which reduced further over time ( $P8_1S1_{1/2}(LT)$ ) (**Figure 39b-c**). Similarly, PAX8 KD alone induced a positive enrichment of the rhabdoid signature, but significance ( $p < 0.05$ ) was only reached when *SMARCB1* was also knocked out (**Figure 39d-e**).



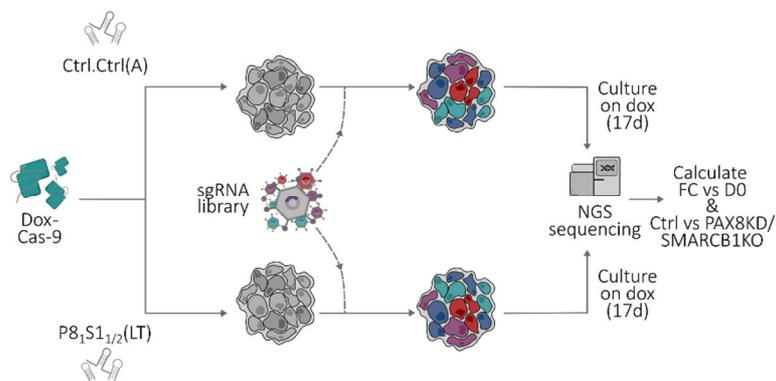
**Figure 39** | (a) Volcano plot of GSEA with cell-type specific transcriptional signatures from mSigDB collection 8, for the comparison Ctrl.Ctrl(A) vs  $P8_1S1_{1/2}(LT)$ . highlighted points (purple/cyan) have a  $p.adjust < 0.05$ . (b) Kidney proximal tubule C4 signature normalised enrichment scores (NES) from GSEA, for Ctrl.Ctrl(A) vs  $P8_1Ctrl(A)$ ,  $P8_1S1_{1/2}(A)$  and  $P8_1S1_{1/2}(LT)$ . (c) GSEA plot of kidney proximal tubule C4 signature for Ctrl.Ctrl(A) vs  $P8_1S1_{1/2}(LT)$ . (d) Rhabdoid signature normalised enrichment scores (NES) from GSEA, for Ctrl.Ctrl(A) vs  $P8_1Ctrl(A)$ ,  $P8_1S1_{1/2}(A)$  and  $P8_1S1_{1/2}(LT)$ . (e) GSEA plot of rhabdoid signature for Ctrl.Ctrl(A) vs  $P8_1S1_{1/2}(LT)$ .

Up and down-regulated signatures derived from my RNA-seq data, kidney proximal tubule signatures, and the rhabdoid signature were also similarly positively and negatively enriched in the *SMARCB1* mutant ccRCC cell lines in the CCLE data set, suggesting that a similar mechanism accounts for the PAX8 inhibition insensitivity in these models (Figure 40).



**Figure 40| (a)** Ridge plot of GSEA result from the comparison of ccRCC CCLE lines, SMARCB1 wild type vs mutant from Fig.10, using the rhabdoid, proximal tubule C3/C4 and up and down-regulated genes from Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT).

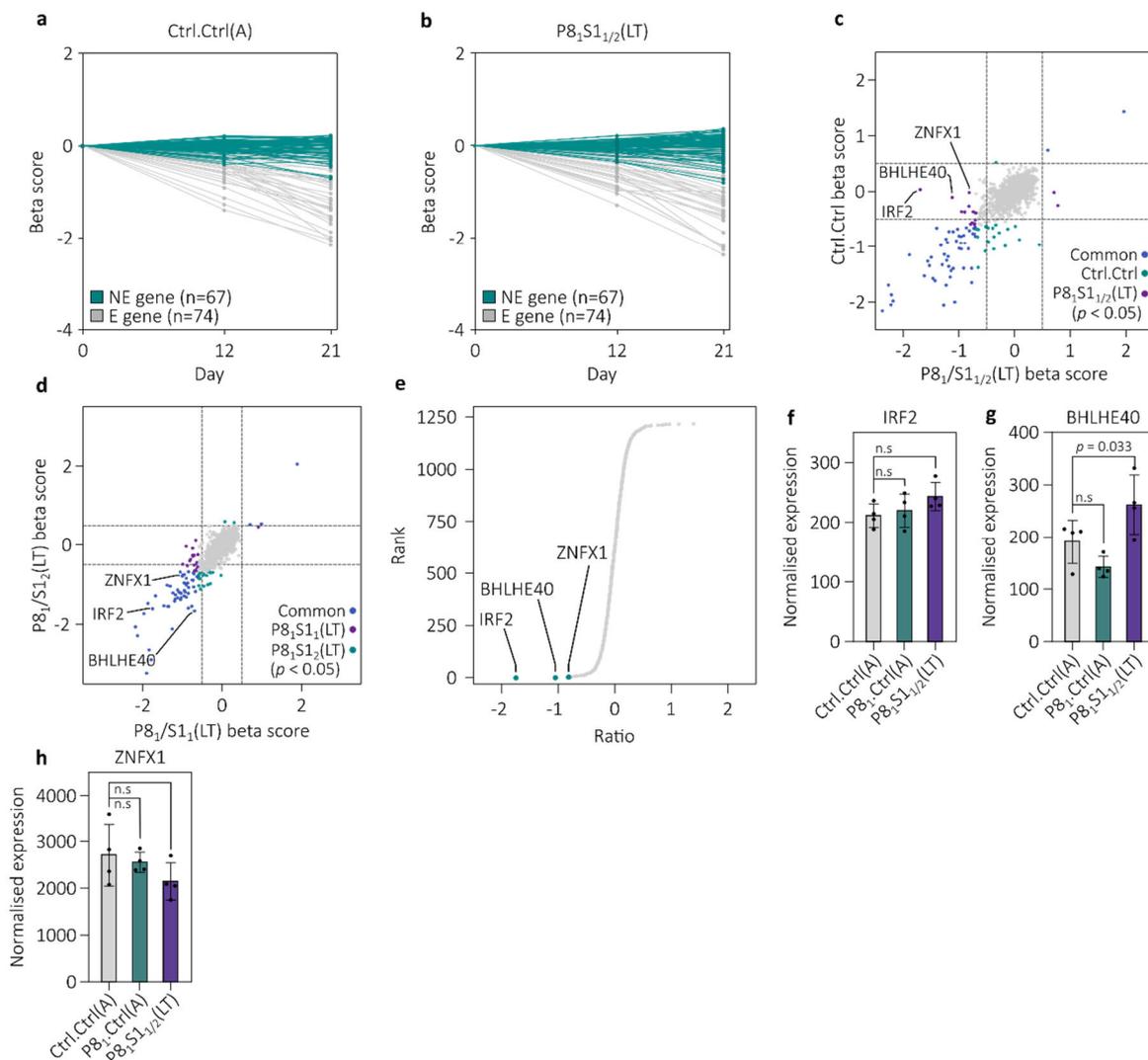
PAX8 inhibition resistance was associated with changes in the cellular lineage state towards rhabdoid de-differentiation, suggesting the possibility that the role of PAX8 in supporting ccRCC growth had been replaced by some other transcriptional lineage factors. We therefore performed a second CRISPR screen targeting known and predicted TFs using the P8<sub>1</sub>S1<sub>1/2</sub>(LT) cell lines (**Figure 41**).



**Figure 41| (a)** A schematic of a pooled CRISPR-Cas9-based loss of function screen using an sgRNA library targeting transcription factors to identify new dependences in P8<sub>1</sub>S1<sub>1/2</sub>(LT) vs Ctrl.Ctrl(A) cells.

The screen passed our quality control measures, essential genes were depleted, and non-essential genes were neither enriched nor depleted (**Figure 42a-b**). We identified three new dependencies which

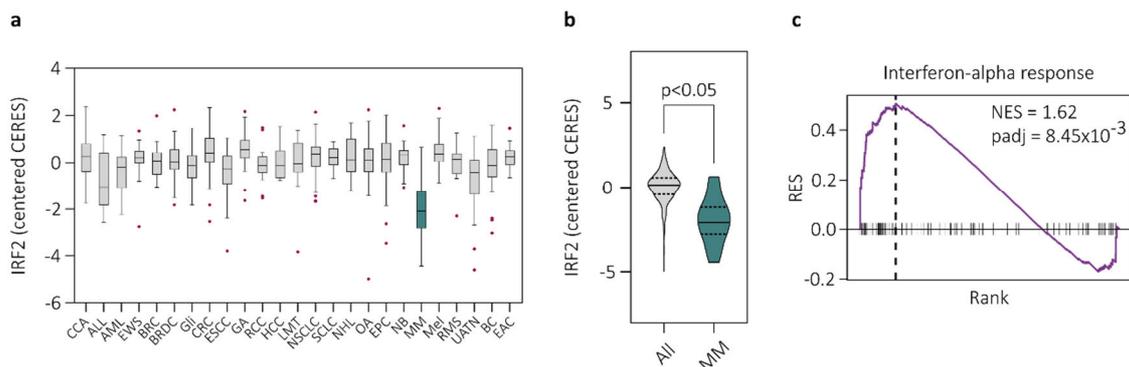
had no phenotype in the control cells: interferon regulatory factor 2 (IRF2), Basic Helix-Loop-Helix E40 (BHLHE40), and zinc finger NFX1-type containing 1 (ZNF1) (**Figure 42c-e**). All of which were expressed in cells prior to SMARCB1 loss (**Figure 42f-h**). IRF2 is a member of a TF family which regulates Toll-like receptor (TLR) signalling, hematopoietic differentiation, and the expression of IFNs and their target genes<sup>337,338</sup>.



**Figure 42** | (a-b) Beta scores of essential (n=67) and non-essential genes (n=74) at time points throughout the screen for the control arm (a) and experimental arm (b). (c-d) Changes in sgRNA abundance over time, measured by calculating beta scores using the top 3 depleted sgRNAs per gene relative to plasmid, from two technical replicates. Highlighted points have a beta score <-0.5 or >0.5

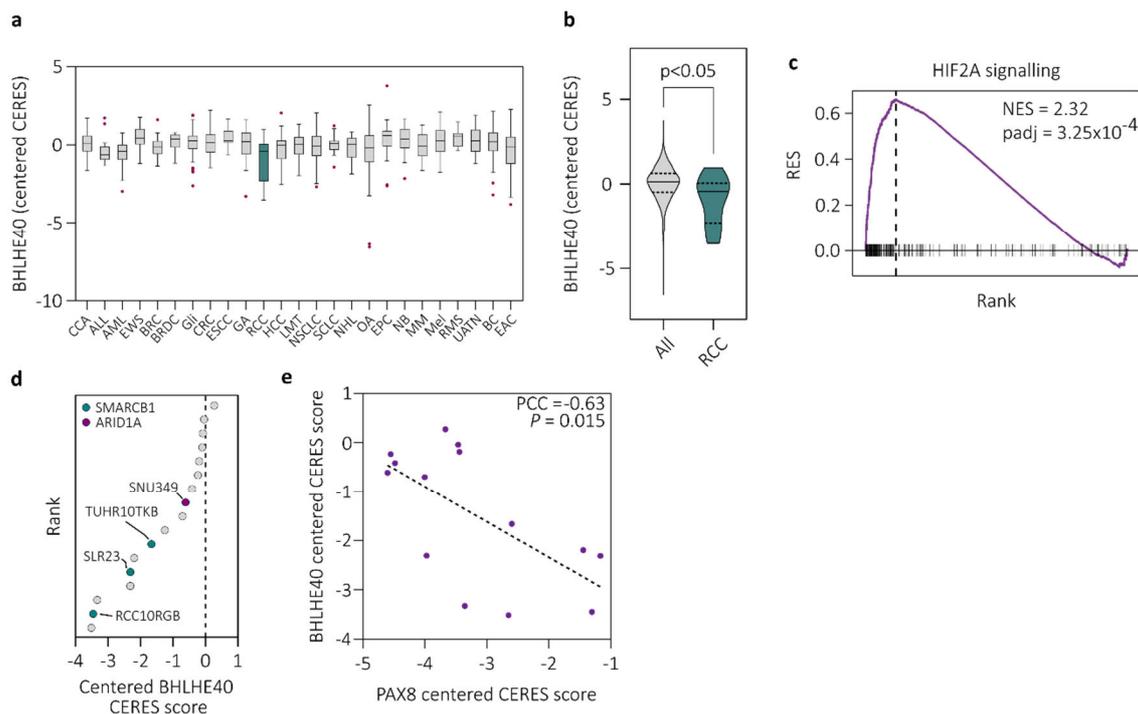
and an adjusted  $p$ -value  $< 0.05$ . (c) The beta scores for the control arm (Ctrl.Ctrl(A)) of the screen versus the pooled experimental arm (P8<sub>1</sub>S1<sub>1/2</sub>(LT)). (d) The beta scores of each experimental arm, P8<sub>1</sub>S1<sub>1</sub>(LT) vs P8<sub>1</sub>S1<sub>2</sub>(LT). (e) Ranked plot of the ratio between the beta scores for the control arm versus the pooled experimental arm (ratio = P8<sub>1</sub>S1<sub>2</sub>(LT)<sub>beta</sub> - Ctrl.Ctrl(A)<sub>beta</sub>). (f-h) Fold change in mRNA expression versus Ctrl.Ctrl cells for the genes (f) IRF2, (g) BHLHE40 and (h) ZNFX1.  $P$ -values were calculated using DEseq2. Error bars are SD.

Similar to the role of PAX8 in renal development and ccRCC, IRF2 plays an important role in cancers originating from the plasma cell lineage and thus the acquired dependence on IRF2 in this context represents the co-option of a regulatory module from another lineage (**Figure 43a-b**). In line with IRF2's role in regulating IFNs, compared to P8<sub>1</sub>Ctrl cells there is a strong increase in the interferon-alpha gene set from the hallmarks collection in P8<sub>1</sub>S1<sub>1/2</sub> cells (**Figure 43c**).



**Figure 43** | (a) Genetic dependency data from the DepMap project for the gene IRF2. Centered CERES dependency scores across 25 lineages, with  $\geq 10$  cell lines per lineage. (b) Genetic dependency data from the DepMap project. IRF2 centred CERES dependency score of multiple myeloma (MM) cell lines ( $n=21$ ) versus cells from all other lineages ( $n=767$ ). Kruskal-Wallis test. (c) GSEA plot of interferon-alpha response signature from mSigDB hallmarks collection, for P8<sub>1</sub>Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). **Abbreviations:-** CCA: cholangiocarcinoma, ALL: acute lymphoblastic leukaemia, AML: acute myeloid leukaemia, EWS: Ewing sarcoma, BRC: breast carcinoma, BRDC: breast ductal carcinoma, Gli: Glioma, CRC: colorectal adenocarcinoma, ESCC: esophageal squamous cell carcinoma, GA: gastric adenocarcinoma, RCC: renal cell carcinoma, HCC: hepatocellular carcinoma, LMT: lung mesothelioma, NSCLC: non-small cell lung cancer, SCLC: small cell lung cancer, NHL: non-Hodgkin lymphoma, OA: ovarian adenocarcinoma, EPC: exocrine pancreatic cancer, NB: neuroblastoma, MM: multiple myeloma, Mel: melanoma, RMS: rhabdomyosarcoma, UATN: upper aerodigestive tract neoplasm, BC: bladder carcinoma, EAC: endometrial adenocarcinoma.

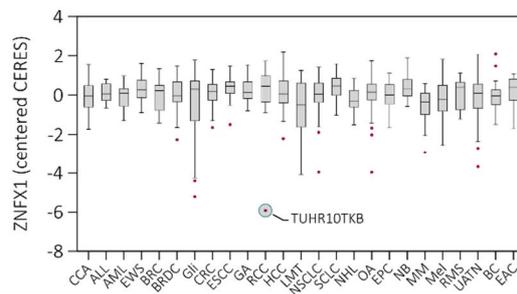
BHLHE40 is a ubiquitously expressed stress-responsive transcription factor that is important in several physiological responses including differentiation, tumorigenesis, and response to hypoxia<sup>339</sup>. The mutation of *VHL* and the stabilisation of HIF2 $\alpha$  protein is a key tumorigenic event in ccRCC, and HIF2 $\alpha$  perturbation RNA-seq has placed BHLHE40 downstream of HIF2A signalling<sup>340</sup>. In line with this, BHLHE40 dependency shows tissue specificity for RCC and there is strong up regulation of HIF2 $\alpha$  signalling in P8<sub>1</sub>S1<sub>1/2</sub> cells compared to P8<sub>1</sub>Ctrl cells (**Figure 44a-c**). In the DepMap cohort approximately half of the *VHL* mutant ccRCC lines are sensitive to BHLHE40 KO, and interestingly, this includes all the *SMARCB1* mutant lines (**Figure 44d**). Further, when considering PAX8 resistant lines with a *SMARCB1* mutation and PAX8 sensitive lines, there is a negative correlation between PAX8 and BHLHE40 dependency (**Figure 44e**).



**Figure 44** | (a) Genetic dependency data from the DepMap project for the gene BHLHE40. Centered CERES dependency scores across 25 lineages, with  $\geq 10$  cell lines per lineage. (b) Genetic dependency data from the DepMap project. BHLHE40 centered CERES dependency score of RCC cell lines ( $n=23$ ) versus cells from all other lineages ( $n=765$ ). Kruskal-Wallis test. (c) GSEA plot of custom HIF2A target genes, for P8<sub>1</sub>Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). See materials and methods for more information. (d) *VHL*-mutant ccRCC DepMap cell lines ranked by BHLHE40 centred CERES dependency score. *SMARCB1*

and ARID1A mutant cell lines from Fig.10. **(e)** Correlation of BHLHE40 and PAX8 centred CERES scores for PAX8 sensitive and SMARCB1 mutant VHL-mutant ccRCC cell lines. **Abbreviations:-** CCA: cholangiocarcinoma, ALL: acute lymphoblastic leukaemia, AML: acute myeloid leukaemia, EWS: Ewing sarcoma, BRC: breast carcinoma, BRDC: breast ductal carcinoma, Gli: Glioma, CRC: colorectal adenocarcinoma, ESCC: esophageal squamous cell carcinoma, GA: gastric adenocarcinoma, RCC: renal cell carcinoma, HCC: hepatocellular carcinoma, LMT: lung mesothelioma, NSCLC: non-small cell lung cancer, SCLC: small cell lung cancer, NHL: non-Hodgkin lymphoma, OA: ovarian adenocarcinoma, EPC: exocrine pancreatic cancer, NB: neuroblastoma, MM: multiple myeloma, Mel: melanoma, RMS: rhabdomyosarcoma, UATN: upper aerodigestive tract neoplasm, BC: bladder carcinoma, EAC: endometrial adenocarcinoma.

ZNFX1 is a ubiquitously expressed, IFN stimulated SF1 helicase capable of detecting viral dsRNA<sup>341</sup>. Unlike IRF2 or BHLHE40, dependence on ZNFX1 is not associated with a particular lineage. Instead, there is a small number of cell lines across multiple lineages which show a strong dependence on ZNFX1, including TUHR10TKB, one of the three *SMARCB1* mutant CCLE ccRCC lines (**Figure 45**).



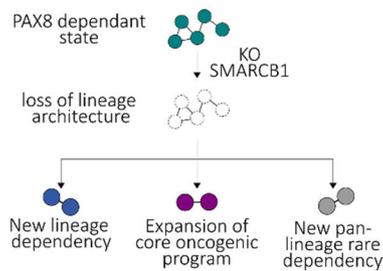
**Figure 45| (a)** Genetic dependency data from the DepMap project. ZNFX1 centered CERES dependency scores across 25 lineages, with  $\geq 10$  cell lines per lineage. **Abbreviations:-** CCA: cholangiocarcinoma, ALL: acute lymphoblastic leukaemia, AML: acute myeloid leukaemia, EWS: Ewing sarcoma, BRC: breast carcinoma, BRDC: breast ductal carcinoma, Gli: Glioma, CRC: colorectal adenocarcinoma, ESCC: esophageal squamous cell carcinoma, GA: gastric adenocarcinoma, RCC: renal cell carcinoma, HCC: hepatocellular carcinoma, LMT: lung mesothelioma, NSCLC: non-small cell lung cancer, SCLC: small cell lung cancer, NHL: non-Hodgkin lymphoma, OA: ovarian adenocarcinoma, EPC: exocrine pancreatic cancer, NB: neuroblastoma, MM: multiple myeloma, Mel: melanoma, RMS: rhabdomyosarcoma, UATN: upper aerodigestive tract neoplasm, BC: bladder carcinoma, EAC: endometrial adenocarcinoma.

### 5.3 Summary

ENCODE DNase I hypersensitivity profiles for 368 samples were grouped into 21 clusters, 20 of which could be annotated based on their sample composition. A collection of cluster-specific peak sets were derived using stringent specificity measures which also maintained a high degree of peak set complexity. Overlap analysis between the ENCODE collection and LA regions from Ctrl.Ctrl(A) versus P8<sub>1</sub>Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) showed an enrichment for the adult kidney cluster peak set, in line with PAX8's role as a key renal lineage factor. Interestingly, SMARCB1 KO in addition to PAX8 KD resulted in the widespread loss of kidney epigenetic identity, whereas PAX8 KD alone triggered a more moderate reduction. HA regions from Ctrl.Ctrl(A) versus P8<sub>1</sub>S1<sub>1/2</sub>(LT) overlapped most strongly with the IPS/progenitor peak set. This overlap was characterised by the gain in discrete modules as opposed to activation of the complete IPS/progenitor program. Instead of the IPS/progenitor cluster, the HA regions from Ctrl.Ctrl(A) versus P8<sub>1</sub>Ctrl(A) were enriched for a fibroblast peak set, specifically associating SMARCB1 loss with the undifferentiated phenotype. GSEA analysis on the accompanying transcriptional data supported the loss of renal identity and the gain of an undifferentiated state when SMARCB1 is lost in conjunction with PAX8. PAX8 KD alone resulted in a similar trend but failed to reach significance for both the downregulation of renal identity and the upregulation of a rhabdoid undifferentiated state. An alternative strategy starting with RNAseq data could have been to implement a network-based approach to identify key regulators of cell-type specific transcriptional programs that show enhanced lineage specificity, but this would likely have missed the more subtle gain in IPS/progenitor signal as we did not observe a complete program switch<sup>206,342,343</sup>.

To understand how the loss of differentiation promoted proliferation after the KD of PAX8, I performed a second loss-of-function genetic screen using a library of sgRNAs targeting TFs. IRF2, ZNFX1, and BHLHE40 were specifically depleted in SMARCB1 KO lines compared to PAX8 sensitive control cells. The newly acquired dependencies fit into three categories: acquisition of a lineage dependency from

another lineage (IRF2 from plasma cells), expansion of the already in place core oncogenic program (VHL-HIF2 $\alpha$ ), and acquisition of a rare pan-cancer dependency (ZNF1) (**Figure 46**).



**Figure 46** | Schematic overview of three SMARCB1 mediated mechanisms to overcome lineage factor inhibition.

In summary, SMARCB1 is a key regulator of the renal cis-regulatory and transcriptional programs, which defines the context in which PAX8 is required for tumour growth. SMARCB1 KO promotes resistance to PAX8 suppression through a mechanism that resembles dedifferentiation that facilitates the co-option of previously dispensable transcriptional regulators.

# 6

## The Discussion

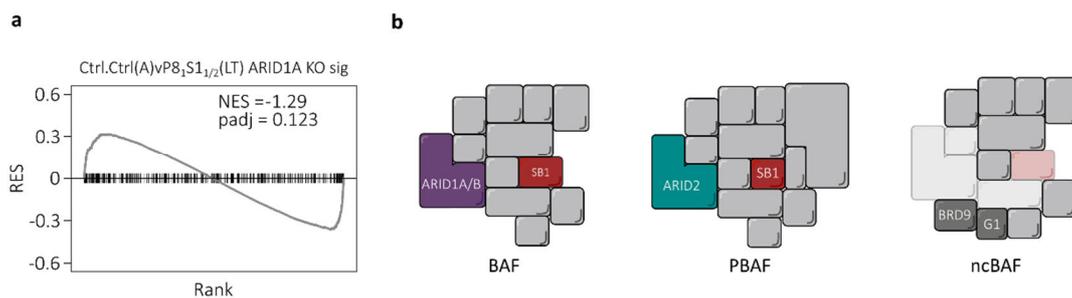
## 6.1 Lineage plasticity as a resistance mechanism

Transcriptional lineage factor dependencies are observed across a range of malignancies and as such are an attractive class of targeted therapies. However, the mechanisms which maintain lineage fidelity during cancer development and the response to long term lineage factor inhibition are poorly understood. Here I demonstrate that SMARCB1, an essential gene and SWI/SNF complex member, conveys resistance to PAX8 inhibition in ccRCC, through a process of dedifferentiation. In this context, SMARCB1 maintains the kidney-specific cis-regulatory program, its inactivation results in the loss of kidney transcriptional and epigenetic identity, altering the cellular context, and removing the requirement for PAX8 mediated signalling. These findings are in line with trans-gene re-introduction experiments in paediatric rhabdoid tumours. In MRT and AT/RT cell lines the exogenous expression of SMARCB1 results in the preferential gain of cis-regulatory elements within the proximity of developmental genes, and the resolution of bivalent promoters, which are an important developmental switch<sup>271,331</sup>. A lineage-tracing study placed the origin of MRTs in the neural crest to mesenchyme developmental trajectory and showed that the re-introduction of SMARCB1 in patient-derived organoids triggered the induction of differentiation towards the mesenchyme<sup>344</sup>. Further, the KD of SMARCB1 in embryonic stem cells (ESCs) impairs neural development in a directed differentiation assay, by reducing accessibility at neural stem cell-specific regions<sup>334</sup>. Taken together, SMARCB1 plays an important role in the maintenance of enhancer programs during development.

Three additional SWI/SNF complex members were enriched in the chromatin screen, BCL7C, BCL7B, and ARID1A. Interestingly, in two recent studies, ARID1A inactivation was linked to resistance to tamoxifen and fulvestrant therapy in ER positive luminal breast cancer<sup>272,301</sup>. In a cohort of ~1,900 breast cancers, mutations in ARID1A were enriched in the hormone and metastatic setting and correlated with worse disease outcome<sup>272,345</sup>. This was causally validated by two independent CRISPR screens which demonstrated enrichment of sgRNAs targeting ARID1A in tamoxifen and fulvestrant

treated MCF7 luminal breast cancer cells<sup>272,301</sup>. ARID1A KO reduced chromatin accessibility at the binding sites of master regulators of the luminal cell fate, enabling a switch from a luminal to basal transcriptional state<sup>301</sup>. Importantly, a similar luminal to basal identity switch was also observed histologically in breast cancer patients who have become refractory to hormone therapy<sup>346</sup>. Mechanistically, ARID1A inactivation may promote resistance through the impaired recruitment of HDAC1 and subsequent deacetylation at particular genomic loci, enabling the binding of bromodomain containing proteins (e.g. BRD4) and the subsequent activation of pro-proliferative programs<sup>272</sup>. In line with this, BET inhibitors show selective efficacy in an ARID1A mutant breast cancer model<sup>272</sup>.

In addition to ARID1A, SMARCB1 was also enriched in both tamoxifen/fulvestrant resistance CRISPR screens, implying that similar mechanisms may govern the dependence of breast and renal epithelial cancers on lineage factor signalling<sup>272,301</sup>. Although SMARCB1 and ARID1A are both SWI/SNF complex members, there is evidence that their loss triggers resistance to lineage therapy by overlapping but distinct mechanisms. Both the inactivation of SMARCB1 in renal cancer and ARID1A in breast cancer led to the loss of tissue-specific epigenetic identity<sup>301</sup>. However, SMARCB1 facilitated the conversion into an undifferentiated rhabdoid-like state, whereas ARID1A seemed to enable the transdifferentiation of luminal cells into a state which transcriptionally resembled basal-like breast cancer derived from myoepithelial cells of the outer layer of the breast duct<sup>301,347</sup>. GSEA for the comparison Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT) using an ARID1A KO signature, showed no statistically significant enrichment, further suggesting that SMARCB1 and ARID1A likely mediate resistance through alternative routes (**Figure 47a**). This may in part be explained by the subunit composition of the SWI/SNF complexes. SMARCB1 is a core component of both BAF and PBAF complexes whereas ARID1A along with ARID1B is specifically found in the BAF complex (**Figure 47b**)<sup>348</sup>.



**Figure 47 | (a)** Schematic of subunit composition of mammalian BAF, PBAF and ncBAF complexes. Purple, cyan and dark grey highlighted subunits have complex specificity, whereas light grey subunits are shared by two or more complexes. Red represents SMARCB1. Schematic adapted from Michel et al.,<sup>314</sup>. **(b)** GSEA plot of ARID1A KO up regulated gene signature (see materials and methods) for Ctrl.Ctrl(A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT). **Abbreviations:-** SB1:SMARCB1, G1:GLTSCR1/1L.

The loss of SMARCB1 inhibits the targeting of both BAF and PBAF complexes to the DNA without destabilising complex stability<sup>271</sup>. This is thought to enable the activity of the ncBAF complex that has been linked to the pathogenesis of RTs and the maintenance of a naïve pluripotent stem cell state<sup>271,314,331,333</sup>. The KO of SMARCB1 in stem cells or the reintroduction of SMARCB1 in RT cell models has been strongly associated with a respective gain or loss of accessibility at regions enriched for the CTCF motif<sup>271,314,331,333,334</sup>. In SMARCB1 null RTs, ncBAF localises to genomic regions enriched for the CTCF motif, and the pharmacological depletion of BRD9 (a ncBAF specifying subunit) inhibits cell growth *in vitro*<sup>314</sup>. In MCF7 breast cancer cells, the inactivation of ARID1A resulted in substantial changes in chromatin accessibility, but there was no enrichment for the CTCF motif<sup>272,301</sup>. This could suggest that SMARCB1 KO may promote resistance through the activity of the ncBAF complex, whereas ARID1A KO likely promotes resistance through the genomic redistribution of still functional canonical PBAF complexes.

In this study, the resistance to PAX8 inhibition in ccRCC cells was concomitant with marked histological changes which resembled NE differentiation, a morphology not commonly seen in ccRCC. Instead, molecular features of NE differentiation are detected in a subset of different cancer types and in

association with advanced disease, including hormone therapy resistant prostate cancer and EGFR inhibition resistant lung cancer<sup>59,76,273</sup>. Recent studies in both lung and prostate cancer, have strongly associated the development of NE dedifferentiation with the concomitant inactivation of RB1 and TP53<sup>59</sup>.

In a *Pten*<sup>-/-</sup> driven mouse model for prostate adenocarcinoma, the addition of an *RB1* KO facilitated the emergence of hormone resistant tumours which exhibited neuroendocrine histology, an up regulation of the stemness related factor SRY-related HMG-box (*Sox2*) and the epigenetic regulator *Ezh2*, and harboured acquired *Trp53* mutations<sup>78</sup>. The upregulation of both *Sox2* and *Ezh2* expression was validated using transcriptomic data from prostate cancer patients, and the treatment with an EZH2 inhibitor was able to resensitise *Trp53/Rb1* KO cells to antiandrogen therapy<sup>78</sup>. In a parallel study published at the same time, the dual knockout of *TP53* and *RB1* in human prostate cancer cells led to the development of antiandrogen therapy resistance, an upregulation of neuroendocrine and basal cell markers, and stemness related factors (e.g. *SOX2*), as well as a downregulation of luminal markers<sup>77</sup>. The luminal to basal switch bears an interesting similarity to acquired tamoxifen resistance in breast cancer, where the underlying molecular drivers are different<sup>301,346</sup>. The reintroduction of functional TP53 and RB1 reversed the widespread molecular changes and restored enzalutamide (antiandrogen therapy) sensitivity, confirming that resistance was mediated by *TP53/RB1* mutations instead of simply the outgrowth of a resistant clone under drug selection<sup>77</sup>. The knock down of SOX2 was also able to restore sensitivity to enzalutamide and reduced expression of basal and neuroendocrine markers, suggesting it is a key driver of castration resistant prostate cancer<sup>77</sup>. Taken together, RB1 and TP53 are key drivers of NE-based androgen therapy resistance in prostate cancer, their inactivation facilitates resistance through the upregulation of *SOX2* possibly mediated through the action of the chromatin modifier EZH2.

Unlike resistance to anti-androgen therapy, the development of NE morphology and resistance to EGFR inhibition in lung cancer requires additional stimuli to *TP53* and *RB1* loss. RB1 KD in NSCLC cells already

harbouring inactivating *TP53* mutations was not sufficient to generate resistance to EGFR therapy (e.g. erlotinib) *in vitro* or *in vivo*<sup>349</sup>. Further, in a leave-one-out analysis, *MYC* and *BCL-2* overexpression, *AKT* overactivation and, *TP53* and *RB1* mutation were required to produce neuroendocrine tumours from normal lung epithelium, whereas *BCL-2* was dispensable for the transformation of prostate epithelium<sup>76</sup>.

Similar to hormone signalling in breast cancer and *PAX8* signalling in RCC, members of the SWI/SNF complex are important mediators of AR signalling in prostate cancer<sup>350</sup>. Both *SMARCE1* and Actin-like (*ACTL*)6A are directly important for the AR-mediated activation of target genes such as kallikrein related peptidase(*KLK*)3, Transmembrane protease serine 2 (*TMPRSS2*), *FKBP5*, and *KLK2*<sup>351–353</sup>. The inhibition of either *SMARCE1* or *ACTL6A* led to the transcriptional downregulation of AR target genes and suppressed cell growth in androgen dependant prostate cancer cells<sup>351–353</sup>. Interestingly, alterations in the SWI/SNF complex subunit composition have also been linked to neuroendocrine differentiation and therapy resistance in prostate cancer. A cohort of ~600 prostate cancer patients with whole-exome sequencing (WES) and mRNA expression data, including 56 CRPC-NE cases, were used to explore the contribution of the SWI/SNF complex to NE differentiation and hormone therapy resistance<sup>354</sup>. Unlike in breast cancer, no recurrent mutations in SWI/SNF complex members were observed. Instead, there were detectable changes in the mRNA expression of discrete complex members specifically in CRPC-NE patients, such as *ACTL6B*, Double PHD Fingers 1 (*DPF1*), *SS18L1*, and *SMARCA4*<sup>354</sup>. *ACTL6B*, *DPF1*, and *SS18L1* are found in a neuron-specific SWI/SNF complex, in line with the neuroendocrine phenotype<sup>355,356</sup>. Further, immunoprecipitation of the SWI/SNF subunit *SMARCC1* followed by mass spectrometry, showed a specific interaction with NK2 homeobox 1 (*NKX2.1*) in CRPC-NE cells compared to adenocarcinoma cells<sup>354</sup>. *NKX2.1* is an important developmental regulator in parts of the brain and is expressed in CRPC-NE patient samples<sup>357,358</sup>. However, the pathogenic contribution of neuronal developmental markers and neural SWI/SNF complex members is unclear. In contrast, there is stronger evidence linking *SMARCA4* activity to disease progression. *SMARCA4* expression is upregulated in both CRPC and NE-CRPC compared to PRAD. In line with this, high *SMARCA4* protein expression correlated

with shorter overall survival, and KD of SMARCA4 inhibited proliferation of LNCaP cells and the androgen resistant C4-2 cells<sup>354</sup>. The exact contribution of upregulated SMARCA4 expression to this process is unclear. High expression of SMARCA4 did not always correspond with the expression of neuronal markers but correlated more closely with SOX2 expression, leading the authors to posit that SMARCA4 may be being incorporated into an embryonic stem cell-specific SWI/SNF complex (esBAF), which includes SMARCA4, ACTL6A, and SMARCC1 subunits but not their paralogs<sup>354,358</sup>. Taken together, these studies provide evidence for the non-genetic reprogramming of the SWI/SNF repertoire during disease progression, changes in the expression of individual subunits can result in significant cellular outcomes, including the development of NE hormone-resistant prostate cancer.

In summary, resistance mechanisms to lineage-targeted therapy seem to converge on SWI/SNF complex members through genetic and non-genetic dysregulation. Changes in the subunit composition and the genomic redistribution of SWI/SNF complexes change the cellular context, removing the need for the original lineage factors and enabling the selection of new transcriptional programs to drive proliferation.

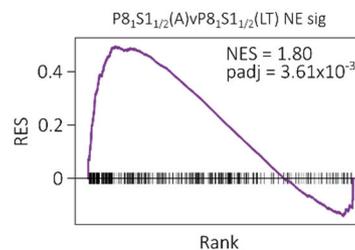
## 6.2 Acquired transcriptional dependencies replace PAX8 signalling

*SMARCB1* KO initially triggers a state of instability that can result in cell death but also acts as a substrate for the selection of new transcriptional modules to support proliferation. The new dependencies experimentally uncovered fall into one of three categories: (1) the co-option of another lineage specific program (IRF2), (2) the expansion of the core oncogenic signalling pathway (BHLHE40) and (3) the acquisition of a rare, lineage agnostic module (ZNF1). Likely the newly acquired modules are not purely stochastic but are rather opportunistic. BHLHE40, ZNF1 and IRF2 are all highly expressed in control cells, therefore, given the correct evolutionary drive and a permissive chromatin context, it is likely they have an increased chance for selection.

The acquisition of new transcriptional dependencies in P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells is analogous to what has been demonstrated in AR insensitive prostate cancer models and more generally in neuroendocrine differentiation. In a prostate cancer model, a CRISPR-cas9 screen identified the gene, conserved chromatin remodeling and assembly factor (*CHD1*), as capable of conferring resistance to enzalutamide treatment<sup>359</sup>. *CHD1* is a chromatin regulator important for maintaining open chromatin in pluripotent stem cells, is one of the most frequently mutated genes in prostate cancer, and the expression of *CHD1* negatively correlates with the outcome of anti-androgen treatment<sup>360–362</sup>. An *in vivo* genetic screen in enzalutamide resistant LNCaP/AR cells with a *CHD1* knockdown, using a library of bioinformatically prioritised TFs, revealed newly acquired transcriptional dependencies<sup>359</sup>. Enzalutamide resistant LNCaP/AR cells with a *CHD1* knockdown had become sensitive to the inactivation of Nuclear Receptor Subfamily 3 Group C Member 1 (*NR3C1*), POU Class 3 Homeobox 2 (*POU3F2*), T-box (*TBX2*), and nuclear receptor subfamily 2 group F member 1 (*NR2F1*), all of which had previously been implicated in anti-androgen therapy resistance<sup>359</sup>. Interestingly, transcriptomic and immunofluorescent profiling of ~20 enzalutamide-resistant xenografts derived from LNCaP/AR *CHD1* KD cells, revealed a heterogeneous expression of the acquired dependencies. Most tumours exhibited an upregulation of *NR3C1*, whereas the upregulation of *NR2F1*, *TBX2*, and *POU3F2* was less consistent and was sometimes observed without concurrent *NR3C1* upregulation<sup>359</sup>. Heterogeneous phenotypes associated with lineage plasticity and acquired resistance to targeted lineage therapy is an emerging theme and will be discussed in the next section.

The concept of acquired transcriptional dependencies has also been more generally associated with NE differentiation. Using an elegant unsupervised PCA based approach, Balanis *et al.* defined a common NE gene signature which they used to identify tumours and cancer cell lines that transcriptionally resembled NE differentiated tumours<sup>273</sup>. This study re-discovered known NE cell lines as well as previously undescribed NE cell lines and tumours. Unexpectedly, the NE cell lines and tumours bore a strong resemblance to haematological malignancies. Using available drug sensitivity and pooled shRNA screen data, the authors confirmed that haematological and NE malignancies showed strongly

overlapping drug and genetic dependency profiles<sup>273</sup>. These data demonstrate a switch from a set of canonical lineage dependencies (i.e. the AR in prostate cancer) to the co-option of new transcriptional programs to support cell growth. Interestingly, the resemblance of NE differentiated tumours to haematological malignancies is reminiscent of the acquired sensitivity to the depletion of IRF2 in P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. Further, the NE transcriptional signature identified by Balanis *et al.* is strongly upregulated in the transition from P8<sub>1</sub>S1<sub>1/2</sub>(A) to P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells (**Figure 48**). These data in combination with the observed histological changes suggest that the SMARCB1-mediated mechanism of PAX8 KD resistance in 786-M1A is similar to the NE transition seen in both lung and prostate cancer.



**Figure 48** | GSEA plot of Neuroendocrine gene signature (see materials and methods) for P8<sub>1</sub>S1<sub>1/2</sub> (A) vs P8<sub>1</sub>S1<sub>1/2</sub>(LT).

### 6.3 Heterogeneity and lineage plasticity

As mentioned, lineage plasticity in response to targeted therapy can lead to variable outcomes. In this study, there were multiple layers of heterogeneity. Firstly, the differing effects of SMARCB1 KO on the development of PAX8 KD resistance in ccRCC cell models. In 786-M1A cells, SMARCB1 inactivation promoted a dedifferentiation-like mechanism enabling the co-option of new transcriptional programs to compensate for PAX8 loss. However, in VHL mutant UOK101 cells, SMARCB1 loss was highly detrimental to the emergence of PAX8 resistant clones. UOK101 P8<sub>1</sub>Ctrl(LT) cells grew faster than the SMARCB1 mutant UOK101 P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells (**Figure 22b**). Further, in the population of UOK101 P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells, there was a selection of escaped *SMARCB1* WT clones, visible as a partial recovery of

the SMARCB1 protein band on the WB (**Figure 22a**). The SMARCB1 escaped UOK101 P8<sub>1</sub>S1<sub>2</sub>(LT) cells also grew faster than UOK101 P8<sub>1</sub>S1<sub>2</sub>(LT) cells in which there was no recovery of SMARCB1 protein (**Figure 22a-b**). These data suggest that in certain contexts SMARCB1 KO is not able to render ccRCC cell lines PAX8 insensitive but instead there are alternative preferential routes to resistance.

Heterogeneity was also seen within the PAX8 insensitive SMARCB1 KO 786-M1A cells. P8<sub>1</sub>S1<sub>1</sub>(LT) cells were highly tumorigenic, whereas most P8<sub>1</sub>S1<sub>2</sub>(LT) grafts failed to give rise to tumours (**Figure 19**). The P8<sub>1</sub>S1<sub>1</sub>(LT) cells also showed a greater loss of renal epigenetic and transcriptional identity and bore a stronger epigenetic resemblance to iPSCs, than their P8<sub>1</sub>S1<sub>2</sub>(LT) counterparts (**Figures 35, 37, 39**). The only difference between P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells is the sgRNA used to target SMARCB1. In the literature, there is evidence for specific hotspot mutations which occur in different domains of SMARCB1 which may have different functional consequences<sup>363,364</sup>. However, both sgRNAs (S1<sub>1/2</sub>) in this experiment recognised sequences in the first exon (and hence the same domain) of the SMARCB1 gene and were both highly efficient at knocking out SMARCB1 protein expression (**Figure 18c**). Further, all nine sgRNA targeting SMARCB1 were enriched in the original PAX8 KD resistance screen and there was no correlation between the most highly enriched constructs in the P8<sub>1</sub> condition versus the P8<sub>2</sub> condition. This suggests that in this model, the genomic location of the induced SMARCB1 mutation is not important, provided it leads to a non-functional protein. Instead, it appears as though there are multiple possible phenotypic outcomes to SMARCB1 KO, which is in agreement with the widespread changes to the chromatin landscape (**Figure 28**). Despite the variable *in vivo* and *in vitro* phenotypes between P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells, they showed highly correlative transcriptional dependencies in the TF CRISPR screen (**Figure 42c**). Given that the largest disparity in the phenotypes between P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells was in their tumorigenic capacity, it would be interesting to repeat the TF screen in an *in vivo* setting to see if this reveals different TF sensitivity profiles between the two SMARCB1 KO clones.

Heterogenous phenotypes associated with lineage plasticity and therapy resistance have been reported in multiple breast and prostate cancer models. As mentioned previously, in a *PTEN*<sup>-/-</sup> driven mouse model of metastatic prostate cancer, the concurrent inactivation of *TP53* and *RB1* leads to NE differentiation and castration resistant tumour growth<sup>78</sup>. The degree of luminal (KRT8) versus neuroendocrine (SYP) markers varied markedly within castration resistant tumours<sup>78</sup>. Rainbow lineage tracing revealed that these tumours were often derived from a single clone, and therefore the observed variability was not due to a polyclonal architecture<sup>78</sup>. Similar variability was seen in a patient-derived CRPC-NE organoid model, immunohistochemical staining for *SOX2*, *SYP*, and various SWI/SNF subunits varied throughout the organoid, suggesting that cells undergoing lineage plasticity cells can exist in various states<sup>354</sup>. Interestingly, the co-expression of specific SWI/SNF subunits correlated with the relative expression of *SYP* and *SOX2* markers, providing further evidence for the importance of specific SWI/SNF complex repertoires in maintaining different cellular states<sup>354</sup>. In support of this observational data, a study in breast cancer by Nagarajan et al. generated two *ARID1A* KO MCF7 luminal breast cancer cell lines which were resistant to tamoxifen treatment<sup>272</sup>. In one *ARID1A* KO clone, a strong shift in H4 acetylation status due to reduced HDAC1 binding was observed. This in turn led to BRD4 recruitment and bromodomain and extra-terminal motif (BET)-dependent growth. The second clone exhibited a similar reduction in HDAC1 recruitment but showed very little change in H4 acetylation status, suggesting an alternative (possibly BRD4 independent) *ARID1A* mediated resistance mechanism<sup>272</sup>.

## 6.4 Targeting PAX8 in ccRCC

Despite ~300 TFs being associated with a disease state, currently only a handful have been successfully drugged<sup>203</sup>. Targeting transcription factors using small molecules has proved very challenging, in part because they are predominantly intrinsically disordered and lack easily druggable binding pockets (like enzymatic active sites)<sup>210,365</sup>. They have two main targetable interfaces, the DNA-protein interaction surface which tends to be convex and highly positively charged, and the protein-protein interaction surface which are typically flatter and do not have deep pockets<sup>210,365</sup>. In contrast targeting nuclear hormone receptors is considerably less challenging because they have a well folded ligand binding domain<sup>365</sup>.

Despite the technical challenges, there are examples where non-hormone receptor TFs have been successfully targeted. For example, TFs often require protein-protein interactions (PPIs) to become transcriptionally active, chemically disrupting these interactions has shown efficacy in certain instances<sup>209</sup>. A good example of this is the HIF2 $\alpha$  inhibitor PT2385 which targets the PASB heterodimerisation domain of HIF2 $\alpha$ , blocking the interaction with its binding partner ARNT<sup>194</sup>. Similar approaches with co-activators/repressors or attempts to destabilised/stabilise the TF depending on the context have also shown efficacy for certain TFs in particular disease settings and are the subject of a number of excellent reviews<sup>202,209,210</sup>. An interesting recent development are bifunctional molecules, known as PROTACS<sup>209</sup>. PROTACS have a ligand that binds to an E3 ligase, which is also attached to a second ligand capable of recognising a specific protein. In this way, they can functionally link a TF to a ubiquitin ligase, triggering ubiquitylation and subsequent proteasome-mediated degradation<sup>366</sup>. PROTACS are interesting because they do not have to functionally inhibit a particular process of a TF (i.e. DNA binding or PPIs) in order to target them for degradation<sup>209</sup>. In theory, this provides more possible binding sites and hence more opportunity for the targeting of new TFs<sup>209,366</sup>. To date, as a proof of principle, PROTACS have largely been used in contexts where there are well validated binding sites

and hence were often already drugged<sup>209</sup>. Going forward there is likely to be a large amount of development around the design and application of PROTACS, which will expand the catalogue of targetable TFs<sup>202,202,209,366</sup>.

The inactivation of PAX8 has a very strong effect on cellular growth both *in vitro* and *in vivo* in a number of ccRCC models, making it an attractive therapeutic target (**Figure 12d, 19**)<sup>367,368</sup>. In line with the hypothesis that lineage programs are fundamental vulnerabilities in the majority of tumour cells despite heterogenous driver mutations, the catalogue of PAX8 sensitive ccRCC cells contains an almost full complement of ccRCC driver gene mutations (e.g. *VHL*, *PBRM1*, *BAP1*, *SETD2*, *KDM5C*, *TP53* and *PTEN*). Given the importance of PAX8 in kidney organogenesis, it would be reasonable to expect acute on-target toxicity to a systemic PAX8 therapy<sup>263,369</sup>. However, the inactivation of PAX8 in a mouse embryo does not lead to embryonic lethality or the malformation of the kidneys<sup>263</sup>. PAX8 and PAX2 co-operate during renal development, and it appears that functional PAX2 can compensate for the loss of PAX8<sup>263</sup>. Therefore, a PAX8 based therapy may have a favourable on-target toxicity profile. Like most targeted treatments, a significant challenge for a PAX8 based therapy is likely to be acquired resistance<sup>265</sup>. In a relatively short time frame (~1-2 months) PAX8 sensitive ccRCC cell lines were able to adapt to PAX8 suppression *in vitro*, without a SMARCB1 mutation (**Figure 18a-d**). Given the kinetics, it is likely that resistance emerged through cellular plasticity and adaption, rather than selection for an acquired somatic alteration. Whether or not a similar adaptation is possible *in vivo*, or requires optimised high growth factor *in vitro* conditions, remains to be tested.

As discussed, there are multiple molecular drivers capable of inducing a state of lineage plasticity in response to targeted therapies<sup>59,370</sup>. However, whilst lineage plasticity engenders therapeutic resistance, it can also expose cells to newly acquired druggable dependencies. For example, an ARID1A mutant, tamoxifen/fulvestrant resistant breast cancer model acquired sensitivity to a BET inhibitor and pan-cancer cells exhibiting a NE transcriptional signature acquired dependence on a BCL2 inhibitor which is an approved therapy to treat chronic lymphoblastic leukemia<sup>273,371</sup>. Perhaps more interesting,

are drugs that can re-sensitise tumours to lineage-targeted therapies, such as EZH2 inhibitor treatment that could re-sensitise Trp53<sup>-/-</sup>Rb1<sup>-/-</sup>CRPC-NE cells to androgen depletion<sup>78</sup>. This raises an interesting concept; an understanding of the molecular logic that governs lineage switching and plasticity in cancer, could aid in the design of therapies which maintain lineage fidelity and can be used in combination with lineage targeted agents. An approach such as this could improve the therapeutic potential of targeting PAX8 by decreasing the likelihood of lineage-resistant clones from emerging. Further, drugs capable of enforcing lineage fidelity may have applications in a range of anti-cancer therapies including chemotherapy, immune therapy, RTK inhibitors, and other targeted therapies<sup>83,372,373</sup>.

## 6.5 Future work

### 6.5.1 Supporting experiments and open questions

In this study, it was demonstrated that SMARCB1 KO could convey resistance to PAX8 inhibition in the RCC cell line 786-M1A. In support of this, I showed that SMARCB1 mutant ccRCC cell lines in the DepMap project were refractory to PAX8 KO and transcriptionally resembled P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. This link would be strengthened further, by inducing PAX8 insensitivity through *SMARCB1* KO in another ccRCC model, for example, LM1 cells. LM1 cells have a typical clear cell phenotype, and so a transition to a state histologically resembling NE differentiation would be particularly striking<sup>374</sup>.

As mentioned previously, the biallelic loss of *VHL* is highly specific to ccRCC and the re-constitution of *VHL* leads to degradation of HIF2A and growth arrest in xenograft models of ccRCC<sup>375</sup>. The active epigenetic and transcriptional programs in renal proximal tubule cells provide a context in which *VHL* mutations are tumorigenic. The KO of *SMARCB1* compromises proximal tubule identity and facilitates re-programming in ccRCC models, which raises an interesting question, are *SMARCB1* KO cells still

sensitive to the reintroduction of VHL? The acquired dependence on the HIF2A target gene *BHLHE40* in P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells would suggest that this may still be the case. However, the dependence of P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells on BHLHE40 is relatively weak, and so it remains possible that even with a *VHL* reintroduction, P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells are still tumorigenic. The VHL re-introduction phenotype only occurs in 786-M1A cells in an *in vivo* setting, therefore, to answer this question a subcutaneous *in vivo* tumour growth assay with Ctrl.Ctrl and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells with either an empty vector or VHL reintroduction would be required.

Lineage plasticity in response to environmental pressures (including therapy) can lead to sensitivity to the inhibition of certain chromatin regulators, such as EZH2 and BRD9<sup>78,307</sup>. It would be interesting to test whether these vulnerabilities are re-producible in P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. Whilst it is possible to choose and validate specific targets and drugs, it could be more fruitful to perform an unbiased genetic screen in Ctrl.Ctrl(A) and P8<sub>1</sub>S1<sub>1/2</sub>(LT) cells. The convergence of drug vulnerabilities on chromatin modifiers could suggest a targeted screening approach, that could even be performed *in vivo*. Alternatively, a totally unbiased genome-wide library targeting ~18000 genes could be used *in vitro*. Either approach may identify specific vulnerabilities for the design of new therapeutics to address lineage fidelity and provide insights into the regulators of states acquired through plasticity.

The marked difference in the tumorigenic potential of P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells is at odds with their highly correlated TF dependency profiles (**Figure 42d**). This could be for several reasons; the critical difference between P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells is not captured in the TF sgRNA library, the difference is masked by redundancy, or the phenotype of differentiating TF dependencies can only be seen in an *in vivo* context. To partially address this question, the TF genetic screen could be performed *in vivo*, and a genome-wide screen could be performed *in vitro*. The question of redundancy is difficult to address in a systematic manner, whilst combinatorial sgRNA libraries do exist, they generally do not account for redundancy that exceeds two factors and the complexity of the library can quickly become

very large<sup>377</sup>. A stratified combinatorial screening approach based on differentially expressed genes in P8<sub>1</sub>S1<sub>1</sub>(LT) and P8<sub>1</sub>S1<sub>2</sub>(LT) cells could help to address this technical limitation.

Resistance to PAX8 suppression without an additional genomic alteration was achieved *in vitro* within a reasonably short time period (**Figure 18a-d**). However, whether a similar adaption is possibly *in vivo* where the growth conditions are arguably more challenging, remains to be seen. To maintain a consistent PAX8 KD over time, escaped cells that had regained PAX8 expression were FACS sorted out of the population based on a fluorescent marker, which is not possible *in vivo*. Instead, a PAX8 KO cell line re-constituted with dox-inducible PAX8 cDNA would be a more suitable system. Following subcutaneous injection, mice could be fed a dox supplemented diet until tumours are established, at which point dox can be removed and the kinetics of adaption can be measured by bioluminescence signal or tumour volume measurements. This model could also be used to test whether acute SMARCB1 KO can convey resistance to PAX8 inhibition *in vivo*.

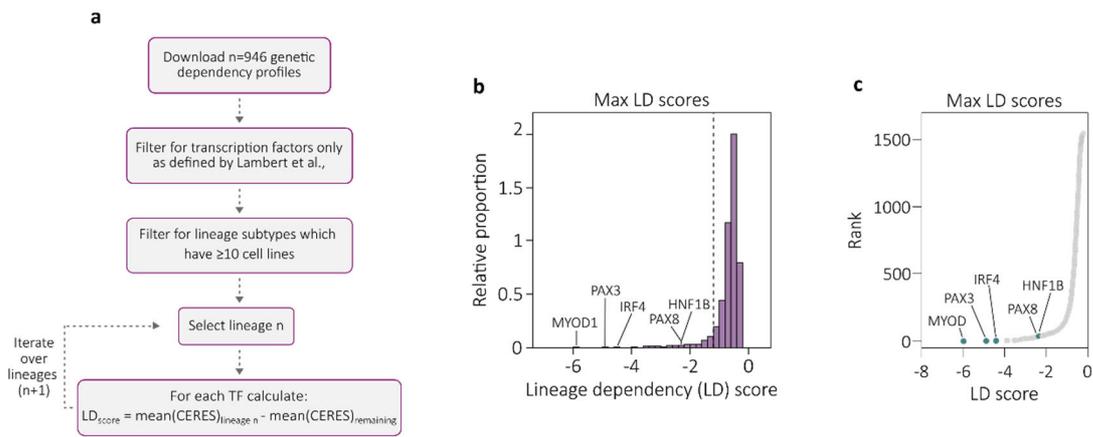
#### 6.5.2 Generalisability of the study - an experimental approach

SMARCB1 was enriched in two independent tamoxifen/fulvestrant resistance screens in a luminal breast cancer cell line<sup>272,301</sup>. These data suggest that SMARCB1 inactivation can convey resistance to lineage factor inhibition in multiple tissue contexts. Therefore, it would be interesting to extend the findings from this study beyond RCC to additional cancer types, both epithelial and non-epithelial. For example, targeting melanocyte inducing transcription factor (MITF) in melanoma, IRF4 in multiple myeloma and HAND2 in neuroblastoma, in conjunction with a SMARCB1 KO.

### 6.5.3 Generalisability of the study - an informatic approach

The occurrence of *de novo* PAX8 KO resistant cell lines, identified in the DepMap project, may suggest a similar pattern is observable in other lineage contexts (**Figure 21**). Identifying additional groups of cell lines that have become decoupled from their lineage-specific transcriptional program would be useful to further probe the importance of the SWI/SNF complex in lineage plasticity and possibly identify additional drivers. In the next section, I will detail an informatic based approach to build a foundation to push the concepts identified in this work forward and potentially generalise them to different cancer types.

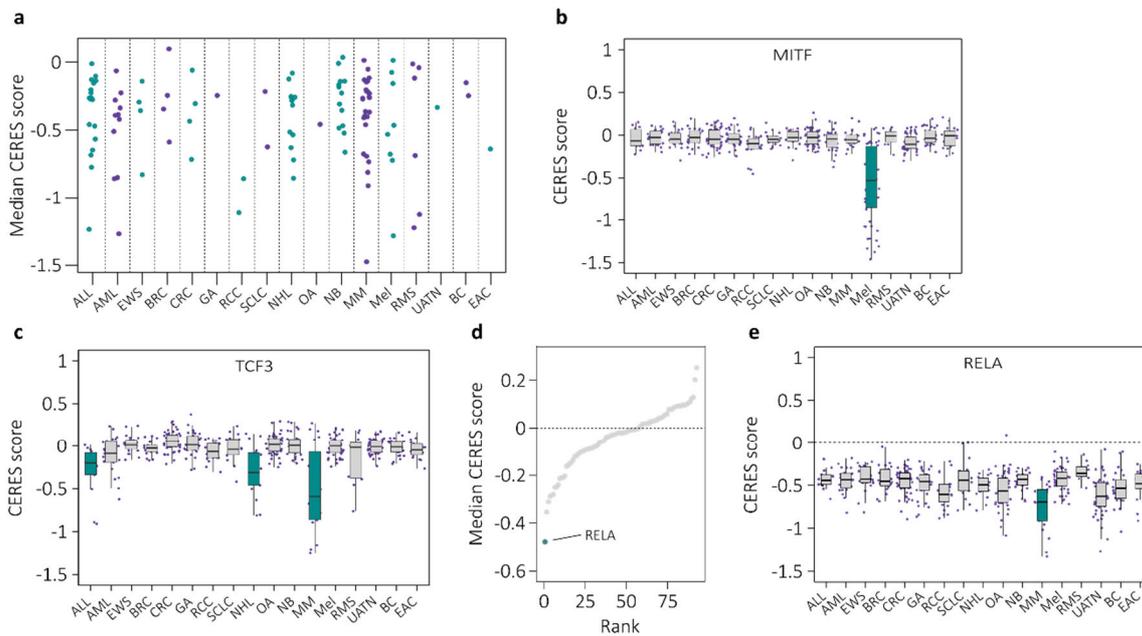
Using the workflow detailed in **Figure 49a**, I made predictions for core regulatory circuitry in 25 lineages represented in the DepMap data. In short, for each lineage I compared the dependency core (CERES score) for each transcription factor against the CERES score for the same TF in cell lines pooled from all other lineages, to create a lineage dependency (LD) score for each TF in each lineage context (see materials and methods). To visualise the distribution of LD scores, I plotted a histogram and a ranked plot of the maximum possible LD<sub>score</sub> (most negative score) for each TF (**Figure 49b-c**). For example, PAX8 and HNF1B had the most negative LD<sub>score</sub> in the RCC lineage context, and so these scores were used. Whereas MYOD1 and PAX3 had the strongest LD score in the rhabdomyosarcoma context, and so for these TFs, those scores were used. As expected, the distribution of the LD<sub>scores</sub> was such, that for most TFs there was no specific dependency in a particular lineage, instead, there was a rare set of TFs which showed very strong specificity (**Figure 49b**).



**Figure 49 | (a)** Workflow for generation of lineage dependency scores (LD) for each TF in each lineage context. Lineage x: selected lineage context, Remaining: cells lines for all other lineages. **(b)** Frequency distribution of the maximum possible LD (i.e. most negative) score for each TF. For example, PAX8 has the strongest lineage specific dependency score in RCC and so the PAX8 LD score in the RCC context is plotted here. **(c)** Ranked plot of the maximum possible LD (i.e. most negative) score for each TF.

To make predictions for lineage-specific TF programs (termed core regulatory circuitry - CRC), for each TF, in each lineage context, I used an  $LD_{score}$  cuff-off ( $< -1.2$ , guided by the distribution of scores in **Figure 49b**) and a  $P$ -value cut-off ( $P < 0.05$ ) (**Figure 50a**). Using this approach, CRC predictions were made for 17 out of 25 lineages. In line with previous reports, the size of the CRC varied greatly between lineages<sup>368</sup>. Hematopoietic lineages (acute lymphoblastic leukaemia -ALL, acute myeloid leukaemia -AML, non-Hodgkin lymphoma -NHL, multiple myeloma -MM) and a neuroectoderm lineage (neuroblastoma -NB) had the largest CRC ranging from 11 to 26 members (**Figure 50a**). Whereas, epithelial cancers, sarcomas, and melanomas tended to have smaller CRC, ranging from 1 to 8 members (**Figure 50a**). The relative cellular dependence on individual members of CRC also varied within lineages. In some instances, cells depended strongly on the lineage factor, for example, PAX8 in RCC, MITF in melanoma, and homozygous transcription factor 3 gene (TCF3) in MM (**Figures 9a, 50b-c**). In contrast, the dependence on TCF3 in ALL and NHL is considerably weaker (**Figure 50c**). Some of the putative CRC members were pan-lineage dependencies that were particularly depleted in certain lineages. These

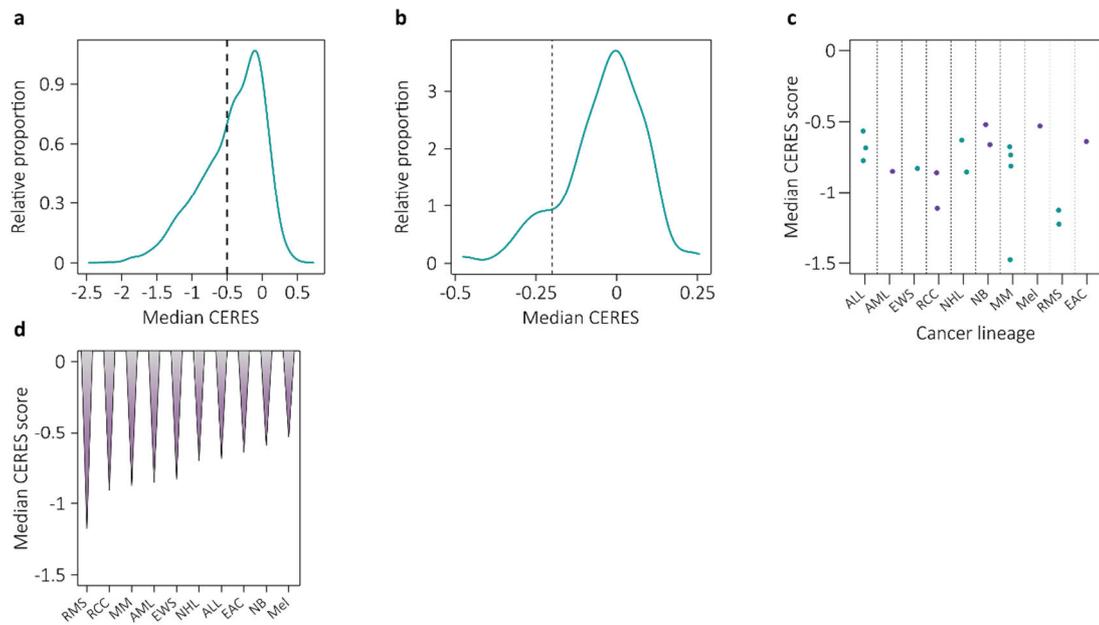
genes had a low median CERES score across all cell lines, the best example of this was *RELA* (Figure 50d). *RELA* was depleted in all 17 lineages but preferentially in MM (Figure 50e).



**Figure 50** | (a) Dot plot of the median CERES score across cell lines in a particular lineage, for putative lineage specific TF dependencies after filtering ( $LD_{score} \leq -1.2$ ,  $P < 0.05$ ). (b) Box plot of CERES scores for the gene *MITF* across all lineages with identified specific dependencies ( $n=17$ ). Highlighted lineage, Mel, is the lineage for which *MITF* was identified as a specific dependency. (c) Box plot of CERES scores for the gene *TCF3* across all lineages with identified specific dependencies ( $n=17$ ). Highlighted lineages – ALL, NHL and MM - are the lineages for which *TCF3* was identified as a specific dependency. (d) A ranked plot of the median CERES scores across all cell lines (across all lineages) for TFs identified as lineage specific. (e) Box plot of CERES scores for the gene *RELA* across all lineages with identified specific dependencies ( $n=17$ ). Highlighted lineage, MM, is the lineage for which *RELA* was identified as a specific dependency. Abbreviations:- ALL: acute lymphoblastic leukaemia, AML: acute myeloid leukaemia, EWS: Ewing sarcoma, BRC: breast carcinoma, CRC: colorectal adenocarcinoma, GA: gastric adenocarcinoma, RCC: renal cell carcinoma, SCLC: small cell lung cancer, NHL: non-Hodgkin lymphoma, OA: ovarian adenocarcinoma, NB: neuroblastoma, MM: multiple myeloma, Mel: melanoma, RMS: rhabdomyosarcoma, UATN: upper aerodigestive tract neoplasm, BC: bladder carcinoma, EAC: endometrial adenocarcinoma.

I decided to apply a second round of filtering on the putative CRC members to focus the analysis on strong cellular dependencies and remove any pan-lineage dependencies, to create lineage networks that were comparable to PAX8 and HNF1B in ccRCC. To filter for strong cellular dependencies, the

distribution of the LD CERES scores did not provide an obvious guide, so I opted for a conservative cut-off; >50% of cell lines within a lineage had to have a CERES score of  $\leq -0.5$  for a respective putative LD (Figure 51a).



**Figure 51** | (a) Frequency distribution of the median CERES scores for each identified lineage dependency from figure 50a in their respective lineage. (b) Frequency distribution of the median CERES scores for each identified lineage dependency from figure 50a across cell lines in their respective lineage. (c) Dot plot of the median CERES score across cell lines in a particular lineage, for prospective lineage specific TF dependencies after three successive rounds of filtering (1)  $LD_{score} \leq -1.2$ ,  $P < 0.05$ , (2)  $\geq 50\%$  of cell lines with the lineage have a CERES score of  $\leq -0.5$  for a prospective LD and (3) the median CERES score of LD across all cell lines  $\geq -0.2$ . (e) Bar plot of the median CERES score across cell lines and LDs in each lineage.

This ensured that strong dependencies were selected whilst accounting for the possibility that there are *de novo* resistant cell lines. To remove pan-lineage dependencies I once again plotted the distribution of median CERES scores across all cell lines for putative CRC members (Figure 51b). A cut-off of median CERES  $\geq -0.2$  separated the bimodal distribution, removing pan-lineage dependencies, like RELA. After filtering I was able to detect strong, lineage-specific dependencies in 10 of the original

25 lineages, including haematological, epithelial and neuroectodermal malignancies, sarcomas and melanomas (Figure 51c; Table 4).

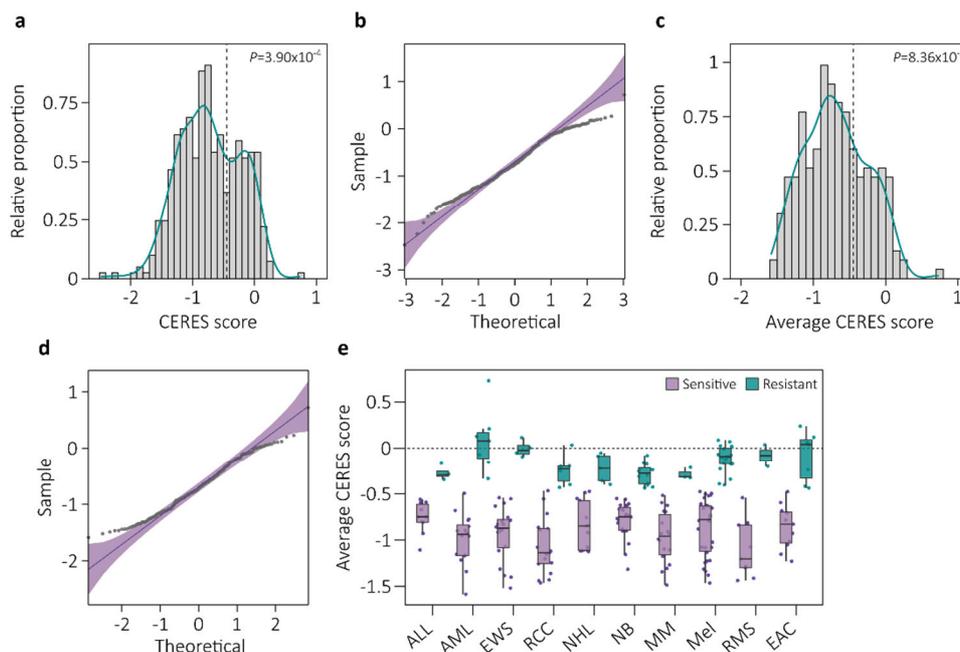
Lineage	Putative CRC
Acute lymphoblastic leukemia	<i>EBF1, PAX5, RUNX1</i>
Acute myeloid leukemia	<i>SPI1</i>
Ewing sarcoma	<i>FLI1</i>
Renal cell carcinoma	<i>HNF1B, PAX8</i>
Non-Hodgkin lymphoma	<i>BCL6, EBF1</i>
Neuroblastoma	<i>ISL1, PHOX2A</i>
Multiple myeloma	<i>IRF4, POU2AF1, PRDM1, TCF3</i>
Melanoma	<i>MITF</i>
Rhabdomyosarcoma	<i>MYOD1, PAX3</i>
Endometrial adenocarcinoma	<i>PAX8</i>

**Table 4** | CRC predictions

The number of CRC members across cancer lineages after filtering was more comparable, ranging from 1 to 4 members (Figure 51c). Ranking of the 10 lineages by how strongly they depended on their respective lineage factors, revealed a substantial range, possibly suggesting that some lineages may respond better to lineage factor targeting (Figure 51d). The most dependent lineages were rhabdomyosarcoma (RMS) and RCC, and the least dependant lineages were melanoma and NB.

Having identified 10 lineages encompassing a range of malignancies with strong and specific transcriptional dependencies, I next sought to uncover examples of *de novo* resistance to lineage factor inhibition. A histogram of cell line dependency scores for their respective LDs revealed a strongly bimodal distribution (Figure 52a). I confirmed the deviation of the distribution of LD CERES scores from a normal Gaussian distribution using a quantile-quantile (Q-Q) plot and the Shapiro-Wilk statistical test (Figure 52a-b). The smaller of the two populations is centred close to a CERES score of 0 and therefore likely represents a small fraction of cell lines that are resistant to lineage factor inhibition. The larger of

the populations, accounting for most cell lines, is centred around a CERES score of -1 (essential gene status) and so represents cell lines sensitive to lineage factor inhibition. The plot from **Figure 52a** considers each lineage factor in each lineage separately, which means for lineages with >1 LD, the same cell line will appear more than once in the distribution. If the LDs are averaged within lineages across the same cell line and the distribution is re-plotted, a similar bimodal distribution is seen (**Figure 52c-d**). However, it does not deviate as strongly from a normal distribution, suggesting that when there are multiple strong LDs, a cell line can become resistant to one whilst maintaining its dependence on others. Nonetheless, in both instances, it is readily possible to identify lineage factor resistant cell lines. Using the distribution of LD CERES scores as a guide, I applied a cut-off (avgCERES score for LD > -0.45) to identify resistant cell lines in each of the 10 lineages (**Figure 52e**). Remarkably, there were examples of lineage-resistant cells in all 10 lineages.



**Figure 52** | **(a)** Frequency distribution of the CERES score for each LD from figure 38e, in each cell line of their respective lineage. **(b)** A Q-Q plot comparing the distribution of the scores from (a) to a theoretical gaussian distribution. The area shaded in grey indicates the 95% confidence interval under the null. **(c)** Frequency distribution of the CERES score for each LD from table 4, in each cell line of their respective lineage. For lineages which contain >1 LD, an average is taken across LDs in each cell line. **(d)** A Q-Q plot comparing the distribution of the scores from (c) to a theoretical gaussian distribution. The area shaded in grey indicates the 95% confidence interval under the null. **(e)** Box plot of the average CERES score of each LD in their respective lineage. Sensitive and resistant lines are identified using the cut off identified in (a) and (c), average CERES > -0.45. Abbreviations:- ALL: acute lymphoblastic leukaemia, AML: acute myeloid leukaemia, EWS: Ewing sarcoma, RCC: renal cell carcinoma, NHL: non-Hodgkin lymphoma, NB: neuroblastoma, MM: multiple myeloma, Mel: melanoma, RMS: rhabdomyosarcoma, EAC: endometrial adenocarcinoma.

As a preliminary analysis to demonstrate the utility of this approach, with the aid of a collaborator (D.S), a permutation-based statistical method (see materials and methods) was applied to look for changes in the genetic dependencies of resistant cell lines. Based on the concepts developed in this project, lineage-resistant cell lines can acquire dependencies on new TFs (i.e. IRF2) and possibly lose dependence on chromatin regulators important for maintaining differentiated transcriptional programs (e.g. SMARCB1 and ARID1A).

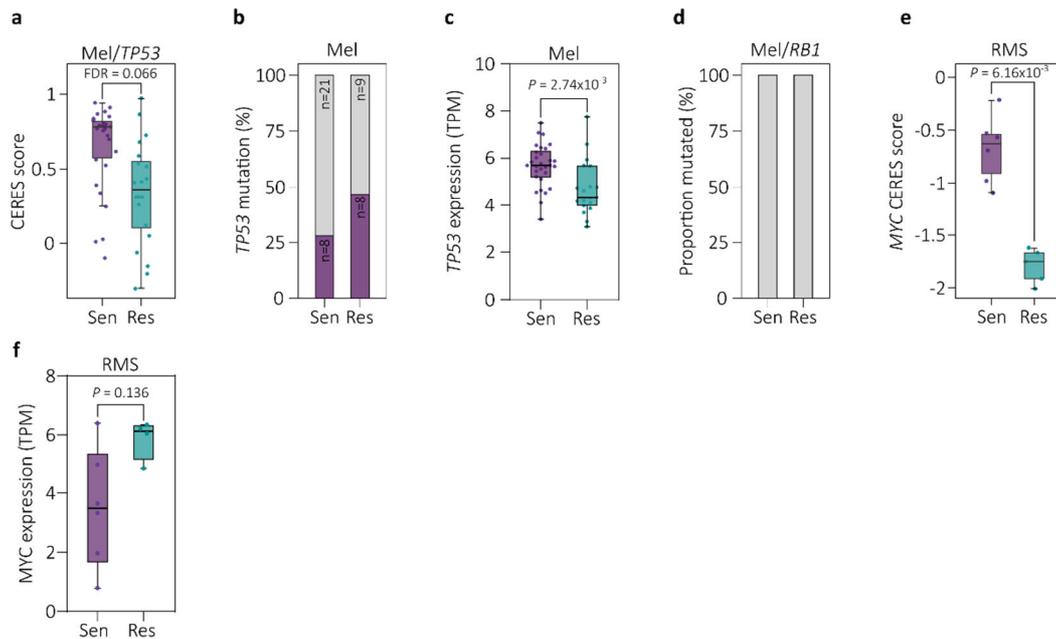
The dependence on LDs defines whether a cell line in a lineage is sensitive or resistant, therefore re-discovering LDs would serve as a sanity check for this analysis. Resistant cell lines should have a statistically significant acquired lack of dependence on LDs. Using a widely accepted FDR cut-off of 0.1, we were able to re-discover 7/19 LDs in 7/10 lineages (**Table 5**).

Lineage	LD	Acquired dependency	Acquired lack of dependency
Acute lymphoblastic leukemia	<i>EBF1</i>	<i>ARID3A, GATA3, ETS1, BCL11B</i>	<i>ZNF143, EBF1</i>
Acute lymphoblastic leukemia	<i>PAX5</i>		
Acute lymphoblastic leukemia	<i>RUNX1</i>		<i>RUNX1</i>
Acute myeloid leukemia	<i>SPI1</i>		<i>SPI1</i>
Ewing sarcoma	<i>FLI1</i>		<i>FLI1</i>
Renal cell carcinoma	<i>HNF1B</i>		<i>HNF1B</i>
Renal cell carcinoma	<i>PAX8</i>	<i>ZNF281, ZNF347</i>	<i>PAX8</i>
Non-Hodgkin lymphoma	<i>BCL6</i>		<i>BCL6</i>
Non-Hodgkin lymphoma	<i>EBF1</i>		<i>PAX5, EBF1</i>
Neuroblastoma	<i>ISL1</i>	<i>OTP</i>	<i>TFAP2B, TBX1, ISL1</i>
Neuroblastoma	<i>PHOX2A</i>		<i>PHOX2A</i>
Multiple myeloma	<i>IRF4</i>		
Multiple myeloma	<i>POU2AF1</i>		
Multiple myeloma	<i>PRDM1</i>		<i>PRDM1</i>
Multiple myeloma	<i>TCF3</i>		<i>TCF3</i>
Melanoma	<i>MITF</i>	<i>TP53</i>	<i>SOX9, NFE2L1, NCOA2, SOX10, TFAP2A, , BPTF, MITF</i>
Rhabdomyosarcoma	<i>MYOD1</i>		
Rhabdomyosarcoma	<i>PAX3</i>	<i>MYC</i>	<i>PAX3</i>
Endometrial adenocarcinoma	<i>PAX8</i>		<i>PAX8</i>

**Table 5** | Results from the lineage resistant acquired dependence / lack of dependence analysis, using a list of TFs, filtered for FDR <0.1. Acknowledgments:- As stated in the Preface, the permutation based statistical test used to identify changes in dependency profiles was performed by D.S.

The low recovery rate of LDs is likely because the number of cell lines within each lineage is not sufficient to account for heterogeneity in LD dependence, therefore we can only detect LDs for which most resistant cell lines have acquired independence. This is supported by the fact that it was possible to recover LDs for all lineages with one LD (4/4), and only 3 out of 6 lineages with multiple LDs. This may suggest that an FDR cut-off of 0.1 is too stringent for a data set of this size. However, to reduce the cut-off would increase the chance of detecting random noise in the data instead of biological signal.

Despite the limitations of the data set size, it was possible to detect instances of both acquired dependencies and lack of dependencies in lineage-resistant cell lines. As an illustrative example of the latter, TP53 KO is enriched in MITF-dependant cells, in line with its role as a TSG, but in MITF-independent cells, the enrichment is less strong (**Figure 53a**).



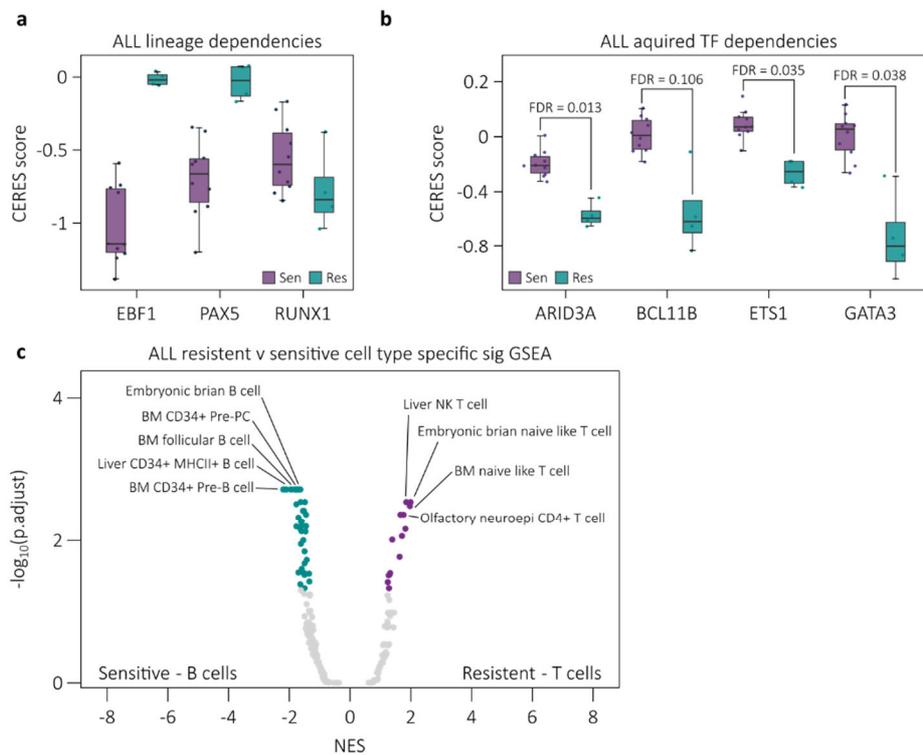
**Figure 53** | (a) Box plot of the CERES score of lineage sensitive and resistant cell lines from the melanoma lineage for TP53 KO. FDR calculated using permutation-based statistics. (b) Stacked bar plot representing the proportion of TP53 mutations in lineage sensitive and resistant cell lines from the melanoma lineage. (c) Box plot of normalised TP53 gene expression (transcripts per million -TPM) in lineage sensitive and resistant cell lines from the melanoma lineage. (d) Stacked bar plot representing the proportion of RB1 mutations in lineage sensitive and resistant cell lines from the melanoma lineage. (e) Box plot of the CERES score of lineage sensitive and resistant cell lines from the rhabdomyosarcoma lineage for MYC KO. FDR calculated using permutation-based statistics. (f) Box plot of normalised MYC gene expression (transcripts per million -TPM) in lineage sensitive and resistant cell lines from the rhabdomyosarcoma lineage. Abbreviations:- Mel: melanoma, RMS: rhabdomyosarcoma.

As mentioned previously, the inactivation of TP53 is an important event in neuroendocrine differentiation and has been shown to increase the efficiency of reprogramming epithelial cells into iPSCs<sup>378–380</sup>. The reduced CERES score for TP53 in MITF-independent melanoma cells suggests it may

already be downregulated or inactivated, thereby facilitating a transition to a MITF independent state. In line with, this there is an increase in the proportion of *TP53* mutations in MITF -independent cells (47% vs 28%) and a reduction in *TP53* mRNA expression (**Figure 53b-c**). Unlike neuroendocrine differentiation and lineage plasticity in prostate cancer, there were no detectable *RB1* mutations concomitant with *TP53* in the lineage resistant melanoma cell lines, possibly suggesting a different reprogramming mechanism (**Figure 53d**).

RMS cells resistant to PAX3 or both PAX3 and MYOD1 KO had an enhanced dependence on MYC compared to lineage-sensitive lines (**Figure 53e**). MYC (c-MYC) has a role in cellular reprogramming, the maintenance of pluripotency, and the control of cell fate decisions<sup>381</sup>. Most famously MYC is one 'Yamanaka factors', the retroviral introduction of OCT3/4, SOX2, KLF4 and MYC can induce the transformation of fibroblasts into induced pluripotent stem cells (iPSCs)<sup>382</sup>. Another MYC family member, N-MYC, is expressed highly in neuroendocrine resistant prostate cancer and the exogenous overexpression of N-Myc in a *Pten*-deficient mouse model induces androgen deprivation therapy resistance and increases the incidence of neuroendocrine differentiation<sup>383–385</sup>. The enhanced dependence on MYC raises the interesting possibility that PAX3/MYOD1-independent cell lines have enhanced MYC activity which promotes lineage factor resistance. In line with this, *MYC* trends towards higher expression in lineage-resistant cell lines (**Figure 53f**). However, the difference does not reach statistical significance, possibly due to the very small sample size and subsequently low statistical power (sensitive lines n= 6, resistant lines n=4). Further, there are two lineage-sensitive cell lines with higher *MYC* expression, suggesting *MYC* expression alone may not be sufficient to promote lineage resistance. It is, therefore, possible that a small sample size cannot account for heterogenous phenotypes associated with higher *MYC* expression. It would be interesting to inhibit PAX3/MYOD1 in the *MYC* high and low expression lineage sensitive lines and compare their propensity to adapt over a longer time frame than the DepMap CRISPR screen (~20 days).

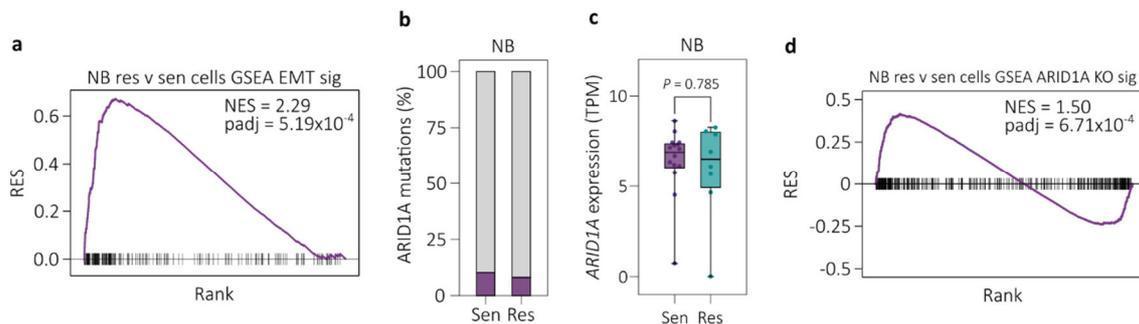
In acute lymphoblastic leukaemia (ALL) a subset of cell lines that were resistant to the inhibition of the lineage factors Early B-Cell Factor 1 (EBF1) and PAX5 but not Runt-related transcription factor (RUNX1), were identified (**Figure 54a**). To replace EBF1 and PAX5, the resistant cell lines had acquired a dependence on the transcriptional output of BAF Chromatin Remodeling Complex Subunit BCL11B (BCL11B), ETS Proto-Oncogene 1 (ETS1), and GATA Binding Protein 3 (GATA3) (**Figure 54b**). Given the challenge of rediscovering LDs, it was remarkable to detect three newly acquired dependencies in ALL. This finding suggested two possible scenarios, either there is a highly probable singular route to lineage independence in ALL or, the resistant and sensitive cell lines represent different subtypes of ALL with related but discrete developmental origins that were not distinguished in the DepMap annotations. To test the second scenario, I performed differential expression analysis with lineage-sensitive and resistant ALL cell lines, coupled with GSEA using the collection of cell-type-specific signatures (C8). The most highly enriched signatures in the resistant cell lines belonged to T cells, whereas the most negatively enriched signatures belonged to B cells (**Figure 54c**). This is in line with literature describing two discrete subtypes of ALL, one derived from a B cell developmental trajectory (B-ALL) and the other from a T cell developmental trajectory (T-ALL)<sup>386</sup>. The ALL lineage resistant (T-ALL) cells are dependent on the T cell-specific TFs GATA3 and BCL11B whereas the sensitive cells are dependent on the B cell-specific factors EBF1 and PAX5<sup>386,387</sup>. Both B-ALL and T-ALL cell lines are sensitive to the inactivation of RUNX1, which is important for the generation of HSCs, a common ancestor in both B and T cell development<sup>386</sup>.



**Fig. 54|** (a) A box plot of the CERES scores for ALL LDs in ALL cell lines, separated into sensitive (purple) and resistant (cyan) cell lines according to figure 39e. (b) A box plot of the CERES scores for transcriptional dependencies specific to the lineage resistant ALL cell lines. (c) Volcano plot of GSEA with cell-type specific transcriptional signatures from mSigDB collection 8, for the comparison ALL lineage sensitive vs ALL lineage resistant cell lines. highlighted points (purple/cyan) have a  $p.adjust < 0.05$ .

In addition to looking for changes in the genetic dependency profiles between resistant and sensitive lines, using mRNA expression data it was also possible to measure changes in the transcriptional programs. As an example, differential expression analysis between lineage resistant and sensitive neuroblastoma cell lines, coupled with GSEA using the hallmarks collection, revealed a very strong upregulation in the EMT signature (**Figure 55a**). Recent molecular characterisation of primary neuroblastoma tumours identified a high-risk molecular subtype with a mesenchymal-like gene expression profile, which strongly overlapped with highly aggressive mesenchymal glioblastoma<sup>388</sup>. In a Zebrafish model, the transition from an adrenergic to mesenchymal state in neuroblastoma was

functionally associated with the mutational inactivation of ARID1A either through a non-synonymous mutation or large chromosomal deletion<sup>389</sup>. However, the lineage-resistant neuroblastoma cells identified in this analysis did not have a higher proportion of ARID1A mutations or a lower expression of ARID1A mRNA, suggesting that in this context there are multiple mechanistic drivers for the mesenchymal transition. (Figure 55b-c).



**Figure 55 | (a)** GSEA plot of the hallmarks EMT signature neuroblastoma lineage sensitive vs lineage resistant cell lines. **(b)** Stacked bar plot representing the proportion of *ARID1A* mutations in lineage sensitive and resistant cell lines from the neuroblastoma lineage. **(c)** Box plot of normalised *ARID1A* gene expression (transcripts per million -TPM) in lineage sensitive and resistant cell lines from the neuroblastoma lineage. **(d)** GSEA plot of *ARID1A* KO up regulated gene signature for neuroblastoma lineage sensitive vs lineage resistant cell lines. Abbreviations:- NB: neuroblastoma.

The identification of ten lineages with specific transcriptional dependencies featuring *de novo* lineage factor resistance provides an interesting base to expand the concepts developed in this study as well as identify additional molecular drivers. The occurrence of *de novo* lineage factor resistance in the absence of lineage factor inhibition is likely the result of lineage plasticity in response to microenvironmental stressors. This is supported by reports of lineage plasticity in response to multiple treatment modalities including chemotherapy, MAPK targeted therapy and immunotherapy, as well as environmental queues such as hypoxia and inflammation<sup>59,68–71,74,75,390–392</sup>. A preliminary analysis of the data set implicated MYC, TP53, and an EMT in independence from lineage factor signalling. Going forward there is a need for a more comprehensive molecular dissection of lineage independence. For

example, the transcriptomic profiling can be expanded to additional lineages beyond NB and include more signatures. Further, a systematic approach for identifying recurrent mutations could be implemented, and other publicly available modalities (e.g. reverse-phase protein array, RNA splicing, DNA methylation, microRNA expression, and drug sensitivity data) can be incorporated into the analysis<sup>303</sup>. Finally, it would be interesting to systematically measure changes in SWI/SNF subunit mRNA and protein expression, to determine if there are re-producible changes in the subunit repertoire during the transition to a lineage independent state, similar to what has been reported in prostate cancer<sup>354</sup>.

7

## Closing Remarks

Transcriptional lineage factors are a prominent class of essential genes in cancer. Beyond targeting NHRs in breast and prostate cancer, the mechanisms that maintain lineage fidelity during carcinogenesis, and whether lineage factor pathways could be broadly exploited for cancer therapy remain poorly understood. In this study, I used clear ccRCC as a model to characterise the mechanisms that underlie lineage factor dependence in cancer. Through CRISPR/Cas9 loss-of-function screening I found that loss of SMARCB1, a member of the SWI/SNF chromatin remodelling complex, can confer an *in vitro growth* advantage and rescued tumorigenesis in ccRCC cells upon inhibition of the essential renal lineage factor PAX8. PAX8 resistant cells (P8<sub>1</sub>S1<sub>1/2</sub>(LT) and P8<sub>1</sub>Ctrl(LT) cells) formed tumours with features of neuroendocrine histology, which is commonly seen in castration-resistant prostate cancers but not in ccRCC. Profiling of the transcriptome and epigenome showed that SMARCB1 inactivation triggered large-scale cis-regulatory changes, a loss of kidney-specific epigenetic identity, acquisition of a cellular state resembling that of RTs, and activation of proliferative programs. The reactivation of proliferative pathways after PAX8 inhibition was achieved through the adoption of new transcriptional dependencies on IRF2, BHLHE40, and ZNFX1. The newly acquired dependencies fit into three categories: acquisition of a lineage dependency from another lineage (IRF2 from plasma cells), expansion of the already in place core oncogenic program (VHL-HIF2 $\alpha$ ), and acquisition of a rare pan-cancer dependency (ZNFX1). Finally, using a large-scale CRISPR/Cas9 screening data set comprising hundreds of cancer cell lines, I identified examples of *de novo* lineage factor resistant cell lines even in the absence of a specific lineage factor targeted therapy. Lineage resistant cell lines showed common changes in dependency profile, such as acquired dependency/lack of dependency on MYC and TP53 respectively. In this context, it is likely that lineage-resistant cancer clones develop in response to environmental or therapeutic stressors. Thus, the principles governing lineage plasticity should be taken into consideration when designing novel lineage factor-targeted and other cancer therapies.

# 8 References

1. Weinberg, R. A. *The Biology of Cancer*. (2013).
2. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell*. **100**, 57–70 (2000).
3. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell*. **144**, 646–674 (2011).
4. Alberts, B. *et al.* Molecular Biology of the Cell. *Molecular Biology of the Cell* (2017).
5. Davies, M. & Samuels, Y. Analysis of the genome to personalize therapy for melanoma. *Oncogene*. **29**, 5545 (2010).
6. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: Weaving a tumorigenic web. *Nature Reviews Cancer*. **11**, 761–774 (2011).
7. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: A changing paradigm. *Nature Reviews Cancer*. **9**, 153–166 (2009).
8. Singh, R., Letai, A. & Sarosiek, K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. *Nature Reviews Molecular Cell Biology*. **20**, 175–193 (2019).
9. Junttila, M. R. & Evan, G. I. P53 a Jack of all trades but master of none. *Nature Reviews Cancer*. **9**, 821–829 (2009).
10. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature*. **502**, 333–339 (2013).
11. Seluanov, A., Gladyshev, V. N., Vijg, J. & Gorbunova, V. Mechanisms of cancer resistance in long-lived mammals. *Nature Reviews Cancer*. **18**, 433–441 (2018).
12. Cagan, A. *et al.* Somatic mutation rates scale with lifespan across mammals. *bioRxiv* (2019)
13. Gomes, N. M. V. *et al.* Comparative biology of mammalian telomeres: Hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*. **10**, 761–768 (2011).
14. Yuan, X., Larsson, C. & Xu, D. Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. *Oncogene*. **38**, 6172–6183 (2019).
15. Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science*. **339**, 957–959 (2013).
16. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science*. **339**, 959–961 (2013).

17. Bell, R. J. A. *et al.* The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*. **348**, 1036–1039 (2015).
18. Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell*. **173**, 611–623 (2018).
19. Valkenburg, K. C., De Groot, A. E. & Pienta, K. J. Targeting the tumour stroma to improve cancer therapy. *Nature Reviews Clinical Oncology*. **15**, 366–381 (2018).
20. Zuazo-Gaztelu, I. & Casanovas, O. Unraveling the role of angiogenesis in cancer ecosystems. *Frontiers in Oncology*. **8**, 248 (2018).
21. George, D. J. & Kaelin, W. G. The von Hippel–Lindau Protein, Vascular Endothelial Growth Factor, and Kidney Cancer. *New England Journal of Medicine*. **349**, 419–421 (2003).
22. Turajlic, S., Larkin, J. & Swanton, C. SnapShot: Renal Cell Carcinoma. *Cell*. **163**, 1556–1556.e1 (2015).
23. Patel, S. A., Rodrigues, P., Wesolowski, L. & Vanharanta, S. Genomic control of metastasis. *British Journal of Cancer*. **124**, 3–12 (2021).
24. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science*. **331**, 1559–1564 (2011).
25. Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: Dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *American Journal of Pathology*. **153**, 865–873 (1998).
26. AF, C., AC, G. & IC, M. Dissemination and growth of cancer cells in metastatic sites. *Nature Reviews Cancer*. **2**, 563–572 (2002).
27. Hall, A. The cytoskeleton and cancer. *Cancer and Metastasis Reviews*. **28**, 5–14 (2009).
28. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine*. **19**, 1423–1437 (2013).
29. Kitamura, T., Qian, B. Z. & Pollard, J. W. Immune cell promotion of metastasis. *Nature Reviews Immunology*. **15**, 73–86 (2015).
30. Patel, S. A. & Vanharanta, S. Epigenetic determinants of metastasis. *Molecular Oncology*. **11**, 79–96 (2017).
31. Vanharanta, S. *et al.* Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nature Medicine*. **19**, 50–56 (2013).
32. De Berardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism. *Science Advances*. **2** (2016).

33. Warburg, O. On the origin of cancer cells. *Science*. **123**, 309–314 (1956).
34. Weinhouse, S., Warburg, O., Burk, D. & Schade, A. L. On respiratory impairment in cancer cells. *Science*. **124**, 267–272 (1956).
35. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends in Biochemical Sciences*. **41**, 211–218 (2016).
36. Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: Glutamine metabolism to cancer therapy. *Nature Reviews Cancer*. **16**, 619–634 (2016).
37. Lane, A. N. & Fan, T. W. M. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research*. **43**, 2466–2485 (2015).
38. Harding, H. P. *et al.* An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Molecular Cell*. **11**, 619–633 (2003).
39. Hosios, A. M. *et al.* Amino Acids Rather than Glucose Account for the Majority of Cell Mass in Proliferating Mammalian Cells. *Developmental Cell*. **36**, 540–549 (2016).
40. Sullivan, L. B., Gui, D. Y. & Van Der Heiden, M. G. Altered metabolite levels in cancer: Implications for tumour biology and cancer therapy. *Nature Reviews Cancer*. **16**, 680–693 (2016).
41. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*. **483**, 474–478 (2012).
42. Fouad, Y. A. & Aanei, C. Revisiting the hallmarks of cancer. *American Journal of Cancer Research*. **7**, 1016–1036 (2017).
43. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell*. **19**, 17–30 (2011).
44. ME, F. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. **18**, 553–567 (2010).
45. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. **529**, 110–114 (2016).
46. Rohle, D. *et al.* An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science*. **340**, 626–630 (2013).
47. Wang, F. *et al.* Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science*. **340**, 622–626 (2013).

48. Gonzalez, H., Hagerling, C. & Werb, Z. Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes and Development*. **32**, 1267–1284 (2018).
49. Waldman, A. D., Fritz, J. M. & Lenardo, M. J. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*. **20**, 651–668 (2020).
50. Burnet, M. Cancer-A Biological Approach I. The Processes of Control. *British Medical Journal*. **1**, 779–786 (1957).
51. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: From immunosurveillance to tumor escape. *Nature Immunology*. **3**, 991–998 (2002).
52. Teng, M. W. L., Swann, J. B., Koebel, C. M., Schreiber, R. D. & Smyth, M. J. Immune-mediated dormancy: an equilibrium with cancer. *Journal of Leukocyte Biology*. **84**, 988–993 (2008).
53. Vajdic, C. M. & Van Leeuwen, M. T. Cancer incidence and risk factors after solid organ transplantation. *International Journal of Cancer*. **125**, 1747–1754 (2009).
54. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell*. **165**, 45–60 (2016).
55. Spranger, S. & Gajewski, T. F. Mechanisms of Tumor Cell-Intrinsic Immune Evasion. *Annual Review of Cancer Biology*. **2**, 213–228 (2018).
56. Malmberg, K. J. Effective immunotherapy against cancer: A question of overcoming immune suppression and immune escape? *Cancer Immunology, Immunotherapy*. **53**, 879–892 (2004).
57. Vinay, D. S. *et al.* Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Seminars in Cancer Biology*. **35**, S185–S198 (2015).
58. Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nature Reviews Molecular Cell Biology*. **20**, 69–84 (2019).
59. Quintanal-Villalonga, Á. *et al.* Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nature Reviews Clinical Oncology*. **17**, 360–371 (2020).
60. Stone, R. C. *et al.* Epithelial-mesenchymal transition in tissue repair and fibrosis. *Cell and Tissue Research*. **365**, 495–506 (2016).
61. Lim, J. & Thiery, J. P. Epithelial-mesenchymal transitions: Insights from development. *Development*. **139**, 3471–3486 (2012).
62. Aiello, N. M. *et al.* EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Developmental Cell*. **45**, 681–695 (2018).

63. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*. **125**, 315–326 (2006).
64. Maruyama, R. *et al.* Epigenetic regulation of cell type-specific expression patterns in the human mammary epithelium. *PLoS Genetics*. **7**, e1001369 (2011).
65. Chaffer, C. L. *et al.* Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell*. **154**, 61 (2013).
66. Rios, A. C. *et al.* Intracлонаl Plasticity in Mammary Tumors Revealed through Large-Scale Single-Cell Resolution 3D Imaging. *Cancer Cell*. **35**, 618-632 (2019).
67. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nature Reviews Cancer*. **18**, 128–134 (2018).
68. Guo, W. *et al.* Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell*. **148**, 1015–1028 (2012).
69. Lim, S. *et al.* SNAI1-Mediated Epithelial-Mesenchymal Transition Confers Chemoresistance and Cellular Plasticity by Regulating Genes Involved in Cell Death and Stem Cell Maintenance. *PLoS One*. **8**, e66558 (2013).
70. Tsoi, J. *et al.* Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. *Cancer Cell*. **33**, 890-904 (2018).
71. Fallahi-Sichani, M. *et al.* Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. *Molecular Systems Biology*. **13**, 905 (2017).
72. Biehs, B. *et al.* A cell identity switch allows residual BCC to survive Hedgehog pathway inhibition. *Nature*. **562**, 429–433 (2018).
73. Sánchez-Danés, A. *et al.* A slow-cycling LGR5 tumour population mediates basal cell carcinoma relapse after therapy. *Nature*. **562**, 434–458 (2018).
74. Landsberg, J. *et al.* Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. *Nature*. **490**, 412–416 (2012).
75. Mehta, A. *et al.* Immunotherapy resistance by inflammation-induced dedifferentiation. *Cancer Discovery*. **8**, 935–943 (2018).
76. Park, J. W. *et al.* Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science*. **362**, 91–95 (2018).
77. Mu, P. *et al.* SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science*. **355**, 84–88 (2017).

78. Ku, S. Y. *et al.* Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science*. **355**, 78–83 (2017).
79. Chen, Z. *et al.* Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13. *Proceedings of the National Academy of Sciences of the United States of America*. **115**, 6810–6815 (2018).
80. De Palma, M., Biziato, D. & Petrova, T. V. Microenvironmental regulation of tumour angiogenesis. *Nature Reviews Cancer*. **17**, 457–474 (2017).
81. Blomberg, O. S., Spagnuolo, L. & De Visser, K. E. Immune regulation of metastasis: Mechanistic insights and therapeutic opportunities. *Disease Models and Mechanisms*. **11** (2018).
82. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science*. **357** (2017).
83. MD, N. *et al.* Combined TP53 and RB1 Loss Promotes Prostate Cancer Resistance to a Spectrum of Therapeutics and Confers Vulnerability to Replication Stress. *Cell Reports*. **31** (2020).
84. Kidney cancer statistics | Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer#heading-One> (2017).
85. Capitanio, U. & Montorsi, F. Renal cancer. *The Lancet*. **387**, 894–906 (2016).
86. American Cancer Society. American Cancer Society - Cancer Statistics. <https://cancerstatisticscenter.cancer.org/#/> (2021).
87. PDQ Adult Treatment Editorial Board. Transitional Cell Cancer of the Renal Pelvis and Ureter Treatment (PDQ®): Health Professional Version. PDQ Cancer Information Summaries (2020).
88. PDQ Adult Treatment Editorial Board. Renal Cell Cancer Treatment (PDQ®)—Health Professional Version. National Cancer Institute. (2021).
89. Chen, F. *et al.* Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma. *Cell Reports*. **14**, 2476–2489 (2016).
90. Hsieh, J. J. *et al.* Renal cell carcinoma. *Nature Reviews Disease Primers*. **3**, 17009 (2017).
91. Brok, J., Treger, T. D., Gooskens, S. L., van den Heuvel-Eibrink, M. M. & Pritchard-Jones, K. Biology and treatment of renal tumours in childhood. *European Journal of Cancer*. **68**, 179–195 (2016).

92. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: A metabolic disease. *Nature Reviews Urology*. **7**, 277–285 (2010).
93. Hakimi, A. A. *et al.* An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell*. **29**, 104–116 (2016).
94. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*. **361**, 594–599 (2018).
95. Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell*. **26**, 319–330 (2014).
96. Frew, I. J. & Moch, H. A Clearer View of the Molecular Complexity of Clear Cell Renal Cell Carcinoma. *Annual Review of Pathology: Mechanisms of Disease*. **10**, 263–289 (2015).
97. Kaelin, W. G. von Hippel-Lindau Disease. *Annual Review of Pathology: Mechanisms of Disease*. **2**, 145–173 (2007).
98. Bakouny, Z. *et al.* Integrative molecular characterization of sarcomatoid and rhabdoid renal cell carcinoma. *Nature Communications*. **12**, 1–14 (2021).
99. Nagwa. Lesson Explainer: Kidney Structure. <https://www.nagwa.com/en/explainers/196146106357/> (2021).
100. Cancer Genome Atlas Research Network *et al.* Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. **499**, 43–49 (2013).
101. Gnarr, J. R. *et al.* Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nature Genetics*. **7**, 85–90 (1994).
102. Hakimi, A. A., Pham, C. G. & Hsieh, J. J. A clear picture of renal cell carcinoma. *Nature Genetics*. **45**, 849–850 (2013).
103. Zhang, H. *et al.* HIF-1 Inhibits Mitochondrial Biogenesis and Cellular Respiration in VHL-Deficient Renal Cell Carcinoma by Repression of C-MYC Activity. *Cancer Cell*. **11**, 407–420 (2007).
104. Shen, C. *et al.* Genetic and functional studies implicate HIF1a as a 14q kidney cancer suppressor gene. *Cancer Discovery*. **1**, 222–235 (2011).
105. Shenoy, N. HIF1 $\alpha$  is not a target of 14q deletion in clear cell renal cancer. *Scientific Reports*. **10**, 1–12 (2020).
106. Gudas, L. J., Fu, L., Minton, D. R., Mongan, N. P. & Nanus, D. M. The role of HIF1 $\alpha$  in renal cell carcinoma tumorigenesis. *Journal of Molecular Medicine*. **92**, 825–836 (2014).
107. Hoefflin, R. *et al.* HIF-1 $\alpha$  and HIF-2 $\alpha$  differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice. *Nature Communications*. **11**, 1–21

- (2020).
108. Yu, T., Tang, B. & Sun, X. Development of inhibitors targeting hypoxia-inducible factor 1 and 2 for cancer therapy. *Yonsei Medical Journal*. **58**, 489–496 (2017).
  109. Cho, H. & Kaelin, W. G. Targeting HIF2 in clear cell renal cell carcinoma. *Cold Spring Harbor Symposia on Quantitative Biology*. **81**, 113–121 (2016).
  110. FDA approves belzutifan for cancers associated with von Hippel-Lindau disease | FDA Drugs. <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-belzutifan-cancers-associated-von-hippel-lindau-disease> (2021)
  111. Kapitsinou, P. P. & Haase, V. H. The VHL tumor suppressor and HIF: insights from genetic studies in mice. *Cell Death Differentiation*. **15**, 650–659 (2008).
  112. Espana-Agusti, J., Warren, A., Chew, S. K., Adams, D. J. & Matakidou, A. Loss of PBRM1 rescues VHL dependent replication stress to promote renal carcinogenesis. *Nature Communications*. **8**, 2026 (2017).
  113. Beroukhi, R. *et al.* Patterns of gene expression and copy-number alterations in von-Hippel Lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Research*. **69**, 4674–4681 (2009).
  114. Dondeti, V. R. *et al.* Integrative genomic analyses of sporadic clear cell renal cell carcinoma define disease subtypes and potential new therapeutic targets. *Cancer Research*. **72**, 112–121 (2012).
  115. Peña-Llopis, S., Christie, A., Xie, X. J. & Brugarolas, J. Cooperation and antagonism among cancer genes: The renal cancer paradigm. *Cancer Research*. **73**, 4173–4179 (2013).
  116. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*. **366**, 883–892 (2012).
  117. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*. **46**, 225–233 (2014).
  118. Shelar, S. *et al.* Biochemical and epigenetic insights into L-2-hydroxyglutarate, a potential therapeutic target in renal cancer. *Clinical Cancer Research*. **24**, 6433–6446 (2018).
  119. Shenoy, N. *et al.* Ascorbic acid-induced TET activation mitigates adverse hydroxymethylcytosine loss in renal cell carcinoma. *Journal of Clinical Investigation*. **129**, 1612–1625 (2019).
  120. Shim, E. H. *et al.* L-2-hydroxyglutarate: An epigenetic modifier and putative oncometabolite in renal cancer. *Cancer Discovery*. **4**, 1290–1298 (2014).
  121. Liu, L. *et al.* Enhancer of zeste homolog 2 (EZH2) promotes tumour cell migration and invasion via epigenetic repression of E-cadherin in renal cell carcinoma. *BJU International*.

- 117**, 351–362 (2016).
122. Duran, A. *et al.* P62 Is a Key Regulator of Nutrient Sensing in the mTORC1 Pathway. *Molecular Cell*. **44**, 134–146 (2011).
  123. Sanchez, D. J. & Simon, M. C. Genetic and metabolic hallmarks of clear cell renal cell carcinoma. *Biochimica et Biophysica Acta - Reviews on Cancer*. **1870**, 23–31 (2018).
  124. Scelo, G. *et al.* Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nature Communications*. **8**, (2017).
  125. Bigot, P. *et al.* Functional characterization of the 12p12.1 renal cancer-susceptibility locus implicates BHLHE41. *Nature Communications*. **7**, (2016).
  126. Purdue, M. P. *et al.* Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nature Genetics*. **43**, 60–65 (2011).
  127. Schödel, J. *et al.* Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nature Genetics*. **44**, 420–425 (2012).
  128. Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Reports*. **23**, 313–326 (2018).
  129. Hirata, H. *et al.* Wnt antagonist DKK1 acts as a tumor suppressor gene that induces apoptosis and inhibits proliferation in human renal cell carcinoma. *International Journal of Cancer*. **128**, 1793–1803 (2011).
  130. Ricketts, C. J., Hill, V. K. & Linehan, W. M. Tumor-specific hypermethylation of epigenetic biomarkers, including SFRP1, predicts for poorer survival in patients from the TCGA kidney renal clear cell carcinoma (KIRC) project. *PLoS One*. **9**, e85621 (2014).
  131. Urakami, S. *et al.* Wnt antagonist family genes as biomarkers for diagnosis, staging, and prognosis of renal cell carcinoma using tumor and serum DNA. *Clinical Cancer Research*. **12**, 6989–6997 (2006).
  132. Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nature Genetics*. **44**, 751–759 (2012).
  133. Mantovani, F., Collavin, L. & Del Sal, G. Mutant p53 as a guardian of the cancer cell. *Cell Death and Differentiation*. **26**, 199–212 (2018).
  134. Turajlic, S. & Swanton, C. TRACERx Renal: Tracking renal cancer evolution through therapy. *Nature Reviews Urology*. **14**, 575–576 (2017).
  135. Turajlic, S., Xu, H., Litchfield, K. & Larkin, J. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*. **173**, 595–610 (2018).

136. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell*. **173**, 581-594 (2018).
137. Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature*. **587**, 126–132 (2020).
138. Wettersten, H. I., Aboud, O. A., Lara, P. N. & Weiss, R. H. Metabolic reprogramming in clear cell renal cell carcinoma. *Nature Reviews Nephrology*. **13**, 410–419 (2017).
139. Li, B. *et al.* Fructose-1,6-bisphosphatase opposes renal carcinoma progression. *Nature*. **513**, 251–255 (2014).
140. Mandriota, S. J. *et al.* HIF activation identifies early lesions in VHL kidneys: Evidence for site-specific tumor suppressor function in the nephron. *Cancer Cell*. **1**, 459–468 (2002).
141. Semenza, G. L. HIF-1: upstream and downstream of cancer metabolism. *Current Opinion in Genetics and Development*. **20**, 51–56 (2010).
142. Semenza, G. L. Regulation of cancer cell metabolism by hypoxia-inducible factor 1. *Seminars in Cancer Biology*. **19**, 12–16 (2009).
143. Kim, J. W., Tchernyshyov, I., Semenza, G. L. & Dang, C. V. HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. *Cell Metabolism*. **3**, 177–185 (2006).
144. Papandreou, I., Cairns, R. A., Fontana, L., Lim, A. L. & Denko, N. C. HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metabolism*. **3**, 187–197 (2006).
145. Valera, V. A., Walter, B. A., Linehan, W. M. & Merino, M. J. Regulatory effects of microRNA-92 (miR-92) on VHL Gene expression and the hypoxic activation of miR-210 in clear cell renal cell carcinoma. *Journal of Cancer*. **2**, 515–526 (2011).
146. White, N. M. A. *et al.* MiRNA profiling in metastatic renal cell carcinoma reveals a tumour-suppressor effect for miR-215. *British Journal of Cancer*. **105**, 1741–1749 (2011).
147. Ivan, M. & Huang, X. MiR-210: Fine-tuning the hypoxic response. *Advances in Experimental Medicine and Biology*. **772**, 205–227 (2014).
148. Pantuck, A. J. *et al.* Prognostic relevance of the mTOR pathway in renal cell carcinoma: Implications for molecular patient selection for targeted therapy. *Cancer*. **109**, 2257–2267 (2007).
149. Robb, V. A., Karbowniczek, M., Klein-Szanto, A. J. & Henske, E. P. Activation of the mTOR Signaling Pathway in Renal Clear Cell Carcinoma. *Journal of Urology*. **177**, 346–352 (2007).
150. Guo, H. *et al.* The PI3K/AKT Pathway and Renal Cell Carcinoma. *Journal of Genetics and Genomics*. **42**, 343–353 (2015).

151. Gebhard, R. L. *et al.* Abnormal cholesterol metabolism in renal clear cell carcinoma. *Journal of Lipid Research*. **28**, 1177–1184 (1987).
152. Qiu, B. *et al.* HIF2 $\alpha$ -dependent lipid storage promotes endoplasmic reticulum homeostasis in clear-cell renal cell carcinoma. *Cancer Discovery*. **5**, 653–667 (2016).
153. Du, W. *et al.* HIF drives lipid deposition and cancer in ccRCC via repression of fatty acid metabolism. *Nature Communications*. **8**, (2017).
154. Sundelin, J. P. *et al.* Increased Expression of the Very Low-Density Lipoprotein Receptor Mediates Lipid Accumulation in Clear-Cell Renal Cell Carcinoma. *PLoS One*. **7**, e48694 (2012).
155. Wettersten, H. I. *et al.* Grade-dependent metabolic reprogramming in kidney cancer revealed by combined proteomics and metabolomics analysis. *Cancer Research*. **75**, 2541–2552 (2015).
156. Perroud, B., Ishimaru, T., Borowsky, A. D. & Weiss, R. H. Grade-dependent proteomics characterization of kidney cancer. *Molecular Cell Proteomics*. **8**, 971–985 (2009).
157. Von Roemeling, C. A. *et al.* Stearoyl-CoA desaturase 1 is a novel molecular therapeutic target for clear cell renal cell carcinoma. *Clinical Cancer Research*. **19**, 2368–2380 (2013).
158. Horiguchi, A. *et al.* Fatty Acid Synthase Over Expression is an Indicator of Tumor Aggressiveness and Poor Prognosis in Renal Cell Carcinoma. *Journal of Urology*. **180**, 1137–1140 (2008).
159. Bensaad, K. *et al.* Fatty acid uptake and lipid storage induced by HIF-1 $\alpha$  contribute to cell growth and survival after hypoxia-reoxygenation. *Cell Reports*. **9**, 349–365 (2014).
160. Platten, M., Nollen, E. A. A., Röhrig, U. F., Fallarino, F. & Opitz, C. A. Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nature Reviews Drug Discovery*. **18**, 379–401 (2019).
161. Trott, J. F. *et al.* Inhibiting tryptophan metabolism enhances interferon therapy in kidney cancer. *Oncotarget*. **7**, 66540–66557 (2016).
162. Routy, J. P., Routy, B., Graziani, G. M. & Mehraj, V. The kynurenine pathway is a double-edged sword in immune-privileged sites and in cancer: Implications for immunotherapy. *International Journal of Tryptophan Research*. **9**, 67–77 (2016).
163. Gameiro, P. A. *et al.* In vivo HIF-mediated reductive carboxylation is regulated by citrate levels and sensitizes VHL-deficient cells to glutamine deprivation. *Cell Metabolism*. **17**, 372–385 (2013).
164. Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature*. **481**, 380–384 (2012).

165. Yoon, C. Y. *et al.* Renal cell carcinoma does not express argininosuccinate synthetase and is highly sensitive to arginine deprivation via arginine deiminase. *International Journal of Cancer*. **120**, 897–905 (2007).
166. Search of: ADI-PEG20 - List Results - ClinicalTrials.gov. <https://clinicaltrials.gov/ct2/results?term=ADI-PEG20+&Search=Search> (2021).
167. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA: A Cancer Journal for Clinicians*. **67**, 93–99 (2017).
168. Renal Cell Cancer Treatment (PDQ)—Health Professional Version | National Cancer Institute. <https://www.cancer.gov/types/kidney/hp/kidney-treatment-pdq> (2021).
169. Early Diagnosis Data Hub | Cancer Research UK. <https://crukcanerintelligence.shinyapps.io/EarlyDiagnosis/> (2018).
170. Mantia, C. M. & McDermott, D. F. Vascular endothelial growth factor and programmed death-1 pathway inhibitors in renal cell carcinoma. *Cancer*. **125**, 4148–4157 (2019).
171. Motzer, R. J. *et al.* Sunitinib versus Interferon Alfa in Metastatic Renal-Cell Carcinoma. *New England Journal of Medicine*. **356**, 115–124 (2007).
172. Sternberg, C. N. *et al.* Pazopanib in locally advanced or metastatic renal cell carcinoma: Results of a randomized phase III trial. *Journal Clinical Oncology*. **28**, 1061–1068 (2010).
173. Motzer, R. J. *et al.* Pazopanib versus Sunitinib in Metastatic Renal-Cell Carcinoma. *New England Journal of Medicine*. **369**, 722–731 (2013).
174. Choueiri, T. K. *et al.* Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. *New England Journal of Medicine*. **373**, 1814–1823 (2015).
175. Choueiri, T. K. *et al.* Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *Lancet Oncology*. **17**, 917–927 (2016).
176. Motzer, R. J. *et al.* Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet*. **372**, 449–456 (2008).
177. Hudes, G. *et al.* Temsirolimus, Interferon Alfa, or Both for Advanced Renal-Cell Carcinoma. *New England Journal of Medicine*. **356**, 2271–2281 (2007).
178. Motzer, R. J. *et al.* Lenvatinib, everolimus, and the combination in patients with metastatic renal cell carcinoma: A randomised, phase 2, open-label, multicentre trial. *Lancet Oncology*. **16**, 1473–1482 (2015).

179. Escudier, B. *et al.* Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet*. **370**, 2103–2111 (2007).
180. Rini, B. I. *et al.* Bevacizumab plus interferon alfa compared with interferon alfa monotherapy in patients with metastatic renal cell carcinoma: CALGB 90206. *Journal Clinical Oncology*. **26**, 5422–5428 (2008).
181. Xu, W., Atkins, M. B. & McDermott, D. F. Checkpoint inhibitor immunotherapy in kidney cancer. *Nature Reviews Urology*. **17**, 137–150 (2020).
182. Motzer, R. J. *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *New England Journal of Medicine*. **373**, 1803–1813 (2015).
183. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune checkpoint blockade: A common denominator approach to cancer therapy. *Cancer Cell*. **27**, 450–461 (2015).
184. Selby, M. J. *et al.* Preclinical development of ipilimumab and nivolumab combination immunotherapy: Mouse tumor models, In vitro functional studies, and cynomolgus macaque toxicology. *PLoS One*. **11**, e0161779 (2016).
185. Curran, M. A., Montalvo, W., Yagita, H. & Allison, J. P. PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 4275–4280 (2010).
186. Yang, Y. *et al.* Comparative Efficacy and Safety of Nivolumab and Nivolumab Plus Ipilimumab in Advanced Cancer: A Systematic Review and Meta-Analysis. *Frontiers in Pharmacology*. **11** (2020).
187. Xu, H. *et al.* Antitumor activity and treatment-related toxicity associated with nivolumab plus ipilimumab in advanced malignancies: A systematic review and meta-analysis. *Frontiers in Pharmacology*. **10** (2019).
188. Motzer, R. J. *et al.* Nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial. *Lancet Oncology*. **20**, 1370–1385 (2019).
189. Motzer, R. J. *et al.* Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *New England Journal Medicine*. **378**, 1277–1290 (2018).
190. Kusmartsev, S. *et al.* Oxidative Stress Regulates Expression of VEGFR1 in Myeloid Cells: Link to Tumor-Induced Immune Suppression in Renal Cell Carcinoma. *Journal of Immunology*. **181**, 346–353 (2008).
191. Rini, B. I. *et al.* Phase III trial of bevacizumab plus interferon alfa versus interferon alfa monotherapy in patients with metastatic renal cell carcinoma: Final results of CALGB 90206. *Journal of Clinical Oncology*. **28**, 2137–2143 (2010).

192. Osada, T. *et al.* The effect of anti-VEGF therapy on immature myeloid cell and dendritic cells in cancer patients. *Cancer Immunology Immunotherapy*. **57**, 1115–1124 (2008).
193. Motzer, R. J. *et al.* Avelumab plus axitinib versus sunitinib in advanced renal cell carcinoma: biomarker analysis of the phase 3 JAVELIN Renal 101 trial. *Nature Medicine*. **26**, 1733–1741 (2020).
194. Chen, W. *et al.* Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature*. **539**, 112–117 (2016).
195. Cho, H. *et al.* On-target efficacy of a HIF-2 $\alpha$  antagonist in preclinical kidney cancer models. *Nature*. **539**, 107–111 (2016).
196. Wallace, E. M. *et al.* A small-molecule antagonist of HIF2 $\alpha$  is efficacious in preclinical models of renal cell carcinoma. *Cancer Research*. **76**, 5491–5500 (2016).
197. Courtney, K. D. *et al.* HIF-2 complex dissociation, target inhibition, and acquired resistance with PT2385, a first-in-class HIF-2 inhibitor, in patients with clear cell renal cell carcinoma. *Clinical Cancer Research*. **26**, 793–803 (2020).
198. Jonasch, E. *et al.* An open-label phase II study to evaluate PT2977 for the treatment of von Hippel-Lindau disease-associated renal cell carcinoma. *Journal of Clinical Oncology*. **37**, 680 (2019).
199. Rini, B. I. *et al.* Results from a phase I expansion cohort of the first-in-class oral HIF-2 $\alpha$  inhibitor PT2385 in combination with nivolumab in patients with previously treated advanced RCC. *Journal of Clinical Oncology*. **37**, 558–558 (2019).
200. Wu, D. *et al.* Bidirectional modulation of HIF-2 activity through chemical ligands. *Nature Chemical Biology*. **15**, 367–376 (2019).
201. Bhagwat, A. S. & Vakoc, C. R. Targeting Transcription Factors in Cancer. *Trends in Cancer*. **1**, 53–65 (2015).
202. Bushweller, J. H. Targeting transcription factors in cancer — from undruggable to reality. *Nature Reviews Cancer*. **19**, 611–624 (2019).
203. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell*. **172**, 650–665 (2018).
204. Cramer, P. Organization and regulation of gene transcription. *Nature*. **573**, 45–54 (2019).
205. Darnell, J. E. Transcription factors as targets for cancer therapy. *Nature Reviews Cancer*. **2**, 740–749 (2002).
206. Paull, E. O. *et al.* A modular master regulator landscape controls cancer transcriptional identity. *Cell*. **184**, 334–351 (2021).

207. Califano, A. & Alvarez, M. J. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nature Reviews Cancer*. **17** 116–130 (2017).
208. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell*. **168**, 629–643 (2017).
209. Henley, M. J. & Koehler, A. N. Advances in targeting ‘undruggable’ transcription factors with small molecules. *Nature Reviews Drug Discovery*. **20**, (2021).
210. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chemistry and Biology*. **21**, 1102–1114 (2014).
211. Peyvandipour, A., Shafi, A., Saberian, N. & Draghici, S. Identification of cell types from single cell data using stable clustering. *Scientific Reports*. **10**, 1–12 (2020).
212. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*. **10**, 252–263 (2009).
213. Lee, T. I. I. & Young, R. A. A. Transcriptional Regulation and Its Misregulation in Disease. *Cell*. **152**, 1237–1251 (2013).
214. Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nature Reviews Genetics*. **14**, 427–439 (2013).
215. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development*. **143**, 1833–1837 (2016).
216. Lorch, Y. & Kornberg, R. D. Chromatin-remodeling for transcription. *Quarterly Reviews of Biophysics*. **50** (2017).
217. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics*. **15**, 272–286 (2014).
218. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics and Development*. **43**, 73–81 (2017).
219. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*. **16**, 167–177 (2015).
220. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nature Reviews Genetics*. **17**, 207–223 (2016).
221. Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell*. **167**, 1188–1200 (2016).

222. Phillips-Cremins, J. E. & Corces, V. G. Chromatin Insulators: Linking Genome Organization to Cellular Function. *Molecular Cell*. **50**, 461–474 (2013).
223. Gibcus, J. H. & Dekker, J. The Hierarchy of the 3D Genome. *Molecular Cell*. **49**, 773–782 (2013).
224. Alarcón, C. *et al.* Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF- $\beta$  Pathways. *Cell*. **139**, 757–769 (2009).
225. Campos, E. I. & Reinberg, D. Histones: Annotating chromatin. *Annual Review of Genetics*. **43**, 559–599 (2009).
226. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*. **17**, 487–500 (2016).
227. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research*. **21**, 381–395 (2011).
228. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology*. **15**, 703–708 (2014).
229. David Allis, C., Caparros, M.-L., Jenuwein, T., Reinberg, D. & Lachner, M. *Epigenetics, Second Edition*. (2015).
230. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature*. **500**, 477–481 (2013).
231. Ghirlando, R. & Felsenfeld, G. CTCF: Making the right connections. *Genes and Development*. **30**, 881–891 (2016).
232. Liu, M. Y., DeNizio, J. E., Schutsky, E. K. & Kohli, R. M. The expanding scope and impact of epigenetic cytosine modifications. *Current Opinion in Chemical Biology*. **33**, 67–73 (2016).
233. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*. **151**, 1417–1430 (2012).
234. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature Genetics*. **45**, 592–601 (2013).
235. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature Genetics*. **45**, 860–867 (2013).
236. Wang, Z. Y. & Chen, Z. Acute promyelocytic leukemia: From highly fatal to highly curable. *Blood*. **111**, 2505–2515 (2008).
237. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell*. **157**, 369–381 (2014).

238. Affer, M. *et al.* Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia*. **28**, 1725–1735 (2014).
239. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell*. **155**, 934–947 (2013).
240. Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes and Development*. **27**, 2648–2662 (2013).
241. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nature Genetics*. **48**, 176–182 (2016).
242. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. **351**, 1454–1458 (2016).
243. Garraway, L. A. & Sellers, W. R. Lineage dependency and lineage-survival oncogenes in human cancer. *Nature Reviews Cancer*. **6**, 593–602 (2006).
244. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*. **167**, 1170–1187 (2016).
245. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell*. **170**, 564–576 (2017).
246. Heinlein, C. A. & Chang, C. Androgen Receptor in Prostate Cancer. *Endocrinology Reviews*. **25**, 276–308 (2004).
247. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature*. **436**, 117–122 (2005).
248. Alluri, P. G., Speers, C. & Chinnaiyan, A. M. Estrogen receptor mutations and their role in breast cancer progression. *Breast Cancer Research*. **16**, 494 (2014).
249. Tan, M. E., Li, J., Xu, H. E., Melcher, K. & Yong, E. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*. **36**, 3–23 (2015).
250. Burris, T. P. *et al.* Nuclear receptors and their selective pharmacologic modulators. *Pharmacological Reviews*. **65**, 710–778 (2013).
251. Pawlak, M., Lefebvre, P. & Staels, B. General Molecular Biology and Architecture of Nuclear Receptors. *Current Topics in Medicinal Chemistry*. **12**, 486–504 (2012).
252. Patel, H. K. & Bihani, T. Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacology and Therapeutics*. **186**, 1–24 (2018).
253. Shiau, A. K. *et al.* The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell*. **95**, 927–937 (1998).

254. Vickman, R. E. *et al.* The role of the androgen receptor in prostate development and benign prostatic hyperplasia: A review. *Asian Journal of Urology*. **7**, 191–202 (2020).
255. Dai, C., Heemers, H. & Sharifi, N. Androgen signaling in prostate cancer. *Cold Spring Harbor Perspectives in Medicine*. **7** (2017).
256. Nevedomskaya, E., Baumgart, S. J. & Haendler, B. Recent advances in prostate cancer treatment and drug discovery. *International Journal of Molecular Sciences*. **19** (2018).
257. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature Genetics*. **49**, 1779–1784 (2017).
258. Dempster, J. M. *et al.* Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nature*. **10**, 1–14 (2019).
259. Rauscher, B., Henkel, L., Heigwer, F. & Boutros, M. Lineage specific core-regulatory circuits determine gene essentiality in cancer cells. *bioRxiv*. (2019).
260. Syafruddin, S. E. *et al.* A KLF6-driven transcriptional network links lipid homeostasis and tumour growth in renal carcinoma. *Nature communications*. **10**, 1152 (2019).
261. Madariaga, L. *et al.* Severe prenatal renal anomalies associated with mutations in HNF1B or PAX2 genes. *Clinical Journal of the American Society of Nephrology*. **8**, 1179–1187 (2013).
262. Clissold, R. L., Hamilton, A. J., Hattersley, A. T., Ellard, S. & Bingham, C. HNF1B-associated renal and extra-renal disease - An expanding clinical spectrum. *Nature Reviews Nephrology*. **11**, 102–112 (2015).
263. Narlis, M., Grote, D., Gaitan, Y., Boualia, S. K. & Bouchard, M. Pax2 and Pax8 Regulate Branching Morphogenesis and Nephron Differentiation in the Developing Kidney. *Journal of the American Society of Nephrology*. **18**, 1121–1129 (2007).
264. Kaminski, M. M. M. *et al.* Direct reprogramming of fibroblasts into renal tubular epithelial cells by defined transcription factors. *Nature Cell Biology*. **18**, 1269–1280 (2016).
265. Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature*. **575**, 299–309 (2019).
266. Ricketts, C. J. & Linehan, W. M. Multi-regional Sequencing Elucidates the Evolution of Clear Cell Renal Cell Carcinoma. *Cell*. **173**, 540–542 (2018).
267. Turajlic, S. & Swanton, C. Implications of cancer evolution for drug development. *Nature Reviews Drug Discovery*. **16**, 441–442 (2017).
268. Mcgranahan, N. & Swanton, C. Leading Edge Review Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. **168**, 613–628 (2017).

269. Fellmann, C. *et al.* An Optimized microRNA Backbone for Effective Single-Copy RNAi. *Cell Reports*. **5**, 1704–1713 (2013).
270. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. **15** (2014).
271. Nakayama, R. T. *et al.* SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nature Genetics*. **49**, 1613–1623 (2017).
272. Nagarajan, S. *et al.* ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response. *Nature Genetics*. **52**, 187–197 (2020).
273. Balanis, N. G. *et al.* Pan-cancer Convergence to a Small-Cell Neuroendocrine Phenotype that Shares Susceptibilities with Hematological Malignancies. *Cancer Cell*. **36**, 17–34 (2019).
274. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*. **16**, 284–287 (2012).
275. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*. **10**, 1213–1218 (2013).
276. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10 (2011).
277. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. **1303**, (2013).
278. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*. **44**, W160–W165 (2016).
279. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*. **9**, (2019).
280. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biology*. **9**, (2008).
281. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*. **38**, 576–589 (2010).
282. Lerdrup, M., Johansen, J. V., Agrawal-Singh, S. & Hansen, K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nature Structural Molecular Biology*. **23**, 349–357 (2016).

283. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*. **20**, 296 (2019).
284. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888-1902 (2019).
285. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*. **36**, 411–420 (2018).
286. Wang, P. *et al.* Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney Development by Single-Cell RNA Sequencing. *Cell Reports*. **24**, 3554-3567 (2018).
287. Lindström, N. O. *et al.* Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis. *Developmental Cell*. **45**, 651-660 (2018).
288. Hochane, M. *et al.* Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS Biology*. **17**, e3000152 (2019).
289. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*. **343**, 80–84 (2014).
290. Pacini, C. *et al.* Integrated cross-study datasets of genetic dependencies in cancer. *Nature Communications*. **12**, 1–14 (2021).
291. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nature Communications*. **9**, 1–13 (2018).
292. Schuster, A. *et al.* RNAi/CRISPR Screens: from a Pool to a Valid Hit. *Trends in biotechnology*. **37**, 38-55 (2019).
293. Perino, M. & Veenstra, G. J. C. Chromatin Control of Developmental Dynamics and Plasticity. *Developmental Cell*. **38**, 610–620 (2016).
294. Cheloufi, S. *et al.* The histone chaperone CAF-1 safeguards somatic cell identity. *Nature*. **528**, 218–224 (2015).
295. Kolundzic, E. *et al.* FACT Sets a Barrier for Cell Fate Reprogramming in *Caenorhabditis elegans* and Human Cells. *Developmental Cell*. **46**, 611-626 (2018).
296. JM, M. *et al.* Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nature Communications*. **9**, (2018).
297. Peretz, L. *et al.* Combined shRNA over CRISPR/cas9 as a methodology to detect off-target effects and a potential compensatory mechanism. *Scientific Reports*. **8**, 1–13 (2018).

298. Lin, A. *et al.* Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Science Translational Medicine*. **11**, 8412 (2019).
299. Mathur, R. & Roberts, C. W. M. SWI/SNF (BAF) Complexes: Guardians of the Epigenome. *Annual Review of Cancer Biology*. **2**, 413–427 (2018).
300. Ye, Y., Chen, X. & Zhang, W. Mammalian SWI/SNF Chromatin Remodeling Complexes in Embryonic Stem Cells: Regulating the Balance Between Pluripotency and Differentiation. *Frontiers in Cell and Developmental Biology*. **8**, 626383 (2021).
301. Xu, G. *et al.* ARID1A determines luminal identity and therapeutic response in estrogen-receptor-positive breast cancer. *Nature Genetics*. **52**, 198–207 (2020).
302. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. **467**, 430–435 (2010).
303. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. **569**, 503–508 (2019).
304. Neigeborn, L. & Carlson, M. Genes Affecting the Regulation of SUC2 Gene Expression by Glucose Repression in *SACCHAROMYCES CEREVISIAE*. *Genetics*. **108**, 845 (1984).
305. Stern, M., Jensen, R. & Herskowitz, I. Five SWI genes are required for expression of the HO gene in yeast. *Journal of Molecular Biology*. **178**, 853–868 (1984).
306. Cairns, B. R., Kim, Y. J., Sayre, M. H., Laurent, B. C. & Kornberg, R. D. A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proceedings of the National Academy of Sciences of the United States of America*. **91**, 1950–1954 (1994).
307. Côté, J., Quinn, J., Workman, J. L. & Peterson, C. L. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*. **265**, 53–60 (1994).
308. Hirschhorn, J. N., Brown, S. A., Clark, C. D. & Winston, F. Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Development*. **6**, 2288–2298 (1992).
309. Laurent, B. C., Treich, I. & Carlson, M. The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Development*. **7**, 583–591 (1993).
310. Pan, J. *et al.* Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. *Cell Systems*. **6**, 555–568 (2018).
311. Alpsoy, A. & Dykhuizen, E. C. Glioma tumor suppressor candidate region gene 1 (GLTSCR1) and its paralog GLTSCR1-like form SWI/SNF chromatin remodeling subcomplexes. *Journal of Biological Chemistry*. **293**, 3892–3903 (2018).

312. Wang, W. *et al.* Diversity and specialization of mammalian SWI/SNF complexes. *Genes Development*. **10**, 2117–2130 (1996).
313. Kaeser, M. D., Aslanian, A., Dong, M. Q., Yates, J. R. & Emerson, B. M. BRD7, a novel PBAF-specific SWI/SNF subunit, is required for target gene activation and repression in embryonic stem cells. *Journal of Biological Chemistry*. **283**, 32254–32263 (2008).
314. Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nature Cell Biology*. **20**, 1410–1420 (2018).
315. Shain, A. H. & Pollack, J. R. The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers. *PLoS One*. **8**, e55119 (2013).
316. Biegel, J. A. *et al.* Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. *Cancer Research*. **59**, 74–79 (1999).
317. Eaton, K. W., Tooke, L. S., Wainwright, L. M., Judkins, A. R. & Biegel, J. A. Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. *Pediatric Blood and Cancer*. **56**, 7–15 (2011).
318. Versteeg, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature*. **394**, 203–206 (1998).
319. Hasselblatt, M. *et al.* High-resolution genomic analysis suggests the absence of recurrent genomic alterations other than SMARCB1 aberrations in atypical teratoid/rhabdoid tumors. *Genes, Chromosomes and Cancer*. **52**, 185–190 (2013).
320. Kieran, M. W. *et al.* Absence of oncogenic canonical pathway mutations in aggressive pediatric rhabdoid tumors. *Pediatric Blood Cancer*. **59**, 1155–1157 (2012).
321. McKenna, E. S. *et al.* Loss of the Epigenetic Tumor Suppressor SNF5 Leads to Cancer without Genomic Instability. *Molecular Cellular Biology*. **28**, 6223–6233 (2008).
322. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. **499**, 214–218 (2013).
323. Lee, R. S. *et al.* A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *Journal of Clinical Investigation*. **122**, 2983–2988 (2012).
324. Abatangelo, L. *et al.* Comparative study of gene set enrichment methods. *BMC Bioinformatics*. **10**, 275 (2009).
325. A, S. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. **102**, 15545–15550 (2005).
326. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews Genetics*. **21**, 292–310

- (2020).
327. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).
  328. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. **464**, 768–772 (2010).
  329. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*. **354**, 769–773 (2016).
  330. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nature Communications*. **8**, 14648 (2017).
  331. Wang, X. *et al.* SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nature Genetics*. **49**, 289–295 (2017).
  332. Janssen, S. M. *et al.* BORIS/CTCF promotes a switch from a proliferative towards an invasive phenotype in melanoma cells. *Cell Death Discovery*. **6**, 1–17 (2020).
  333. Gatchalian, J. *et al.* A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nature Communications*. **9** (2018).
  334. Langer, L. F., Ward, J. M. & Archer, T. K. Tumor suppressor SMARCB1 suppresses super-enhancers to govern hESC lineage determination. *Elife*. **8** (2019).
  335. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science*. **362** (2018).
  336. Lake, B. B. *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nature Communications*. **10** (2019).
  337. Taniguchi, T., Ogasawara, K., Takaoka, A. & Tanaka, N. IRF family of transcription factors as regulators of host defense. *Annual Review of Immunology*. **19**, 623–655 (2001).
  338. Zhao, G. N., Jiang, D. S. & Li, H. Interferon regulatory factors: At the crossroads of immunity, metabolism, and disease. *Biochim. Biophys. Acta - Molecular Basis of Disease*. **1852**, 365–378 (2015).
  339. Ow, J. R., Tan, Y. H., Jin, Y., Bahirvani, A. G. & Taneja, R. Stra13 and Sharp-1, the non-grouchy regulators of development and disease. *Current Topics in Developmental Biology*. **110**, 317–338 (2014).
  340. Persson, C. U. *et al.* ARNT-dependent HIF-2 transcriptional activity is not sufficient to regulate downstream target genes in neuroblastoma. *Experimental Cell Research*. **388** (2020).

341. Wang, Y. *et al.* Mitochondria-localised ZNF1 functions as a dsRNA sensor to initiate antiviral responses through MAVS. *Nature Cell Biology*. **21**, 1346–1356 (2019).
342. Rackham, O. J. L. *et al.* A predictive computational framework for direct reprogramming between human cell types. *Nature Genetics*. **48**, 331–335 (2016).
343. Cahan, P. *et al.* CellNet: Network biology applied to stem cell engineering. *Cell*. **158**, 903–915 (2014).
344. Custers, L. *et al.* Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nature Communications*. **12**, 1–11 (2021).
345. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell*. **34**, 427–438 (2018).
346. Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nature Reviews Cancer*. **9**, 631–643 (2009).
347. Bertucci, F., Finetti, P. & Birnbaum, D. Basal Breast Cancer: A Complex and Deadly Molecular Subtype. *Current Molecular Medicine*. **12**, 96–110 (2012).
348. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*. **175**, 1272–1288 (2018).
349. Niederst, M. J. *et al.* RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nature Communications*. **6** (2015).
350. Kukkonen, K. *et al.* Chromatin and epigenetic dysregulation of prostate cancer development, progression, and therapeutic response. *Cancers*. **13** (2021).
351. Link, K. A. *et al.* BAF57 Governs Androgen Receptor Action and Androgen-Dependent Proliferation through SWI/SNF. *Molecular Cellular Biology*. **25**, 2200–2215 (2005).
352. Link, K. A. *et al.* Targeting the BAF57 SWI/SNF subunit in prostate cancer: A novel platform to control androgen receptor activity. *Cancer Research*. **68**, 4551–4558 (2008).
353. Jin, M. L., Kim, Y. W. & Jeong, K. W. BAF53A regulates androgen receptor-mediated gene expression and proliferation in LNCaP cells. *Biochemical and Biophysical Research Communications*. **505**, 618–623 (2018).
354. Cyrta, J. *et al.* Role of specialized composition of SWI/SNF complexes in prostate cancer lineage plasticity. *Nature Communications*. **11**, 1–16 (2020).
355. Lessard, J. *et al.* An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron*. **55**, 201–215 (2007).

356. Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature*. **460**, 642–646 (2009).
357. Labrecque, M. P. *et al.* Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *Journal of Clinical Investigation*. **129**, 4492–4505 (2019).
358. Minocha, S. & Herr, W. Cortical and Commissural Defects Upon HCF-1 Loss in Nkx2.1-Derived Embryonic Neurons and Glia. *Developmental Neurobiology*. **79**, 578–595 (2019).
359. Zhang, Z. *et al.* Loss of CHD1 Promotes Heterogeneous Mechanisms of Resistance to AR-Targeted Therapy via Chromatin Dysregulation. *Cancer Cell*. **37**, 584-598 (2020).
360. Rodrigues, L. U. *et al.* Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. *Cancer Research*. **75**, 1021–1034 (2015).
361. Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature*. **460**, 863–868 (2009).
362. Abeshouse, A. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. **163**, 1011–1025 (2015).
363. Valencia, A. M. *et al.* Recurrent SMARCB1 Mutations Reveal a Nucleosome Acidic Patch Interaction Site That Potentiates mSWI/SNF Complex Chromatin Remodeling. *Cell*. **179**, 1342-1356 (2019).
364. McBride, M. J. *et al.* The nucleosome acidic patch and H2A ubiquitination underlie mSWI/SNF recruitment in synovial sarcoma. *Nature Structural & Molecular Biology*. **27**, 836–845 (2020).
365. Liu, J. *et al.* Intrinsic disorder in transcription factors. *Biochemistry*. **45**, 6873–6888 (2006).
366. Schapira, M., Calabrese, M. F., Bullock, A. N. & Crews, C. M. Targeted protein degradation: expanding the toolbox. *Nature Reviews Drug Discovery*. **18**, 949–963 (2019).
367. Bleu, M. *et al.* PAX8 activates metabolic genes via enhancer elements in Renal Cell Carcinoma. *Nature Communications*. **10**, 3739 (2019).
368. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell*. **170**, 577-592 (2017).
369. Bouchard, M., Souabni, A., Mandler, M., Neubüser, A. & Busslinger, M. Nephric lineage specification by Pax2 and Pax8. *Genes Developmental*. **16**, 2958–2970 (2002).
370. Le Magnen, C., Shen, M. M. & Abate-Shen, C. Lineage Plasticity in Cancer Progression and Treatment. *Nature Reviews Clinical Oncology*. **17**, 360–371 (2017).

371. Ding, L. *et al.* Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell*. **173**, 305-320 (2018).
372. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. *Nature Reviews Clinical Oncology*. **14**, 611–629 (2017).
373. Kudo-Saito, C., Shirako, H., Takeuchi, T. & Kawakami, Y. Cancer Metastasis Is Accelerated through Immunosuppression during Snail-Induced EMT of Cancer Cells. *Cancer Cell*. **15**, 195–206 (2009).
374. Kinouchi, T., Kotake, T., Mori, Y. & Abe, T. Human renal cell carcinoma: Establishment and characterization of a new cell line (OS-RC-2). *In Vitro Cellular & Developmental Biology*. **21**, 195–199 (1985).
375. Gossage, L., Eisen, T. & Maher, E. R. VHL, the story of a tumour suppressor gene. *Nature Reviews Cancer*. **15**, 55–64 (2015).
376. Ku, S. Y. *et al.* Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science*. **355**, 78–83 (2017).
377. Zhou, P. *et al.* A Three-Way Combinatorial CRISPR Screen for Analyzing Interactions among Druggable Targets. *Cell Reports*. **32** (2020).
378. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature*. **460**, 1132–1135 (2009).
379. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature*. **460**, 1140–1144 (2009).
380. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature*. **460**, 1149–1153 (2009).
381. Chappell, J. & Dalton, S. Roles for MYC in the establishment and maintenance of pluripotency. *Cold Spring Harbor Perspectives in Medicine*. **3** (2013).
382. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. **126**, 663–676 (2006).
383. Dardenne, E. *et al.* N-Myc Induces an EZH2-Mediated Transcriptional Program Driving Neuroendocrine Prostate Cancer. *Cancer Cell*. **30**, 563–577 (2016).
384. Beltran, H. *et al.* Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discovery*. **1**, 487–495 (2011).
385. Berger, A. *et al.* N-Myc-mediated epigenetic reprogramming drives lineage plasticity in advanced prostate cancer. *Journal of Clinical Investigation*. **129**, 3924–3940 (2019).

386. Teitell, M. A. & Pandolfi, P. P. Molecular genetics of acute lymphoblastic leukemia. *Annual Review of Pathology: Mechanisms of Disease*. **4**, 175–198 (2009).
387. Ha, V. L. *et al.* The T-ALL related gene BCL11B regulates the initial stages of human T-cell differentiation. *Leukemia*. **31**, 2503–2514 (2017).
388. Rajbhandari, P. *et al.* Cross-cohort analysis identifies a TEAD4–MYCN positive feedback loop as the core regulatory element of high-risk neuroblastoma. *Cancer Discovery*. **8**, 582–599 (2018).
389. Shi, H. *et al.* ARID1A loss in neuroblastoma promotes the adrenergic-to-mesenchymal transition by regulating enhancer-mediated gene expression. *Science Advances*. **6**, 3440 (2020).
390. A cell identity switch allows residual BCC to survive Hedgehog pathway inhibition. *Nature*. **562**, 429–433 (2018).
391. A slow-cycling LGR5 tumour population mediates basal cell carcinoma relapse after therapy. *Nature*. **562**, 434–438 (2018).
392. Batie, M. *et al.* Hypoxia induces rapid changes to histone methylation and reprograms chromatin. *Science*. **363**, 1222–1226 (2019).

SMARC-be-done...

