Investigation of RUNX1
in Normal Prostate and Prostate Cancer

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the faculty of Biology, Medicine and Health

2021

Renaud Mével

Cancer Research UK Manchester Institute

Table of contents

Word count: 67,668 words.

List of figures

List of tables

Supplementary data

Supplementary data files can be found at:

https://www.dropbox.com/sh/m2u42ez1hqve93p/AADQFr3PmuJx0YChEBAygaUSa?dl=0

Supplementary data 1. Clinical annotations of Manchester's PCa TMA.

Supplementary data 2. Clinical annotations of Manchester's RP TMA.

Supplementary data 3. Clinical annotations of the BPH samples.

Supplementary data 4. Sample metadata of the Adult scRNA-seq dataset.

Supplementary data 5. Sample metadata of the UGS explant scRNA-seq dataset.

Supplementary data 6. Lists of differential expressed genes.

Supplementary data 7. Gene lists used for the analysis of gene set activity.

List of abbreviations

Abbreviation	Meaning
3D	3-Dimensions
4-OHT	4-Hydroxytamoxifen
ADT	Androgen Deprivation Therapy
AGM	Aorta-Gonad-Mesonephros
ALL	Acute Lymphoblastic Leukaemia
AMACR	Alpha-Methylacyl-CoA Racemase
AML	Acute Myeloid Leukaemia
Amp	Ampullary gland
AP	Anterior Prostate
AR	Androgen Receptor
ATAC	Assay for Transposase-Accessible Chromatin
AUC	Area Under the Curve
AWERB	Animal Welfare and Ethics Review Body
BCR	Biochemical Recurrence
BMP	Bone Morphogenetic Protein
BP	Biological Processes
BPH	Benign Prostatic Hyperplasia
BrdU	5-bromo-2-deoxyuridine
BSA	Bovine Serum Albumin
C	Cluster
CARBs	Castration-Resistant Bmi1 expressing cells
CARLIN	CRISPR array repair lineage tracing
CARNs	Castration-Resistant Nkx3-1 expressing cells
Cas	Castrated
CBFβ	Core-Binding Factor subunit β
ccRC	clear cell Renal Cell Carcinoma
CD	Cluster of Differentiation
CDH1	Cadherin E
CFUe	Erythroid Colony-Forming Unit
CLP	Common Lymphoid Progenitor
CODEX	CO-Detection by indEXing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRUK	Cancer-Research UK
cT	Clinical Tumour stage
D	Day
DAPI	4',6-diamidino-2-phenylindole
DHT	Dihydrotestosterone
DKFZ	Deutsches Krebsforschungszentrum
DLP	Dorso-Lateral Prostate
DMEM	Dulbecco's Modified Eagle's Medium
dNTP	Deoxyribose Nucleoside Tri-Phosphate
DP	Dorsal Prostate

E	Embryonic
EGF	Epidermal Growth Factor
EHT	Endothelial-to-Haematopoietic Transition
EMMA	European Mouse Mutant Archive
EMT	Epithelial-to-Mesenchymal Transition
EPCAM	Epithelial cell adhesion molecule
ER	Estrogen Receptor
ER1	Epitope Retrieval 1
ER2	Epitope Retrieval 2
ERAD	Endoplasmic-reticulum-associated protein degradation
ERT2	Tamoxifen-inducible Cre-estrogen receptor
EYFP	Enhanced Yellow Fluorescent Protein
F	Flox
FACS	Fluorescence-Activated Cell sorting
FBS	Foetal Bovine Serum
FCS	Foetal Calf Serum
FFPE	Formalin-Fixed Paraffin-Embedded
FITC	Fluorescein isothiocyanate
FSC	Forward Side Scatter
G	Gleason
GEM	Gel Beads in EMulsion
GEMM	Genetically Engineered Mouse Model
GFP	Green Fluorescent Protein
GGG	Gleason Grade Group
GMP	Granulocyte/Monocyte Progenitor
GO	Gene Ontology
GS	Gleason Score
GUCR	Genito-Urinary Cancer Research group
H&E	Haematoxylin and Eosin
HCA	Human Cell Atlas
HF	Hair Follicles
HFSCs	Hair Follicles Stem Cells
Hi	High
HMWCK	High Molecular Weight Cytokeratins
HRP	Horseradish Peroxidase
HSC	Haematopoietic Stem Cells
HSPCs	Haematopoietic Stem and Progenitor Cells
ID	Identifier
IF	Immunofluorescence
IHC	Immunohistochemistry
Int	Intact
ISH	*In situ* hybridisation
ISUP	International Society of Urological Pathology
K	Keratin
KO	Knock-out
L/D	Live/Dead

LGR	Leucine-rich repeat-containing G-protein coupled receptor
LH	Luteinizing Hormone
LHRH	Luteinising Hormone Release Hormone
LMO	Lipid-Modified Oligonucleotide
LMPP	Lymphoid-primed Multipotent Progenitor
LN	Lymph Node
LNM	Lymph Node Metastasis
Lo	Low
LP	Lateral Prostate
LSL	loxP-STOP-loxP cassette
LTS	Luminal TROP2 SCA-1
Lum	Luminal
M	Metastasis
MkP	Megakaryocyte Progenitor
mCRPC	metastatic Castration-Resistant Prostate Cancer
MDS	Myelodysplastic Syndrome
mESC	mouse Embryonic Stem Cells
MI	Manchester Institute
MOI	Multiplicity of Infection
MPP	Multipotent Progenitor
MRI	Magnetic Resonance Imaging
mTmG	membrane-targeted TdTomato / membrane-targeted GFP
NGS	Next-Generation Sequencing
NLS	Nuclear Localisation Signal
NP-40	Nonidet P-40
nt	Nucleotide
OFC	Organoid Formation Capacity
P	Postnatal
P1	Distal promoter
P2	Proximal promoter
PAGA	Partition-based graph abstraction
panCK	pan-Cytokeratin
PAP	Prostatic Acid Phosphatase
Pb	Probasin
PBS	Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline 0.05% Tween-20
PC	Principal Components
PCa	Prostate Cancer
PCR	Polymerase Chain Reaction
PDXs	Patient-Derived Xenografts
PEI	Polyethylenimine
PET	positron emission tomography
PFA	Paraformaldehyde
PIN	Prostatic Intraepithelial Neoplasia
PLCs	Proximal Luminal Cells
PreCFUe	Pre-erythroid Colony Forming Unit

PreMegE	Pre-Megakaryocyte/Erythroid progenitor
PrU	Peri-urethral
PSA	Prostate Specific Antigen
PSCA	Prostate Stem Cell Antigen
PSE	P1-*Runx1*:GFP reporter line
PSN	P2-*Runx1*:RFP reporter line
pT	Pathological Tumour stage
PTEN	Phosphatase and tensin homolog
QIBC	Quantitative Image-Based Cytometry
QC	Quality Control
RFP	Red Fluorescent Protein
RGF BME 2	Reduced Growth Factor Basement Membrane Extract Type 2
Rgn	Regenerated
RHD	Runt Homology Domain
RNA-seq	RNA-sequencing
RP	Radical prostatectomy
RT-qPCR	Real-Time quantitative Polymerase Chain Reaction
RUNX	Runt-related transcription factor
SCA-1	Stem Cell Antigen-1
scRNA-seq	single-cell RNA-sequencing
SNP	Single-Nucleotide Polymorphism
SSC	Side Scatter
SV	Seminal Vesicles
Tam	Tamoxifen
TBS	Tris-Buffered Saline
TBS-T	Tris-Buffered Saline 0.05% Tween-20
TCGA	The Cancer Genome Atlas
TF	Transcription Factor
TGF-β	Transforming Growth Factor β
TMA	Tissue Micro-Array
TRAMP	Transgenic Adenocarcinoma of the Prostate
TRUS	Transrectal ultrasound
TRUSBx	Transrectal ultrasound guided needle core biopsies
TSA	Tyramide Signal Amplification
TURP	Trans-Urethral Resection of the Prostate
UGE	Urogenital Epithelium
UGS	Urogenital Sinus
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Index
UTR	Untranslated Region
VP	Ventral Prostate
WNT	Wingless/Integrated pathway
WT	Wild-Type

Abstract

Prostate cancer (PCa) is the most commonly diagnosed cancer in men, and it remains challenging to accurately stratify patients diagnosed with indolent and aggressive forms of PCa, such as castration-resistant PCa. The lack of prognosis biomarkers represents therefore a central challenge to avoid a worsening problem of over diagnosis and treatment. In the recent years, there has been a growing body of literature involving the master regulator of haematopoiesis RUNX1 in non-haematopoietic cancers, including hormone-regulated epithelial tissues, suggesting that this transcription factor could be broadly implicated in the biology and pathology of epithelial tissues. These intriguing new findings led us to propose that RUNX1 plays a role in normal prostate homeostasis and prostate tumorigenesis.

To explore this hypothesis, RUNX1 expression was characterised in the mouse prostate. RUNX1 was expressed in a specific subpopulation of proximal luminal cells (PLCs), enriched in the peri-urethral region of the developing and adult prostate, distinct from the previously identified NKX3.1+ luminal castration resistant cells. Functionally, RUNX1+ PLCs are not committed to the secretory function and display facultative stem cell behaviour in organoid assays. Single-cell RNA-sequencing profiling and genetic lineage tracing revealed that RUNX1+ PLCs are unaffected by androgen deprivation, and do not contribute to the regeneration of the distal luminal compartments. Furthermore, a transcriptionally similar RUNX1+ population emerges at the onset of embryonic prostate specification to populate the proximal region of the ducts. Thus, RUNX1+ PLCs constitutes an intrinsic castration-resistant and self-sustained lineage that emerges early during prostate development.

Finally, the expression of RUNX1 was investigated in patient samples to evaluate whether RUNX1 could be exploited as a clinical marker of PCa. In benign prostate glands, RUNX1 was low in secretory luminal cells but higher in atrophic epithelia. RUNX1 was rarely expressed in tumours, but a limited number of high-grade tumours exhibited very high RUNX1 levels. However, overall survival analyses revealed poor prognostic performance of RUNX1 used as a single marker or in combination with NKX3.1.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

I would like to begin by thanking Professor Georges Lacaud for giving me the opportunity to undertake my PhD under your supervision. I am particularly grateful for your continuous guidance and support, transmitting your enthusiasm to approach new techniques and models, and for allowing me to attend fascinating conferences. I am also thankful for the trust and freedom you have given me to explore many different paths, making this experience truly worthwhile.

I would like to convey my thanks to Dr Valerie Kouskoff for your help and guidance at the start of my PhD. I also want to thank all the - past and present - members of the Stem Cell Biology and Stem Cell Haematopoiesis groups, for being a such a great bunch of people to work with. I wish to especially thank Zaki for introducing me to the R world, as well as Michael, Julia, Anne, Divya, Neo, Rahima, Josh, Ewan, Ahmad, Roshana and Kiran for sharing your expertise with me, helping out when I needed it, discussing science and beyond, and all your support! Thanks also to Thomas for all your help. You have all contributed to make this experience so enjoyable and truly memorable.

I would like to thank my co-supervisor Dr Esther Baena for providing regular help and constructive recommendations throughout the course of my PhD, as well as all the members of the Prostate Oncobiology group - past and present - including Adnan and Valentina. Specials thanks go to Ivana for being tremendously helpful with *in vivo* experiments, as well as many other aspects of my PhD, it has been a great adventure to share from start to finish.

I also wish to thank members of the biobank and the GUCR group, including Professor Noel Clarke, Dr Mick Brown and Claire Hart for providing access to human patient samples and clinical annotations. Special thanks go to Dr. Ashwin Sachdeva for your assistance in dealing with clinical databases, discussing multiplexing, and

sharing your clinical knowledge and experience with me. Thanks also to Dr Pedro Oliveira for helping out with histopathological annotations of human samples.

I would like to thank and express my appreciation to all core facilities of CRUK MI, including the Histology, MBCF, Advanced Imaging, Flow-cytometry, BRU and Incubator teams. I would particularly like to thank Garry and Caron for teaching me a bunch of techniques and for always being available to troubleshoot experiments. I also thank Emma, Usman, Catherine, Isabel, Heather, Wolfie, John, Jeff, Toni, Jen, Kim and Lauren. I could not have managed without your assistance.

I wish to express my indebted gratitude to Professor Karen Blyth for being such a fantastic collaborator. You have always been welcoming when I visited the Beatson Institute, were consistently encouraging me, and immensely supportive with all my experimental plans. In addition, I would like to thank Susan Mason and Laura Galbraith for being incredibly helpful with experiments conducted in the Beatson Institute.

I would like to take this opportunity to thank my parents, for always being there for me, believing in me, as well as for showing unconditional love and support throughout my studies and in all my endeavours.

I am eternally grateful to Alice for brightening my days and having made this journey so much easier than I could have ever wished for. Your love, support, patience and understanding have helped me stay focus and motivated regardless of how (un)successful things turned out in the lab. Finally, I am immensely thankful to Sofia for making the write up of this thesis seem such a pleasant task!

Publications

Mevel, R., Draper, J. E., Lie-a-Ling, M., Kouskoff, V., & Lacaud, G. (2019). RUNX transcription factors: Orchestrators of development. *Development*, 146(17), dev148296. https://doi.org/10.1242/dev.148296

Lie-A-Ling, M.*, **Mevel, R.***, Patel, R., Blyth, K., Baena, E., Kouskoff, V., & Lacaud, G. (2020). RUNX1 Dosage in Development and Cancer. *Molecules and Cells*, 43(2), 126–138. https://doi.org/10.14348/molcells.2019.0301
* equal contribution

Mevel, R., Steiner, I., Mason, S., Galbraith, L. C., Patel, R., Fadlullah, M. Z., Ahmad, I., Leung, H. Y., Oliveira, P., Blyth, K., Baena, E., & Lacaud, G. (2020). RUNX1 marks a luminal castration-resistant lineage established at the onset of prostate development. *ELife*, 9, e60225. https://doi.org/10.7554/eLife.60225

Chapter 1 Introduction

1.1. The prostate gland

1.1.1. Anatomy

The prostate is an exocrine accessory gland of the male mammalian urogenital system. In men, the prostate is often described as a cone-shaped walnut-sized gland, weighing 10 to 20 grams on average (Leissner and Tisell, 1979). Anatomically, it is located at the neck of the urinary bladder and surrounding the urethra. The human prostate is enclosed by an outer layer of fibromuscular tissue in continuity with the prostatic stroma, also known as prostatic “capsule” (Ayala et al., 1989).

John McNeal divided the prostate in four zones based on variations in incidence of prostate diseases, as well as subtle histological differences: the central zone, the transitional zone, the peripheral zone, and the anterior fibromuscular stroma (McNeal, 1981; Selman, 2011) (Figure 1.1). The peripheral zone occupies the majority of the prostate volume (~70%) and is the predominant site of prostate adenocarcinomas and inflammation (De Marzo et al., 2007). The transitional zone accounts for only 5-10% of the prostate volume in young adults, but it is frequently implicated in benign prostatic hyperplasia (BPH), a non-malignant enlargement of the prostate (McNeal et al., 1988). The central zone surrounds the ejaculatory ducts and makes up ~25% of the prostate. The fibromuscular stroma is situated in the anterior part of the gland and composed of a thick non-glandular tissue.

In addition to these anatomical differences, recent work has confirmed the existence of minor variation in cell type composition between the different zones (Henry et al., 2018). Along these lines, tumours arising in the different prostate zones express distinct gene signatures (Sinnott et al., 2015). Interestingly, tumour aggressiveness also appears to vary according to the zonal anatomy of the prostate. Tumours of the transitional zone are often more indolent than peripheral zone tumours (Lee et al., 2015; Shannon et al., 2003). In contrast, tumours of the central

zone are particularly rare but thought to be particularly aggressive (Cohen et al., 2008).

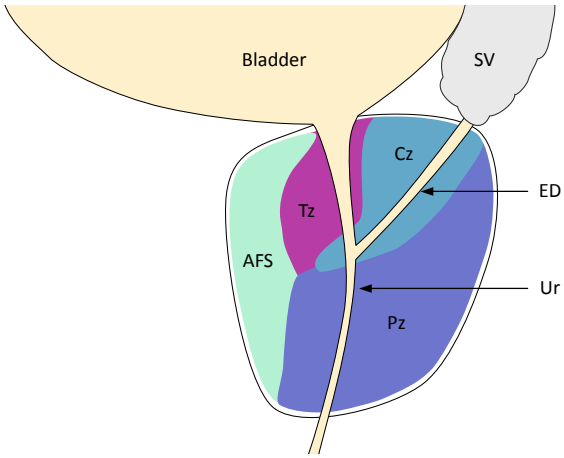

Figure 1.1. Zonal anatomy of the human prostate.
Cz: Central zone; Tz: Transitional zone; Pz: Peripheral zone; ED: Ejaculatory duct; SV: Seminal vesicle; AFS: Anterior fibromuscular stroma.

1.1.2. Functions

The main function of the prostate gland consists in the production of a slightly alkaline seminal fluid by a complex ductal network of secretory epithelial cells (Verze et al., 2016). Prostatic fluid accounts for about 30% of the semen's volume and drains into the prostatic urethra. There, it merges with secretions produced by the seminal vesicles transported via the ejaculatory ducts. In addition to its secretory function, an important role of the prostate is carried out by the dense fibromuscular tissue, which contracts to expel semen during ejaculation.

Prostate secretions contain a variety of electrolytes such as high levels of zinc (Kavanagh, 1985), and strong proteolytic enzymes important for the liquefaction of seminal coagulum, including prostatic acid phosphatase (PAP), fibrinolysin and prostate specific antigen (PSA). PSA is an androgen-regulated glycoprotein, secreted by glandular prostate epithelial cells into the prostatic lumen. It cleaves semenogelins into lower molecular weights proteins to prevent the formation of a gel matrix (Balk et al., 2003; Lilja et al., 1987). Alkanisation of semen by prostatic secretions and subsequent liquefaction of seminal coagulum are both essential to

sustain and promote spermatozoa survival and motility in the female genital tract during fertilisation (Verze et al., 2016).

1.1.3. Histology

The prostate is composed of epithelial and stromal compartments. The mature glandular prostate epithelium consists in three main epithelial cell types: luminal, basal and neuroendocrine cells (Shen and Abate-Shen, 2010) (Figure 1.2). Luminal cells form a continuous layer of polarised columnar cells around the gland lumen and produce prostatic secretions. These cells are characterised by the expression of cytoplasmic keratin (K) 8 and 18, as well as strong levels of nuclear androgen receptor (AR). The supportive layer of basal cells is located between luminal cells and the surrounding stroma, and marked by the expression of K5, K14, and p63. Neuroendocrine cells are far less abundant than basal and luminal cells, and they can be identified by the expression of synaptophysin and chromogranin-A, but their precise role in the prostate remain unknown (Abrahamsson, 1999). The stroma is composed of several cell types, including smooth muscle cells, fibroblasts, myofibroblasts, nervous fibres, vascular, perivascular and haematopoietic cells (Toivanen and Shen, 2017; Wegner et al., 2017).

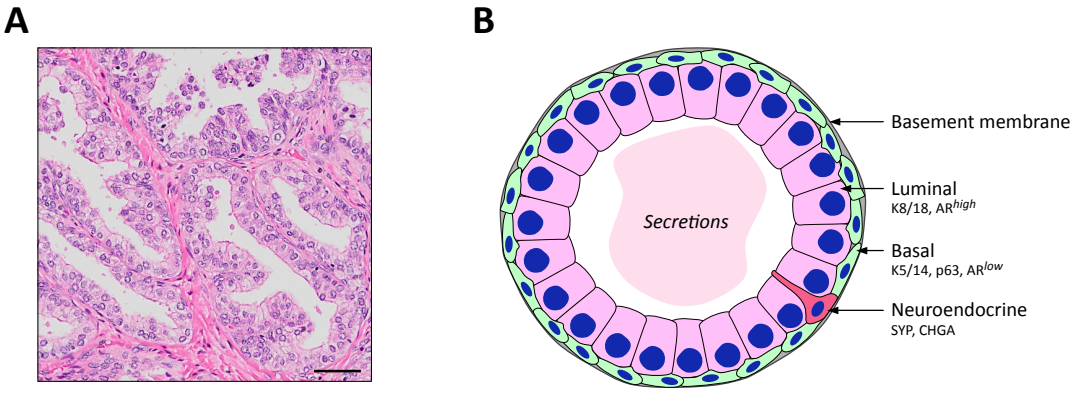

Figure 1.2. Structure of the prostate epithelial gland.
(A) Representative H&E staining of human prostate glands. (B) The normal prostate gland is composed of basal cells characterised by the expression of K5/14, p63 and low levels of AR, while luminal cells express mainly K8/18 and high levels of AR. Rare neuroendocrine cells expressing Synaptophysin (SYP) or Chromogranin A (CHGA) can occasionally be observed. Glands are surrounded by a basement membrane constituted of laminins and collagen.

1.1.4. Development

The development of the prostate starts during embryogenesis and is completed during puberty with the formation of a mature gland. In humans, testosterone production by foetal testes has been detected around 6 weeks of gestation (Siiteri and Wilson, 1974). Prostate development initiates between 9 and 10 weeks of gestation from the epithelium of the urogenital sinus (UGS), an endoderm derived tubular and transient structure located between the bladder at its rostral end and the penile urethra on the opposite side (Dauge et al., 1986; Timms, 2008). At this stage, the urogenital epithelium (UGE) invaginates into the surrounding mesenchyme to generate the first prostate buds. While the precise steps leading to prostate budding emergence remain unclear, androgens present in both epithelial and mesenchymal compartments are thought to play a central role (Toivanen and Shen, 2017). Interestingly, experiments performed *in vitro* indicate that developing mouse female UGS can differentiate into prostatic structures upon androgen stimulation, highlighting the central role played by male androgens (Keil et al., 2012). During early developmental stages, the epithelium is mainly composed of poorly differentiated epithelial cells often co-expressing the classical basal and luminal markers (Cunha et al., 2018). Occasional neuroendocrine cells have also been observed from 13 weeks of gestation, but these are believed to derive from the neural crest rather than the urogenital epithelium (Szczyrba et al., 2017). Nascent epithelial buds elongate and branch into the mesenchyme in response to paracrine signals from mesenchymal AR signalling (Cunha, 1975). Following branching morphogenesis, ductal canalisation of solid epithelial cords gradually gives rise to prostatic lumens, which is accompanied with differentiation in functional basal and luminal lineages (Marker et al., 2003). In humans, prostate development is mostly complete by birth, and undergoes a final growth episode in response to a surge in testosterone at the time of puberty (Prins and Lindgren, 2015).

1.1.5. Comparison between the human and mouse prostate

The mouse is the most used animal model to study mammalian organogenesis and human malignancies, including prostate morphogenesis and diseases. It is therefore important to understand similarities and differences between these species. In mice, prostate development follows similar steps than in humans. Prostate budding starting around embryonic day (E) 16.5, however, most branching morphogenesis and growth occur postnatally to be completed by the end of the puberty (Bhatia-Gaur et al., 1999; Sugimura et al., 1986a; Tika et al., 2019). Anatomically, the gross appearance of the mouse prostate is more heterogeneous than the human prostate. While the human prostate has been subdivided in zones which are adjacent to one another (see section 1.1.1), the mouse prostate is organised in four distinct pairs of lobes surrounding the urethra: anterior (AP), dorsal (DP), lateral (LP) and ventral (VP) prostate. The anterior lobe is the largest lobe, nearly attached to the seminal vesicles. The lateral lobe is located between the ventral and the dorsal lobes and often found intermingled with these as a result of their proximity. The dorsal and lateral lobes are often combined and referred to as the dorsolateral prostate (DLP).

Histologically, the mouse and human prostates have a similar ductal architecture, but the mouse prostate contains proportionally less basal and neuroendocrine cells compared with luminal cells (Shappell et al., 2004). In mice, each lobe has distinct branching patterns and slight histological differences (Figure 1.3). The AP shows typical papillary infolding, with a relatively flat layer of luminal cuboidal and columnar epithelial cells lining lumens filled with homogenous eosinophilic secretion. Secretory cells have large eosinophilic and granular cytoplasms, with nuclei centrally to basally located. The epithelium of the DP has a similar appearance to the AP, although it is organised in smaller acini, with less pronounced infolding, and occasional tufting patterns. The LP epithelium consists of a flat layer of cuboidal and columnar luminal cells rarely infolded. LP acini are of different sizes and contain eosinophilic secretions. Nuclei of secretory LP cells are generally basally located within lightly granular cytoplasms. The VP has less

infolding compared with other lobes, though frequent epithelial tufting can be observed. VP acini are lined by cuboidal and tall columnar luminal cells with basally located nuclei. In contrast to other lobes, VP glands are filled with paler secretions. The fibromuscular layer surrounding glands of the AP and DP is slightly more important than for the VP and LP. However, the overall thin stromal layer contrasts with the abundant contractile fibromuscular tissue surrounding human prostate glands. Finally, a distinct non-prostatic lobular structure termed ampullary gland surrounds mouse vas deferens while strongly resembling prostate tissue. In contrast to proper prostate lobes, lumens of the ampullary gland present eosinophilic secretions with a “swiss cheese” appearance (Shappell et al., 2004), and this structure is generally ignored in the context of prostate research.

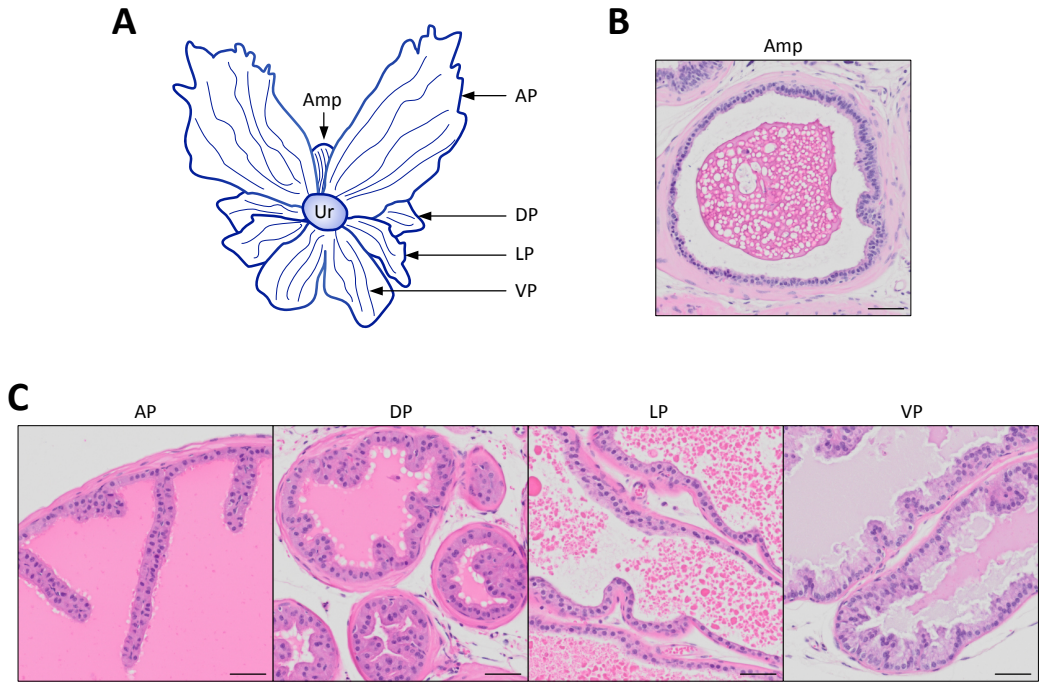

Figure 1.3. Lobular anatomy of the mouse prostate.
(A) Schematic of the mouse prostate lobes. AP: Anterior Prostate, DP: Dorsal Prostate, LP: Lateral Prostate, VP: Ventral Prostate, Amp: Ampullary gland, Ur: Urethra. (B-C) H&E staining representative of the histology of (B) the ampullary gland, and (C) individual mouse prostate lobes as in A. Scale bar: 50 µm.

1.1.6. Androgens

Androgens are sex steroid hormones that regulate the development of the prostate and other primary male sex organs, as well as male secondary sex

characteristics. In men, about 95% of androgens are synthesised in the testes, while the remainder is produced by adrenal glands. Synthesis of testosterone, the primary androgen, is tightly regulated by the hypothalamic–pituitary–gonadal axis. At low testosterone levels, pulses of luteinising hormone release hormone (LHRH) are released by the hypothalamus, which subsequently stimulates the pituitary gland to release luteinizing hormone (LH). The latter stimulates testosterone synthesis and release by the Leydig cells within the testes. Increasing levels of testosterone induce a negative feedback loop to suppress LHRH and LH release by the hypothalamus and pituitary respectively.

In the prostate, testosterone is reduced into 5-alpha-dihydrotestosterone (DHT) by the enzyme 5-alpha reductase (Marks, 2004). DHT is a considerably more potent AR agonist than testosterone and allows the maintenance of a high concentration of androgens in the prostate to sustain its function, including luminal cell activity. Mechanistically, testosterone binds AR in the cytoplasm either directly or after reduction into DHT, which in turns induces AR translocation to the nucleus. AR acts as a DNA-binding transcription factor to regulate gene expression in a ligand-dependent manner to initiate male sexual development and differentiation (Tan et al., 2015). Thus, AR plays major roles in the development and homeostasis of the prostate, and deregulation of AR are frequently observed in prostate malignancies (see section 1.2.6).

1.2. Prostate cancer

1.2.1. Incidence and outcomes

Prostate cancer (PCa) is the most commonly diagnosed cancer in the UK, with about 47,500 new cases and more than 11,500 deaths every year (Prostate Cancer UK, “Facts and figures”, 2018). Over the last decades, the incidence of PCa has been rising, in part due to the development and availability of PSA screening, the global increase in life expectancy as well as environmental and lifestyle factors such as obesity and low physical activity (Pernar et al., 2018). Despite a high incidence rate, survival outcomes are improving, and currently, about 80% of PCa patients will

survive their disease for at least 10 years (Cancer Research UK, 2020). But while early diagnosis and treatment is associated with good prognosis, late stage metastatic PCa remains a deadly disease with a mortality rate of ~30% after 5 years (SEER Cancer Statistics Review, 1975-2016).

1.2.2. Risk factors

Ageing is one of the strongest risk factors associated with PCa. In fact, the association between increasing age and the risk of developing PCa is stronger than for any other cancer, with an odds ratio of 0.2 up to 49 years old, 1.8 between the age of 50 and 59, 4.7 between the age of 60 and 69 and 8.2 above 70 years old (Siegel et al., 2020). While PCa is mainly diagnosed in men above the age of 50, numerous studies have identified healthy young men with histologic foci of PCa leading to the assumption that cancer could initiate at an early age and remain latent (Sakr et al., 1996; Salinas et al., 2014). Ageing is also associated with the development of additional prostate diseases which may influence the development of PCa, such as prostate inflammation (De Marzo et al., 2007).

Hereditary PCa represents only a minor proportion of all PCa cases. Yet, family history has been associated with increased risk of PCa diagnosis (Bratt et al., 2016; Kiciński et al., 2011). The most studied example of germline mutations in PCa affect *BRCA1* and *BRCA2* which have both been associated with increased risk of developing PCa (Carter et al., 2019; Pritchard et al., 2016; Taylor et al., 2019). *BRCA1*/*2* carriers are associated with more aggressive diseases (Castro et al., 2013), and *BRCA2* mutation carriers present a two to five times higher risk of PCa than the remainder male population (Nyberg et al., 2020). Other germline mutations have been implicated in PCa, such as Homeobox B13 (*HOXB13*) (Ewing et al., 2012), *ATM* (Na et al., 2017) Checkpoint kinase 2 (*CHECK2*) (Seppälä et al., 2003).

Ethnicity and geography are also associated with variations in PCa incidence and mortality. At the geographic level, PCa incidence is higher in Oceania, Northern America and Europe compared with Africa and Asia (Culp et al., 2020). In contrast, African American and Caribbean men have a higher incidence and mortality rate compared with Caucasians and Asians. Overall, it is difficult to evaluate whether

biological differences between these populations truly account for variations in PCa outcomes over factors such as diet, socioeconomic backgrounds and access to care (Khani et al., 2014).

1.2.3. The development of prostate cancer

PCa is thought to develop in a stepwise manner, via the emergence of a pre-cancerous but non-malignant lesion known as prostatic intra-epithelial neoplasia (PIN) (Shen and Abate-Shen, 2010). Histologically, PIN is characterised by a moderate reduction in basal cells number, luminal epithelial hyperplasia, nuclear enlargement and atypia, and cytoplasmic hyperchromasia (Bostwick, 1989; McNeal and Bostwick, 1986). These lesions may remain indolent for several years or decades, before progressing to high-grade PIN lesions marked by an increase in proliferation but without invasion of the basement membrane. High-grade PIN lesions are usually observed in the peripheral zone of the prostate and can further evolve in invasive PCa.

Localised low and intermediate risk PCa are usually treated using radical therapies such as radical prostatectomy and radiotherapy with favourable long-term outcomes (Litwin and Tan, 2017). However, more advanced forms of PCa including metastatic PCa are managed with androgen-deprivation therapy (ADT) which relies on androgen-dependency of prostate cancer cells to reduce tumour burden. While initially effective, ADT is associated with elevated risks of progression to lethal hormone-refractory or castration resistant prostate cancer (CRPC). The main stages of progression from a normal epithelium towards metastatic castration resistant PCa are summarised in Figure 1.4.

The heterogeneity and multifocality of PCa renders challenging the initial diagnosis and classification into different histopathological subtypes and has direct consequences on subsequent clinical management of the disease (Boutros et al., 2015; Cancer Genome Atlas Research Network, 2015; Carm et al., 2019; Tomlins et al., 2015). Most PCa exhibit a luminal phenotype, with the expression of high levels of AR, and are classified as adenocarcinomas. Prostate adenocarcinomas have a luminal phenotype and can be confirmed by the absence of basal cells and

associated markers (p63, K5/14) and an increase of alpha-methylacyl-CoA racemase (AMACR) expression (Humphrey, 2007; Luo et al., 2002). Other rare forms of PCa include ductal adenocarcinoma, mucinous carcinomas, signet ring carcinomas, and neuroendocrine carcinomas (Grignon, 2004). This classification relies mainly on histopathological evaluation, but recent and on-going genomic studies aim at identifying molecularly defined subtypes of PCa in order to better stratify patient outcomes and treatment responses (Arora and Barbieri, 2018; Boutros et al., 2015; Cancer Genome Atlas Research Network, 2015; Tomlins et al., 2015).

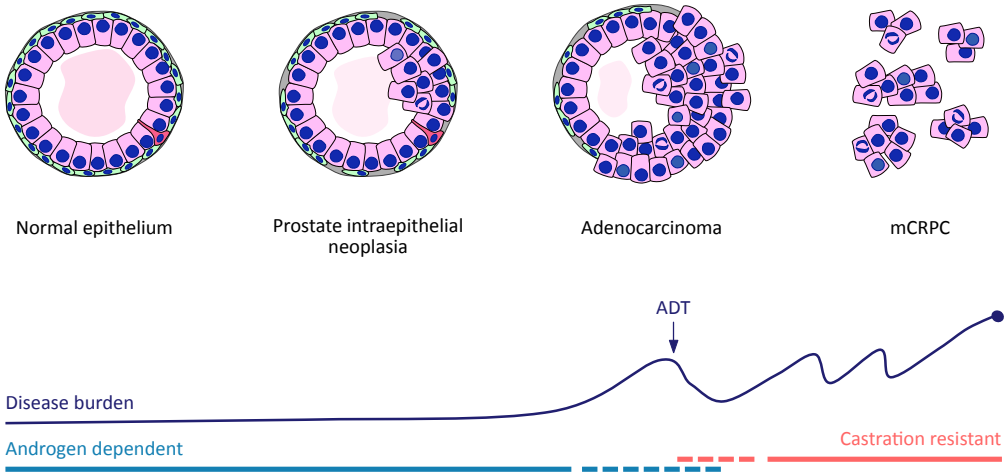

Figure 1.4. Principal stages of PCa development.
Prostatic Intraepithelial Neoplasia (PIN) state consists of abnormal prostatic glands and is considered a precursor for most adenocarcinoma of the prostate. During carcinogenesis, glands become gradually less differentiated, with a loss of the basal layer, an increase in mitotic cells, focal hyperplasia, and hyperchromatic pleomorphic nuclei. Prostate adenocarcinoma can invade through the basement membrane, allowing subsequent dissemination. Initially dependent on androgens, PCa can be treated with ADT, but ultimately progress towards a highly aggressive metastatic castration-resistant disease (mCRPC).

1.2.4. Diagnostic

The diagnostic of patients with suspected PCa usually involves a combination of the following tests: digital rectal examination, PSA test, imaging and prostate biopsy.

1.2.4.1. Digital rectal examination

Digital rectal examination is commonly performed to analyse the posterior aspect of the peripheral zone by palpation. But this test remains subjective and meta-analyses showed that its routine use may be questionable due to poor performance (Jones et al., 2018; Naji et al., 2018), highlighting the need to complement the diagnosis with additional tests.

1.2.4.2. PSA screening

PSA screening is currently the widest test used to diagnose PCa, despite controversial specificity and reliability. PSA production and secretion by luminal cells is often increased in PCa, and the rise in serum PSA levels can be directly measured in peripheral circulation due to the breakage of the basement membrane in disorganised cancerous glands (Balk et al., 2003). Its development has led to an increased number of patients diagnosed with PCa, with the vast majority being asymptomatic at diagnosis. However, PSA screening does not allow a clear distinction between latent and more aggressive forms of PCa, and the mortality benefit of early detection appears to be insignificant (Ilic et al., 2018; Martin et al., 2018; Schröder et al., 2014).

1.2.4.3. Imaging

In the recent years, multi-parametric (mp) magnetic resonance imaging (MRI) has been increasingly used both pre- and post-biopsy to improve the accuracy of diagnosis of clinically relevant PCa (Stabile et al., 2020). However, similar to the previously described tests, mpMRI has a number of limitations, including the cost-effectiveness, and variable accuracy and reproducibility. In particular, this approach can miss a significant proportion of clinically relevant lesions (Johnson et al., 2019; Purysko et al., 2019). Beyond mpMRI, computed tomography and bone scintigraphy are classically used in patients with suspicion of metastases. But prostate-specific membrane antigen (PSMA) positron emission tomography (PET)-CT is emerging as a promising alternative for high-risk patients to identify nodal or distant metastatic disease (Hofman et al., 2020).

1.2.4.4. Biopsy

PCa diagnosis is usually confirmed following histopathological evaluation of prostate biopsy, obtained either via transrectal ultrasound (TRUS)-guided biopsy or via a transperineal route (Lomas and Ahmed, 2020). During the biopsy procedure, several pieces of tissues are sampled either randomly or in a targeted area defined following abnormal mpMRI. The complex glandular architecture of the prostate, and the multifocality of PCa renders particularly challenging the detection of PCa in an untargeted first-time biopsy. To illustrate this aspect, the chances of detecting PCa using 6, 12 and 21 biopsy cores for diagnostic were 32.5%, 40.4%, and 43.3% respectively (Ploussard et al., 2014). Another study showed that about a quarter of patients who received untargeted TRUS-guided biopsy had a cancer subsequently diagnosed via targeted biopsies (Ahmed et al., 2017). Of note, patients with suspected PCa but negative biopsies are actively monitored for PSA levels and repeat biopsies.

1.2.5. Grading and risk stratification

1.2.5.1. Gleason grading

More than 30 years ago, Donald F. Gleason established the most commonly used grading system in pathology to evaluate the prognosis of PCa (Gleason and Mellinger, 1974). The International Society of Urological Pathology (ISUP) updated criteria defining the “Gleason score” in 2005 (Epstein et al., 2005), 2014 (Epstein et al., 2016a), and introduced the concept of “Gleason grade groups” in 2016 (Epstein et al., 2016b) to better reflect disease prognosis. Biopsies are attributed two grades between 1 and 5, representative of the two predominant patterns. High grades correspond to cells with poorly differentiated morphology and more advanced cancers. The two Gleason grades are then added, resulting in the Gleason score. Depending on the Gleason score, Gleason grade groups are attributed according to ISUP’s criteria (Table 1.1).

Table 1.1. ISUP's Gleason grade groups (GGG).

ISUP's GGG	Gleason score	Definition
Grade 1	2-6	Only individual discrete well-formed glands
Grade 2	3+4 = 7	Predominantly well-formed glands with lesser component of poorly formed/fused/cribriform glands
Grade 3	4+3 = 7	Predominantly poorly formed/fused/cribriform glands with lesser component of well-formed glands
Grade 4	4+4 = 8	Only poorly formed/fused/cribriform glands
	3+5 = 8	Predominantly well-formed glands and lesser component lacking glands
	5+3 = 8	Predominantly lacking glands and lesser component of well-formed glands
Grade 5	9-10	Lack of gland formation (or with necrosis) with or without poorly formed/fused/cribriform glands

(Epstein et al., 2016b)

1.2.5.2. TNM classification

The TNM classification system for PCa follows the guidelines of the American Joint Committee on Cancer Prostate Staging and is used to characterise disease stage and plan clinical management (Table 1.2).

Table 1.2. TNM classification system for PCa.

Clinical T (cT): Primary tumour stage	
Tx	Cannot be assessed
T0	No evidence of primary tumour
T1	Clinically inapparent tumour that is not palpable
T1a	Tumour incidental histologic finding in 5% or less of tissue resected
T1b	Tumour incidental histologic finding in more than 5% of tissue resected
T1c	Tumour identified by needle biopsy found in one or both sides, but not palpable
T2	Tumour is palpable and confined within prostate
T2a	Tumour involves one-half of one side or less
T2b	Tumour involves more than one-half of one side but not both sides
T2c	Tumour involves both sides
T3	Extraprostatic tumour that is not fixed or does not invade adjacent structures
T3a	Extraprostatic extension (unilateral or bilateral)
T3b	Tumour invades seminal vesicle(s)
T4	Tumour is fixed or invades adjacent structures (other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall)
Pathological T (pT): Primary tumour stage (determined after radical prostatectomy)	
T2	Organ confined
T3	Extraprostatic extension
T3a	Extraprostatic extension (unilateral or bilateral) or microscopic invasion of bladder neck
T3b	Tumour invades seminal vesicle(s)
T4	Tumour is fixed or invades adjacent structures (other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall)

N: Regional lymph nodal status	
Nx	Cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis
M: Distant metastasis	
Mx	Cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
	M1a: non regional lymph nodes
	M1b: bone(s)
	M1c: other site(s)

(Buyyounouski et al., 2017)

Briefly, T1 corresponds to clinically invisible or not palpable tumours, for example diagnosed after needle biopsy or upon transurethral resection of the prostate (TURP) for obstructive lower urinary tract symptoms. T2 are confined to the prostate, T3 extend into the capsule, and T4 invade adjacent structures. Other components of the classification system correspond to the presence of metastases, either in regional lymph nodes (N) or distant organs (M).

1.2.5.3. Risk classification

Previously described features can be combined into PCa risk groups to stratify patients into low, intermediate, and high risk of progression or biochemical recurrence. Several methods have been proposed over the years, and the widely used guidelines of the European Association of Urology are summarised below (Table 1.3).

Table 1.3. Risk classification of PCa.

	Localised PCa		Locally advanced PCa
Low risk	Intermediate risk	High risk	
PSA < 10 ng/mL	PSA 10-20 ng/mL	PSA > 20 ng/mL	Any PSA
ISUP's GGG 1	ISUP's GGG 2/3	ISUP's GGG 4/5	Any ISUP's GGG
cT1-T2a	cT2b	cT2c	cT3-4 or N1

(Mottet et al., 2017)

1.2.6. Clinical management of prostate cancer

The clinical management of PCa depends on a combination of factors which notably includes disease stage and risk, patient characteristics and patient

preferences (Litwin and Tan, 2017). Low risk patients generally follow a conservative management plan which includes watchful waiting or active surveillance. Patients diagnosed with localised or locally advanced disease are offered a radical form of treatment such as surgical excision of the prostate (radical prostatectomy) with or without pelvic lymph node dissection, or radiotherapy. In most advanced cases, including metastatic patients, these regimens can be either supplemented or replaced with hormonal therapy and chemotherapy.

Initial treatment of PCa is assessed by regular follow-ups, which aim at managing any possible treatment side-effects and monitor disease progression. After radical prostatectomy, serum PSA levels should remain below the detection limit for at least 6 weeks, and any measurable PSA is thought to be associated with increased risk of disease progression due to the presence of residual prostate tissue post-surgery (Preisser et al., 2019). Based on follow-up PSA measurements, biochemical recurrence (BCR) is classically defined as two consecutive PSA tests above 0.2 ng/mL for radical prostatectomy (Freedland et al., 2003). This definition varies in case of radiotherapy due to slower decline in PSA levels after treatment. In this case, BCR is defined as an increase of 2 ng/mL above the lowest PSA test post-treatment according to the Phoenix criteria (Abramowitz et al., 2008). While first-line treatments show relative efficacy, disease often recurs and progresses to more aggressive forms, including the incurable and lethal metastatic CRPC. Clinically, the emergence of CRPC is often associated with a rising PSA in an environment where testosterone levels are castrate, following maximal androgen blockade (Gillessen et al., 2015).

Overall, one of the main issues resides in the difficulty, at the stage of diagnosis, to discriminate between aggressive diseases from more indolent forms of PCa that will never progress beyond the prostate gland. Indeed, misdiagnosis often leads to over treatment, which can cause unnecessary harm to patients due of treatment side effects such as incontinence and erectile dysfunction (Antonarakis and Armstrong, 2011; Daskivich et al., 2011; Loeb et al., 2014). The lack of accurate prognosis markers represents therefore a major challenge to improve patient stratification.

1.2.7. Androgen deprivation therapy and castration resistance

In the 1940's, Charles Huggins demonstrated the effect of androgen deprivation via removal of the testes, causing regression of PCa (Huggins C, 1941), which built-up the basis for hormone therapy. Castration can be achieved either surgically via orchiectomy, or medically using ADT drugs. Surgical castration results in permanent androgen withdrawal and less adverse side effects compared with medical castration (Sun et al., 2016), but it is also associated with physiological and cosmetic issues. Thus, medical castration is the predominant mode of ADT used nowadays in western countries.

The first ADT compound developed was a LHRH agonist developed by Andrew Schally (Tolis et al., 1982). Since then, other molecules have been used to achieve chemical castration using alternative LHRH agonists (e.g., leuprorelin) or antagonists (e.g., degeralix), as well as nonsteroidal anti-androgens (e.g., bicatulamide) to target AR. Second-generation anti-androgens have then been developed, including abiraterone to inhibit androgenic steroid synthesis (Potter et al., 1995), and enzalutamide to selectively prevent AR nuclear translocation (Pritchard et al., 2016). In particular, enzalutamide has recently shown promising effect in advanced and metastatic PCa (Davis et al., 2019).

Patients on ADT often remain in long-term remission, but the development of CRPC is inevitable. Interestingly, AR alterations are mainly found in advanced diseases or after the use of targeted therapies, suggesting a role in resistance to ADT (Barbieri et al., 2012; Grasso et al., 2012). These include amplification and overexpression of AR, mutations or expression of constitutively active AR splice variants such as AR-V7 (Chen et al., 2004; Linja et al., 2001; Robinson et al., 2015; Sharp et al., 2019; Watson et al., 2015). Castration resistance has also been proposed to be present intrinsically in the normal untransformed prostate with the existence of specific intrinsically castration resistant cell populations (Barros-Silva et al., 2018; Kwon et al., 2016; Wang et al., 2009; Wang et al., 2020a; Yoo et al., 2016). Thus, there is currently a major clinical and biological focus to decipher how, why

and when does PCa becomes castration-resistant, with the hope that this ultimately help improve PCa therapy.

1.2.8. Molecular alterations in prostate cancer

Major advances in sequencing technologies have recently facilitated the identification of molecular alterations present in PCa patient samples. In 2015, the Cancer Genome Atlas Research Network (TCGA) study described seven main molecular subtypes of PCa based upon the identification of ETS fusions or mutations in speckle-type POZ protein (*SPOP*), forkhead box A1 (*FOXA1*), and Isocitrate dehydrogenase 1 (*IDH1*) (Cancer Genome Atlas Research Network, 2015). While long-term outcome data were not available to conduct meaningful analyses, these alterations did not correlate with Gleason score, highlighting heterogeneity between histopathological phenotypes and the genomic landscape of PCa. A brief snapshot of frequently observed genomic alterations in PCa is given below.

The homeobox protein NKX3.1 is an androgen-regulated transcription factor frequently found altered in PCa, either through loss of heterozygosity or down regulation via promoter methylation (Abate-Shen et al., 2008). In normal development, its expression marks one of the earliest stages of prostate organogenesis, and its absence leads to defects in structure and functions, including reduced prostatic ductal branching and hormones secretions (Bhatia-Gaur et al., 1999). Wang et al. have reported the existence of a specific luminal prostate population of NKX3.1 expressing cells with enhanced regenerative potential in surgically castrated mouse models (Wang et al., 2009b). Numerous studies have also reported alterations of phosphatase and tensin homolog (*PTEN*) in PCa with alterations in ~40% of primary PCa and almost all metastatic tumours (Mitchell and Neal, 2015; Taylor et al., 2010). *PTEN* loss is also associated with progression towards castration resistance (Mulholland et al., 2011) and found altered in half of CRPC patients (Jamaspishvili et al., 2018). Cooperative mechanisms between *Pten* and *Nkx3-1* loss of function were also described in mouse models (Kim et al., 2002).

Chromosomal rearrangements activating members of the ETS family of transcription factors (*ERG*, *ETV1*, *ETV4*, *FLI1*) are extremely frequent and found in a

significant fraction of prostate carcinomas, in both primary tumours and metastasis (Cancer Genome Atlas Research Network, 2015; Clark and Cooper, 2009). In particular, the *TMPRSS2-ERG* translocation results in a fusion gene expressing the N-term truncated ERG protein under the control of the androgen-dependent promoter of *TMPRSS2* (Tomlins et al., 2005). The functional consequences of such translocations are not fully understood but have been shown to impair prostate epithelial cell differentiation and support aggressive tumours (Baena et al., 2013; Shen and Abate-Shen, 2010; Sun et al., 2008; Zong et al., 2009). Moreover, cooperative mechanisms between *TMPRSS2-ERG* translocation and activation of the PI3K pathway give rise to high-grade PIN and carcinoma in mice (Squire, 2009).

Other common molecular alterations found in PCa include *AR*, *CDKN1B*, *PIK3CA*, *RB1*, *TP53*, *MYC*, *HER2/neu* and *SCR* (Abida et al., 2019; Mitchell and Neal, 2015). Recurrent mutations were reported in the *SPOP* gene, leading to increasing genomic instability through deregulation of DNA repair mechanisms (Barbieri et al., 2012; Wang et al., 2020b). Overexpression of the histone methyltransferase *EZH2*, a member of the polycomb complex, was associated with aggressive and metastatic disease (Varambally et al., 2002).

Overall, PCa has been shown to be highly heterogeneous disease, with the identification of multiple genomic and epigenomic alterations, and correlations with cancer progression and aggressiveness remain unclear. A plethora of additional genetic changes have been reported in PCa, but these are not described in this thesis. Current efforts aim at better characterising and stratifying the remarkable heterogeneity observed in prostate tumours.

1.2.9. Models of prostate cancer

1.2.9.1. Cell lines

There is a lack of cellular models to capture the heterogeneity of PCa. Among more than 1,000 established cancer cell lines, less than ten originate from PCa or PCa metastasis. The LNCaP cell line was established from a metastatic lesion of human prostatic adenocarcinoma (Horoszewicz et al., 1983). LNCaP cells are androgen responsive, express K8/18, carry the *ETV1* translocation and have loss of

PTEN. VCaP cells (Korenchuk et al., 2001), androgen-responsive as well, were derived from a vertebral metastatic lesion, and exhibit amplification of *AR*, *TP53* gain of function mutation, and bear the ERG translocation. VCaP cells express high levels of PSA, K8/18 and prostatic acid phosphatase. The 22Rv1 cell line (Sramkoski et al., 1999) was derived from a xenograft serially propagated in mice after castration-induced regression and relapse of the parental androgen-dependent human prostatic carcinoma xenograft. The line expresses the luminal markers K8/18 and has been used as a model of CRPC in several studies, though AR remains highly expressed. Of note, this line has been shown to produce high titres of human retrovirus XMRV (Xenotropic Murine Leukaemia Virus-related) suggesting that any experimental results would need to be carefully considered (Knouf et al., 2009). The two most tumorigenic and metastatic PCa cell lines PC-3 (Kaighn et al., 1979) and DU145 (Stone et al., 1978) are castration-resistant, and express Vimentin and K5/8/18. PC-3 were isolated from bone metastasis of advanced PCa and does not express *PSA* and *PTEN*. DU145, less aggressive, was derived from a brain metastasis and do not express PSA but have *TP53* gain of function mutation. Finally, the RWPE-1 (Bello et al., 1997) is a human prostate epithelial cell line derived from a histologically normal adult and immortalized with the human papilloma virus 18. It expresses *AR* and is commonly used as a model of non-malignant prostate cells.

1.2.9.2. Organoids

In addition to cell lines, organoids have emerged as an exciting alternative to study cancer behaviours, and create cellular models recapitulating organs heterogeneity and organisation (Clevers, 2016; Fatehullah et al., 2016). Despite the technical difficulty to establish and maintain such 3D structures in culture, methods have recently been published to generate organoids from mouse primary cells as well as human primary and metastatic PCa cells (Chua et al., 2014; Drost et al., 2016; Gao et al., 2014; Karthaus et al., 2014). These prostate organoid culture protocols allow the expansion of both basal and luminal lineages, express various levels of key prostate regulators such as AR, NKX3.1 and p63 and harbour different genetic backgrounds. Organoids can be genetically manipulated to study the

functional relevance of specific genes (Pietrzak et al., 2020; Wang et al., 2020a), or used to assess the differentiation and plasticity potential of cell populations in both normal (Barros-Silva et al., 2018; Karthaus et al., 2020; Kwon et al., 2016) and cancerous (Agarwal et al., 2015; Park et al., 2016) contexts. For example, organoids derived from the *Pten*/*Tp53* mouse model of PCa have allowed the identification of distinct tumour luminal subpopulations with specific stem/progenitors' potentials (Agarwal et al., 2015). This technology can also be used to evaluate drug responsiveness on multiple models and provide a more comprehensive tool than the currently available cell lines (Corrò et al., 2020). However, the absence of a rich stroma and vasculature limits the extent of the conclusions that can be drawn from these models. To date, co-culture between prostate organoids and other components of the stroma such as fibroblasts remain anecdotical (Kwon et al., 2019; Wei et al., 2019). To circumvent this issue, mature organoids can also be transplanted into recipient mice to provide a more complex microenvironment (Park et al., 2016).

1.2.9.3. Mouse models of prostate cancer

Although the mouse and human prostates exhibit some differences (see section 1.1.5), the use of xenograft and genetically engineered mouse models (GEMM) have contributed to improving the current understanding of the complex biology of both the normal prostate and PCa.

Xenograft models of PCa have been established using human cancer cell lines such as LNCaP, VCaP, PC-3 or DU145 implanted subcutaneously or orthotopically into the mouse prostate (Wu et al., 2013). These models have been of particular interest for chemotherapeutic approaches and studying molecular mechanisms of PCa. However, they are limited by the genetic backgrounds of PCa cell lines, the heterologous microenvironment, and the absence of fully functional recipient's immune system. Alternatively, xenografts can also be transplanted directly from patient tissue into immune-deficient mice after surgery, and without prior *in vitro* culture. Although these PCa patient-derived xenografts (PDXs) can serve as powerful models for pre-clinical studies of targeted therapies (Lin et al., 2014) and

give valuable insights into the biology of patient-specific tumours (Hidalgo et al., 2014), PCa PDXs remain challenging to grow and expand with an overall poor success rate from primary tumours (Elbadawy et al., 2020; Joshi et al., 2020; Palanisamy et al., 2020).

The first GEM models of PCa were generated using transgenes overexpressing viral oncogenes. The TRAMP mouse (Transgenic Adenocarcinoma of the Prostate) carries a rat *Probasin* (*Pb*) promoter to drive the expression of the SV40 large T and small t antigens in prostate cells (Greenberg et al., 1995). Even though TRAMP tumours are highly aggressive, they are not associated with specific genetic changes implicated in PCa, which limits the utility of this model. In the second generation of GEM models, loss-of-function mutations in genes of interest were created to evaluate their implication in PCa initiation and progression. The most widely used models carry null mutations of *Nkx3-1* and *Pten* either alone or in combination. Later, models were improved by restricting the presence of the desired mutations to the prostate epithelium, using tissue specific expression of the *Cre* recombinase such as the *Probasin*-Cre4 (*Pb-Cre4*) transgene (Wu et al., 2001) or *Nkx3-1-Cre* knock-in alleles (Stanfel et al., 2006). A notable limitation of these models is that *Cre* expression can be detected during organogenesis since it depends entirely on endogenous promoter activity. To address this limitation, inducible conditional mouse models have then been generated to control the spatiotemporal occurrence of any gene overexpression or deletion in specific tissues, for example using tamoxifen-inducible *Cre* recombinase (Luchman et al., 2008; Wang et al., 2009b). This approach is however limited by the number of prostate-specific promoters available to drive *Cre* expression, and possible leakiness of the system (Stifter and Greter, 2020). While modelling PCa in mice has been successfully achieved using mouse models, all of the current promoters to drive *Cre* expression are androgen-responsive elements, and therefore may be limited for the study of castration-resistant disease (Shen and Abate-Shen, 2010).

1.3. The RUNX family of transcription factors

[This section has been adapted from (Mevel et al., 2019)]

The RUNX family of transcription factors orchestrate various developmental and cellular processes, such as cell proliferation, differentiation, and cell lineage specification. The RUNX genes were named after the discovery of the developmental regulatory gene 'runt', which was found essential for early embryonic segmentation after being identified in a mutagenesis screen for the development of Drosophila melanogaster (Gergen and Butler, 1988; Nüsslein-Volhard and Wieschaus, 1980). RUNX genes have been described in the majority of sequenced metazoan genomes, with single copies of a RUNX gene present in most bilaterians, and at least three genes in insects and vertebrates (Grasso et al., 2012). The mammalian RUNX transcription factors consist of RUNX1, RUNX2 and RUNX3.

1.3.1. Conservation and transcriptional regulation of the RUNX genes

The genomic architecture of the three mammalian RUNX genes is very similar (Figure 1.5), and highly conserved across metazoans. In vertebrates, all RUNX genes contain two alternative promoters: a distal P1 promoter and a proximal P2 promoter, which is thought to represent the primordial promoter (Levanon and Groner, 2004; Rennert et al., 2003). The main protein isoforms encoded by these transcripts differ in their N-terminal amino acid sequences; distal isoforms are usually longer and always begin with the MAS(D/N)S amino acid sequence, whereas proximal isoforms begin with the MRIPV motif (Bangsow et al., 2001; Miyoshi et al., 1995; Xiao et al., 1998). The 500 million years of conservation of this dual promoter structure for the three RUNX genes in vertebrates suggests specific — but currently unclear — functions for each promoter or for their respective transcripts (Levanon and Groner, 2004; Rennert et al., 2003). Indeed, differential activity of the two promoters has been reported across a wide range of developmental stages and tissues, suggesting divergent context-specific requirements (Bangsow et al., 2001; Bee et al., 2010; Chung et al., 2007; Draper et al., 2016; Harada et al., 1999; Liu et al., 2011; Rini and Calabi, 2001; Sroczynska et al., 2009).

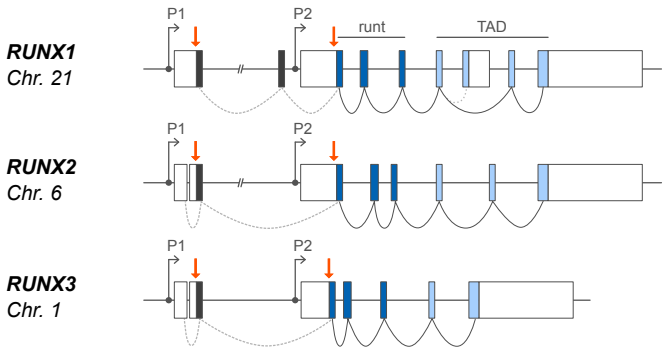

Figure 1.5. Structure of the RUNX genes.
Conserved genomic structure of human *RUNX1* (Entrez Gene ID: 861), *RUNX2* (Entrez Gene ID: 860), and *RUNX3* (Entrez Gene ID: 864). The initiator ATG codons of P1 and P2 derived transcripts are indicated by orange arrows. Exons coding for runt and transactivation domains (TAD) are shown in dark and light blue respectively, and untranslated regions (UTRs) are in white. Dominant splicing is indicated by a continuous line, and some alternative splicing is shown by dotted line.

Alternative splicing events add further diversity to RUNX transcripts originating from the P1 and P2 promoters. Currently, the exact physiological role — if any — of the different isoforms is unknown. Interestingly, all RUNX genes express a panel of isoforms with increased or reduced transactivation activities (Bae et al., 1994; Bangsow et al., 2001; Geoffroy et al., 1998; Jin et al., 2004; Puig-Kröger et al., 2010; Telfer and Rothenberg, 2001; Terry et al., 2004). Also, sequence variations occurring in the 5′ and 3′ untranslated regions (UTR) of the multiple RUNX transcripts affect the stability and translation efficiency of the different RUNX mRNAs. P1-derived transcripts, generally bearing a shorter 5′ UTR, have been shown to direct cap-mediated translation more efficiently than the P2-derived 5′ UTR (Pozner et al., 2000). In addition, miRNAs can regulate RUNX transcripts by targeting isoform-specific 3′UTRs (Ben-Ami et al., 2009; Zhang et al., 2011).

1.3.2. Structural homologies between the RUNX proteins

The defining component of the RUNX proteins is the presence of the highly conserved runt homology domain (RHD): a 128-amino-acid sequence located near the N-terminus (Figure 1.6) (Rennert et al., 2003). The RHD is essential first, for binding to the DNA at the consensus RUNX motif “PyGPyGGTPy” (Kamachi et al.,

1990; Wang and Speck, 1992) second, for protein-protein interactions (Lilly et al., 2016; Nagata et al., 1999), and third, for the nuclear localisation of the RUNX factors (Michaud et al., 2002; Telfer et al., 2004). The C-terminus is less conserved and contains an activation domain, an inhibitory domain and a five amino acid C-terminal motif (VWRPY in most cases), known as the recruitment signal for Groucho/TLE (transducin-like enhancer of split) family of co-repressors (Levanon et al., 1998; Seo et al., 2012; Yarmus et al., 2006). Within the transactivation domain, RUNX proteins also bear a conserved nuclear matrix-targeting signal sequence, which plays a role in the regulation of RUNX activity and nuclear localisation (Zaidi et al., 2001; Zeng et al., 1998).

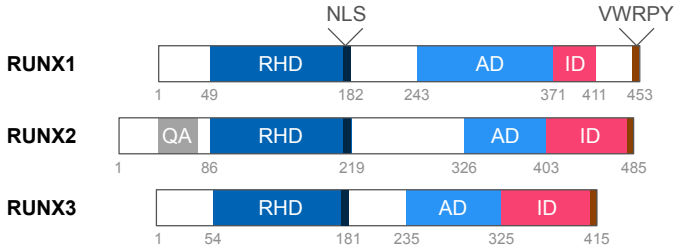

Figure 1.6. Structure of RUNX proteins.
Schematic representation of human RUNX1 (NP_001001890), RUNX2 (NP_001265407), and RUNX3 (NP_004341) protein structures and domains, including the conserved Runt Homology Domain (RHD), Nuclear Localisation Signal (NLS), trans-activation (AD) and inhibitory (ID) domains, C-terminal Groucho/TLE binding site (VWRPY), and the RUNX2-specific glutamine/alanine-rich (QA) sequence.

1.3.3. Mechanism of action

RUNX transcription factors are part of a heterodimeric complex formed from the dimerization between the α subunit RUNX and its main partner, the core binding factor subunit β (CBFβ) (Figure 1.7) (Kamachi et al., 1990; Ogawa et al., 1993; Wang et al., 1993). CBFβ acts as a non-DNA binding regulatory element, which allosterically increases the DNA-binding affinity and stability of the complex by interacting with RUNX through the RHD (Bravo et al., 2001; Huang et al., 2001; Tang et al., 2000; Yan et al., 2004). Of note, CBFβ-independent functions of RUNX have also been suggested in the literature (Bresciani et al., 2014), including a non-transcriptional role for RUNX1 and RUNX3 in the Fanconi Anemia DNA repair

pathway (Tay et al., 2018; Wang et al., 2014). Usually regarded as weak transcription factors by themselves, RUNX proteins might have the potential to act as “pioneer” transcription factors, which are able to engage condensed chromatin to facilitate its opening and promote the recruitment of other transcriptional regulators (Zaret and Carroll, 2011). Indeed, this seems to be the case for RUNX1 in haematopoiesis (Hoogenkamp et al., 2009; Lichtinger et al., 2012), and RUNX3 in cell-cycle progression (Lee et al., 2019), however it remains to be determined to which extent these observations could apply to other settings and RUNX proteins. Importantly, RUNX factors can function as both transcriptional activators and repressors. For instance, RUNX1 is able to drive PU.1/Spi1 expression in myeloid and B cells, while it represses the same gene in T cells and megakaryocytes (Huang et al., 2008). This gene- and context-specific ambivalence stems from the myriad of RUNX/CBFβ interactors.

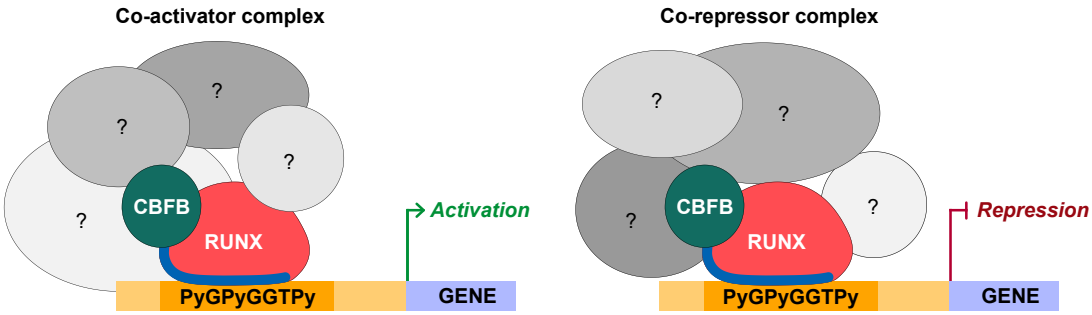

Figure 1.7. Mechanism of action of RUNX transcription factors.
RUNX factors can act as transcriptional activators or repressors depending on binding partners and post-translational modifications. CBFβ dimerises with the runt binding domain to enhance DNA-binding to the RUNX consensus motif. The runt domain is indicated in blue.

1.3.4. Post-translational modifications

RUNX factors are actively regulated at the post-translational level to enable fine tuning of their transcriptional potency, stability, and localization (Figure 1.8) through several signalling pathways. RUNX transcriptional activity is generally associated with its acetylation or interaction with chromatin modifiers such as p300/CBP or MOZ (Jin et al., 2004; Kitabayashi et al., 2001; Pelletier et al., 2002;

Wang et al., 2011a). Furthermore, serine and threonine phosphorylation by diverse signalling cascades can enhance RUNX activity in a variety of cellular contexts (Aikawa et al., 2006; Guo and Friedman, 2011; Imai et al., 2004; Kim et al., 2008; Zhang et al., 2008). Methylations of RUNX factors play ambivalent roles, as they can increase DNA binding (Zhao et al., 2008) or favour transcriptional repression depending on the cellular context (Herglotz et al., 2013; Vu et al., 2013). Repressive activity of the RUNX proteins is further controlled by tyrosine phosphorylation (Goh et al., 2010; Huang et al., 2012), association with histone deacetylases (HDACs) (Jeon et al., 2006; Lutterbach et al., 2000), and interaction with transcriptional regulators, such as SIN3A and Groucho/TLE family members (Imai et al., 1998; Imai et al., 2004; Levanon et al., 1998; Zhao et al., 2008). Since both acetylation of non-histone proteins and ubiquitination occurs at lysine residues, HDACs can reduce the transcriptional activity of RUNX by promoting their ubiquitination (Jin et al., 2004). Indeed, the stability of the RUNX factors is tightly regulated by ubiquitin-mediated degradation and depends on a combination of other PTMs. Additionally, the CDC20 subunit of the anaphase-promoting complex (APC) can regulate the level of RUNX proteins during cell cycle progression (Biggs et al., 2006). Finally, the stability of RUNX has also been reported to be decreased by SUMOylations (Kim et al., 2014a). While these PTMs have mostly been studied in isolation, it is evident that RUNX proteins are regulated by complex crosstalk between these modifications, which represent a major level of regulation of RUNX transcriptional activity. Importantly, deregulations of these subtle regulatory mechanisms have been shown to play important roles in the context of disease (Ito et al., 2015).

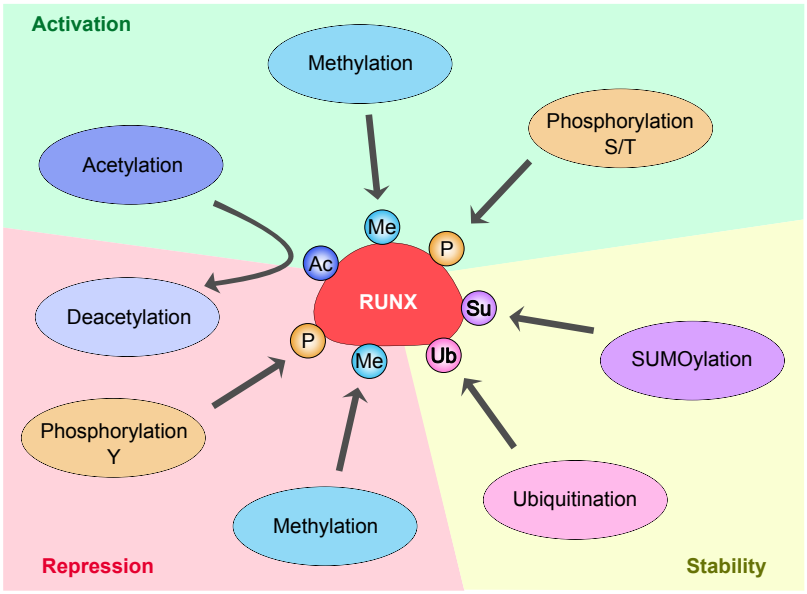

Figure 1.8. Post-translational modifications affecting RUNX proteins activity.
S, Serine; T, Threonine; Y, Tyrosine.

1.3.5. Auto-regulation and functional redundancy

Analyses of the promoters of each of the RUNX members has revealed the presence of consensus RUNX binding sites, raising the possibility of auto-regulatory loops (Ghozi et al., 1996), which was originally confirmed for *Runx2* in the context of bone formation (Drissi et al., 2000; Ducy et al., 1999). Similarly, *Runx1* has since been reported to regulate its own expression both *in vitro* (Pimanda et al., 2007) and *in vivo* (Martinez et al., 2016). Cases of cross-talk between RUNX factors have also been described, whereby RUNX3 regulates RUNX1 expression level by repressing its transcription in human B lymphoid cell lines (Brady et al., 2009; Spender et al., 2005).

The extent to which RUNX factors play redundant roles remains largely unknown. A few studies have demonstrated that the RUNX proteins can substitute each other in specific instances. For example, it has been shown using an *in vitro* co-culture system of the murine embryonic para-aortic splanchnopleural region that ectopic expression of *Runx2* or *Runx3* is able to rescue the haematopoietic defects caused by *Runx1* deficiency (Goyama et al., 2004).

Similarly, replacement of RUNX1 C-terminus by the equivalent portions of RUNX2 or RUNX3 rescues *Runx1*-null embryonic lethal phenotype (Fukushima-Nakase, 2005). Likewise, premature senescence was induced in murine embryonic fibroblasts by individually overexpressing either of the three RUNX genes (Kilbey et al., 2007). Studies in *Runx1*/*Runx3* double knockout (DKO) mice have exposed functional redundancy between these genes in the Fanconi Anemia DNA repair pathway, independent of their transcriptional role (Tay et al., 2018; Wang et al., 2014). Compensation between the RUNX genes mean that assessment of redundancy during tumorigenesis is important (Kamikubo, 2018). Although transcription factors are difficult to target with pharmaceuticals, pan-inhibitors of RUNX factors have shown promising anti-tumour effects in pre-clinical settings by small-molecule inhibition of RUNX/CBFβ complexes (Illendula et al., 2016), or via Pyrrole-imidazole polyamides, which selectively target the consensus RUNX motif in chromatin to prevent transcription (Morita et al., 2017a).

The studies described above, as well as the strong structural homologies and potential auto- and cross-regulations, suggest that RUNX proteins could compensate for each other. While new technologies such as single cell RNA transcriptomics are likely to reveal further co-expression in defined cellular compartments, the specific spatiotemporal expression patterns of the RUNX genes is thought to explain, at least partially, their non-redundant functions and requirements in several developmental processes (Levanon and Groner, 2004; Levanon et al., 2001). However, even in the context of co-expression such as in teeth development, the comparison of the phenotypes observed in *Runx2*/*Runx3* DKO mice with those of single *Runx2* knockouts showed no obvious compensations between the two factors (Wang et al., 2005a). In other contexts, partial, but not complete, redundancy has been reported. Only a certain degree of redundancy has been observed between *Runx2* and *Runx3* during chondrocyte development using single and double knockout mouse models (Yoshida, 2004), and partial redundancy has also been reported during lacrimal gland development (Voronov et al., 2013).

Taken together, while these studies reveal possible compensation between the RUNX genes in defined contexts, they also highlight their crucial and non-

redundant functions that are partly, but not exclusively, associated with their intricate spatiotemporal regulation.

1.4. Developmental functions of RUNX1

[This section has been adapted from (Mevel et al., 2019)]

The functions of the RUNX family members have initially been uncovered using knockout mice. Deletion of each of the three RUNX proteins has severe consequences on survival; *Runx1* knockout is embryonically lethal, *Runx2* display neonatal lethality, and *Runx3* knockouts have mixed survival rates depending on the model and strain used. These models have initially revealed that RUNX1 is essential for definitive haematopoiesis (Okuda et al., 1996; Wang et al., 1996), RUNX2 is critical for skeletal development (Ducy et al., 1997; Komori et al., 1997; Otto et al., 1997), and RUNX3 plays an important role in neurogenesis (Inoue et al., 2002; Levanon et al., 2002). Later, stage- and tissue-specific conditional knockouts have started to uncover a myriad of additional roles for RUNX proteins in other tissues, which were previously concealed by the severity of the developmental defects in the complete knockout models (Mevel et al., 2019). The following section presents an overview of the most thoroughly examined roles and requirements of RUNX1 in the haematopoietic system, hair follicles, mammary glands, as well as repair and regeneration processes of other tissues.

1.4.1. Haematopoietic system

Of the three RUNX proteins, RUNX1 is generally considered as a master regulator of haematopoiesis due to its critical role in the ontogeny of the whole haematopoietic system while RUNX3 has important functions in the lymphocyte and myeloid lineages.

1.4.1.1. Embryonic haematopoiesis

The vertebrate haematopoietic system is established through three main successive waves of blood cell generation. The first two waves take place in the extra-embryonic yolk sac and produce sequentially primitive erythrocytes, then

erythro-myeloid and lymphoid progenitors (Costa et al., 2012; Frame et al., 2016; Lux et al., 2008; McGrath et al., 2015; Palis et al., 1999; Yoshimoto et al., 2011; Yoshimoto et al., 2012). The third wave takes place in the intra-embryonic aorta-gonad-mesonephros (AGM) region and generates the haematopoietic stem cells (HSCs) that will sustain the haematopoietic system during adulthood (Medvinsky and Dzierzak, 1996) (Figure 1.9 A). Both stem and progenitor haematopoietic cells (HSPCs) arise from hemogenic endothelial cells through an endothelial to haematopoietic transition (EHT) (Figure 1.9 B) (Bertrand et al., 2010; Boisset et al., 2010; Eilken et al., 2009; Jaffredo et al., 1998; Kissa and Herbomel, 2010; Lam et al., 2010; Lancrin et al., 2009; Zovein et al., 2008). HSPCs then move to the foetal liver, where they further expand and mature before seeding the bone marrow that serves as the main haematopoietic organ during adulthood (Mikkola and Orkin, 2006).

Runx1 is expressed in all these sites of *de novo* blood cell emergence, as well as in all the haematopoietic cells, with the exception of mature erythrocytes (Lacaud et al., 2002; Lorsbach et al., 2004; North et al., 2002; North et al., 2004; Sroczynska et al., 2009; Stefanska et al., 2017; Swiers et al., 2013; Zeigler et al., 2006). Disruption of *Runx1* results in the complete absence of all haematopoietic cells other than primitive erythroid cells (Lacaud et al., 2002; Okuda et al., 1996). RUNX1 is critical for both initiation and completion of EHT through epigenetically silencing the endothelial program (Lancrin et al., 2012; Liakhovitskaia et al., 2014; Lie-A-Ling et al., 2014; Thambyrajah et al., 2016; Tober et al., 2013), and redistributing haematopoietic transcription factor binding (Lichtinger et al., 2012), to form a stable epigenetic state where RUNX1 is then dispensable (Chen et al., 2009; Lancrin et al., 2009; North et al., 1999) (Figure 1.9 B). The precise modulation of RUNX1 levels and activity are essential for the efficiency and the correct timing of haematopoietic progenitor emergence (Cai et al., 2000; Eliades et al., 2016; Eliades et al., 2016; Lacaud et al., 2004; Lie-A-Ling et al., 2018).

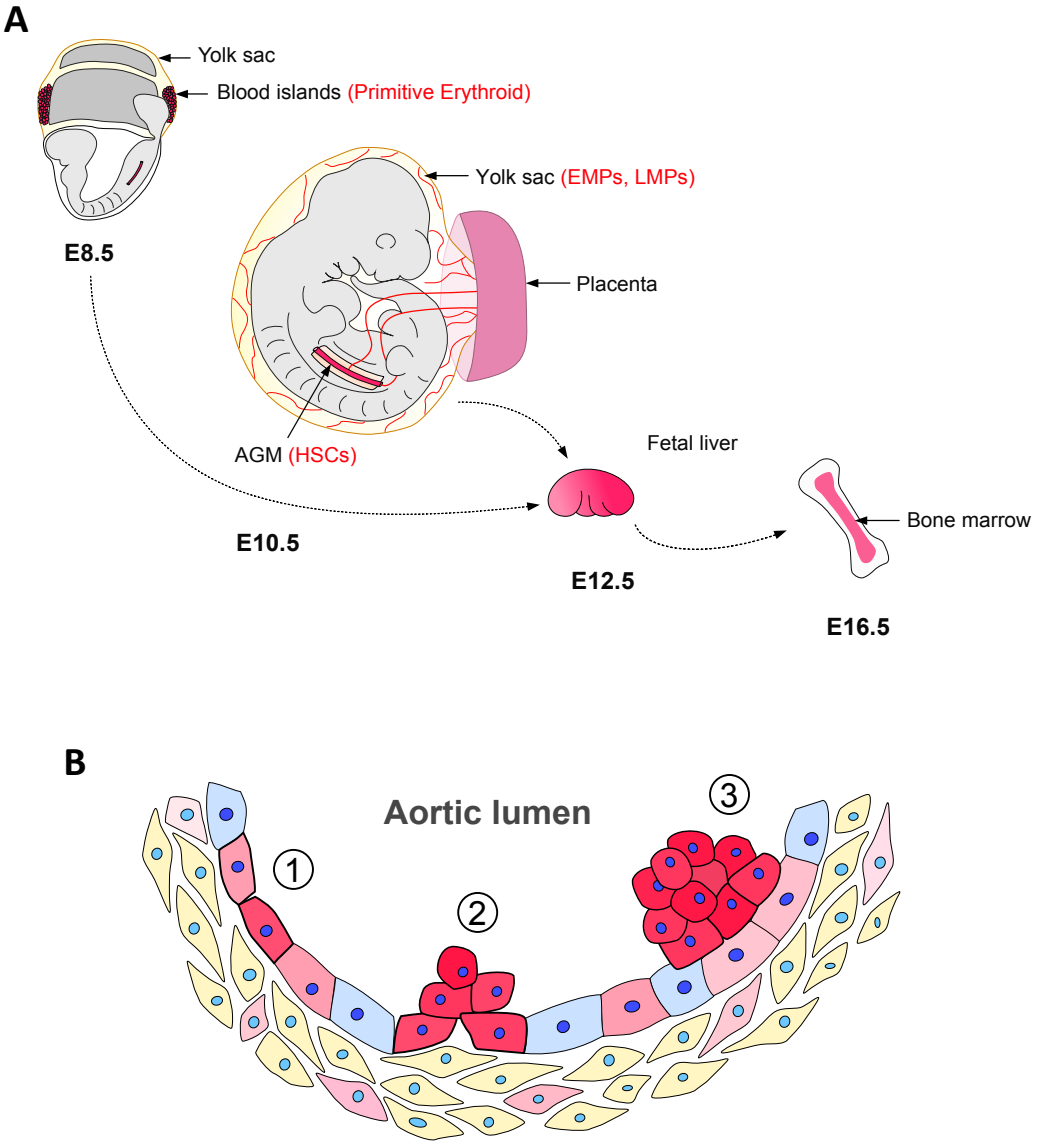

Figure 1.9. RUNX1 and embryonic haematopoiesis.
(A) Timeline of embryonic blood development. RUNX1 is expressed at all the different sites of haematopoietic development during the different haematopoietic waves. HSCs, Haematopoietic Stem Cells; EMPs, Erythro-Myeloid Progenitors; LMPs, Lymphoid-Myeloid Progenitors. (B) RUNX1 is critical for the endothelial-to-haematopoietic transition (EHT). RUNX1 is expressed in endothelial cells of the ventral wall of the mouse dorsal aorta and in mesenchymal cells situated below (pink cells). A rare subset of endothelial cells, the hemogenic endothelial cells (red cells), are committed to forming blood cells (1). RUNX1 is required for these cells to undergo morphological changes and bud from the endothelial lining into the lumen of the dorsal aorta (2). These precursors further proliferate and differentiate into mature blood cells in intra-aortic clusters.

1.4.1.2. Adult haematopoiesis

Runx1 is broadly expressed in most adult blood cells (Lorsbach et al., 2004; North et al., 2004) (Figure 1.10). However, despite this broad expression, *Runx1* appears partially dispensable in adult haematopoiesis as indicated by the absence of lethality when deleted either from the onset of HSC development at E11.5 using the *Vav1*-Cre mouse line (Chen et al., 2009) or in established adult haematopoiesis with the pan haematopoietic inducible *Mx1*-Cre system (Cai et al., 2011; Growney et al., 2005; Ichikawa et al., 2004; Jacob et al., 2010; Putz et al., 2006). *Runx1* deletion in adult mice was shown to result either in expansion (Growney et al., 2005; Ichikawa et al., 2004) or exhaustion of phenotypic HSC (Jacob et al., 2010). A more recent study suggests that deregulation of expression of HSC cell surface markers in the absence of RUNX1 might explain some of these conflictual results, and that loss of *Runx1* does not substantially alter the frequency of functional long-term repopulating HSCs based on bone marrow chimerism (Cai et al., 2011; Growney et al., 2005; Ichikawa et al., 2004; Jacob et al., 2010) (Figure 1.10). *Runx1* deletion causes significant expansion of the entire bone marrow haematopoietic stem and progenitor cell compartment and decreased HSC apoptosis and ribosome biogenesis, which have been proposed to contribute to preleukemic states (Cai et al., 2015) (Figure 1.10).

Haematopoietic deletion of *Runx1* in adult mice also exhibit decreased B and T cell numbers, and lower platelet counts (Growney et al., 2005; Ichikawa et al., 2004). Although RUNX1 is downregulated in mature erythrocytes and platelets, *in vivo* and *in vitro* studies demonstrated a key role for RUNX1 in balancing specification of platelet-producing megakaryocytic lineage commitment through multiple interactions such as AP-1, p300, GATA and ETS transcription factors (Elagib, 2003; Pencovich et al., 2011; Pencovich et al., 2013). Enforced *RUNX1* expression in the K562 human cell line promotes megakaryopoiesis by directly repressing the erythroid lineage (Kuvardina et al., 2015). Surprisingly, the loss of *Runx1* causes also an increased commitment to the megakaryocytic lineage (Behrens et al., 2016), perhaps due to the properties of the different RUNX1 isoforms. Of note, RUNX1 is

also involved in various immune cell subsets (Mevel et al., 2019; Voon et al., 2015), but these will not be discussed in this thesis.

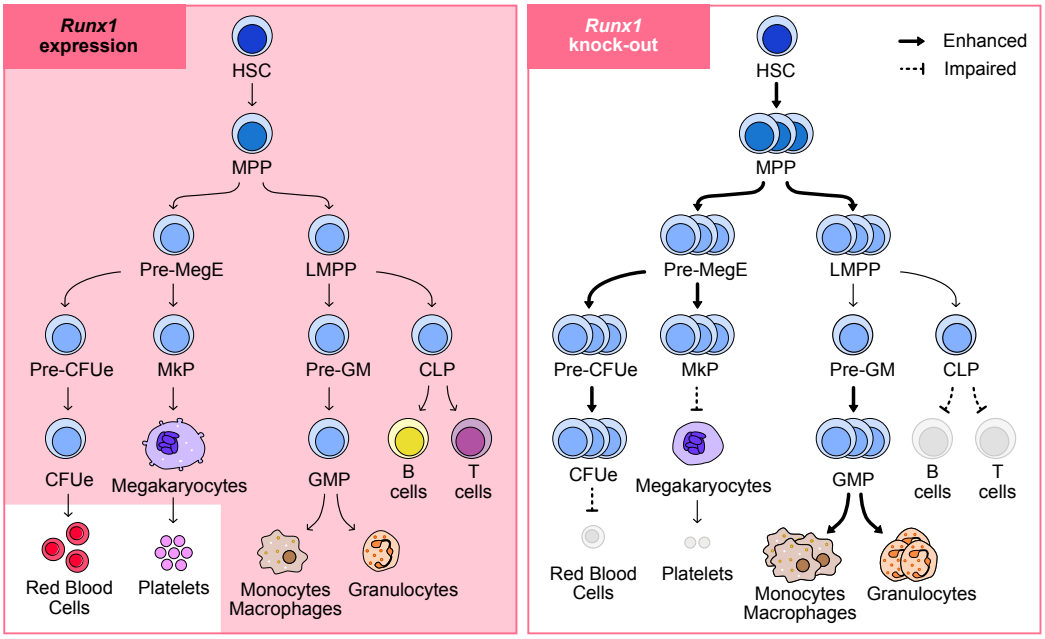

Figure 1.10. RUNX1 expression and impact of knockout mice on adult haematopoiesis. (Left) Expression of RUNX1 in the different haematopoietic populations of the adult haematopoietic system. (Right) Impact of haematopoietic *Runx1* knockout on the frequencies of these populations. Enhanced processes are indicated by green arrows, impaired processes are indicated by red bars. HSC, Haematopoietic Stem Cell; MPP, Multipotent Progenitor; LMPP, Lymphoid-primed Multipotent Progenitor; PreGM, Pre-Granulocyte/Monocyte progenitor; GMP, Granulocyte/Monocyte Progenitor; CLP, Common Lymphoid Progenitor; PreMegE, Pre-Megakaryocyte/Erythroid progenitor; PreCFUe, Pre-erythroid Colony-Forming Unit; CFUe, erythroid Colony-Forming unit; MkP, Megakaryocyte Progenitor.

1.4.1.3. Runx1 isoforms in haematopoiesis

RUNX1 isoforms display well-defined structural differences and distinct developmental expression patterns. During embryogenesis, the emergence is in part sustained by the sequential activation of *Runx1* promoters, with proximal P2 transcripts detected during EHT, while distal P1 transcripts are only found once haematopoietic commitment is completed (Bee et al., 2009; Challen and Goodell, 2010; Fujita et al., 2001; Sroczynska et al., 2009; Zambidis et al., 2005). Accordingly, the activity of the P2 proximal promoter is critical for blood cell emergence whereas

abrogation of P1 distal transcripts results in more subtle defects(Lam et al., 2009; Mukai et al., 2012; Pozner et al., 2007; Sroczynska et al., 2009).

In adult haematopoiesis, the P1-derived RUNX1c is dominant, whereas P2-derived RUNX1b is confined to progenitor subsets of the granulocyte/macrophage, lymphoid lineages and megakaryocytes (Bee et al., 2009; Draper et al., 2016; Telfer and Rothenberg, 2001). Downregulation of *Runx1b* is a prerequisite for terminal differentiation of these lineages, except megakaryocytes. In myeloid cells, *Runx1b* expression correlates with increased proliferation and colony-forming unit-culture activity in the bipotential pre-megakaryocyte/erythroid (PreMegE) progenitor (Draper et al., 2016). In P1-null mice, inactivation of Runx1c results in lineage-specific defects, reminiscent of total *Runx1* deficiency (Bee et al., 2010; Draper et al., 2016; Draper et al., 2017). In these mice, PreMegE progenitors produce more erythroid, and fewer megakaryocyte, progenitors (Draper et al., 2017). Unlike in complete *Runx1* knockout models (Growney et al., 2005; Ichikawa et al., 2004), *Runx1c*-null megakaryocytes differentiate and produce proplatelets, suggesting that RUNX1c specifies megakaryocytes, whereas RUNX1b drives the maturation of committed megakaryocytes. In a human B lymphoblastoid cell line infected with Epstein–Barr virus, RUNX1c – but not RUNX1b – inhibits growth (Brady et al., 2013). Finally, a shorter isoform designated RUNX1a, transcribed from the P2 promoter and lacking the transactivation domain, has been identified in humans (Komeno et al., 2014). RUNX1a is proposed to act as a dominant negative (Levanon et al., 2001), that enhances haematopoietic commitment (Ran et al., 2013) and increases HSPC renewal (Tsuzuki and Seto, 2012; Tsuzuki et al., 2007).

1.4.2. Hair follicles and epidermis

RUNX proteins, and in particular RUNX1, participate in hair follicles (HF) morphogenesis and maintenance. During development, *Runx1* is expressed in both the HF epithelium and the surrounding mesenchyme (Levanon et al., 2001; Osorio et al., 2011; Raveh et al., 2006). In the forming epithelium, disruption of *Runx1* expression delays HF development (Osorio et al., 2011). Loss of *Runx1* in mesenchymal cells does not initially impact early HF development, but

subsequently leads to the emergence of defective hair follicle stem cells (HFSCs) precursors, which differentiate preferentially in enlarged sebaceous cysts instead of healthy hair bulbs (Osorio et al., 2011). After follicular morphogenesis, hair growth cycle starts postnatally; *Runx1* is expressed in specific HF compartments and absent in the surrounding mesenchyme (Osorio et al., 2011; Raveh et al., 2006). During the hair growth phase (anagen), *Runx1* is broadly found in bulge cells. Loss of *Runx1* impairs HFSCs self-renewal and delays entry into anagen (Hoi et al., 2010; Osorio et al., 2011). Conversely, ectopic expression of *Runx1* during anagen initiates hair degeneration (Lee et al., 2014). During the resting phase (telogen), *Runx1*-null HFSCs are able to exit quiescence either with time (Hoi et al., 2010; Osorio et al., 2011) or following injury (Osorio et al., 2008), indicating that *Runx1* is dispensable for this process. Like in other tissues, *Runx1* dosage has an important role in the regulation of skin epithelial cell fate. While low *Runx1* expression in bulge stem cells enhances self-renewal (Hoi et al., 2010), higher RUNX1 levels promotes their transition towards early progenitor hair germ cells (Lee et al., 2014). Mechanistically, RUNX1 orchestrates HF specification and maturation by modulating Wnt signalling (Osorio et al., 2011), and regulates HFSCs proliferation in a P21-dependent manner (Hoi et al., 2010; Lee et al., 2013). Additionally, RUNX1 has been implicated in the lipid metabolism of skin epithelial cells by regulating fatty acid production (Jain et al., 2018). RUNX1 is also expressed in mouse keratinocytes, where it collaborates with p63 to regulate the balance between proliferation and differentiation (Masse et al., 2012; Qu et al., 2018).

1.4.3. Mammary gland

Specific roles of RUNX proteins in the development and functions of the mammary gland are also starting to emerge. The mammary gland is generated from the embryonic mammary placode and develops mainly after birth into a branched network of collecting ducts and tubes. The mammary gland undergoes further dynamic changes, greatly affected by hormone levels during oestrous cycles, pregnancy and lactation. The epithelium of the mammary gland is composed of two distinct cell types forming a bi-layer structure, in which the alveolar luminal

secretory cells, responsible for milk production, are surrounded by basal or myoepithelial cells (Inman et al., 2015).

All three RUNX genes are expressed at different levels within the mouse mammary epithelium (Blyth et al., 2010). *Runx1* expression is higher than *Runx2*, while *Runx3* expression is barely detectable (McDonald et al., 2014; Owens et al., 2014; van Bragt et al., 2014). RUNX1 and RUNX2 levels are higher in the basal than in the luminal compartment (Kendrick et al., 2008; McDonald et al., 2014; van Bragt et al., 2014), and RUNX1 is completely absent in the alveolar luminal cells (van Bragt et al., 2014). The expression of the RUNX genes appears to fluctuate extensively; *Runx1* is highly expressed in the epithelium of virgin females and post-lactation, but it gradually decreases throughout pregnancy to reach its lowest levels in late pregnancy and lactation (Blyth et al., 2010; van Bragt et al., 2014). This has been linked to the extensive tissue remodelling that takes place during pregnancy, which results in the large expansion of alveolar luminal cells that do not express *Runx1* (Inman et al., 2015; van Bragt et al., 2014). In particular, *Runx1* regulates the fate of the oestrogen receptor positive luminal lineage *in vivo*, where it represses the alveolar transcription factor *Elf5* and promotes expression of a more mature luminal transcriptional program. Indeed, deletion of *Runx1* results in a reduction of mature luminal cells (van Bragt et al., 2014). *In vitro*, 3D morphogenesis studies utilising the non-tumorigenic basal-like MCF10A cell line have shown that *Runx1* is essential to promote the differentiation of acinar structures into ductal and lobular tissue (Sokol et al., 2015; Wang et al., 2011b). Together, these results suggest that RUNX1 is important for the differentiation of luminal cells.

1.4.4. RUNX1 in repair and regeneration

RUNX factors have been implicated in the biology of several other tissues that are particularly subject to frequent repair and regeneration processes. High *Runx1* expression has been reported in denervated skeletal muscles (Zhu et al., 1994), and *in vivo* deletion of *Runx1* in skeletal muscle (*Mck*-Cre) has revealed a role for RUNX1 in protecting denervated myofibers from excessive atrophy, autophagy and muscle wasting (Wang et al., 2005b). Interestingly, while *Runx1* is not expressed in naïve

developing or adult striated muscle, its expression peaks following myopathic damage and is proposed to regulate the balance of myoblast proliferation and differentiation during muscle regeneration (Umansky et al., 2015).

Both RUNX1 mRNA and protein levels increase in cardiomyocytes following myocardial infarction, where RUNX1 modulates calcium uptake and contractile functions. Interestingly, this study also indicated that injured *Runx1*-deficient mice are protected from the adverse effects of cardiac remodelling (McCarroll et al., 2018).

In the lacrimal gland, which secretes the aqueous layer of the tear film, both *Runx1* and *Runx3* expression is increased during tissue regeneration after inflammation-induced lacrimal gland damage. Furthermore, deletion of *Runx1* is associated with impaired epithelial development of the gland (Voronov et al., 2013).

1.5. Implication of RUNX1 in tumorigenesis

[This section has been adapted from (Lie-A-Ling et al., 2020)]

RUNX genes are associated with hallmarks of cancer development, including proliferation and epithelial-to-mesenchymal transition, and have oncogenic or tumour-suppressive functions. Mutations in all three *RUNX* genes, and *CBFβ*, are frequently identified in cancers and their roles are being actively investigated (Blyth et al., 2005; Chuang et al., 2013; Groner, 2017; Ito et al., 2015).

1.5.1. RUNX1 in leukaemia

RUNX1 (formerly AML1; Acute Myeloid Leukaemia 1) was initially discovered in humans as part of the t(8;21) translocation (Miyoshi et al., 1991). Then, other chromosomal translocations involving *RUNX1* – such as *RUNX1-ETV6*, *RUNX1-EVI1* – have been described as risk factor of leukaemia, especially when associated with additional cooperatives oncogenic mutations (Sood et al., 2017). For example, the t(8;21) generates an in frame RUNX1-ETO fusion protein that can bind to RUNX1 target genes through the Runt domain. Stability of the fusion protein is enhanced,

and able to bind to multiple other transcriptional regulators via the four nervy homology regions (NHR1-4) of ETO (Peterson and Zhang, 2004). This transcriptional complex was shown to promote the self-renewal capacity of haematopoietic stem cells, but to be insufficient to induce leukaemia on its own (Higuchi et al., 2002). Overall, *RUNX1* has been found to be implicated in more than 50 chromosomal translocations leading to paediatric acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS) (Bellissimo and Speck, 2017; Sood et al., 2017). A number of mono and bi-allelic somatic mutations of *RUNX1* have been documented in MDS, AML, ALL and chronic myelomonocytic leukaemia (CMML) (Osato, 2004). *RUNX1* has been associated with hereditary disease with the presence of germline mono-allelic mutations causing familial platelet disorder with predisposition to AML (FPD/AML) (Song et al., 1999). In addition to these genomic alterations, high levels of RUNX1 mRNA have been reported in AML, T cell-ALL (T-ALL) and B cell-ALL (B-ALL) (Sun et al., 2019). Interestingly, the presence of a functional wild-type *RUNX1* appears to be required for the survival of specific leukaemia, including *RUNX1-ETO* (t8;21), *CBFB-SMMHC* (inv16), *MLL-AF9* and *CBFB-MYH11* (inv16) translocation leukaemia (Ben-Ami et al., 2013; Goyama et al., 2013; Hyde et al., 2015). Depletion of RUNX1 has been shown to lead to compensation by the other RUNX family members (Morita et al., 2017a; Morita et al., 2017b), but pan-inhibitors of RUNX factors have shown promising anti-tumour effects in pre-clinical settings by small-molecule inhibition of RUNX/CBFβ complexes (Illendula et al., 2016), or via Pyrrole-imidazole polyamides, which selectively target the consensus RUNX motif in chromatin to prevent transcription (Morita et al., 2017a).

1.5.2. Emerging roles of RUNX1 in solid cancers

A growing body of research suggests a role for RUNX1 in various non-haematopoietic tissues of epithelial origin. Indeed, high throughput next-generation sequencing has revealed relatively high frequencies of genomic alterations of *RUNX1*, and *CBFB* in solid cancers (Blyth et al., 2005; Ito et al., 2015), albeit to lower levels than in leukaemia (Figure 1.11). Interestingly, while it is yet to

be fully determined to what extent these alterations contribute to tumour biology, mutations of *RUNX1* have been associated with loss of function (van Bragt et al., 2014). Beyond the presence of these mutations, earlier studies identified *RUNX1* mRNA as part of a 17-gene signature associated with metastasis in a panel of adenocarcinomas, including breast and prostate cancers, with its expression inversely correlating with tumour aggressiveness (Ramaswamy et al., 2003). Overall, under- and over-expression of endogenous *RUNX1* has been found in several solid tumours, suggesting that it is broadly implicated in the biology and pathology of epithelial tissues (Blyth et al., 2005; Ito et al., 2015; Scheitz et al., 2012).

Figure 1.11. Meta-analysis of RUNX1 alterations in the TCGA PanCancer atlas.
(A) Frequency of RUNX1 genomic alterations across the TCGA PanCancer atlas (cBioPortal). Cancers with no alterations were excluded. Cancers affecting the haematopoietic system are coloured in pink, hormone related cancers in blue, cancers of soft tissues in green, and other epithelial cancers in grey. (B) Proportion of RUNX1 amplification, homozygous deletion, fusion and mutation in cancers affecting the haematopoietic system, hormone

related cancers, and additional epithelial cancers. Soft tissue cancers were excluded from these analyses due to the small number of patients affected.

1.5.3. RUNX1 in female-related cancers

The role of RUNX1 in solid tumorigenesis has been best studied in female-related cancers, including tumours of the mammary gland (Riggio and Blyth, 2017). Several studies involving large patient cohorts have identified recurrent *CBFβ* and *RUNX1* mutations (Banerji et al., 2012; Cancer Genome Atlas Network, 2012; Ellis et al., 2012; Kas et al., 2017; Nik-Zainal et al., 2016; Pereira et al., 2016). At the protein level, high-grade primary breast tumours also displayed generally reduced levels of RUNX1 compared to low/mid-grade tumours (Kadota et al., 2010). These observations have led to the hypothesis that RUNX1 could have a tumour suppressor function. The proliferation of ER+ breast cancer cells was increased upon RUNX1 knockdown, which led to oestrogen-mediated AXIN1 suppression and enhanced β-catenin activation (Chimge et al., 2016). In agreement with a tumour suppressor role, a link has emerged between RUNX1 and suppression of the epithelial-to-mesenchymal transition (EMT) process. Indeed, downregulation of RUNX1 in the normal mammary epithelial cell line MCF10A was sufficient to induce hyperproliferation and abnormal morphogenesis (Wang et al., 2011b). The morphological changes observed upon RUNX1 knockdown were characteristic of an EMT and associated with the activation of Transforming Growth Factor (TGF)-β and Wingless/Integrated pathway (Wnt) signalling pathways (Hong et al., 2017). Both RUNX1 and RUNX3 were also shown to prevent the induction of YAP-mediated EMT in this same cell line (Kulkarni et al., 2018). Likewise, the RUNX1-CBFβ complex was able to prevent the migration potential of the ER+ breast cancer cell line MCF7 in an ER-dependent manner (Pegg et al., 2019). The emerging role of RUNX1 in EMT was not unexpected considering its well documented role in EHT, a process often referred to as ‘EMT-like’ (Hamidi and Sheng, 2018; Monteiro et al., 2016). However, while RUNX1 is critical for the induction of EHT during haematopoietic development, it appears to act as a gatekeeper of EMT in breast cancer cells.

In contrast to its putative tumour suppressive functions, RUNX1 is also believed to be associated with oncogenic roles. Indeed, higher *RUNX1* mRNA levels were found in the triple-negative breast cancer subgroup (Karn et al., 2011; Rody et al., 2011). This was later corroborated by a strong correlation between high RUNX1 protein levels and poor prognosis in triple-negative and ER-negative breast cancers (Ferrari et al., 2014). Increased expression of RUNX1 was also associated with disease progression in patient samples and in the MMTV-PyMT mouse model. Interestingly, the invasiveness of the cells isolated from this mouse model could be repressed by knocking-down *Runx1* expression (Browne et al., 2015), suggesting that its role in EMT may be context-dependent.

In addition to breast cancer, overexpression of RUNX1 was correlated with overexpression of p21WAF1/CIP1 in invasive endometrioid carcinoma, where it was suggested to play a role in promoting myometrial infiltration (Planagumà et al., 2004; Planagumà et al., 2006). In this respect, Doll and colleagues found that ectopic overexpression of *RUNX1* in the endometrial cancer cell line HEC1A was associated with the establishment of distant metastasis (Doll et al., 2009). High levels of RUNX1 were also reported in human epithelial ovarian tumours, and its knockdown in the SKOV-3 cell line led to a decrease in proliferation, migration, and invasion (Keita et al., 2013).

1.5.4. RUNX1 in prostate cancer

Although less substantial than in female-related cancers, there is accumulating evidence for a potential role of RUNX1 in prostate cancer. Single-Nucleotide Polymorphisms (SNPs) within the *RUNX1* gene – such as the rs2253319 polymorphism – were associated with an increased risk of prostate cancer progression and metastasis (Huang et al., 2011). *RUNX1* was also found amplified in a significant proportion of neuroendocrine castration-resistant prostate cancer (Beltran et al., 2016). However, the biological relevance of these alterations, if any, remains unknown. Contrasting studies looking at RUNX1 expression in prostate cancer have reported that *RUNX1* mRNA increases with pathological stage (Yeh et al., 2009), while protein levels have been reported to be decreased in advanced

forms of the disease (Takayama et al., 2015). Interestingly, the links between RUNX1 and hormones reported in breast cancer (Riggio and Blyth, 2017) seem to extend to the prostate gland which is particularly rich in androgens. In *Nkx3-1*/*Pten* mutant mice, prolonged exposure to reduced androgens levels resulted in prostate *tumours* with up-regulated *Runx1* (Banach-Petrosky et al., 2007). *RUNX1* has also been shown to be a downstream target of AR signalling and is thought to play divergent roles in AR-dependent and castration-resistant prostate cancer cell lines (Takayama et al., 2015). With regards to the growing importance of stroma-cancer interactions, downregulation of RUNX1 expression in mesenchymal stem cells was shown to reduce their proliferative potential in response to TGF-β, before their differentiation into prostate cancer-associated myofibroblasts (Kim et al., 2014b).

1.5.5. Other tissues

RUNX1 has been linked to skin cancer in mice, where its activated expression during chemically induced skin carcinogenesis was proposed to be oncogenic (Hoi et al., 2010). In line with this, loss of *RUNX1* impaired the proliferation of human oral and skin squamous cell carcinoma cell lines (Scheitz et al., 2012).

RUNX1 has also been linked with tumours of the gastrointestinal tract, where it was found to be frequently downregulated (Miyagawa et al., 2006; Sakakura et al., 2005). In conditional mouse models, *Runx1* deletion is sufficient to induce intestinal tumorigenesis (Fijneman et al., 2012). In gastric cancer cell lines, both the knockdown of RUNX1 and its therapeutic inhibition resulted in reduced tumorigenic potential via suppression of the ErbB2/HER2 signalling pathway (Mitsuda et al., 2018).

Finally, the previously noted emerging link between RUNX1 and EMT has also been documented in colorectal cancer (Li et al., 2019), and renal fibrosis (Zhou et al., 2018) in which RUNX1 acts as an inducer of EMT. Increased expression of RUNX1 was also predictive of poor prognosis in patients diagnosed with clear cell renal cell carcinoma (ccRCC) (Fu et al., 2019; Rooney et al., 2020). Functionally, deletion of RUNX1 in ccRCC decreased tumour burden and viability both *in vivo* and *in vitro* (Rooney et al., 2020).

1.6. Project aims

PCa is the most commonly diagnosed cancer in men, and while its incidence continues to rise, there is a need to identify biomarkers to improve patient stratification and clinical management in early-stage PCa patients. The transcription factor RUNX1 has been mainly studied in the context of developmental and adult haematopoiesis as well as haematological malignancies. However, roles for RUNX1 have recently been described outside the haematopoietic system, suggesting a broader implication than initially thought. While the contribution of RUNX1 to epithelial tissue biology and pathology remains less documented than in the haematopoietic setting, increasing evidence suggest that RUNX1 is involved in hormone-regulated epithelia. On this basis, I hypothesised that RUNX1 plays a role in normal prostate homeostasis and prostate cancer.

To test this hypothesis, the main aims of my PhD were as follows:

- Characterise the expression pattern of RUNX1 and the biological potential of RUNX1+ cells in the normal adult mouse prostate
- Evaluate the impact of androgen deprivation on RUNX1 expression and RUNX1+ cells
- Investigate the dynamics of RUNX1 expression and the developmental origin of RUNX1+ cells during prostate organogenesis
- Characterise the expression of RUNX1 in the human prostate and determine whether RUNX1 levels correlates with specific clinical outcomes in patient samples of prostate cancer

Chapter 2 Material and methods

[This section has been adapted from (Mevel et al., 2020)]

2.1. Animal studies

2.1.1. Animal husbandry

Animal experiments were approved by the Animal Welfare and Ethics Review Body (AWERB) of the Cancer Research UK Manchester Institute and conducted according to the UK Home Office Project Licence (PPL 70/8580). Genetic lineage-tracing experiments were performed at the Beatson Biological Services Unit (PPL 70/8645 & P5EE22AEE) and approved by the University of Glasgow AWERB. Mice were maintained in purpose-built facility in a 12-hour light/dark cycle with continual access to food and water.

2.1.2. Animal breeding

2.1.2.1. Mouse lines

Immunocompetent wild-type ICR (CD-1) mice were purchased from Envigo. Isoform specific *Runx1* reporter mouse models P1-*Runx1*:GFP (PSE) and P2-*Runx1*:RFP (PSN) have been described previously (Draper et al., 2018; Sroczynska et al., 2009), and maintained on a ICR (CD-1) background. The C57Bl/6J *Runx1*$^{mER\text{-}Cre\text{-}mER}$ line (Samokhvalov et al., 2007) was provided by RIKEN (Japan). The *Rosa26*^{CreERT2} line was generated previously (Ventura et al., 2007) and obtained from Prof Angeliki Malliri. The C57Bl/6J *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}tdRFP}$ line (Luche et al., 2007) were acquired from the European Mouse Mutant Archive (EMMA). For lineage-tracing experiments, *Runx1*$^{mER\text{-}Cre\text{-}mER}$ and *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}tdRFP}$ mice were housed and crossed in the Beatson Biological Services Unit to generate *Runx1*^{CreER} *Rosa26*$^{LSL\text{-}RFP}$ mice. The maintenance of these lines was organised by Susan Mason, Dr Laura Galbraith and Prof Karen Blyth. The *Runx1*$^{flox/flox}$ line was generated previously and obtained from Prof Nancy Speck (Growney et al., 2005), with loxP sites flanking the Exon 4 which encodes the *runt* DNA binding domain. The *Probasin*-Cre4 (*Pb-Cre4*)

line (Wu et al., 2001), *Rosa26*$^{\text{loxP-STOP-loxP-EYFP}}$ line (Ye et al., 2003), and the *Pten*$^{\text{flox}}$ line (Lesche et al., 2002) were generated previously and kindly provided by the laboratory of Dr Esther Baena at the CRUK MI. The *Trp53*$^{\text{flox}}$ line (JAX stock #008462) was generated previously (Marino et al., 2000) and kindly provided by the laboratory of Prof Nic Jones at the CRUK MI. For all transgenic lines, routine genotyping was undertaken at weaning (3 weeks of age) by automated PCR genotyping (Transnetyx). All mice were maintained on a mixed background. Except for the *Runx1*$^{\textit{Cre}\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ line, all breeding requests were carried out by the CRUK MI Transgenic Breeding Unit.

2.1.2.2. Generation of prostate specific Runx1 mutant mouse models

Three different mouse lines were generated for the purpose of this work by taking advantage of previously described transgenic mouse models (see section 2.1.2.1): PBRX, PRYP, PBTR. Of note, experiments carried out using these mice are not described in this thesis, and only referred to in chapter's discussions.

- PBRX (*Pb-Cre4* ; *Runx1*$^{\text{flox}}$; *Rosa26*$^{\text{loxP-STOP-loxP-EYFP}}$) was generated to study the effect of *Runx1* deletion in the prostate driven by the Pb-Cre4 transgene, by crossing *Runx1*$^{\text{flox}}$ onto *Rosa26*$^{\text{loxP-STOP-loxP-EYFP}}$ mice followed by *Pb-Cre4* mice. Of note, only males were used to breed the *Pb-Cre4* transgene due to misexpression in the oocytes.
- PRYP (*Runx1*$^{\text{flox/flox}}$; *Trp53*$^{\text{flox/flox}}$; *Pten*$^{\text{flox/flox}}$; *Rosa26*$^{\text{loxP-STOP-loxP-EYFP}}$) were generated to then be crossed onto PBRX to generate an intermediate step to obtain PBTR. First, *Runx1*$^{\text{flox/flox}}$ were crossed onto *Trp53*$^{\text{flox/flox}}$ lines. These were subsequently crossed onto the *Pten*$^{\text{flox/flox}}$; *Rosa26*$^{\text{loxP-STOP-loxP-EYFP}}$ line and bred to full homozygosity.
- PBTR (*Pb-Cre4* ; *Runx1*$^{\text{flox/flox or flox/WT or WT/WT}}$; *Pten*$^{\text{flox/WT}}$; *Trp53*$^{\text{flox/WT}}$; *Rosa26*$^{\text{loxP-STOP-loxP-EYFP}}$) were generated by crossing PBRX males onto PRYP females to study the impact of *Runx1* loss in a mouse model of prostate tumorigenesis.

2.1.2.2. Timed matings

Timed mating experiments were set up by the CRUK MI Transgenic Breeding Unit. Vaginal plug detection was considered as embryonic day (E) 0.5. Pregnant females were culled at specific time points and the uterine horn containing embryos placed in Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 10% Foetal Calf Serum (FCS; Sigma), 2 mM L-Glutamine (Invitrogen), 50 µg/mL penicillin/streptomycin (Invitrogen) and kept at 4˚C prior dissections.

2.1.3. Surgical procedures

All animal procedures were performed on adult males at least 7 weeks of age. Surgical castrations in CRUK MI (Manchester, UK) were carried out by Ivana Steiner. Surgical procedures for *in vivo* genetic lineage-tracing experiments were done in CRUK Beatson Institute (Glasgow, UK) and performed by Susan Mason and Laura Galbraith. Surgical castrations were carried out under aseptic conditions. In brief, an incision was made in the scrotum of the mouse to access the testicles. The tunica of each testicle was incised in order to pull out testis, vas deferens, and surrounding testicular fat pad, which were then cut by cauterisation. The scrotum incision was then closed using absorbable Vicryl suture. For prostate regeneration assays, testosterone pellets (Belma Technologies) were implanted subcutaneously. For *in vivo* genetic lineage-tracing experiments, tamoxifen (Sigma, #T5648) was resuspended in ethanol and diluted in corn oil at a concentration of 10 mg/mL and administered via intra-peritoneal injections daily for 4 consecutive days using the following regimen: 3 mg, 2 mg, 2 mg, 2 mg. For BrdU (5-bromo-2-deoxyuridine) labelling experiments, 250 µL doses of pre-made Cell Proliferation Labelling Reagent (GE Healthcare, #RPN201) were administered via intra-peritoneal injections daily for 3 consecutive days.

2.1.4. Tissue harvesting

Mice were sacrificed by cervical dislocation (Schedule 1). The abdomen was sprayed with 70% ethanol and an incision was made through the skin and

peritoneum to expose abdominal contents. The entire urogenital system was removed, including the bladder, seminal vesicles, urethra and prostate, and placed in a Petri dish containing a solution of phosphate buffered saline (PBS). The next part of the dissection was performed under a dissecting microscope using fine instruments to remove the surrounding fat tissue and vas deferens. To isolate the prostate, the seminal vesicles were first carefully cut at their base, followed by the bladder. The prostate was then isolated from the urethra and further cleaned from any remaining fat tissue. The ampullary gland, which is located between the two anterior lobes, was carefully excised, and disposed. To isolate individual lobes, the anterior lobes were separated from the remaining fraction. Finally, the ventral lobes were isolated from the dorsolateral lobes. Individual lobes were placed in 1.5 mL tubes, containing either dissociation media or formalin.

2.2. Human samples

2.2.1. Ethical approval

Sections of the human prostate cancer tissue micro-arrays (TMAs) were obtained from the MCRC Biobank under the MCRC Biobank Project reference 15_GELA_01. Slides from human prostatectomy specimens were obtained from the MCRC Biobank under the MCRC Biobank Project reference 16_RIMA_06 Amendment 6.

2.2.2. Human Prostate Cancer Tissue Micro-Arrays

2.2.2.1. General description

Tissue Micro-Arrays were previously constructed using 1 mm cores by the Genito-Urinary Cancer Research group (GUCR, University of Manchester) and the CRUK MI Histology Core Facility. Two cohorts of samples were used, originally consisting of 2,474 cores from 343 patients, and comprising samples from patients who underwent transurethral resection of the prostate (“TURP” cohort) or transrectal ultrasound guided needle core biopsies (“biopsy” cohort) between 1994 and 2004 in Salford (Table 2.1). Serial TMA tissue sections were reviewed by Dr

Jonathan Shanks, Consultant Urological Histopathologist at The Christie Hospital. All cores were categorised by histopathological type as either 'normal-adjacent core' or 'tumour core' and referred to as 'normal' or 'tumour' thereafter.

Table 2.1. Composition of the PCa TMA cohorts.

Cohort	Number of TMA blocks	Number of cores	
		Normal-adjacent	Tumour
TURP	25	979	1034
Biopsy	8	10	451

2.2.2.2. Clinical annotation

Survival data and additional clinical annotations were updated and reviewed in 2017 by the GUCR group (University of Manchester). The complete clinical annotations are available in supplementary data 1. The following clinically relevant parameters were used for analysis:

- Anonymised patient ID
- Date of birth
- Date of diagnosis
- Age at diagnosis (in years)
- Gleason Score (GS)
- Clinical tumour stage at diagnosis (cT1-4, Tx)
- PSA at diagnosis (ng/mL)
- Metastasis stage at diagnosis (M0, M1, Mx)
- Vital status at last follow-up (alive or dead)
- Time to last follow-up or death (in months)
- Cause of death

Patients were further categorised into groups according to Gleason, PSA, Clinical T stage variables as follows:

- ISUP's Gleason Grade Group: according to the 2016 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grade Group

of Prostatic Carcinoma (Epstein et al., 2016b): GGG1 (GS ≤6), GGG2 (GS 3+4), GGG3 (GS 4+3), GGG4 (GS 8), and GGG5 (GS 9-10)

- PSA Group: 0-0.9 ng/mL, 1-19.9 ng/mL, > 20 ng/mL
- Clinical T (cT) Group: T1+T2, T3+T4
- Risk Group: low risk (GGG1, PSA 0-9.9, T1, M0), intermediate risk (GGG2-3, PSA 10-19.9, T3-4, M0), high risk (GGG4-5, PSA >20, T3-4, M0)

2.2.2.3. TMA maps

TMA maps consisting of anonymised patient identifiers (IDs) and core histopathological annotations ('normal' or 'tumour') were generated by the GUCR group. This core-level database was then linked to the annotated clinical database using patient IDs. Cores lacking associated patient clinical data were excluded from the analysis.

2.2.3. Radical Prostatectomy TMA

A TMA was constructed from patients who underwent radical prostatectomy (RP). Patients with varying Gleason grades, with and without the presence of lymph node metastases (LNM) were included in the construction (Table 2.2). The complete clinical annotations are available in supplementary data 2.

Table 2.2. Characteristics of the patients included in the RP TMA.

Lymph Node status at surgery	Diagnostic Gleason Grade Group	Diagnostic PSA (mean ng/mL - IQR)
LN 0 (n = 9)	2 (n = 2)	18.0 - 8.0
	3 (n = 3)	11.3 - 4.9
	4 (n = 1)	16.0 - 0.0
	5 (n = 3)	11.2 - 6.5
LN 1 (n = 11)	2 (n = 3)	14.7 - 9.0
	3 (n = 4)	13.1 - 6.9
	5 (n = 4)	10.4 - 4.3

IQR: inter-quartile range.

A pathologist (Dr. Pedro Oliveira, Consultant Urological Histopathologist at The Christie Hospital) selected multiple areas of normal-adjacent and tumour tissue

from resected RP specimens. At least 3 replicate cores were punched from each annotated region, resulting in a total of 169 cores. The pathologist annotated histologically distinct regions as either “Normal”, “Atrophy” or “Inflammatory (Infl.) Atrophy”, and tumour glands were scored using the Gleason grading system: Gleason (G) 3, 4 or 5. LNM tissue samples were not included in the original TMA construction and stained separately on whole tissue sections.

2.2.4. Human Benign Prostatic Hyperplasia samples

Formalin-fixed paraffin-embedded prostate blocks from patients undergoing Trans-Urethral Resection of the Prostate (TURP) were obtained from the GUCR group (University of Manchester). Sixteen blocks were selected from patients with varying sizes of Benign Prostatic Hyperplasia (BPH) nodules (range: 3-56 g, median: 28 g). The complete clinical annotations are available in supplementary data 3.

2.2.5. Human Radical Prostatectomy specimens

Formalin-fixed paraffin-embedded whole prostate blocks from patients undergoing prostatectomy for prostate cancer were obtained from the MCRC Biobank. All samples were assessed by Dr. Pedro Oliveira. Seven blocks were selected from patients having minimal tumour involvement (< 10%) and low Gleason scores (3+3 or 3+4) to study normal prostate epithelia. Regions corresponding to the following pathological features were annotated on serial H&E sections: ‘Normal Secretory’ epithelium, ‘BPH’, ‘Cystic atrophy’ and ‘Inflammatory atrophy’.

2.3. Cell culture methods

2.3.1. Cell lines

Cells were maintained in standard culture conditions, at 37°C, 5% CO2 and 95% humidity. Human Embryonic Kidney 293 (293T), MDA-MB-231 and VCaP cells were cultured in phenol-red free Dulbecco’s Modified Eagle’s Medium (Sigma) supplemented with 10% FCS (Sigma), 2 mM L-Glutamine (Invitrogen), 50 µg/mL

penicillin/streptomycin (Invitrogen). LNCaP, DU145 and Jurkat cells were maintained in phenol-red free RPMI-1640 Medium (Sigma) supplemented with 10% FCS, 2 mM L-Glutamine, 50 µg/ml penicillin/streptomycin. LCR cells were grown in phenol-red free RPMI-1640 Medium supplemented with 10% charcoal-stripped FBS (#10611235; Fischer Scientific), 2 mM L-Glutamine, 50 µg/mL penicillin/streptomycin. PC-3 and PC-3/m cells were maintained in Nutrient Mixture F-12 Ham medium (Sigma) supplemented with 10% FCS, 2 mM L-Glutamine, 50 µg/mL penicillin/streptomycin. RWPE-1 cells were cultured in Keratinocyte-SFM medium (Thermo Fischer Scientific; #17005-042). Cell line authentication and mycoplasma testing was performed by the Molecular Biology Core Facility of the CRUK MI.

2.3.2. Lentiviral production and purification

A third generation HIV derived lentiviral vector packaging system was used to produce lentiviruses. It consists in three plasmids (pMDG.2, pRSV-Rev and pMDL) encoding essential viral proteins. 293T cells were transfected using the polyethylenimine (PEI) method at ~60% confluence in T75 flasks. 48 µL of PEI were diluted in 1 mL of DMEM 10% FCS and mixed with 1 mL of DMEM 10% FCS containing the following quantities of each plasmid: 2.6 µg of pMDG.2, 1.8 µg of pRSV-Rev, 4.8 µg of pMDL, and 7.4 µg of the LT3GEPIR vector containing the respective shRNA sequence. The mixture was vortexed for 1 min and incubated for 20 min at RT before addition to the 293T cells. The media was replaced 12h later day with 8 mL of antibiotic-free culture medium. After 24h and 48h, the supernatant containing lentiviruses was collected, filtered through 0.45-µm filters, and immediately used to transduce recipient cells. PC-3 cells were incubated with the virus for 4h, and 2h for the RWPE-1. Cells were amplified and selected in puromycin (1-2 µg/mL) 48h after infection and for a minimum of 2 weeks. Induction of shRNA expression was carried out by addition of doxycycline (1 µg/mL) and controlled by GFP co-expression.

2.3.3. Isolation of primary mouse prostate cells

All dissections were performed under a stereo microscope in sterile PBS. Dissociated murine prostate cells were obtained by digesting pre-minced prostate tissue for 1h at 37˚C in digestive medium prepared in prepared in ADMEM/F12 (Gibco) and containing 1 mg/mL Collagenase Type I (ThermoFischer Scientific, #17018029), 1 mg/mL Dispase II (ThermoFischer Scientific, #17105041), 10% Fetal Bovine Serum (Gibco), 1% Penicillin-Streptomycin-Glutamine (Sigma), and 10 µM Y-27632 dyhydrochloride (Chemdea, #CD0141). For embryonic urogenital sinuses (UGS), dissociation time was reduced to 30 min. Single cells were obtained after an additional 10 min incubation in TrypLE (Gibco) at 37˚C before mechanical dissociation with a syringe and needle (25G). Cells were then filtered through a 50 µm cell strainer.

2.3.4. Primary mouse prostate organoid cultures

In vitro organoid formation assays were performed as described in Drost et al., 2016. Single cells were resuspended in 40 µL drops of phenol red-free Cultrex Reduced Growth Factor Basement Membrane Extract Type 2 (RGF BME 2, Amsbio, #3533-005-02), and seeded in CellCarrier-96 Ultra Microplates (PerkinElmer, #6055302). Defined organoid culture medium was prepared using Advanced DMEM/F-12 (Gibco), supplemented with 10 mM Hepes (Sigma), Gutamax (Gibco), penicillin/streptomycin (Sigma), B27 (Life Technologies, 17504-044), 50 mg/mL Epidermal Growth Factor (EGF, PeproTech, #AF-100-15), 500 ng/mL R-spondin 1 (R&D Systems, #4645-RS), 100 ng/mL Noggin (R&D Systems, #6057-NG), 10 mM Y-27632 dyhydrochloride (Chemdea, #CD0141), 200 nM A83-01 (Tocris Bioscience, #2939), 1.25 mM N-Acetylcystein (Sigma), and 1 nM DHT (Sigma #730637). Medium was refreshed every 2-3 days, and organoid cultures were scored after 7 days.

For adenoviral transductions, Ad5CMVCre-eGFP (#VVC-U of Iowa-1174) or Ad5CMVeGFP (#VVC-U of Iowa-4) adenoviral vectors were obtained from the Gene Transfer Core of the University of Iowa. Viral particles were added at a Multiplicity

of Infection (MOI) of 10 in a 50 µL cell suspension and incubated for 1h at 37 °C. Before plating, viral particles were washed out and cells were re-suspended in ~10 µL of culture medium before addition of RGF BME 2 and culture medium.

2.3.5. UGS explant cultures

UGS explant cultures were performed as described in Kruithof-de Julio et al., 2013. Briefly, E15.5 embryos were obtained from timed matings. UGS were isolated from embryos and cultured using a Durapore Membrane Filter 0.65 µm (#DVPP02500) placed on a stainless-steel mesh for up to 7 days in Ham's F-12/DMEM (Gibco) supplemented with Insulin-Transferrin-Sodium Selenite Supplement (Roche) and 10 µM DHT (Sigma). Media were renewed every 2-3 days. For lineage-tracing experiments, tamoxifen-induced labelling was performed using 0.5 µM 4-hydroxytamoxifen (Sigma, #T176).

2.4. Fluorescence-Activated Cell sorting (FACS)

2.4.1. Sample staining and preparation

Single cell suspensions were kept in Advanced DMEM/F-12 (Gibco) containing 5% FBS (Foetal Bovine Serum) supplemented with 10 µM Y-27632. Cells were incubated for 10 min using unconjugated anti-mouse CD16/32 antibody (Biolegend, C93, #101301) at 4˚C prior to labelling with specific fluorochrome-labelled antibodies for at least 15 min in the dark at 4˚C. Details of FACS reagents and antibodies are listed in table 2.3. Cells were then washed and filtered through a 50 µm cell strainer prior to acquisition. Hoechst 33258 (1 µg/mL, Sigma) or Sytox blue (1 µL/mL, ThermoFischer Scientific) were used as viability stains.

Table 2.3. List of antibodies used for FACS.

Target	Provider	Reference	Clone	Fluorophore
CD16/32	Biolegend	101301	93	none (Fc block)
CD45	Biolegend	62-0451-82	30-F11	SB436
EPCAM	Biolegend	118225	G8.8	BV421
EPCAM	Biolegend	118214	G8.8	APC
CD49f	Biolegend	313606	GoH3	FITC

CD49f	Biolegend	313616	GoH3	APC
SCA-1 (Ly-6A/E)	Biolegend	108108	D7	PE
CD24	BD Biosciences	744470	M1/69	BV786
TROP-2	R&D	FAB1122F	-	FITC

2.4.2. Sample acquisition

Single-cell suspensions were analysed on a Fortessa (BD Biosciences) and sorts were performed on a FACSAriaIII (BD Biosciences) with the help of the Flow-Cytometry Core Facility. FACS data were analysed using FlowJo software (BD Life Sciences). An example of the gating strategy used for analysing or sorting mouse prostate epithelial cells from individual prostate lobes is shown in figure 2.1.

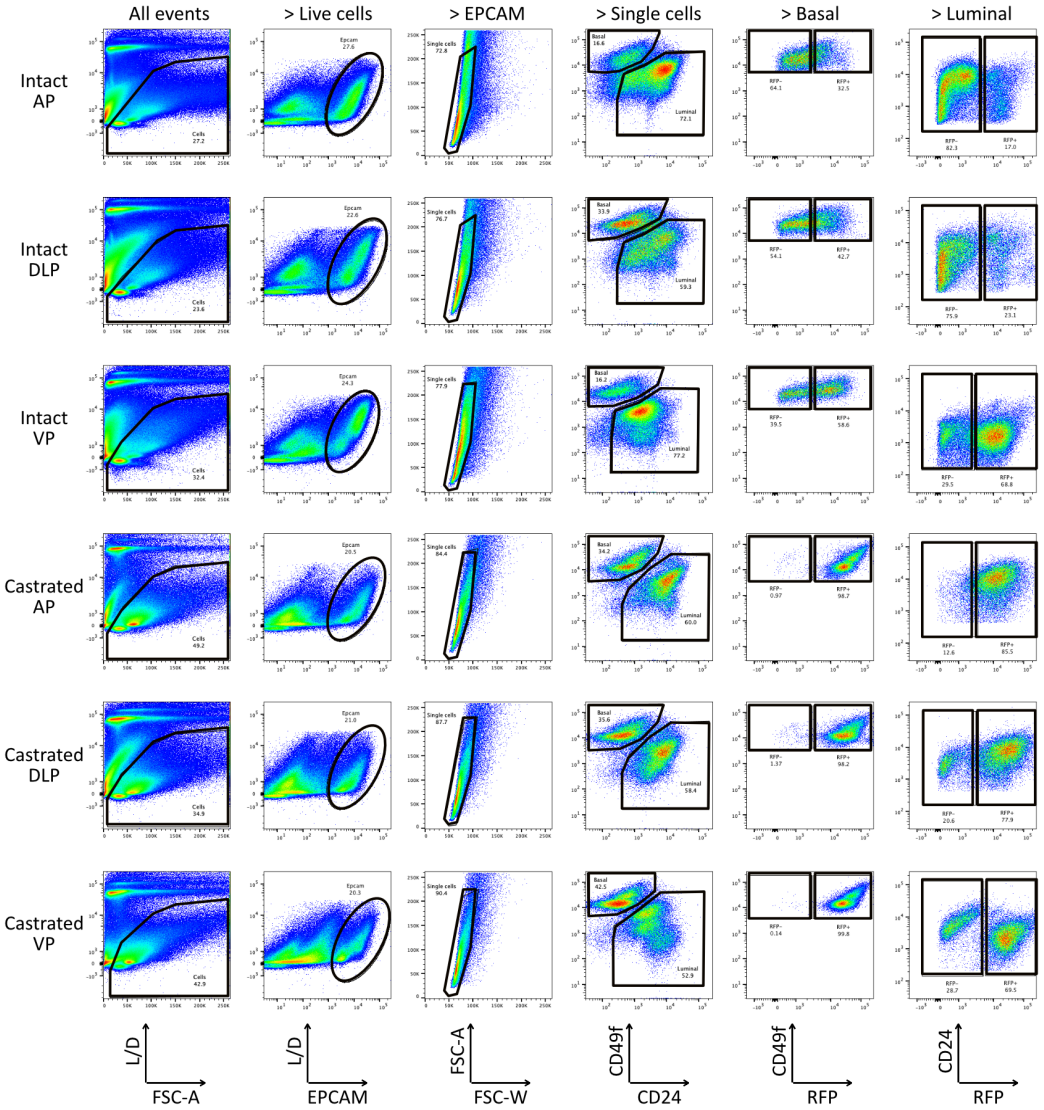

Figure 2.1. Example gating strategy to sort or analyse mouse prostate epithelial cells.

Examples are shown for cells isolated from individual lobes of intact and castrated mice. AP: Anterior Prostate, DLP: Dorsolateral Prostate. VP: Ventral Prostate. L/D: Live/Dead marker.

2.5. Molecular and biochemical assays

2.5.1. Genotyping Polymerase Chain Reaction (PCR)

For DNA extraction, cells were first incubated at 95˚C for 8 min. Then, 0.8 mg/mL of Proteinase K (Sigma) diluted in 0.1X PBS was added to each sample followed by incubation at 55˚C for 30 min, and further heat-inactivation at 96˚C for 8 min. For mouse *Runx1* genotyping PCR, the following reaction was set up in a final volume of 20 µL 10 µl GoTaq Green Master mix (Promega), 0.8 µL U17 (10 µM), 0.8 µL U18 (10 µM stock), 1.6 µL AN94 (10 µM stock), 6.8 µl extracted DNA. The PCR conditions were as follows: 2 min denaturation at 95°C; 35 cycles of 95°C x 30 sec, 60°C x 30 sec, and 72°C x 60 sec; and 5 min final extension at 72°C. The following primers were used: U17 5'-GAGTCCCAGCTGTCAATTCC-3', U18 5'-GGTGATGGTCAGAGTGAAGC-3', AN94 5'-CCAAGATAGTCCTTAACGGTCG-3'.
PCR products were loaded onto an agarose gel 0.8% to check the approximate size of the amplified fragments. Expected bands were ~450 bp for Floxed alleles, ~380 bp for WT and ~280 bp for excised alleles.

2.5.2. Real-Time Quantitative Polymerase Chain Reaction (RT-qPCR)

RNA was isolated using the RNeasy Micro Kit according to the manufacturer's instructions (Qiagen). Reverse transcription of 1 µg of extracted RNA was carried out using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fischer Scientific). Subsequently, 0.3 µL of synthesised cDNA was used in the RT-qPCR reaction prepared with SYBR Green Master Mix (Life Technologies) or TaqMan Universal PCR Master Mix (Applied Biosystems) and performed in a QuantStudio 5 Real-Time PCR System (Thermo Fischer Scientific). The reaction was set up in a total volume of 10 µL using 0.1 µL of the appropriate Exigon Universal Roche Probe, 0.1 µM of each primer (Sigma) and 5 µL of the relevant master mix. Oligonucleotides

used for real-time PCR are listed in table 2.4. Gene expression was calculated relative to *GAPDH* using the ΔΔCt method.

Table 2.4. List of primers used for RT-qPCR.

Primer ID	Target	Sequence (5' to 3')
BX93	Human *GAPDH* forward	CCTGCACCACCAACTGCTTA
BX94	Human *GAPDH* reverse	CCATCACGCCACAGTTTCC
BX91	Human *RUNX1* forward	ACTCGGCTGAGCTGAGAAATG
BX92	Human *RUNX1* reverse	GACTTGCGGTGGGTTTGTG
BX95	Human *RUNX2* forward	CGGCCCTCCCTGAACTCT
BX96	Human *RUNX2* reverse	TGCCTGCCTGGGGTCTGTA
BY03	Human *RUNX3* forward	TCTGTAAGGCCCAAAGTGGGTA
BY04	Human *RUNX3* reverse	ACCTCAGCATGACAATATGTCACAA

2.5.3. Cloning of shRNA sequences

97-mer oligonucleotides containing the respective shRNA guide sequences (Table 2.5) were designed according to a previously described method (Fellmann et al., 2013).

Table 2.5. Guide sequences of the different shRNA used to target human *RUNX1.*

shRNA ID	Target	Sequence 5' to 3' (shRNA ID)
shRUNX1-1	*RUNX1* (3'UTR)	TGCCTTGAAATACCTGTTTCTT (TRCN0000013658)
shRUNX1-5	*RUNX1* (exon 5)	TTTGTGAAGACAGTGATGGTCA (RUNX1_861_5673)
shRenilla	*Renilla Luciferase*	TAGATAAGCATTATAATTCCTA (Renilla 713)

Oligonucleotides were PCR amplified using primers BL38 miRE-Xho-fw (TGAACTCGAGAAGGTATATTGCTGTTGACAGTGAGCG) and BL39 miRE-EcoOligo-rev (TCTCGAATTCTAGCCCCTTGAAGTCCGAGGCAGTAGGC), 0.05 ng oligonucleotide template, and the High Fidelity Phusion (Thermo Fisher Scientific). The PCR was run in a total volume of 50 µL containing 10 µL of 5X Phusion HF buffer (Thermo Fischer Scientific), 1 µL dNTP (10 mM each), 2 µL primer mix (10 µM each), and 1 µL Phusion High Fidelity DNA polymerase (Thermo Fischer Scientific). Cycling parameters were 95°C for 5 min; 30 cycles of 95°C for 25 sec, 56°C for 25 sec, and 72°C for 30 sec; 72°C for 5 min. Following amplification, the 125 nt shRNA fragments were cloned into the LT3GEPIR vector, containing a lentiviral backbone

compatible with the third generation HIV derived lentiviral packaging system (Fellmann et al., 2013). Each PCR were digested at 37°C for 2h in a final volume of 20 µL containing 16 µL of amplified inserts, 1 µL of EcoRI and XhoI enzymes and 2 µL of 10X Cutsmart buffer (NEB). Digested inserts were purified using the QIAquick PCR Purification Kit (Qiagen). In parallel, 5 µg of vector were digested for 2h at 37°C using 2 µL of the same enzymes and 2 µL of 10X Cutsmart buffer in a final volume of 20 µL. The digested vector was dephosphorylated by addition of 1 µL alkaline phosphatase (Roche) to the mix and incubated for an extra 30 min at 37°C. The digested vector was gel purified using the QIAquick Gel Extraction Kit (Qiagen). Ligations were carried out using the T4 Ligase (Roche) using 20 ng of purified vector and 200 ng of purified inserts. Ligation products were finally transformed using Oneshot TOP10 chemically competent bacteria (Invitrogen) according to standard procedures. Final vectors were sequenced using the primer BN31 miRseq5 (TGTTTGAATGAGGCTTCAGTAC).

2.5.4. Western blotting

Cells harvested for western blotting were washed with cold PBS and lysed using NP-40 lysis buffer containing 150 mM NaCl, 50mM Tris-HCl pH 8, 1% of Nonidet P-40 (Sigma), in the presence of Protease Inhibitor Cocktail (Sigma) and incubated for 30 min at 4°C. The lysate was then centrifuged at 16,000g for 10 min, and the supernatant was saved. Whole protein extracts were aliquoted and stored at -80°C. Protein concentrations were determined using either the BCA protein assay reagent kit (Thermo Scientific) or the Bradford Protein Assay (Bio-Rad). Prior loading, protein preparations were incubated for 10 min at 70°C in gel-loading buffer (NuPage LDS sample buffer and NuPage reducing agent; Invitrogen). Proteins were separated using Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS PAGE) with NuPAGE 4-12% Bis-Tris gels in XCell SureLock™ Mini-Cell Electrophoresis System containing MES buffer (Invitrogen). The SeeBluePlus2 (Invitrogen) marker was used as a molecular weight indicator. Proteins were transferred onto a nitrocellulose membrane using the iBlot® Gel Transfer Stacks Nitrocellulose (Invitrogen) for 13 min. Transferred proteins were visualised by

reversible Ponceau S (Sigma) staining and then washed in TBS-T (TBS-Tween 20, 0.05%). Membranes were blocked with TBS-T 5% BSA (Sigma) for 1 hour at RT and blotted with primary antibodies diluted in blocking buffer for 2h at RT. Following 3 washes in TBS-T (10 min each), membranes were incubated with the appropriate amount of horseradish peroxidase-coupled secondary antibodies. Alternatively, immunoblotting was performed on an iBind™ Western Device (Thermo Fischer Scientific) according to the manufacturer's protocol. The membranes were developed using the Amersham ECLTM Prime Western Blotting Detection Reagent (GE Healthcare), according to the manufacturer's protocol. Films were exposed and developed in a MAS automated developing machine. All antibodies used for western blotting are listed in table 2.6.

Table 2.6. List of primary antibodies used for western blots.

Target	Provider	Reference	Raised specie	Targeted specie	Dilution
RUNX1	Cell Signaling	4336	Rabbit	Human	1:1,000
α-TUBULIN	Cell Signaling	2144	Rabbit	Human	1:2,000
β-ACTIN	Cell Signaling	3700	Rabbit	Human	1:2,000
GAPDH	Cell Signaling	2118	Rabbit	Human	1:2,000

2.6. Histopathological methods

2.6.1. Sample preparation

Tissues were harvested and fixed in 10% buffered formalin for 24h and preserved in 70% ethanol for longer storage. Samples were processed by the staff of the CRUK MI Histology Core Facility. Fixed tissues were processed using standard procedures and embedded in paraffin. Formalin-fixed paraffin-embedded (FFPE) sections (4 μm) were cut and dried overnight at 37°C. For frozen samples, tissues were fixed with 4% paraformaldehyde (PFA, ThermoFisher Scientific, #28906) between 30 min to 24h depending on the size of the tissues. Freezing and sectioning were performed by the CRUK MI Histology Core Facility.

2.6.2. Haematoxylin & Eosin stainings

Haematoxylin & Eosin (H&E) staining was performed by the CRUK MI Histology Core Facility. Sections of 4 µm were cut from FFPE blocks and incubated 3 x 5 min in xylene to remove paraffin. Slides were subsequently re-hydrated in decreasing concentration of ethanol: 3 x 1 min 100% ethanol, followed by 1 min in 90% ethanol and 1 min in 70% ethanol. Slides were then washed in water for 1 min before a 3 min incubation in Shandon Gill Hematoxylin 2 (Thermoshandon, #6765008). Slides were further washed in water for 1 min before incubation in acetic acid (ThermoFisher Scientific, #10394970) for 30 sec. After a 1 min wash in water, slides were incubated in alkaline water for 1 min, washed in water for 1 min and stained with Shandon Eosin-Y (Thermoshandon, #6766008) for 1 min. After a final wash in water for 1 min, slides were dehydrated in ethanol: 15 sec in 70% ethanol, 15 sec in 90% ethanol and 3 times 30 sec in 100% ethanol followed by 1 min in xylene. Pertex (CellPath, #SEA-0100-00A) was used for coverslipping.

2.6.3. Chromogenic stainings

Chromogenic stainings were performed on the Leica BOND-III, BOND-Max or BOND-RX autostainers (Leica Biosystems) with the BOND polymer refine detection system (Leica Biosystems, #DS9800). Briefly, sections were dewaxed, and rehydrated, and endogenous peroxidase activity was quenched by 10 min pre-treatment with 3% hydrogen peroxide diluted in TBS-T (Tris-Buffered Saline 0.05% Tween-20). Bond Wash Solution was used between each step. Following on-board heat-induced epitope retrieval with citrate buffer (pH 6.0) for 20 min, sections were incubated for 10 min with 10% Casein (Vector Laboratories) diluted in TBS-T. Sections were then incubated 30 min with primary antibody diluted in the BOND Primary Antibody Diluent (Leica Biosystem, #AR9352). Following additional washes, sections were incubated with for an additional 15 min with HRP labelled secondary antibodies. Detection was carried out by incubation with the mixed DAB Refine reagent (66 mM 3,3'- Diaminobenzidine tetrahydrochloride hydrate, ≤ 0.1% (v/v) hydrogen peroxide and < 0.1% haematoxylin) for 10 min and then washed in water

for 5 min. Finally, sections were dehydrated in increasing concentration of ethanol (15 sec in 70% ethanol, 15 sec in 90% ethanol and 3 times 30 sec in 100% ethanol), followed by 3 x 1 min washes in xylene. Pertex was used for coverslipping. The list of primary and secondary antibodies used is available respectively in table 2.7 and table 2.8.

2.6.4. Immunofluorescent stainings

Multiplexed immunofluorescent stainings of FFPE sections were performed on an automated Leica BOND RX platform using the Opal multiplexing workflow (PerkinElmer). In brief, sections were dewaxed, and rehydrated, and endogenous peroxidase activity was quenched by 10 min pre-treatment with 3% hydrogen peroxide diluted in TBS-T (Tris-Buffered Saline 0.05% Tween-20). Following on-board heat-induced epitope retrieval with citrate buffer (pH 6.0) for 20 min, sections were incubated for 10 min with 10% Casein (Vector Laboratories) diluted in TBS-T. Each staining cycle included a primary antibody diluted in the BOND Primary Antibody Diluent (Leica Biosystem, #AR9352) for a 30 min incubation, followed by buffer washes, and 30 min incubation with HRP labelled secondary antibodies. After further washes, the Tyramide labelled with a fluorophore (Opal 520, Opal 570 or Opal 650, PerkinElmer) was added for a final 10 min. Subsequent antibody stainings were performed by repeating the same procedure, separated by heat-mediated antibody denaturation using citrate buffer (pH 6.0) for 5 min at 95°C. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI, Sigma) and slides were sealed using ProLong Gold Antifade Mountant (ThermoFischer Scientific). The list of primary and secondary antibodies used is available respectively in table 2.7 and table 2.8.

2.6.5. *In situ* hybridisation

In situ hybridisation (ISH) to detect *Nupr1* (ACD, LS 2.5 Mm-Nupr1 #434818) was done using the Multiplex Fluorescent detection kit (ACD) on the automated Leica BOND RX platform following the manufacturer's instructions. Pre-treatment

was done using an EDTA based pH 9.0 epitope retrieval solution for 15 min at 88°C followed by 10 min protease incubation. After ISH, antibody staining was carried out using an anti-RFP antibody for 1h detected with EnVision HRP anti-rabbit secondary (Agilent) followed by incubation with Tyramide-conjugated Opal 570 (PerkinElmer) as described above. Anti-CDH1 antibody was applied for 1h and detected using an anti-goat Alexa Fluor 647 secondary antibody (ThermoFischer Scientific, #A-21447). The list of primary and secondary antibodies used is available respectively in table 2.7 and table 2.8.

2.6.6. Immunofluorescent staining of frozen sections

Tissue sections were thawed and blocked for 1h in PBS supplemented with 10% FCS, 0.05% Tween20 (Sigma) and 10% of goat serum (Dako). After 3 washes in PBS, sections were incubated at 4°C overnight with primary antibodies diluted in blocking solution, followed by 3 washes of 15 minutes in PBS supplemented with 0.05% Tween (PBS-T). Subsequent incubation with appropriate secondary antibodies diluted in PBS-T was performed for 1h at RT and followed by 3 washes in PBS. Nuclei were counterstained with DAPI (Sigma) and slides were sealed using ProLong Gold Antifade Mountant (ThermoFischer Scientific). Alternatively, tissues were mounted using the Prolong Gold anti-fade medium with DAPI (ThermoFischer Scientific). The list of primary and secondary antibodies used is available respectively in table 2.7 and table 2.8.

2.6.7. List of primary and secondary antibodies

Table 2.7. List of primary antibodies used for immunohistochemistry.

Target	Provider	Reference	Raised specie	Targeted specie	Dilution
RUNX1	Cell Signaling	8529	Rabbit	Mouse	1:200
RUNX1	Cell Signaling	4336	Rabbit	Human	1:200
RUNX1/2/3	Abcam	92336	Rabbit	Mouse	1:200
NKX3.1	Athenaes	AES-0314	Rabbit	Mouse	1:200
NKX3.1	Cell Signaling	83700	Rabbit	Human	1:200
TP63	Cell Signaling	M3562	Rabbit	Mouse / Human	1:400
AR	Abcam	Ab133273	Rabbit	Mouse / Human	1:400
BMI1	Cell Signaling	6964	Rabbit	Mouse	1:200

TROP-2	R&D	AF1122	Goat	Mouse	1:200
LY6D	Proteintech	17361-1-AP	Rabbit	Mouse	1:200
SCA-1	R&D	AF1226	Goat	Mouse	1:200
CDH1	Thermo Fischer	AF748	Goat	Mouse / Human	1:400
K5	Abcam	ab52635	Rabbit	Mouse / Human	1:400
K8	Abcam	ab53280	Rabbit	Mouse / Human	1:400
K4	Abcam	Ab9004	Mouse	Mouse	1:100
K7	Abcam	Ab9021	Mouse	Mouse / Human	1:400
HMWCK	Dako (Agilent)	GA05161-2	Mouse	Human	1:1,000
panCK	Sigma	C2931	Mouse	Human	1:10,000
BrdU	Abcam	ab6326	Rat	Any	1:400
Ki67	Abcam	ab15580	Rabbit	Mouse / Human	1:800
RFP	Rockland	600-402-379	Rabbit	Any	1:400
RFP	MBL	PM005	Rabbit	Any	1:200
GFP	MBL	598	Rabbit	Any	1:200

Table 2.8. List of secondary antibodies used for immunohistochemistry.

Antibody	Provider	Reference
EnVision+/HRP Anti-Rabbit	Dako (Agilent)	K4003
EnVision+/HRP Anti-Rabbit	Dako (Agilent)	K4001
ImmPRESS HRP Anti-Goat	Vector Laboratories	MP-7405
ImmPRESS HRP Anti-Rat	Vector Laboratories	MP-7444
Goat anti-Rabbit IgG 488	ThermoFischer Scientific	A11034
Goat anti-Rabbit IgG 555	ThermoFischer Scientific	A21430
Goat anti-Rabbit IgG 647	ThermoFischer Scientific	A32733
Goat anti-Mouse IgG 488	ThermoFischer Scientific	A-11029
Goat anti-Mouse IgG 555	ThermoFischer Scientific	A-21424
Goat anti-Mouse IgG 647	ThermoFischer Scientific	A-32728
Donkey anti-Goat IgG 647	ThermoFischer Scientific	A-21447

2.6.8. Whole-mount immunofluorescent staining

Whole-mount staining was adapted from Yokomizo et al., 2012. Organoids were fixed directly in 96-well plates using 4% paraformaldehyde for 1h at 4°C. After 3 washes of 5 min in PBS, organoids were incubated in PBS-BST, containing PBS, 1% milk, 1% Bovine Serum Albumin (BSA), 10% goat serum (Agilent, #X090710), 0.4% Triton X-100. Pre-conjugated primary antibodies, K5 Alexa Fluor 647 (#ab193895, Abcam) and K8 Alexa Fluor 488 (#ab192467, Abcam) were diluted at 1/400 in PBS-BST and incubated with the organoids overnight at 4°C on a rocking platform. After

3 washes of 1h in PBS-BST at 4°C, organoids were stained with DAPI at 2 µg/mL diluted in PBS-BST and incubated for another 30 min at 4°C on a rocking platform. Images were acquired on an Opera Phenix High Content Screening System (Perkin Elmer) using the 10x air and 20x water lenses.

2.6.9. Image acquisition and processing

Whole-slide images were acquired either on a Leica SCN400 histology scanner (chromogenic only), a Leica Aperio Versa 200 slide scanner (chromogenic and immunofluorescence), or an Olympus VS120 slide scanner (chromogenic and immunofluorescence). Confocal images were acquired using a Leica TCS SP8 confocal microscope and LAS X Leica software. Images of immunofluorescent stainings were captured using either a low-light microscope (Leica) and the Metamorph imaging software, or a Leica MZ FLIII microscope. Images of chromogenic stainings were captured using an EVOS XL Core Cell Imaging System (ThermoFischer Scientific) or extracted from whole-slide images. QuPath v0.2 (Bankhead et al., 2017) was used to extract representative high-quality raw images of selected areas from whole slide images using the 'Send region to ImageJ' tool. Images were systematically processed with ImageJ (NIH Image, Maryland, USA) for optimal visualisation.

2.7. Image analysis

2.7.1. Whole slide fluorescent images

Whole-slide images were analysed using QuPath v0.2 (Bankhead et al., 2017). First, annotations were drawn manually to select areas of interest. Nuclear detection was achieved using the "cell detection" module on the DAPI channel, and parameters were optimised for each individual batch of stainings. A classifier was then trained for each batch of images using the random forest algorithm, to detect the epithelial layers based on either CDH1 or K5/K8 stainings. Single-cell intensity measurements were exported as comma-separated values files and imported into R (3.6.3) for downstream analysis and visualisation. For Quantitative Imaged-Based

Cytometry (QIBC), single-cell intensity measurements were log10 transformed and plotted using the 'geom_hex' function of the ggplot2 R package. Thresholds were set manually based on the distribution of the intensity measurements and cross-checking the raw images, in order to discriminate between 'positive/high' and 'negative/low' cells.

2.7.2. Tissue Micro-Arrays

TMA analyses were performed using HALO (Indica Labs Inc.). First, the 'TMA' module of HALO was used to segment TMA slides according to each corresponding TMA map, thereby allowing the assignment of a core to a patient identifier and its "normal"/"tumour" status. Each core was manually evaluated on the software to exclude regions or entire cores containing either of the following artefacts: folded tissue, crushed tissue, imaging artefacts. Then, the same procedure as described above (section 2.6.1) was used for the extraction of single-cell intensity measurements: an epithelial classifier was trained based on pan-cytokeratin (panCK) staining, and the 'Highplex FL' module was used for both nuclei detection and intensity measurements. Single-cell intensity measurements exported from HALO as comma-separated values files and imported into R (3.6.3) for downstream analyses and visualisation. Survival analyses were carried out using R packages 'survival' (version) and 'survminer' (version).

For survival analyses, a first threshold was set to classify each cell or nucleus as 'positive' or 'negative' for each marker tested. Then, a second threshold corresponding to the percentage of 'positive' cells per patient was determined using quantiles to classify patients into 'high' and 'low' groups. A Kaplan-Meier survival curve was plotted to evaluate the predictive potential of the biomarker for the chosen thresholds. These values were then used in multivariate analyses (Cox proportional hazards regression model) using the 'coxph()' function of the 'survival' package to evaluate the robustness of the predictions with additional clinical variables. Categorical marker data (High and Low) was adjusted for age, log10-transformed PSA at diagnosis, ISUP Gleason Garde Group, Metastasis stage, Tumour stage and cohort of patient (TURP and biopsy).

2.7.3. Organoids classification

Quantitative analysis of whole-mount immunofluorescent stainings was performed using the Harmony software (PerkinElmer) to classify organoids into three types based on the expression of basal (K5) and luminal (K8) markers. Briefly, maximum projection images of stained and imaged (10x lens) wells were generated. The sum of all fluorescent channels was used to identify organoids as objects. Objects smaller than 30 µm and larger than 250 µm were excluded. Mean fluorescence intensity of individual objects were calculated and stored in a table. Single object fluorescent data were imported into R (v3.6.3), and organoids were classified into the following categories based on mean fluorescence intensity: unipotent "basal-like" Keratin 5+ (K5+), unipotent "luminal-like" Keratin 8+ (K8+), or multipotent (K5+ K8+).

2.8. Next Generation Sequencing data analysis

2.8.1. Single-cell RNA-sequencing

2.8.1.1. Sample preparation

A detailed description of the samples, replicates, and the corresponding cellular populations used for each sequencing run is provided in table 2.9.

Table 2.9. Description of the samples used for scRNA-seq.

Run	Sample	# of replicates	Sorted population	Estimated cell #
Adult Run 1	Intact	4	Int AP RFP+	9615
			Int AP RFP-	16000
			Int DLP RFP+	392
			Int DLP RFP-	1601
			Int VP RFP+	6745
			Int VP RFP-	5434
	Castrated	4	Cas RFP+	3327
			Cas RFP-	733
Adult Run 2	Intact	3	Int AP RFP+	8212
			Int AP RFP-	16044
			Int DLP RFP+	6293
			Int DLP RFP-	11442
			Int VP RFP+	15538

			Int VP RFP-	12239
	Castrated	5	Cas AP RFP+	20000
			Cas AP RFP-	4306
			Cas DLP RFP+	20000
			Cas DLP RFP-	1754
			Cas VP RFP+	20000
			Cas VP RFP-	2345
Adult Run 3	Intact	4	Int AP RFP+	4513
			Int AP RFP-	8000
			Int DLP RFP+	9129
			Int DLPRFP-	16000
			Int VP RFP+	6000
			Int VP RFP-	6292
UGS explants	Day 0	8	EPCAM+	11000
	Day 1	10	EPCAM+	10000
	Day 3	7	EPCAM+	20000
	Day 6	8	EPCAM+	20000

Int: Intact, Cas: Castrated, AP: Anterior Prostate, DLP: Dorsolateral Prostate. VP: Ventral Prostate

For the adult mouse prostate dataset, AP, DLP, and VP lobes were micro dissected and pooled from P2-*Runx1*:RFP reporter mice after dissociation. Single live EPCAM+ cells from RFP+ and RFP- fractions of each lobes were sorted separately, containing a mix of CD49f-high basal and CD24-high luminal cells. For the UGS explant culture dataset, the middle regions of the explants were micro dissected to enrich for prostatic branching events and pooled by time point after dissociation. Single live EPCAM+ cells were sorted for each independent time point. Complete sample metadata information is available in supplementary data 4 and 5 respectively for the Adult and UGS explant datasets.

2.8.1.2. Sample multiplexing

Individually sorted populations were multiplexed using the MULTI-seq protocol (McGinnis et al., 2019b). Reagents were kindly provided by Dr. Zev Gartner. In brief, after sorting, cells were washed once in cold serum- and BSA-free PBS. A lipid-modified DNA oligonucleotide and a specific sample barcode oligonucleotide were then mixed and added to the cells at a final concentration of 200 nM each, and incubated in cold PBS for 5 min. Each individual sample to be

multiplexed received an independent sample barcode. Next, a common lipid-modified co-anchor was added at 200 nM to each sample to stabilize the membrane bound barcodes. After an additional 5 min incubation on ice, cells were washed 2 times with PBS containing 1% FBS 1% BSA in order to quench unbound barcodes. Samples were then pooled together and washed once with PBS 1% FBS 1% BSA. After cell counting, cells were loaded in a Chromium Single Cell 3′ GEM Library & Gel Bead Kit v3 (10x Genomics). Cell encapsulation, library preparation, and sequencing were performed by the CRUK MI Molecular Biology Core Facility.

2.8.1.3. Library preparation, sequencing and pre-processing

Gene expression (cDNA) libraries were prepared by the CRUK MI Molecular Biology Core Facility according to the manufacturer's protocol. MULTI-seq barcode libraries were separated from the cDNA libraries during the first round of size selection, and PCR amplified prior to sequencing according to the MULTI-seq library preparation protocol (McGinnis et al., 2019b). Figure 2.2 shows examples of the quantified cDNA and MULTI-seq libraries before and after amplification for the 3 runs of the adult dataset.

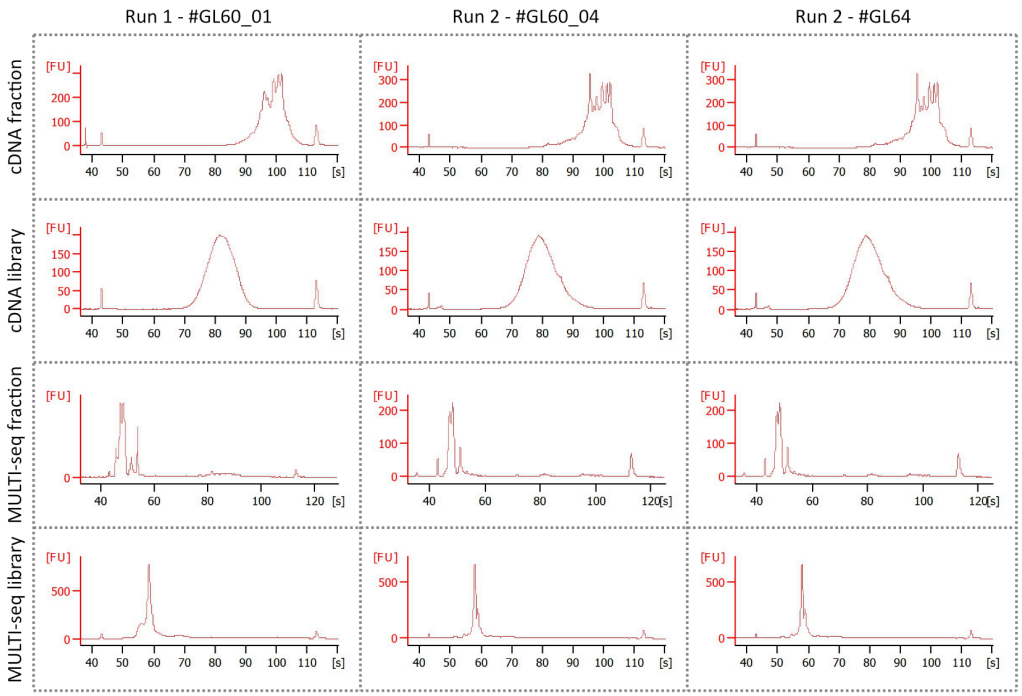

Figure 2.2. Quality control of the library preparation before and after amplification.

Bioanalyzer tracks showing the abundance of different fragment sizes of the cDNA and MULTI-seq fractions for the 3 experimental runs. Data generated by the CRUK MI Molecular Biology Core Facility.

For the adult mouse prostate dataset, cDNA libraries of 'run 1' and 'run 2' were sequenced on Illumina NovaSeq 6000 System, and 'run 3' was sequenced on Illumina HiSeq 2500. The UGS mouse prostate explant run was also sequenced on Illumina HiSeq 2500. Sequencing data of cDNA libraries were processed using Cellranger v3.1.0 and mapped onto mm10 mouse reference genome. Pre-processing of the MULTI-seq library fastq files was performed using the 'deMULTIplex' (v1.0.2) R package (https://github.com/chris-mcginnis-ucsf/MULTI-seq) to generate a sample barcode UMI count matrix. Detailed quality control metrics of each sequencing run are provided in table 2.10.

Table 2.10. Quality control metrics of each sequencing run.

	Adult			UGS
	Run 1: GL60_01	Run 2: GL60_04	Run 3: GL64	GL60_03
# of cells before QC	2612	6034	3028	7474
# of cells after recovered QC	1891	4398	2243	5731*
# of after tags classification	777	2186	1673	5935*
# of cells after DoubletFinder	752	2091	-	-
Final # of cells	752	2091	1673	5731

* Tags classification performed before QC

2.8.1.4. Adult mouse prostate dataset

Quality control and barcode demultiplexing of individual runs.

Each run was pre-processed individually prior data integration. Cellranger outputs were loaded into the R package Seurat (v3.1.5). Cells were kept if they had more than 750 detected genes, less than 7500 UMIs and less than 10% mitochondrial transcripts. Sample barcodes were demultiplexed using the 'HTODemux()' function implemented in Seurat. Briefly, a negative binomial distribution was used to estimate the background levels based on k-means clustering of the barcode normalised counts. Barcodes with values above the 99% quantile were considered 'positive' for a given sample. Cells positive for more than one barcode were considered as 'doublets'. Doublets and negative cells were excluded for all

downstream analyses. Thresholds were empirically adjusted to remove additional cells with possible ambiguous classification (table 2.11). Of note, in both 'run 1' and 'run 2', a large number of cells were classified 'negative' due to the failed labelling of 'Bar3' (corresponding to 'Intact DLP RFP+' sample). For these runs, we used DoubletFinder (McGinnis et al., 2019a) to remove predicted doublets missed out as a consequence of the failed labelling of 'Bar3'. After classification, barcodes were represented in UMAP space to confirm the purity of the barcode assignment obtained for each sample (Figure 5.3). We obtained a total of 4,499 cells from 3 independent experiments.

Table 2.11. Barcodes classification thresholds.

Run	Sample / Population	HTODemux threshold (read counts)	Empiric threshold (read counts)	Final cell #
Adult Run 1	Int AP RFP+	32	35	402
	Int AP RFP-	122	15	169
	Int DLP RFP+	Failed labelling	Failed labelling	Failed labelling
	Int DLP RFP-	46	50	67
	Int VP RFP+	19	30	78
	Int VP RFP-	22	30	19
	Cas RFP+	6	25	16
	Cas RFP-	5	25	1
Adult Run 2	Int AP RFP+	33	35	426
	Int AP RFP-	70	75	205
	Int DLP RFP+	Failed labelling	Failed labelling	Failed labelling
	Int DLP RFP-	75	80	112
	Int VP RFP+	93	100	265
	Int VP RFP-	148	150	147
	Cas AP RFP+	30	35	395
	Cas AP RFP-	17	25	65
	Cas DLP RFP+	24	25	218
	Cas DLP RFP-	12	20	33
	Cas VP RFP+	19	25	209
	Cas VP RFP-	17	25	16
Adult Run 3	Int AP RFP+	36	-	387
	Int AP RFP-	205	-	237
	Int DLP RFP+	61	-	406
	Int DLP RFP-	312	-	316
	Int VP RFP+	57	-	197
	Int VP RFP-	44	-	130

UGS explants	Day 0	20	50	798
	Day 1	29	50	940
	Day 3	55	60	1303
	Day 6	373	375	2690

Int: Intact, Cas: Castrated, AP: Anterior Prostate, DLP: Dorsolateral Prostate. VP: Ventral Prostate

Integration, low dimensional embedding and clustering.

Data aggregation was performed according to the standard integration procedure implemented in Seurat. In brief, each dataset was log normalised, and 3000 variable features were initially computed using the 'vst' method. For integration, 2000 features and 50 dimensions were used as anchors. Integrated data were scaled, and the first 50 principal components (PC) were calculated for downstream analyses. Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) was used for visualisation. Graph-based louvain clustering was performed on a shared nearest neighbour graph constructed using 20 nearest neighbours for every cell, and a resolution of 0.4, which gave a reasonable segmentation of the data (Figure 5.4). Extensive exploration of each cluster based on known marker genes was then carried out to subset prostate epithelial cells.

Analysis of prostate epithelial populations.

The same dimension reduction approach described above was performed on the selected prostate epithelial clusters, using a resolution of 0.3 for graph-based louvain clustering. We annotated 1 large population of basal cells by merging 3 subclusters highly expressing *Krt5*, *Krt14* and *Trp63* as we did not discuss the heterogeneity of the basal compartment in this study (Figure 5.6 B). We annotated the different luminal clusters expressing higher levels of *Cd26*/*Dpp4*, *Cd24a*, *Krt8* and *Krt18*, as Lum-A, Lum-B, Lum-C, Lum-D, Lum-E and Lum-F. Several genes specifically marked each cluster, including *Sbp*/*Spink1* in Lum-A, *Tgm4* in Lum-B, *Msmb* in Lum-C, *Psca*/*Krt4* in Lum-D, *Basp1*/*Lpl* in Lum-E, and *Crym* in Lum-F (Figure 5.6). Data were then imported in Scanpy (v1.4.6) to infer lineage relationships between cellular populations via partition-based graph abstraction (PAGA) implemented in the tl.paga function (Wolf et al., 2019). Briefly, a single cell

neighborhood graph (n_neighbors = 50) was computed using the integrated principal components previously calculated in Seurat. PAGA was generated based on our annotated clusters. The final UMAP representation was generated using PAGA-initialised positions to better preserve the global topology of the data. All final data visualizations were generated in R.

Differential gene expression analysis and gene ontology.

Differential gene expression analyses between clusters were performed using the MAST method (Finak et al., 2015) implemented in Seurat within the 'FindAllMarkers' and 'FindMarkers' functions. Testing was limited to genes detected in at least 25% of the tested populations (min.pct = 0.25) and showing at least ± 0.25 log fold change difference (logfc.threshold = 0.25). The 'g:GOSt' function of the gprofiler2 R package was used to perform functional enrichment analysis on gene ontology terms (GO:BP, biological processes). Genes showing at least 0.50 log fold change enrichment in the group tested were kept. Lists of differentially expressed genes are provided in supplementary data 6.

2.8.1.5. UGS explant cultures dataset

A similar strategy was applied for the analysis of the UGS explant culture dataset, with some alterations described in the sections below.

Quality control and barcode demultiplexing.

Cells were kept if they had more than 1000 detected genes, and less than 7.5% mitochondrial transcripts. Barcode classification was performed as above, using the 90% quantile in 'HTODemux' (Table 2.11). We obtained a total of 5,122 cells that passed quality control from the 4 time points.

Low dimensional embedding and clustering.

The first 50 principal components and 20 neighbours were used for UMAP visualisation. Graph-based clustering was done using a resolution parameter of 0.3. We noticed a strong effect of cell cycle using cell cycles genes defined in Tirosh et

al., 2016. This was particularly evident using the 'CellCycleScoring' function implemented in Seurat (Figure 6.8 A). To minimize the impact of cell cycle on downstream analyses, the cell cycle scores were regressed out during data scaling. We identified 6 main clusters, that we annotated based on the expression of several marker genes (Figure 6.8 C, E). We identified 2 clusters of developing mesonephric derivatives (*Hoxb7*, *Wfdc2*, *Gata3*, *Sox17*, *Pax2*, *Pax8*, *Lhx1*), 1 cluster of developing bladder urothelium (*Upk3a*, *Foxq1*, *Plaur*, *Krt7*, *Krt20*), 1 cluster of mesenchymal cells (*Vim*, *Col3a1*, *Col1a1*, *Pdgfra*, *Zeb1*) and 1 cluster corresponding to the developing prostatic epithelium (*Epcam*, *Krt8*, *Krt5*, *Krt14*, *Krt15*, *Shh*, *Hoxb13*, *Hoxd13*, *Nkx3-1*). We also identified one cluster largely associated with hypoxia and cellular stress ontologies (Figure 6.8 D).

Analysis of the developing prostatic epithelium.

The same dimension reduction approach was initially applied on the developing prostatic cluster. After graph-based clustering using a resolution of 0.5, 10 clusters were identified and visualised via UMAP (Figure 6.9). We computed diffusion components using 'runDiffusionMap' (ncomponents = 20, k = 20) implemented in the scater (v1.14.6) R package. We found the small cluster C9 to be largely diverging from the remainder fraction in diffusion space, therefore it was excluded for downstream analysis (Figure 6.9 E). We then imported the data in Scanpy and used the first 10 diffusion components to compute a neighbourhood graph (n_neighbors = 20) which was used for PAGA. We finally computed a force-direct layout (ForceAtlas2) using PAGA-initialised positions.

Analysis of gene set activity.

Gene signatures were generated from the list of differentially expressed genes by keeping those showing at least 0.50 log fold change enrichment in each given group. Lists of differentially expressed genes are provided in supplementary data 6. Gene lists were used as custom gene sets (Supplementary data 7) in the AUCell Aibar et al., 2017 R package (v1.8.0). Briefly, AUCell uses the Area Under the Curve to evaluate the enrichment of a given gene set in each cell, in a ranking based

manner. It outputs an AUC score for each individual cell, which is used to explore the relative expression of the signature. Per cell AUC scores of each signature were overlaid on the dimension reduction layout and plotted as box-plots to visualize enrichments across the different cellular subsets.

2.8.1.6. Data availability

Raw sequencing files and processed gene expression matrices of scRNA-seq datasets have been deposited in the NCBI Gene Expression Omnibus under the accession number GSE151944. The processed datasets for both mouse adult prostate and UGS prostate explant cultures can be accessed via a searchable R Shiny application available at http://shiny.cruk.manchester.ac.uk/pscapp/. All code used to process data and generate figures is available on a public GitHub repository at https://github.com/glacaud/prostate-scRNAseq.

2.8.2. Mining of publicly available datasets

The cBioPortal database (https://www.cbioportal.org/) was used to collect information about the nature and frequency of genomic alterations affecting *RUNX1*, *RUNX2* and *RUNX3* in all independent Prostate Adenocarcinoma studies available using the following query: “HOMDEL HETLOSS GAIN AMP MUT”. Figures of “OncoPrints” and “Cancer Types Summary” were downloaded via the online interface. RNA-seq data for *RUNX1* and matching clinical data were downloaded via the cBioPortal database for both the TCGA (Cancer Genome Atlas Research Network, 2015) and DKFZ (Gerhauser et al., 2018) studies. Based on the expression levels of RUNX1, the upper quartile (75th percentile) was defined as the *RUNX1*-Hi group and compared with remaining *RUNX1*-Lo fraction. Survival analyses were performed as described in section 2.7.2.

2.9. Statistical analysis

Statistical analyses were performed using Graphpad/Prism (v8.4.2). Where possible, data are represented as mean ±SD. Unless otherwise specified, two-tailed unpaired t-tests were used to compare means between two groups. Statistical

significance was set at $p < 0.05$. For animal model studies, no statistical method was used to pre-determine the sample size. No randomization or blinding was used for *in vivo* studies.

Chapter 3 Characterisation of *Runx1* expressing cells in the normal mouse prostate

[This section has been adapted from (Mevel et al., 2020)]

3.1. Introduction

The mouse prostate is organised in pairs of lobes, including the anterior prostate (AP), dorso-lateral prostate (DLP) and ventral prostate (VP). Each lobe has distinct branching patterns and histopathological characteristics, and they are thought to contribute differently to the physiological function of the prostate (Berquin et al., 2005; Cunha et al., 1987; Shappell et al., 2004; Thomson and Marker, 2006). The differentiated glandular epithelium of the prostate is composed of three epithelial cell types: luminal, basal and rare neuroendocrine cells. These have mainly been defined based on their anatomical locations, histological characteristics, as well as the expression of a small subset of markers (Shen and Abate-Shen, 2010; Toivanen and Shen, 2017; Wang et al., 2001). Luminal cells form a layer of polarised tall columnar cells producing prostatic secretions and rely on androgen signalling. Luminal cells are characterised by high levels of K8/18 and AR, while basal cells express high molecular weight K5/14 and p63.

In adult mice, lineage-tracing experiments performed using canonical basal (*Krt5*, *Krt14*) and luminal (*Krt8*, *Krt8*) drivers suggest that basal and luminal lineages are largely self-sustained by the presence of unipotent basal and luminal progenitors (Choi et al., 2012; Lu et al., 2013; Tika et al., 2019; Tika et al., 2019; Wang et al., 2013). However, the precise characterisation of these progenitors remains elusive and controversial (Toivanen and Shen, 2017; Zhang et al., 2018). In light of the complex architecture of the prostate, several studies have proposed that different epithelial progenitor populations reside at specific locations. Indeed, in addition to its lobular pattern, the mouse prostate can be segmented on a proximal-distal axis, ranging from the periurethral region to the tip of the ducts respectively (McNeal, 1981). Early studies have reported regional differences in the

levels of DNA synthetic activity, showing that distal ducts contain a greater proportion of proliferating cells following androgen-induced prostatic regeneration (Sugimura et al., 1986b). A recent report demonstrated the spatiotemporal restricted localisation of multipotent progenitors at the distal tips of the ducts during postnatal development, switching to unipotency during puberty (Tika et al., 2019). The prostate specific transcription factor *Nkx3-1* has been shown to mark a population of luminal NKX3.1 expressing cells (CARNs) with enhanced regenerative potential (Wang et al., 2009b), and to be more enriched at distal locations during early developmental stages (Kruithof-de Julio et al., 2013). Aged mice were also shown to contain an increased proportion of TROP2+ luminal progenitor cells, located in ridges protruding into the lumen of distal ducts (Crowell et al., 2019). In contrast to these reports indicating the presence of more regenerative cells at distal locations, label-retention studies using bromodeoxyuridine (BrdU) suggested that that slow-cycling progenitors located in the proximal region retained increased proliferative and differentiation capacities (Tsujimura et al., 2002). This is further supported by serial *in vivo* transplantation assays showing the greater progenitor potential of cells from both the urethra and the proximal regions (Goto et al., 2006). The proximal region was shown to be enriched in cells expressing SCA-1 and with high *in vivo* proliferative potential (Burger et al., 2005). More precisely, SCA-1 marks a luminal progenitor population capable of forming differentiated prostate glands using an *in vivo* prostate regeneration assay. SCA-1+ luminal cells express low levels of NKX3.1 and are located in the proximal region of the ducts (Kwon et al., 2016). Likewise, TROP2 (*Tacstd2*) was found enriched in the periurethral region, in basal and luminal cells with progenitor characteristics, and low levels of *Nkx3-1* expression (Goldstein et al., 2008). Taken together, these controversial findings suggest that several progenitor populations seem to co-exist at various locations in the prostate epithelium and contribute differently to the development and the regeneration of the organ in a context-dependent manner.

RUNX transcription factors (TF) are master regulators of lineage commitment and cell fate (Mevel et al., 2019). RUNX1 is essential for the ontogeny of the

hematopoietic system and alterations of RUNX1 have been associated with a broad spectrum of haematological malignancies. As discussed in section 1.4 and 1.5, increasing evidence also implicates RUNX1 in the biology and pathology of hormone-associated epithelia (Lie-A-Ling et al., 2020; Riggio and Blyth, 2017; Scheitz and Tumbar, 2013), including breast (Browne et al., 2015; Chimge et al., 2016; Ferrari et al., 2014; van Bragt et al., 2014), uterine (Planagumà et al., 2004; Planagumà et al., 2006), ovarian (Keita et al., 2013), and prostate cancers (Banach-Petrosky et al., 2007; Scheitz et al., 2012; Takayama et al., 2015). In light of the prevalence of hormone signalling as well as the glandular nature of the mammary gland, parallels are often made between the mammary and prostate glands. RUNX1 has previously been found to be expressed in the basal and luminal compartments of the mammary epithelium, but not in secretory alveolar cells (Blyth et al., 2010; van Bragt et al., 2014). Mechanistically, RUNX1 was found to repress the alveolar transcription factor *Elf5*, and to regulate the expression of mature luminal genes involved in ER transcriptional program (van Bragt et al., 2014). Furthermore, *Runx1* was shown to promote ductal and lobular differentiation using the basal-like MCF10A cell line (Sokol et al., 2015; Wang et al., 2011b). While these results suggest a role for RUNX1 in mammary luminal differentiation, it also questions whether similar observations would apply to the prostate.

Despite the emerging literature involving RUNX1 in non-hematopoietic tissues, including reports of RUNX1 in PCa, its expression in the normal prostate gland has not been explored. To better understand the possible contribution of RUNX1 to PCa, an initial key question to address was whether RUNX1 is expressed in the normal adult prostate epithelium under homeostatic conditions, and if so, what would be its precise expression pattern. Therefore, in this Chapter, the spatial distribution of RUNX1 was thoroughly investigated in the different lobes and epithelial cell types of the adult mouse prostate. Next, taking advantage of isoform specific reporter mouse models of *Runx1* expression, RUNX1^{+} cells were purified, and their *ex vivo* biological potential evaluated in prostate organoids assays. Finally, the functional requirement of *Runx1* was evaluated in these prostate organoids.

3.2. Results

3.2.1. Validation of mouse anti-RUNX1 antibodies

To characterise the expression pattern of *Runx1* in adult mouse prostate, several anti-RUNX1 antibodies were tested by immunohistochemistry using *Runx1* WT and *Runx1* KO mouse prostate organoids derived from the prostate epithelial (EPCAM+) fraction of *Runx1*$^{flox/flox}$ *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}YFP}$ mice (as described in section 3.2.6). The best 2 candidates are shown in Figure 3.1. While both antibodies displayed strong and specific nuclear staining in *Runx1* WT organoids, the D4A6 antibody clone (#8526, Cell Signalling) appeared to be the most specific, with only rare and faint unspecific cytoplasmic stains in *Runx1* KO organoids. In contrast, the RUNX1/2/3 antibody (#92336, Abcam) had evident reminiscent nuclear staining, suggesting either unspecific binding, or binding to another RUNX family member.

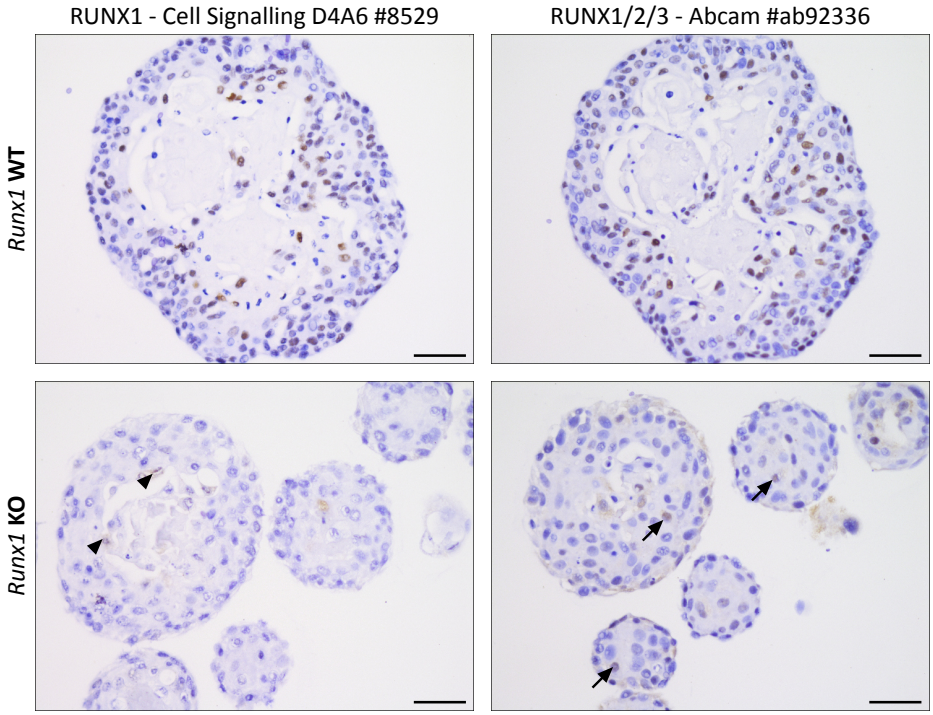

Figure 3.1. Validation of RUNX1 antibodies by chromogenic immunohistochemistry. Validation of anti-RUNX1 antibodies by chromogenic immunostaining on serial sections from *Runx1* WT and *Runx1* KO mouse prostate organoids. Arrows indicate reminiscent nuclear staining in *Runx1* KO organoids, arrowheads indicate faint cytoplasmic staining in *Runx1* KO organoids. Scale bars: 50 µm.

3.2.2. RUNX1 is enriched in the proximal region of the prostate

Using the D4A6 antibody clone, the expression of RUNX1 was evaluated in the mouse urogenital system, including the prostate, urethra, bladder, seminal vesicles and ampullary gland (Figure 3.2). In the prostate, RUNX1 was highly expressed in focal areas of the proximal - or periurethral - AP, in particular in cells lining the luminal side of the ducts, as well as some basal cells (Figure 3.2 A). In the distal AP and DLP, RUNX1 staining was occasionally found in the basal layer but not in luminal cells, except in the proximal region (data not shown). In the VP, both basal and luminal RUNX1+ cells were detected. In light of the specific expression of RUNX1 in the proximal region of the prostate, the proximal and distal AP were analysed separately throughout this chapter. Unless otherwise specified, the DLP and VP were kept as a whole due to the smaller areas of their peri-urethral region. In the other tissues, RUNX1 was strongly expressed in the urethral and bladder epithelia (Figure 3.2 B). Only rare RUNX1+ cells were found in seminal vesicles epithelia, and none were found in the ampullary gland epithelium. Of note, RUNX1 expression was not exclusively constrained to the epithelium, and also observed in the stromal compartment of urogenital tissues (Figure 3.2, chevron arrows).

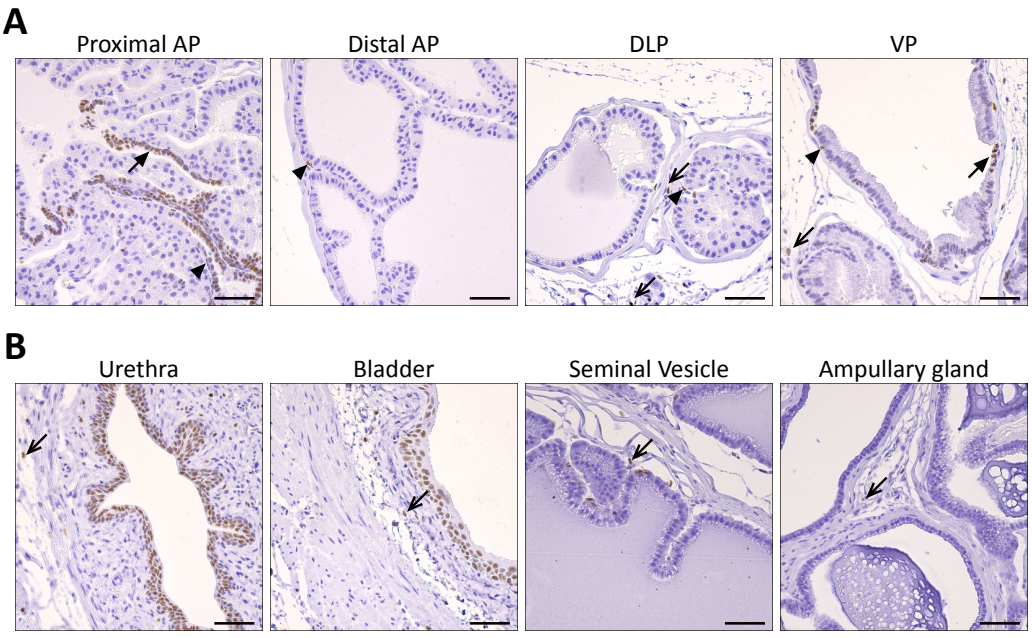

Figure 3.2. RUNX1 expression in the mouse urogenital system.

(A) Immunostaining of RUNX1 at different location of the mouse prostate: proximal AP, distal AP, DLP and VP. (B) Immunostaining of RUNX1 in the urethra, bladder, seminal vesicle and ampullary gland. Arrows indicate RUNX1+ luminal cells, arrowheads indicate RUNX1+ basal cells, chevron arrows indicate RUNX1+ stromal cells. Scale bars: 50 µm.

Chromogenic immunostainings of tissue sections are informative to characterise protein expression *in situ* with the preservation of the spatial context. However, it remains difficult to evaluate co-localisation with additional markers. To circumvent this issue, the phenotypic profile of RUNX1+ cells was further characterised by multiplexed immunofluorescence in combination with classical prostate markers, including K5, K8, p63 and AR (Figure 3.3).

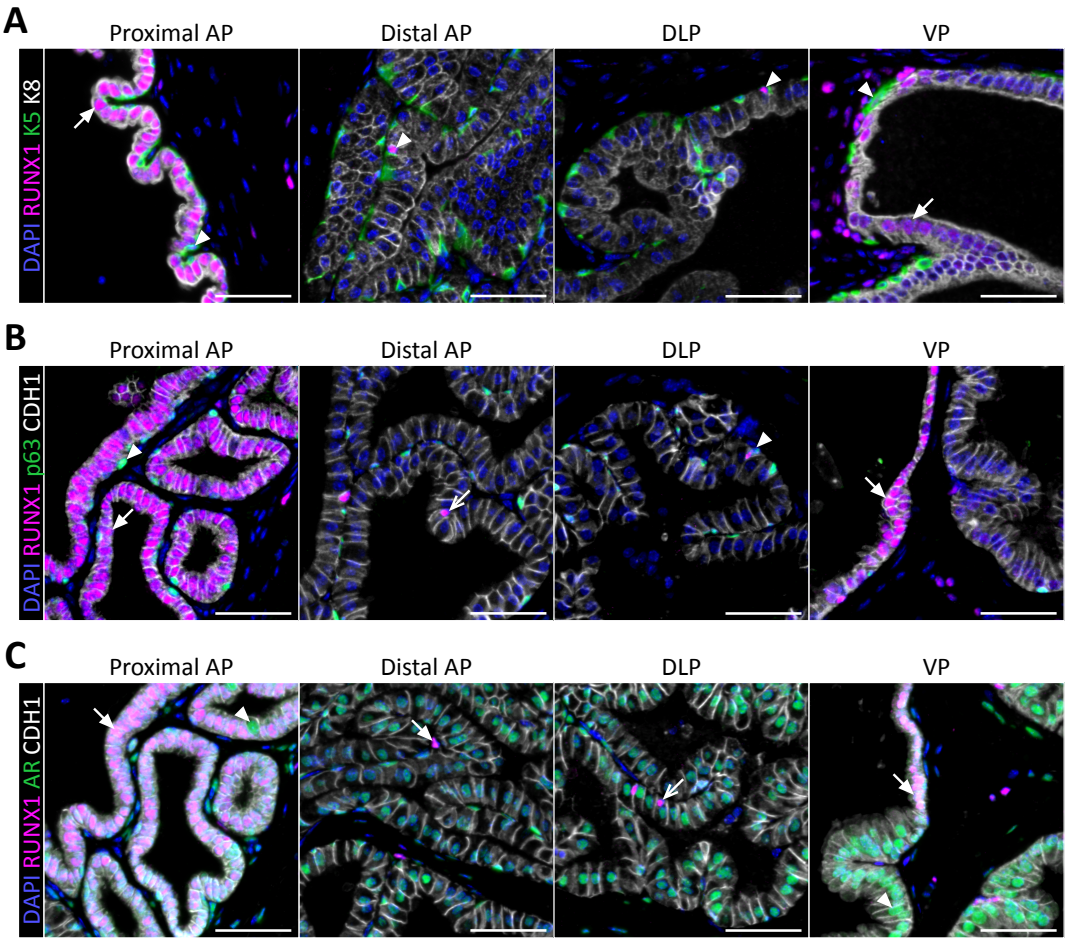

Figure 3.3. Phenotypic characterisation of RUNX1+ cells in the mouse prostate.
(A) Co-immunostaining of RUNX1, K5 and K8 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ K8+ luminal cells, arrowheads indicate RUNX1+ K5+ basal cells.

(B) Co-immunostaining of RUNX1, p63 and CDH1 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ p63- luminal cells, arrowheads indicate RUNX1+ p63+ basal cells, chevron arrows indicate RUNX1+ p63- basal cells.
(C) Co-immunostaining of RUNX1, AR and CDH1 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ AR+ cells, arrowheads indicate RUNX1- AR+ cells, chevron arrows indicate RUNX1+ AR- cells. Scale bars: 50 µm.

In line with previous observations, RUNX1 was specifically highly expressed in luminal K8+ cells of the proximal AP (Figure 3.3 A). High RUNX1 levels were also observed in K8+ AR+ luminal cells of the proximal DLP and proximal VP (data not shown). As expected, these luminal cells expressed AR but not p63 (Figure 3.3 B, C). Luminal RUNX1+ K8+ AR+ cells were also frequently observed in the VP, but rarely in distal locations of the AP or DLP. As already observed with chromogenic staining, RUNX1 expression was detected in basal cells throughout the prostate. Also, while the majority of RUNX1+ basal cells were K5+, not all co-expressed p63 or AR, suggesting some degree of heterogeneity between RUNX1+ and RUNX1- basal cells. Overall, the unique expression pattern of RUNX1 in the prostate indicates that RUNX1 is differentially expressed in distinct subpopulations of epithelial prostate cells.

3.2.3. RUNX1 and NKX3.1 mark distinct luminal lineages

NKX3.1 is a master regulator of prostate development broadly expressed in prostate luminal cells (Abate-Shen et al., 2008; Bhatia-Gaur et al., 1999; Wang et al., 2009b). To further characterise the spatial distribution and heterogeneous expression pattern of RUNX1 in the mouse prostate, tissue sections were co-stained for RUNX1 and NKX3.1 (Figure 3.4). Strikingly, RUNX1 and NKX3.1 had a largely mutually exclusive expression pattern, with a sharp transition from RUNX1+ NKX3.1- to RUNX1- NKX3.1+ luminal cells in the proximal region of the prostate (Figure 3.4 A). This was further confirmed using quantitative image-based cytometry (QIBC), showing a distinct RUNX1+ NKX3.1- population in proximal regions, absent in distal areas (Figure 3.4 B). RUNX1+ NKX3.1- basal cells were also found throughout the prostate epithelium. Unlike the AP and DLP, a fraction of VP

luminal cells co-expressed RUNX1 and NKX3.1 and may represent another luminal subset. Together, these results suggest that RUNX1 marks a subset of proximal luminal cells, referred to as PLCs thereafter, distinct from the abundant NKX3.1+ luminal population lining the rest of the prostate epithelium.

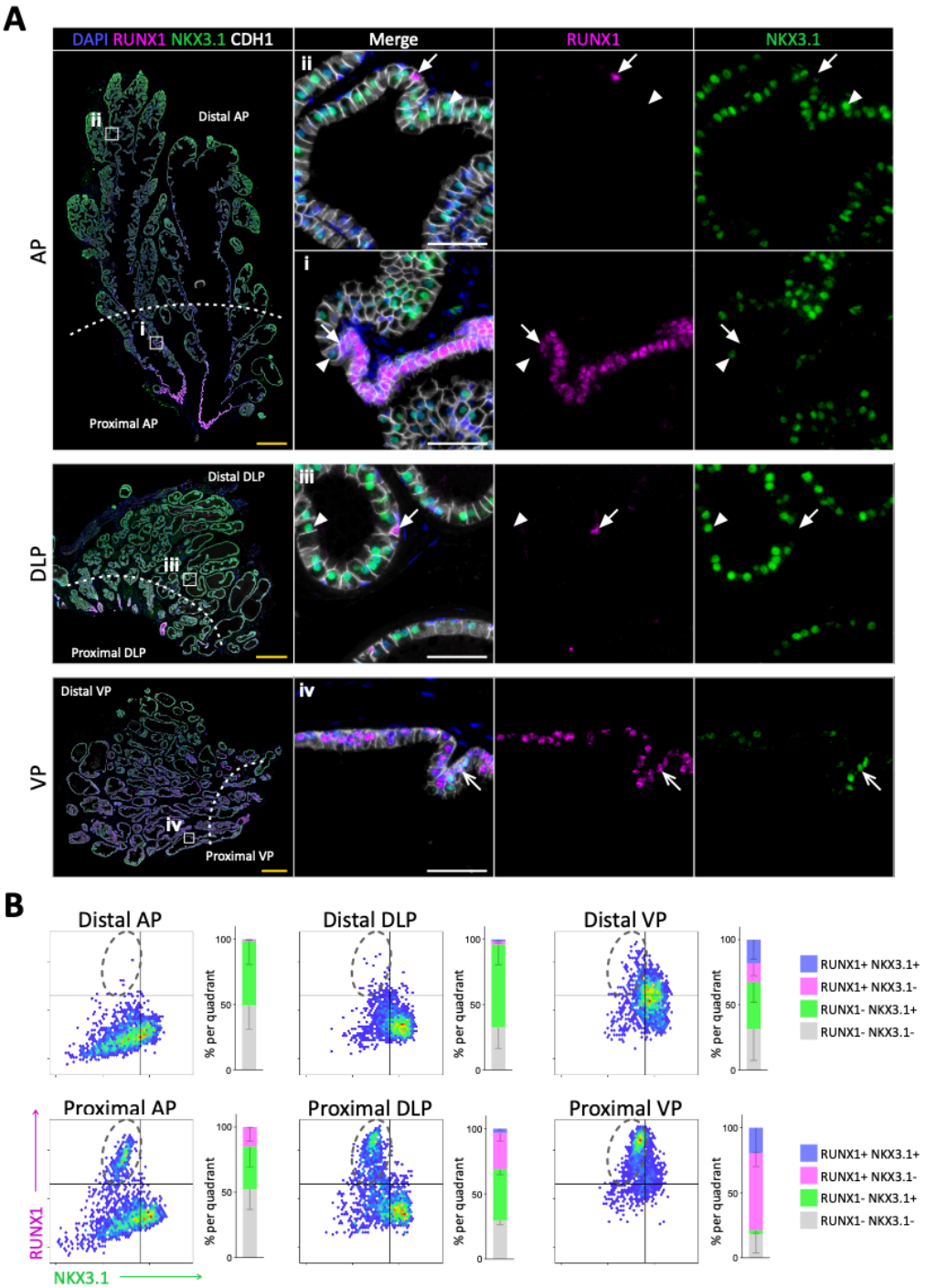

Figure 3.4. RUNX1 is mutually exclusive with NKX3.1.
(A) Co-immunostaining of RUNX1, NKX3.1 and CDH1 in the mouse AP, DLP, and VP. Higher magnification images of (i) proximal AP, (ii) distal AP, (iii) distal DLP and (iv) distal VP are

shown. Arrows indicate RUNX1+ NKX3.1- cells, arrowheads show RUNX1- NKX3.1+ cells, chevron arrows indicate RUNX1+ NKX3.1+ cells. Scale bars: 500 µm (yellow) and 50 µm (white).
(B) Quantification of RUNX1 and NKX3.1 nuclear intensity (log10 transformed) in CDH1+ epithelial cells by QIBC in proximal and distal AP, DLP and VP. n = 6-8 mice. Dashed circles highlight the RUNX1+ NKX3.1- population enriched in the proximal region of each prostate lobe.

Examination of mouse prostate tissue sections using haematoxylin and eosin (H&E) staining revealed that luminal cells located in the proximal region had a unique histological profile. PLCs had a compact organisation, intense nuclear haematoxylin staining, and increased nuclear-to-cytoplasmic ratio (Figure 3.5). This observation was common to all prostate lobes (representative example of the AP display in Figure 3.5). In contrast, distal luminal cells had a large cytoplasm with intense pink eosin staining, likely to reflect their principal secretory function. Therefore, the secretory function of the prostate seems to be mediated by NKX3.1+ luminal cells found throughout the distal prostate. Instead, RUNX1+ PLCs appear to have a non-secretory phenotype.

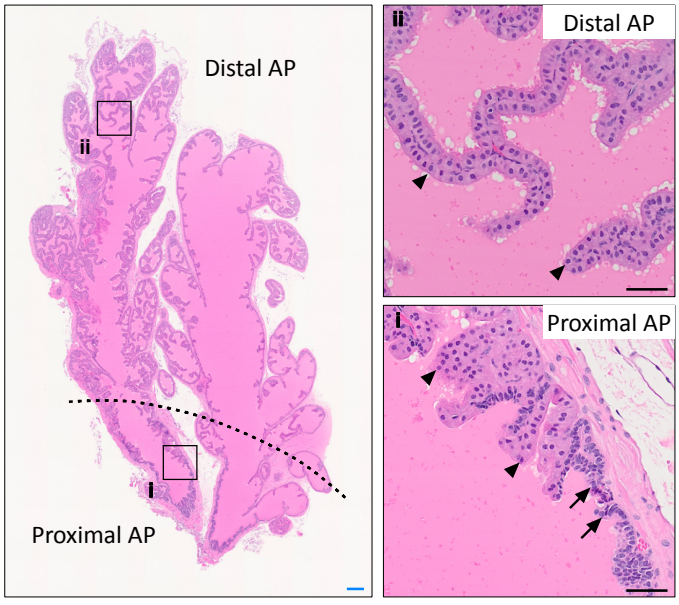

Figure 3.5. RUNX1+ proximal luminal cells have a non-secretory phenotype.
H&E staining of the mouse AP. Higher magnification images of (i) proximal AP and (ii) distal regions are shown. Arrows show typical proximal luminal cells with condensed nuclei and

cytoplasm, arrowheads show typical distal secretory luminal cells with enlarged cytoplasm and strong pink eosinophilic staining. Scale bars: 200 μm (blue) and 50 μm (black).

Several genes, including *Tacstd2*/*Trop2* (Crowell et al., 2019; Goldstein et al., 2008), *Ly6a*/*Sca-1* (Kwon et al., 2016; Kwon et al., 2020), *Keratin 7* (Ceder et al., 2017) and *Bmi1* (Yoo et al., 2016), have previously been reported to be expressed in the proximal region of the prostate, and to mark putative subsets of prostate stem/progenitor cells. The expression of these candidates was therefore evaluated in combination with RUNX1. While all 4 markers co-expressed RUNX1 in PLCs, none of these had a completely overlapping pattern with RUNX1 (Figure 3.6). TROP2, for example, was found expressed in RUNX1- cells of the distal AP, DLP and VP (Figure 3.6 A). Similarly, RUNX1 was expressed in cells of the DLP and VP which did not co-localise with SCA-1 (Figure 3.6 B) or K7 (Figure 3.6 C). In contrast with a previous report (Yoo et al., 2016), BMI1 was broadly expressed throughout the prostatic epithelium, in all lobes of both proximal and distal regions (Figure 3.6 D). Yet not all RUNX1+ cells co-expressed BMI1 and vice-versa.

Overall, RUNX1 was consistently found highly expressed in NKX3.1- PLCs which do not appear committed to the secretory function of the prostate and co-express previously described prostate stem/progenitor markers.
These results therefore highlight the heterogeneity of the prostate epithelia and suggest that RUNX1+ PLCs represent a phenotypically and functionally distinct luminal cell type.

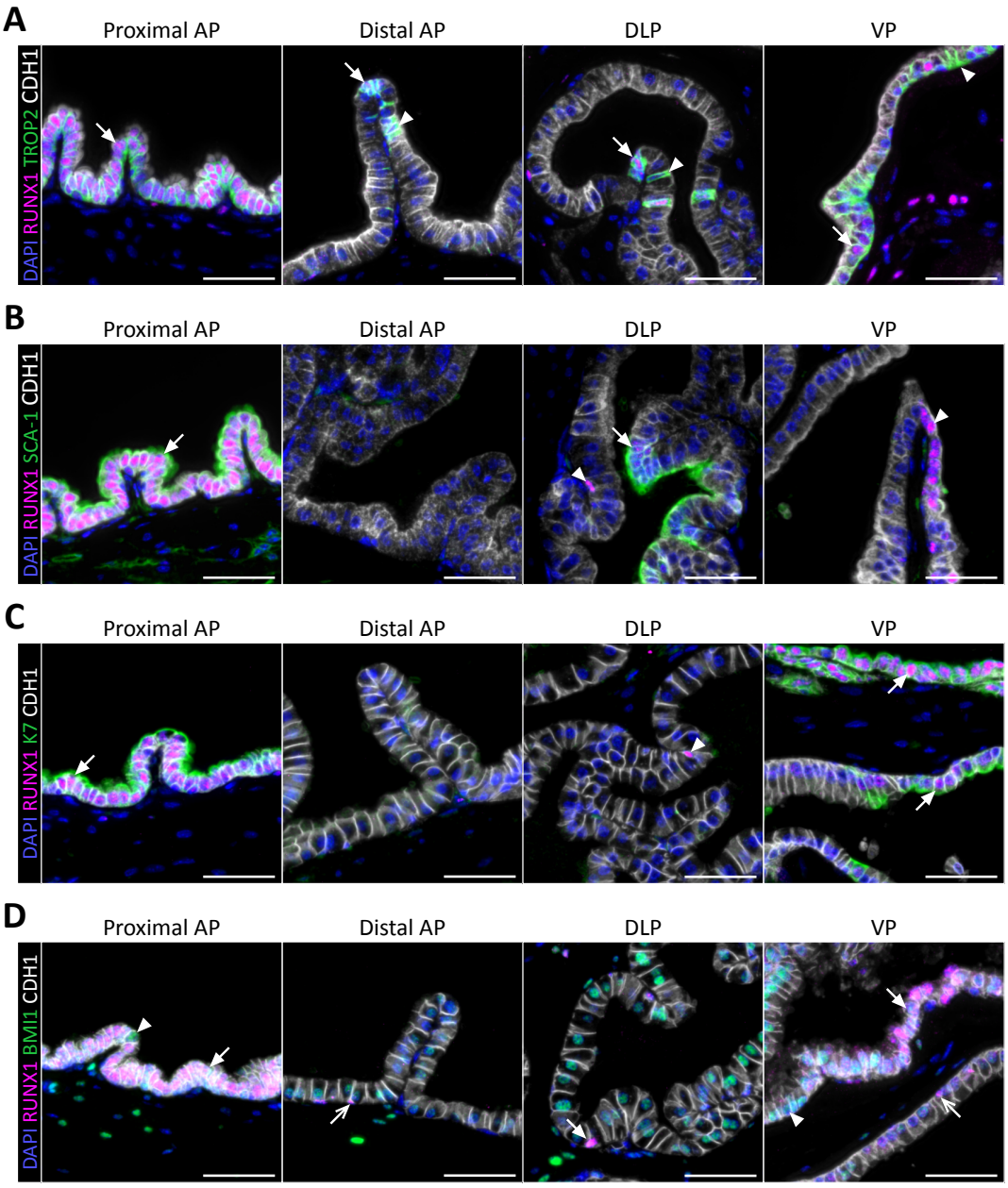

Figure 3.6. Phenotypic characterisation of RUNX1+ cells: TROP2, SCA-1, K7, BMI1.
(A) Co-immunostaining of RUNX1, TROP2 and CDH1 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ TROP2+ cells, arrowheads indicate RUNX1- TROP2+ cells.
(B) Co-immunostaining of RUNX1, SCA-1 and CDH1 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ SCA-1+ cells, arrowheads indicate RUNX1+ SCA-1- cells.
(C) Co-immunostaining of RUNX1, K7 and CDH1 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ K7+ cells, arrowheads indicate RUNX1+ K7- cells.
(D) Co-immunostaining of RUNX1, BMI1 and CDH1 in the proximal AP, distal AP, DLP and VP. Arrows indicate RUNX1+ AR+ cells, arrowheads indicate RUNX1- BMI1+ cells, chevron arrows indicate RUNX1+ BMI1- cells.
Scale bars: 50 µm.

3.2.4. *Runx1* expression in epithelial prostate cells is mediated by the P2 promoter

Runx1 expression is controlled by two promoters, the distal P1 and proximal P2 promoters, that respectively drive the expression of the *Runx1c* and the *Runx1b* isoform (Figure 3.7 A) (Mevel et al., 2019). In order to study the stem/progenitor properties of *Runx1* expressing cells *ex vivo*, I took advantage of previously developed isoform-specific fluorescent reporter mouse models of *Runx1* to prospectively isolate populations of interest (Draper et al., 2018; Sroczynska et al., 2009). *Runx1* expression in the prostate epithelium was exclusively mediated by the proximal P2 promoter, in up to 30% of all EPCAM+ prostate cells, whereas distal P1 promoter activity could not be detected by flow-cytometry (Figure 3.7 B). This observation was further confirmed *in situ* by immunofluorescence, with strong RFP expression in the proximal AP of P2-*Runx1*:RFP prostate epithelial cells, but only rare GFP signal in the stromal compartment of P1-*Runx1*:GFP prostates (Figure 3.7 C).

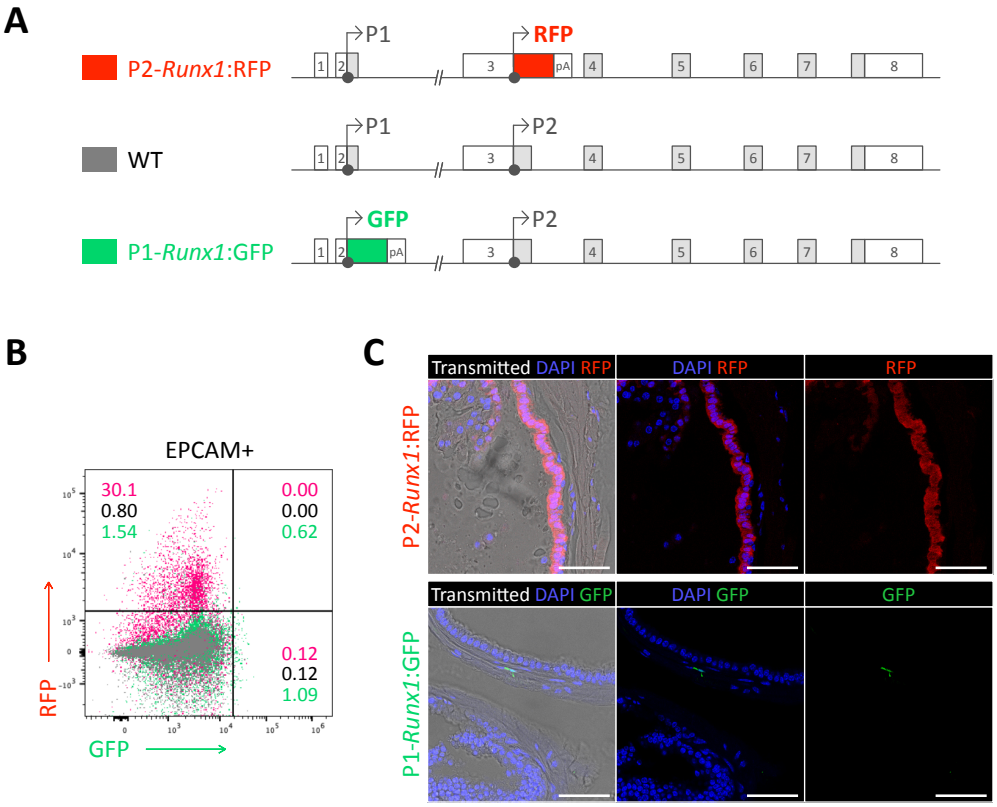

Figure 3.7. Runx1 expression is mediated by the P2 promoter in the mouse prostate epithelium.
(A) Schematic diagrams of the P2-*Runx1*:RFP reporter allele (top), *Runx1* WT allele (middle), and P1-*Runx1*:GFP reporter allele (bottom). In P2-*Runx1*:RFP or P1-*Runx1*:GFP mice, expression of RFP or GFP is directed by *Runx1* proximal P2 or distal P1 promoter respectively. WT *Runx1* expression is directed by the remaining WT allele.
(B) Flow-cytometry analysis of the epithelial EPCAM+ fraction of WT, P2-*Runx1*:RFP and P1-*Runx1*:GFP mice. Percentages are indicated in each quadrant gates.
(C) Confocal images indicating P2-*Runx1*:RFP signal in the glandular epithelium (top), while P1-*Runx1*:GFP (bottom) activity is detected in the surrounding stroma. Scale bar: 50 μm.

Following tissue dissection and dissociation to single cell suspensions, prostate epithelial cells can be analysed by flow-cytometry using CD49f and CD24 to enrich for basal (EPCAM+ CD49f^{high}) and luminal (EPCAM+ CD24^{high}) lineages (Figure 3.8 A) (Karthaus et al., 2014; Lawson et al., 2007). Using the P2-*Runx1*:RFP reporter mouse model, the expression of *Runx1* was extensively characterised by flow-cytometry in the proximal and distal AP, DLP and VP (Figure 3.8 B-D). RFP was detected in a large proportion of the basal fraction of the proximal AP (39.9 ± 5.8%), DLP (56.2 ± 7.1%) and VP (75.2 ± 18.0%) and less abundant in basal cells of the distal AP (16.7± 4.1%). In line with previous results, RFP was more abundant in the proximal AP (17.2 ± 7.0%) than the distal AP (5.4 ± 2.9%) (Figure 3.8 B). RFP was also found in ~15.4 ± 8.4% of DLP luminal cells (Figure 3.8 C), as well as the majority of the luminal compartment of the VP (63.7 ± 10.5%) (Figure 3.8 D). Given the high sensitivity of flow-cytometers, it is worthwhile noting that RFP expression in the basal fraction resembled more of a gradient, as opposed to more clearly defined binomial distributions of RFP expression in the luminal compartment (Figure 3.8 B-D).

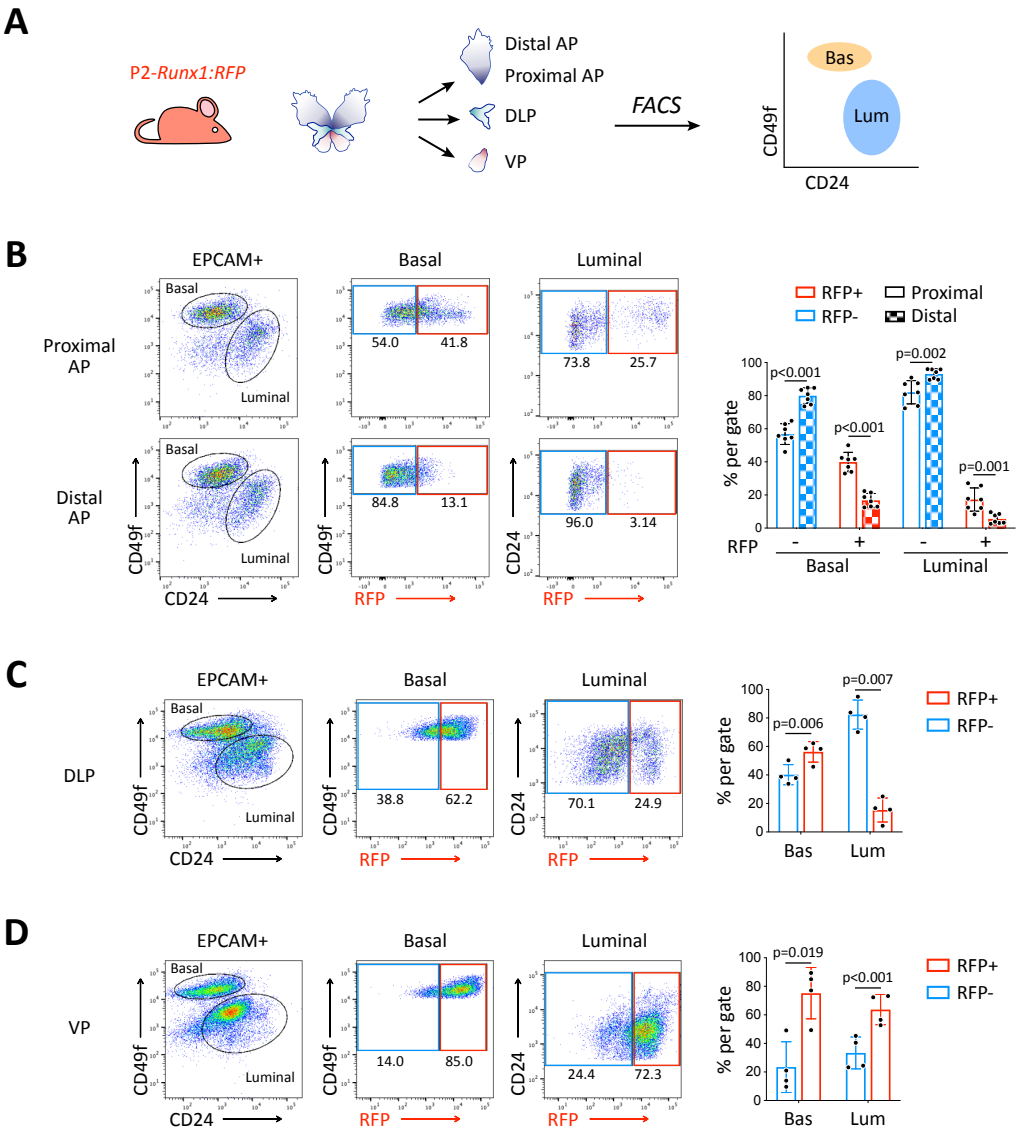

Figure 3.8. Flow-cytometry profiling of the prostate basal and luminal fractions of P2-*Runx1*:RFP mouse reporters.

(A) Experimental strategy to characterise the expression of RFP as a surrogate of *Runx1* in P2-*Runx1*:RFP mouse reporters by flow-cytometry.

(B-D) Flow-cytometry analysis of P2-*Runx1*:RFP mice, and corresponding quantification of the percentages of RFP+ and RFP- cells in the basal and luminal fractions of (A, n = 7 mice) the proximal and distal AP, (B, n = 4 mice) DLP, and (C, n = 4 mice) VP.

3.2.5. RUNX1+ proximal luminal cells have an enhanced Organoid Forming Capacity

To study the stem/progenitor properties of *Runx1* expressing cells, the P2-*Runx1*:RFP reporter mouse line was used to isolate RUNX1 positive (RFP+) and negative (RFP-) epithelial cells from the basal and luminal compartments of all three

pairs of prostate lobes. Sorted cells were plated in a 3D matrix to evaluate their regenerative potential *ex vivo* in organoid culture assays (Drost et al., 2016) (Figure 3.9 A). Strikingly, proximal RFP+ luminal cells of the AP consistently displayed higher Organoid Formation Capacity (OFC) than the RFP- fraction (Figure 3.9 B). Luminal RFP+ sorted cells from the DLP and VP also had a greater OFC than RFP- counterparts. Conversely, RFP+ and RFP- luminal cells isolated from the distal AP had a similar OFC. Also, no significant differences in OFC were observed between basal enriched RFP+ and RFP- subsets.

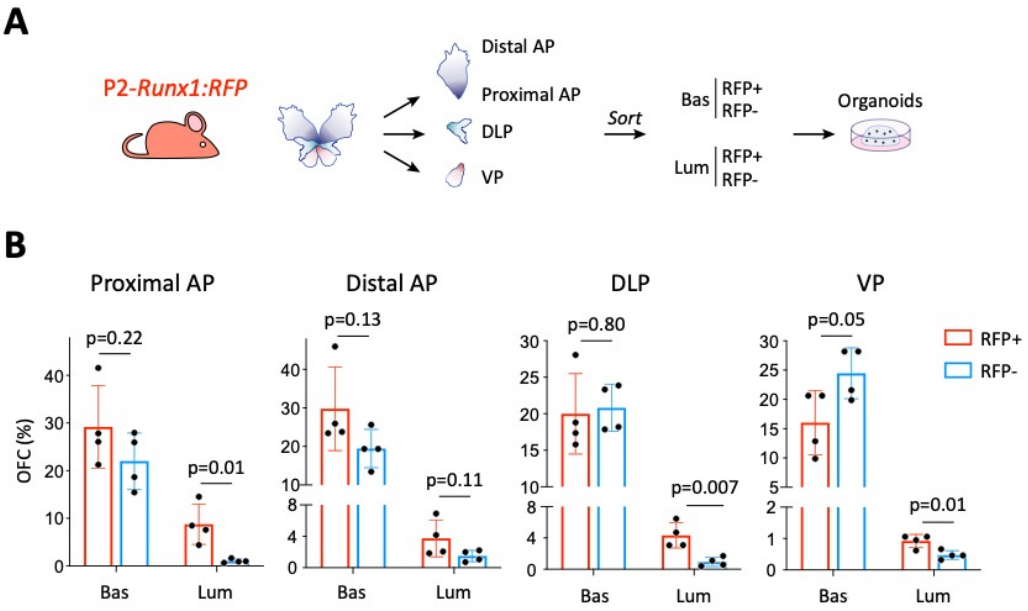

Figure 3.9. Organoid Forming Capacity of RUNX1+ and RUNX1- basal and luminal cells.
(A) Experimental strategy to grow organoids from sorted RFP+ and RFP- cells from the basal (EPCAM+ CD49f^{high}) and luminal (EPCAM+ CD24^{high}) lineages of P2-*Runx1*:RFP mouse reporters.
(B) Organoid Forming Capacity (OFC) of RFP+ and RFP- basal and luminal sorted cells from the proximal AP, distal AP, DLP and VP after 7 days in culture. n = 4 mice.

All mature organoids expressed RFP, including organoids emerging from RFP- sorted populations, suggesting that *Runx1* is actively expressed during their formation (Figure 3.10 A). The culture conditions allowed the growth of 3D structures having a predominantly 'solid' morphology, as well as less frequent 'hollow' ones (Figure 3.10 A-B). Organoids derived from RFP+ luminal subsets, had more frequently a hollow phenotype compared with the RFP- fractions (Figure 3.10

B). This was notably the case for organoids derived from the proximal AP (74.6 ± 12.3% hollow from luminal RFP+ cells, against 21.4 ± 10.2% hollow from luminal RFP- cells).

To further characterise their lineage potential, organoids were stained using whole-mount immunofluorescence with classical basal (K5) and luminal (K8) markers. After 3D imaging, maximum projection images were batch quantified (description of the method in section 2.7.3) to classify organoids into three types based on the expression of these lineage markers: unipotent "basal-like" Keratin 5+ (K5+), unipotent "luminal-like" Keratin 8+ (K8+), or multipotent K5+ K8+ (Figure 3.10 C-D). Interestingly, AP proximal luminal RFP+ derived organoids were predominantly unipotent K8+, while the remainder fractions mainly generated larger unipotent K5+ organoids (Figure 3.10 D). Few multipotent K5+ K8+ organoids were also identified in nearly all populations. The estimated 2D organoid area measurement was also extracted from the image analysis pipeline. Out of all populations analysed, the AP proximal luminal RFP+ derived organoids were significantly smaller than the RFP- counterpart (Figure 3.10 E). Most of the time, those smaller organoids corresponded mainly to the hollow and unipotent "luminal-like" K8+.

Taken together, these results show that RFP+ (RUNX1+) prostate luminal cells have a greater organoid forming potential. In particular, RFP+ cells isolated from the proximal region of the prostate - with the example of the AP here - have a particular predisposition to form unipotent K8+ hollow organoids. Overall, these observations suggest a lineage bias of PLCs towards the luminal identity.

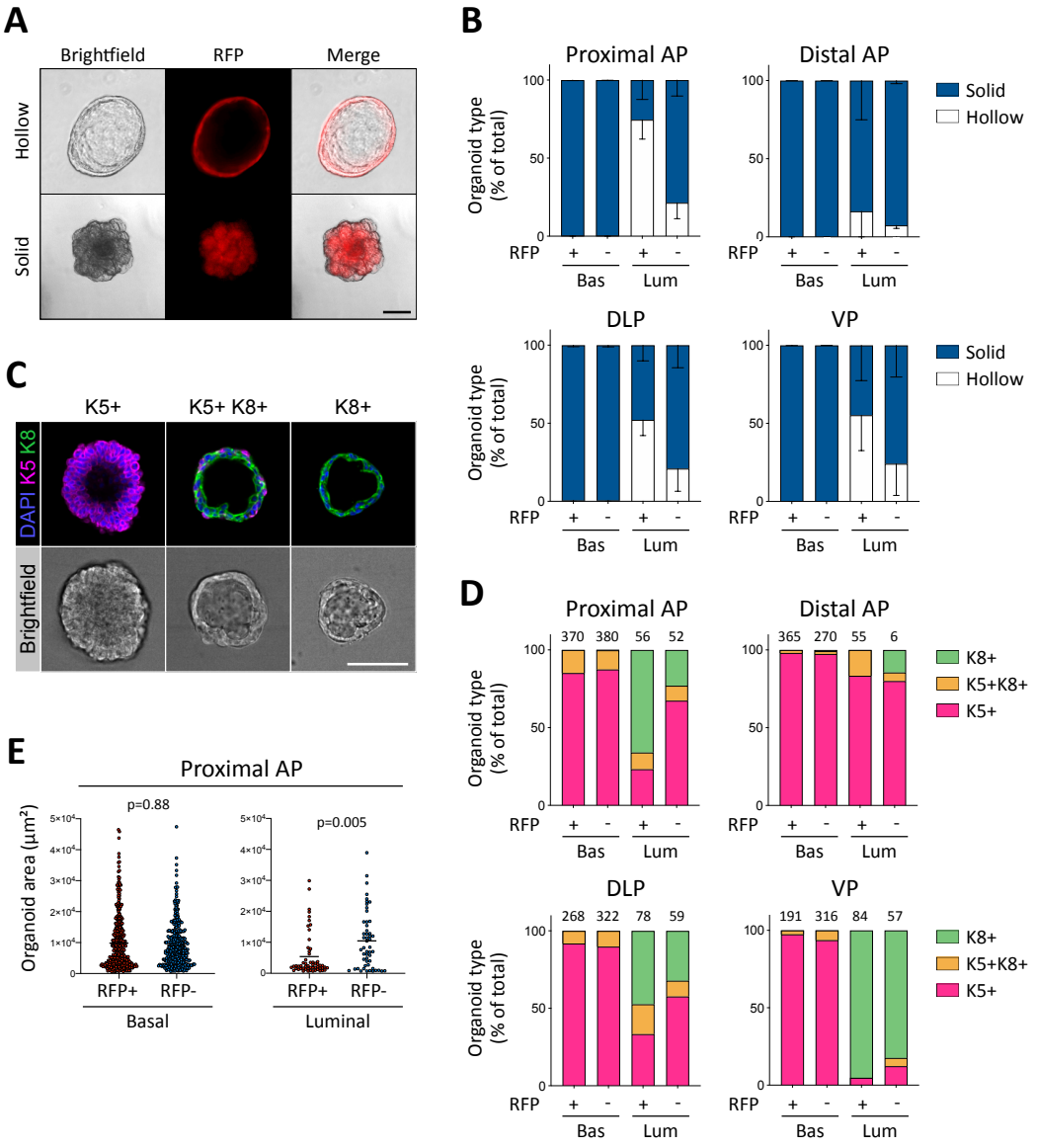

Figure 3.10. Characterisation of the prostate organoids.

(A) Examples of mature 'Hollow' and 'Solid' organoids after 7 days in culture.

(B) Quantification of organoid morphologies (grey: solid, white: hollow) determined by brightfield assessment. n = 4 mice.

(C) Whole-mount immunostaining of unipotent K5+, unipotent K8+ or multipotent K5+ K8+ organoids.

(D) Quantification of the type of organoids (pink: unipotent K5+, green: unipotent K8+, orange: multipotent K5+ K8+ organoids) characterised by whole-mount immunostaining. n = 2 mice.

(E) Quantification of the organoid area estimated from maximum projection of stacked images from basal (left) and luminal (right) fractions of the proximal AP. Numbers of organoids quantified are shown above the graph. n = 2 mice.

Scale bars: 50 µm.

3.2.6. *Runx1* is dispensable for organoid formation

To further understand the contribution of RUNX1 to prostate organoid formation, its genomic deletion was induced *in vitro* using cells derived from the whole prostate of *Runx1*$^{flox/flox}$ *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}YFP}$ mice (Figure 3.11 A). Following prostate dissociation and enrichment of epithelial EPCAM+ cells, deletion of *Runx1* was performed via adenoviral *Cre*-mediated recombination of floxed alleles around its Exon 4 and reported by the concomitant deletion of poly(A) signal repeats at the *Rosa26* locus (STOP cassette), which subsequently activates YFP transcription and expression (Figure 3.11 B, C). After four to five days in culture, successfully infected cells were sorted on the basis of YFP expression and replated at similar densities to evaluate their ability to form *de novo* organoids. In addition to YFP expression (Figure 3.11 C), successful recombination of *Runx1* floxed alleles was verified by genotyping PCR (Figure 3.11 D). Out of 3 independent experiments, and despite overall variations in efficiency between experiments, no significant differences of OFC were observed in the presence or absence of *Runx1* (Figure 3.11 E). Despite its broad expression in organoid cultures (Figure 3.10), these data suggest that *Runx1* is dispensable for mouse prostate organoid formation.

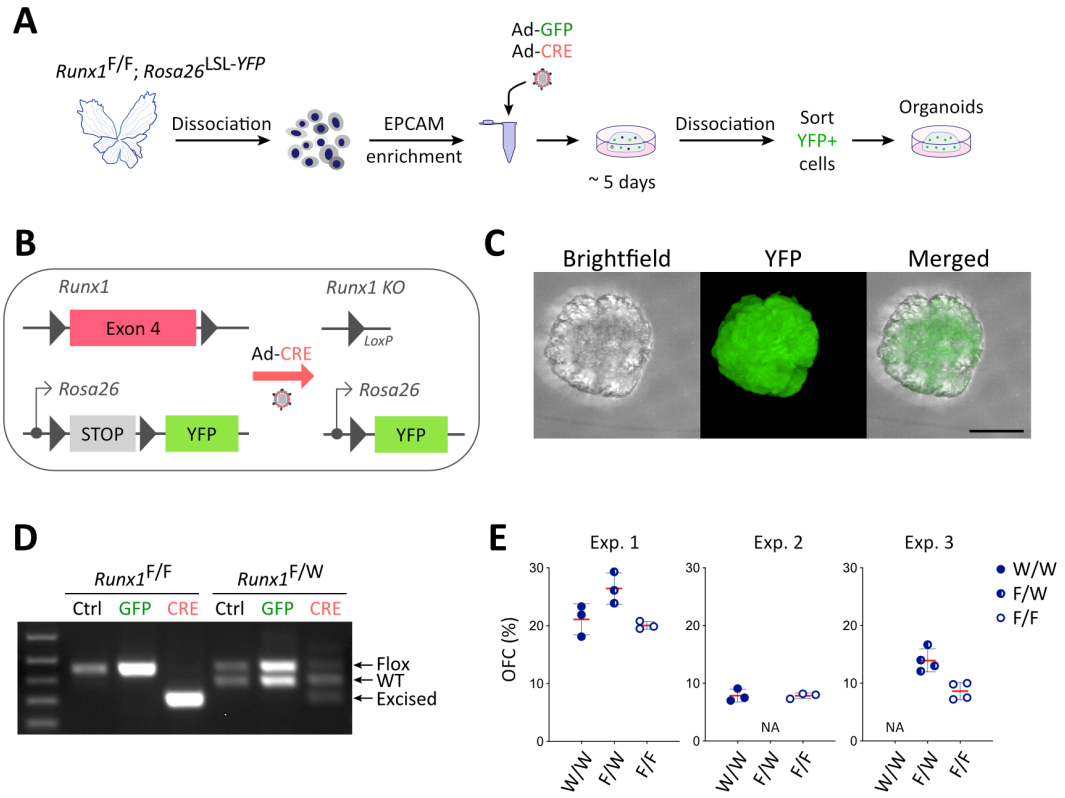

Figure 3.11. Evaluation of the requirement of *Runx1* in mouse prostate organoid cultures.
(**A**) Experimental strategy to generate mouse prostate organoids with *Runx1* deletion *ex vivo*. Cells were isolated from *Runx1*$^{WT/WT}$ / *Runx1*$^{flox/WT}$ / *Runx1*$^{flox/flox}$ *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}YFP}$ mice, enriched for EPCAM+ cells using magnetic beads, and infected with an adenovirus expressing either GFP alone (Ad-GFP), or GFP and a *Cre* recombinase (Ad-CRE). After 5 days in culture, YFP positive cells were sorted and replated for downstream analyses.
(B) Expression of the *Cre* recombinase drives recombination of floxed sequences to knockout *Runx1* and concomitantly activate YFP expression from the *Rosa26* locus.
(C) Maximum projection image of YFP+ organoids following adenoviral *Cre*-mediated activation of YFP expression. Scale bar: 50 µm.
(D) *Runx1* genotyping PCR to control appropriate recombination of floxed sequences after Ad-GFP or Ad-CRE infection.
(E) Organoid Forming Capacity of mouse prostate epithelial cells with *Runx1*$^{WT/WT}$ / *Runx1*$^{flox/WT}$ / *Runx1*$^{flox/flox}$ genotypes. Technical replicates of n = 3 independent experiments are shown. W = wild-type, F = flox.

In light of the high RUNX1 expression levels in the luminal layer of the proximal prostate, as well as the enhanced clonogenicity of this subset, *Runx1* was hypothesised to have a specific requirement in this cellular subset. However, PLCs represent only a small fraction of the total prostate, and therefore the effect of *Runx1* loss, if any, might be masked using the whole dissociated prostate. To

address this hypothesis, a sorting strategy was adopted to enrich for a CD24^{high} luminal population with high expression levels of the cell surface markers TROP2 and SCA-1 (Figure 3.12 A), termed LTS thereafter (Luminal TROP2 SCA-1), previously shown to co-express RUNX1 in the proximal AP (see Figure 3.6 A, B). LTS cells were isolated from prostates of *Runx1*$^{flox/flox}$ *Rosa26*^{CreERT2} *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}YFP}$ mice (Figure 3.12 B). In this model, the *Cre* recombinase is fused to a modified oestrogen receptor (*Cre*ERT2), preventing it from entering the nucleus (Feil et al., 1997). Genomic deletion of *Runx1* was induced *ex vivo* in the presence of 4-hydroxytamoxifen, which acts as an oestrogen receptor agonist, binds to *Cre*ERT2 and promotes its translocation to the nucleus where it can recombine floxed target sites at the *Runx1* and *Rosa26* loci.

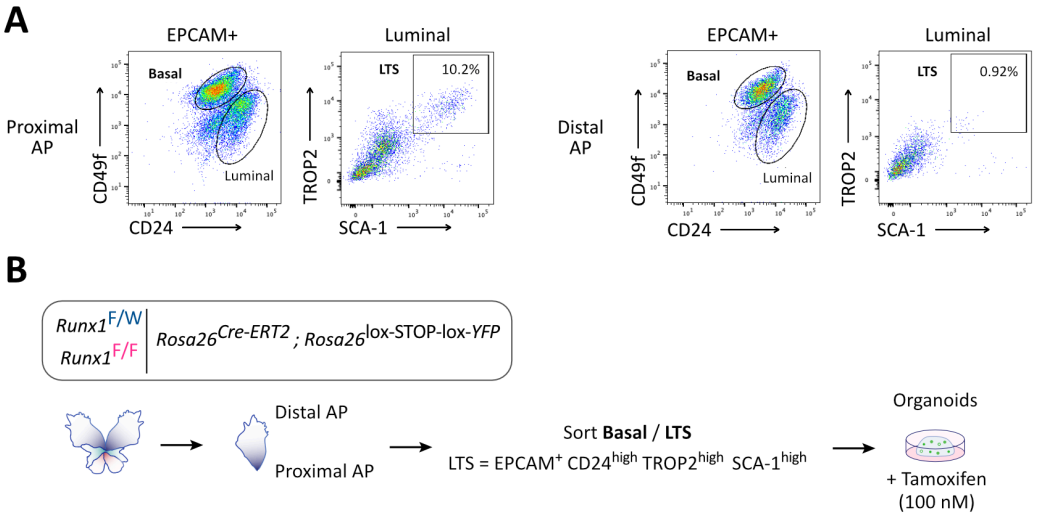

Figure 3.12. Strategy to delete *Runx1* in organoids derived from proximal luminal cells.
(A) Representative gating strategy to isolate basal and luminal LTS (CD24^{high} TROP2^{high} SCA-1^{high}) cells from proximal (left) and distal (right) AP.
(B) Experimental strategy to induce the deletion of *Runx1* in proximal and distal, basal (CD49f) and luminal LTS (CD24^{high} TROP2^{high} SCA-1^{high}) cells in mouse prostate organoid cultures.

Concentrations of 4-hydroxytamoxifen ranging from 1 µM to 25 nM were initially tested using *Runx1*$^{flox/flox}$; *Rosa26*^{CreERT2} ; *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}YFP}$ derived prostate cells, in order to minimise toxicity and maximise targeting efficiency. Toxicity was only observed at 1 µM, and 50 nM were sufficient for all mature organoids to be YFP+ (data not shown), therefore the use of 100 nM seemed

reasonable. Proximal and distal luminal cells having a similar immunophenotypic profile were also sorted for comparison and basal cells (CD49f^{high}) were used as internal control. Overall, no growth difference was found between *Runx1* WT and *Runx1* KO basal cells sorted from the proximal and distal AP (Figure 3.13 A). Of note, a strong difference in OFC was observed between *Runx1* WT and *Runx1* KO proximal basal cells of Exp. 2, but this result was not reproduced in Exp. 1, suggesting a possible technical mistake. More importantly, no growth difference was observed between *Runx1* WT and *Runx1* KO luminal cells, sorted from both proximal and distal regions (Figure 3.13 A). Furthermore, brightfield assessment of the type of organoids also failed to exhibit any morphological differences upon *Runx1* deletion (Figure 3.13 B). Overall, these results invalidate the hypothesis that *Runx1* is required for the *ex vivo* regenerative potential of proximal and distal CD24^{high} TROP2^{high} SCA-1^{high} (LTS) luminal cells in prostate organoid cultures.

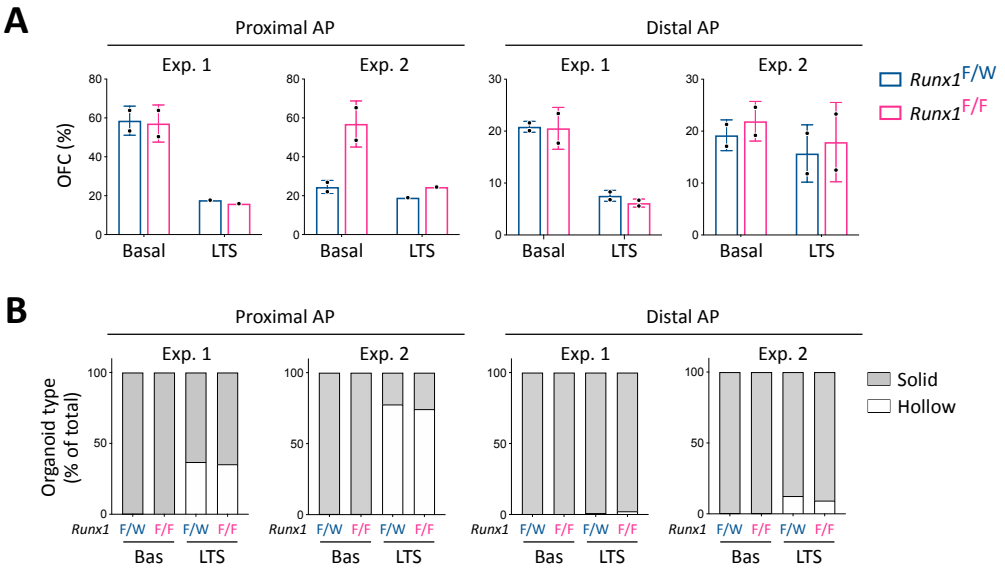

Figure 3.13. Evaluation of the requirement of *Runx1* in proximal luminal cells.
(A) Organoid Forming Capacity of mouse prostate epithelial cells with *Runx1*$^{WT/WT}$ / *Runx1*$^{flox/flox}$ genotypes from either basal or LTS populations. Technical replicates of n = 2 independent experiments are shown. W = wild-type , F = flox.
(B) Quantification of organoid morphologies (grey: solid, white: hollow) determined by brightfield assessment. n = 2 independent experiments.

3.3. Discussion

3.3.1. Summary

Beyond the haematopoietic system, RUNX1 has been shown to play major roles in the development and homeostasis of numerous tissues (Mevel et al., 2019). However, despite the documented importance of RUNX TFs and reports of RUNX1 in PCa (see section 1.5), neither its role nor expression in the normal mouse prostate gland had been reported. As such, the investigations conducted and presented in this Chapter aimed at addressing this knowledge gap. Several experimental approaches were used, combining *in situ* immunostainings of RUNX1, flow cytometry profiling of *Runx1* reporter mouse models, as well as *ex vivo* prostate organoid cultures of RUNX1+ and RUNX1- cells. Collectively, the findings presented here demonstrate that RUNX1 marks a specific subset of proximal luminal cells, with an expression pattern distinct from NKX3.1 otherwise widely expressed in the remainder luminal fraction. Functionally, RUNX1+ PLCs do not appear to be committed to the secretory function of the prostate and display stem cell behaviour in organoid assays with a lineage bias towards the luminal identity. These results highlight the existence of strong intrinsic heterogeneity within the luminal compartment of proximal and distal regions.

3.3.2. RUNX1 marks a specific subset of proximal luminal cells

After careful validation of a specific anti-RUNX1 antibody, the expression of RUNX1 was investigated in adult mouse urogenital systems by chromogenic and multiplexed immunofluorescent stainings. Coupled with whole-slide imaging, this approach allowed the extensive characterisation of RUNX1+ cells location and phenotype. While RUNX1+ basal cells were scattered throughout the glands, RUNX1+ luminal cells were mainly restricted to the proximal/peri-urethral region of the prostate. These RUNX1+ PLCs had a singular morphology with an increased nuclear-to-cytoplasmic ratio compared to the majority of the luminal cells present in the distal prostate, suggesting that they are not contributing to prostatic secretions. For these reasons, subsequent investigations were carried out with a

focus on the proximal prostate, and by analysing the proximal AP and distal AP separately as a representative example of proximal-to-distal axis. Indeed, the proximal region of the prostate has been the focus of many studies and proposed to be enriched in cells with stem/progenitor properties. Accordingly, RUNX1+ PLCs were found to co-express TROP2, SCA-1, K7 and BMI1, which have previously been shown to mark cells with enhanced stem/progenitor potential (Crowell et al., 2019; Goldstein et al., 2008; Kwon et al., 2016; Tsujimura et al., 2002; Yoo et al., 2016). Also, these cells did not express NKX3.1, a master regulator of prostate epithelial cells, otherwise expressed in the majority of the distal luminal secretory epithelium. These findings suggest that RUNX1 marks a functionally and phenotypically distinct luminal lineage from NKX3.1+ luminal cells.

3.3.3. *Runx1* expression is mediated by the P2 promoter

In addition to histopathological analyses, the expression of *Runx1* was studied using isoform specific fluorescent reporter mouse lines. These models had originally been developed to study the differential isoform usage of *Runx1* taking place during the ontogeny and homeostasis of the haematopoietic system (Mevel et al., 2019). In these models, a fluorescent protein (either GFP or RFP) is used as a surrogate of *Runx1* expression via the transcription of one of its two alleles, while the remaining wild-type allele expresses endogenous *Runx1*. Since the dosage of *Runx1* have been shown to be important (Lie-A-Ling et al., 2020), *Runx1* heterozygosity in these mice could possibly have consequences on biological interpretations. However, no abnormal defects were reported during their initial characterisations (Draper et al., 2018; Sroczynska et al., 2009). Likewise, no abnormalities were observed in the prostates of either reporter lines compared to age-matched wild-type controls. Also, the expression of GFP/RFP may not exactly reflect the levels of *Runx1* expression at the time of detection, due to possible differences between mRNA processing and protein synthesis, stability, and degradation rates. Nevertheless, the analysis of these fluorescent reporter lines revealed that *Runx1* expression in prostate epithelial cells is exclusively mediated by the proximal P2 promoter, which is responsible for the transcription of the *Runx1b* isoform. Comparatively to RUNX1

in situ immunostainings, flow-cytometry profiling of the basal (EPCAM+ CD49f^{high}) and luminal (EPCAM+ CD24^{high}) compartments of P2-*Runx1*:RFP reporters highlighted strong RUNX1 levels in PLCs, as well as broad expression in basal cells and in the majority of the VP. The P2-*Runx1*:RFP reporter line was then used to further study the biology of RUNX1+ cells, including the isolation of RFP+ (RUNX1+) cells by FACS for *ex vivo* organoid cultures. The fact that *Runx1b* was the only isoform of *Runx1* expressed in mouse prostate epithelial cells contrasts with the documented differential promoter usage in the haematopoietic system, in which the P1-derived *Runx1c* isoform is dominant (Bee et al., 2009; Draper et al., 2016; Telfer and Rothenberg, 2001). Since the RUNX1c isoform has been shown to have stronger binding on certain genes (Telfer and Rothenberg, 2001), and given that RUNX1 is a master regulator of haematopoietic cell fate, it is tempting to speculate that the expression and activity of RUNX1b may be more appropriate in a non-haematopoietic tissue. Hence, beyond the mouse prostate, the expression of RUNX1b and RUNX1c in other non-haematopoietic tissues should be explored.

3.3.4. RUNX1+ PLCs exhibit higher organoid forming capacity than RUNX1- luminal cells

Functionally, RUNX1+ PLCs exhibited a greater organoid forming potential compared to the remaining luminal fractions. This was consistent with previous reports isolating similar proximal populations using different markers such as SCA-1, TROP2 or CD26, both in *ex vivo* organoid cultures and/or *in vivo* transplantation assays (Crowley et al., 2020; Goldstein et al., 2008; Guo et al., 2020; Karthaus et al., 2020; Kwon et al., 2016). Furthermore, RUNX1+ PLCs predominantly formed unipotent K8+ hollow organoids, demonstrating their preferential commitment to the luminal lineage. Since these cells do not appear to be committed to the secretory function of the prostate, it is tempting to speculate that they act as a latent niche of 'facultative' stem cells (Clevers and Watt, 2018), primed to generate a structured prostatic epithelium under defined conditions. However, it is worth noting that the majority of the unipotent K8+ hollow organoids derived from RUNX1+ PLCs were smaller in size compared to organoids derived from other basal

and luminal populations. Therefore, while RUNX1+ PLCs retain a specific organoid forming potential, these cells seem to have an overall limited, or slower, regenerative capacity in organoid cultures. This result is somewhat in contradiction with a previous study reporting that organoids derived from TROP2+ luminal cells formed larger organoids than TROP2- luminal cells, albeit using the whole dissociated prostate (Crowell et al., 2019). Since they did not analyse proximal and distal regions separately, and because not all distal TROP2+ luminal cells co-express RUNX1, it is conceivable that these distal TROP2+ luminal cells were responsible for the formation of larger organoids.

3.3.5. *Runx1* is dispensable for organoid formation

Having demonstrated the specific organoid formation potential of RUNX1+ PLCs, and broad expression of RUNX1 during organoid formation, *Runx1* was hypothesised to be required for the formation of these structures. However, two experimental strategies aiming at generating *Runx1* KO organoids failed to exhibit any phenotypic differences. The first method consisted in the induction of *Cre* mediated recombination of floxed sequenced via adenoviral delivery in whole EPCAM+ cells, and the second approach was based on tamoxifen mediated *Cre* activation specifically after the isolation of proximal and distal LTS (Luminal TROP2+ SCA-1+) cells. In the first set up, the use of adenoviruses resulted initially in the death of a significant fraction of the cells, and only partial infection rates (~30%, data not shown). This required subsequent enrichment of successfully transduced cells by sorting and replating YFP+ cells. It is possible that this initial delay before the beginning of the clonogenic assay would have reduced the likelihood to observe even a small effect upon *Runx1* deletion. In the second set up, no initial delay was added between tissue dissociation and the start of the organoid culture, tamoxifen mediated *Cre* activation was performed under non-toxic conditions, and subsets of cells expressing the highest levels of RUNX1 *in situ* were specifically isolated. In this context, the absence of phenotypic difference between *Runx1* WT and *Runx1* KO organoids clearly indicates that *Runx1* is dispensable for mouse prostate organoid formation.

While the absence of phenotype upon *Runx1* deletion was evident and reproducible, a putative role for *Runx1* during prostate organoid growth cannot be entirely ruled out. The prostate organoid culture medium contains several cytokines and chemical compounds used to modulate key intracellular signalling pathways in order to promote organoid growth and differentiation. For example, DHT is crucial for AR signalling, Noggin is used to inhibit Bone Morphogenetic Protein (BMP) signalling and promote prostate morphogenesis (Cook et al., 2007), R-spondin1 stimulates Lgr/Wnt signalling (Binnerts et al., 2007; de Lau et al., 2011), A83-01 serves to inhibit TGF-β pathway and prevent a proliferative block (Ding et al., 2011), and EGF has pleiotropic effects on intracellular signalling pathways. Since it is well established that *Runx1* is involved in the regulation of these signalling pathways (Chuang et al., 2013; Goyama et al., 2015; Ito et al., 2015), it is possible that its role, if any, would be compensated upon its loss. Further work is therefore required to iteratively remove individual, or multiple, component(s) of the media, in the context of *Runx1* WT or KO to assess possible phenotypic consequences. Along these lines, other RUNX factors, in particular *Runx2*, may have redundant roles and compensate for *Runx1* loss. It would then be interesting to disrupt both genes to further interrogate their contribution during organoid formation. It is also worth noting that the *ex vivo* organoid culture conditions may not sufficiently reflect normal prostate differentiation processes. Indeed, the majority of organoids had a 'basal-like' phenotype and were predominantly K5+, which suggest that the culture conditions are not optimal to promote and/or sustain luminal growth. Therefore, beyond the *ex vivo* organoid culture system, it would be of interest to evaluate the putative role of *Runx1* in alternative regenerative contexts, such as *in vivo* transplantations in the sub-renal capsule, a gold standard prostate regeneration assay to assess stem cell phenotypes.

Chapter 4 Tracing *Runx1* expressing cells in castration-regeneration assays

[This section has been adapted from (Mevel et al., 2020)]

4.1. Introduction

In PCa patients, androgen deprivation therapy (ADT) is frequently used to target androgen-dependent tumour cells. However, the emergence of castration-resistant PCa cells is a major complication of this treatment (Shen and Abate-Shen, 2010; Zhang et al., 2018). Understanding the origin and identity of these castration-resistant cells both in the normal prostate and PCa represents, therefore, an essential step towards improving PCa treatments. Interestingly, there is contrasting evidence that the castration-resistant phenotype can be acquired *de novo* by transformed prostate cells (Carceles-Cordon et al., 2020; Watson et al., 2015) and/or that it can be present intrinsically in the normal prostate (Barros-Silva et al., 2018, 2012; Qin et al., 2012; Yoo et al., 2016; Zhang et al., 2008; Zong and Goldstein, 2013).

In rodents, androgen-deprivation can be modelled by surgical castration which leads to prostate regression and enriches for castration-resistant cells (Toivanen and Shen, 2017; Zhang et al., 2018). This process is accompanied by the death of luminal androgen-dependent cells and a small proportion of basal cells (English et al., 1987; Sugimura et al., 1986b). However, not all luminal cells disappear during this process. Therefore, surgical castration-induced prostate involution has proven an invaluable tool to identify progenitor castration-resistant cell populations, characterised by their ability to survive in the absence of androgens, and to fully regenerate an intact adult prostate after re-administration of testosterone. The existence of castration-resistant and regenerative properties within specific subpopulations has been convincingly demonstrated with the help of lineage tracing mouse models allowing long term cell fate tracking. This has notably been the case for CAstration-Resistant *Nkx3-1* (CARNs) expressing cells (Wang et al.,

2009b), CAstration-Resistant *Bmi1* (CARBs) expressing cells (Yoo et al., 2016), SOX2+ cells (McAuley et al., 2019) and LGR5+ cells (Wang et al., 2015). Further to these lineage-tracing studies, other populations have been shown to be enriched in the castrated prostate and to possess an enhanced regenerative potential *ex vivo,* such as CD117+ cells (Leong et al., 2008), SCA-1+ cells (Kwon et al., 2016), TROP2+ cells (Goldstein et al., 2008), slow-cycling label-retaining cells (Ceder et al., 2017; Tsujimura et al., 2002; Zhang et al., 2018), LY6D+ cells (Barros-Silva et al., 2018). These studies highlight a high degree of phenotypic and functional heterogeneity within the adult prostate epithelium.

In the previous Chapter, RUNX1 was shown to be expressed in specific subsets of the adult mouse prostate epithelium. In particular, a subpopulation of RUNX1+ proximal luminal cells displayed facultative stem cell behaviour in organoid assays. Throughout the normal prostate epithelium, RUNX1 displayed a partially overlapping expression pattern with some of the above cited markers, including TROP2, SCA-1, K7 and BMI1. In contrast, RUNX1+ proximal luminal cells were largely distinct from NKX3.1+ cells. These observations suggest that RUNX1 may be marking a distinct luminal lineage from the abundant distal NKX3.1+ luminal population. To gain further insights into both the dynamics of *Runx1* expression and the nature of RUNX1+ prostate epithelial cells, surgical castration experiments were performed using wild-type and P2-*Runx1*:RFP reporter mice. Next, the castration-resistant and regenerative properties of RUNX1+ cells were evaluated in castration-regeneration using a *Runx1* driven lineage tracing mouse model.

4.2. Results

4.2.1. RUNX1 expressing cells are enriched in the castrated prostate epithelium

To track changes in *Runx1* expression following androgen withdrawal, P2-*Runx1*:RFP mice were castrated and tissue harvested ≥ 4 weeks post-surgery (Figure 4.1 A). Overall, castration reduced the size and cellularity of all mouse prostate

lobes by ~3 to 6 times (Figure 4.1 B). Strikingly, while intact prostates contained 22.8 ± 6.0% RFP+ epithelial cells, their frequency increased to 87 ± 6.0% following castration (Figure 4.1 C). After castration, high RUNX1 levels were no longer restricted to the proximal region of the prostate, and RFP was detected in virtually all basal cells of the AP, DLP and VP, as well as more than 75% of the luminal layer (Figure 4.1 D, E). These results indicate that RUNX1 is expressed in the majority in castration-resistant prostate epithelial cells.

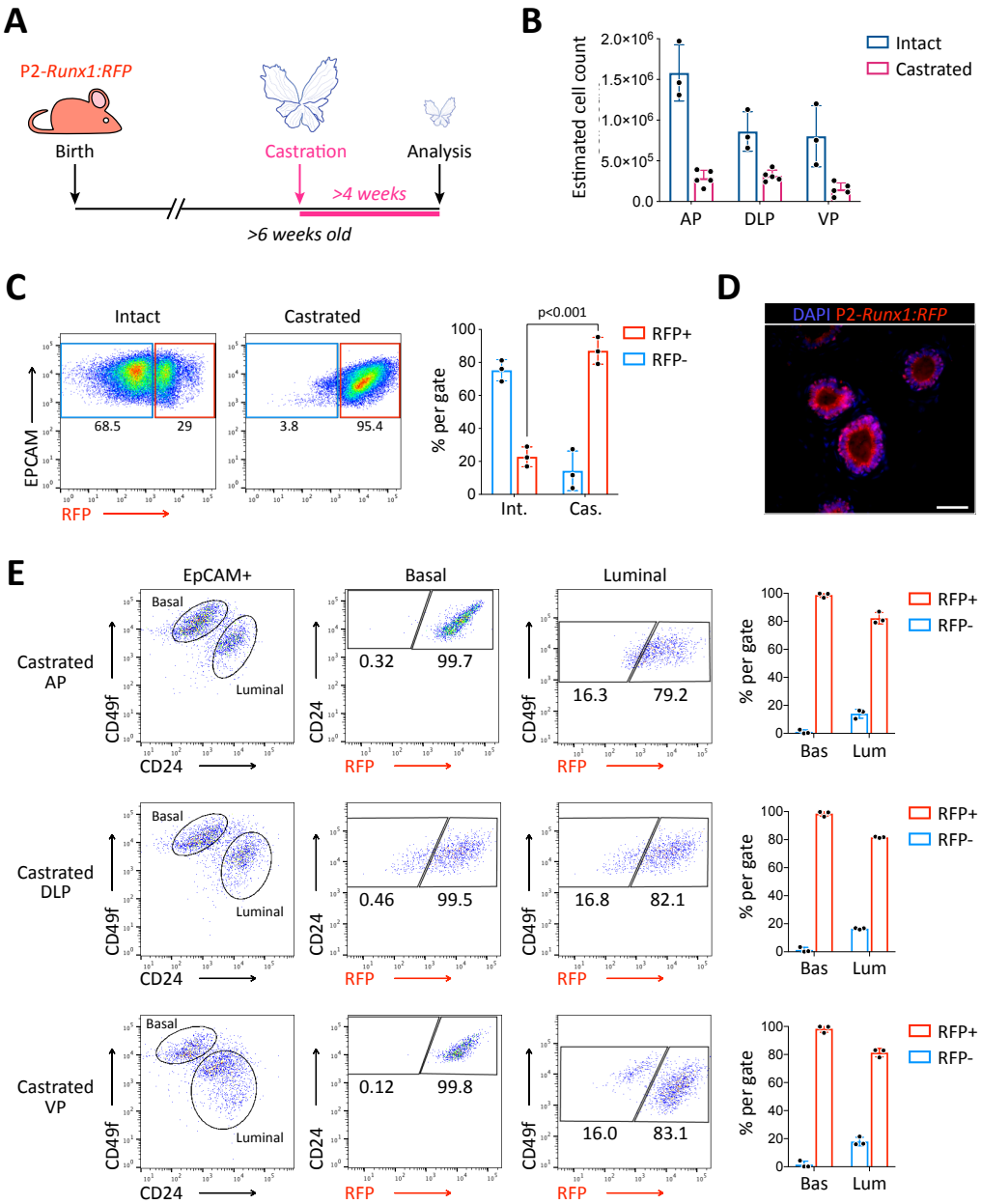

Figure 4.1. Surgical castration of P2-*Runx1*:RFP reporter mice.

(A) P2-*Runx1*:RFP reporter mice were surgically castrated between 6 to 12 weeks of age and analysed at least 4 weeks post-castration.
(B) Estimated absolute number of EPCAM+ epithelial cells in the AP, DLP and VP of intact (n = 3) and castrated (n = 5) mice. Cell counts were estimated after dissection and dissociation to single cells. Only live nucleated cells (DRAQ5+ SYTOX-) were counted.
(C) Flow-cytometry analysis and corresponding quantification of the proportion of RFP+ and RFP- cells in the EPCAM+ fraction of intact and castrated prostates of P2-*Runx1*:RFP mice. n = 3 mice per group. Int: Intact, Cas: Castrated.
(D) Endogenous RFP expression in the castrated P2-*Runx1*:RFP distal DLP prostate. Scale bar: 50 µm.
(E) Flow cytometry analysis and corresponding quantifications of the basal and luminal EPCAM+ fraction from AP, DLP and VP lobes of castrated P2-*Runx1*:RFP mice (n = 3).

To further characterise the nature of castration-resistant RUNX1+ cells, tissue sections were co-stained with classical prostate markers, including K5, K8 and AR (Figure 4.2).

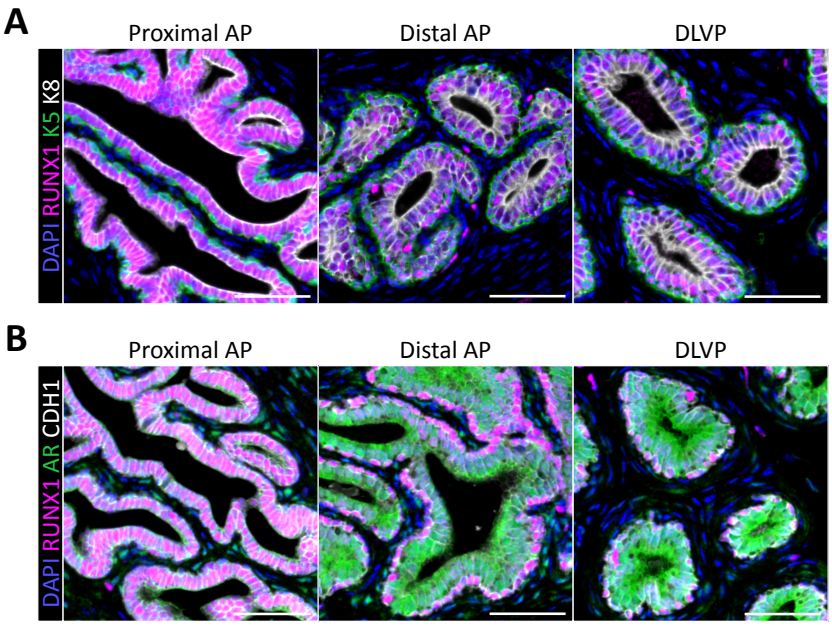

Figure 4.2. Phenotypic characterisation of RUNX1 expression in the castrated prostate epithelium.
(A) Co-immunostaining of RUNX1, K5 and K8 in the proximal AP, distal AP and DLVP.
(B) Co-immunostaining of RUNX1, AR and CDH1 in the proximal AP, distal AP and DLVP.
Scale bars: 50 µm (white).

As expected, the ratio of luminal-to-basal cells reduced after castration compared to the intact prostate (Figure 4.2 A and Figure 3.3 A), likely due to the greater

proportion of luminal cells being lost as a consequence of androgen-deprivation. Also, RUNX1 had stronger expression levels in K8+ PLCs, while those of the distal AP and DLVP (DLP and VP combined) had comparatively lower RUNX1 intensities (Figure 4.2 A). Interestingly, AR was broadly expressed in the castrated prostate, and frequently found to exhibit unusual cytoplasmic staining in addition to its common nuclear localisation (Figure 4.2 B). Also, AR expression seemed to be lower in PLCs compared to luminal cells of the distal AP and DLVP. These results suggest that distinct cellular subsets may respond differently to androgen deprivation.

Several castration-resistant luminal populations have been identified in mice (Barros-Silva et al., 2018; Kwon et al., 2016; McAuley et al., 2019; Tsujimura et al., 2002; Wang et al., 2009b; Wang et al., 2015, 5; Yoo et al., 2016), including rare CARNs. Accordingly, low, but detectable, levels of NKX3.1 were observed in some luminal cells post-castration (Figure 4.3). However, only occasional RUNX1+ NKX3.1+ luminal cells were found in the distal regions of the castrated prostate, indicating that RUNX1 and NKX3.1 mark mainly distinct castration-resistant populations. Importantly, the clear transition from RUNX1+ to NKX3.1+ cells identified in the proximal luminal layer of intact mice seemed conserved after castration.

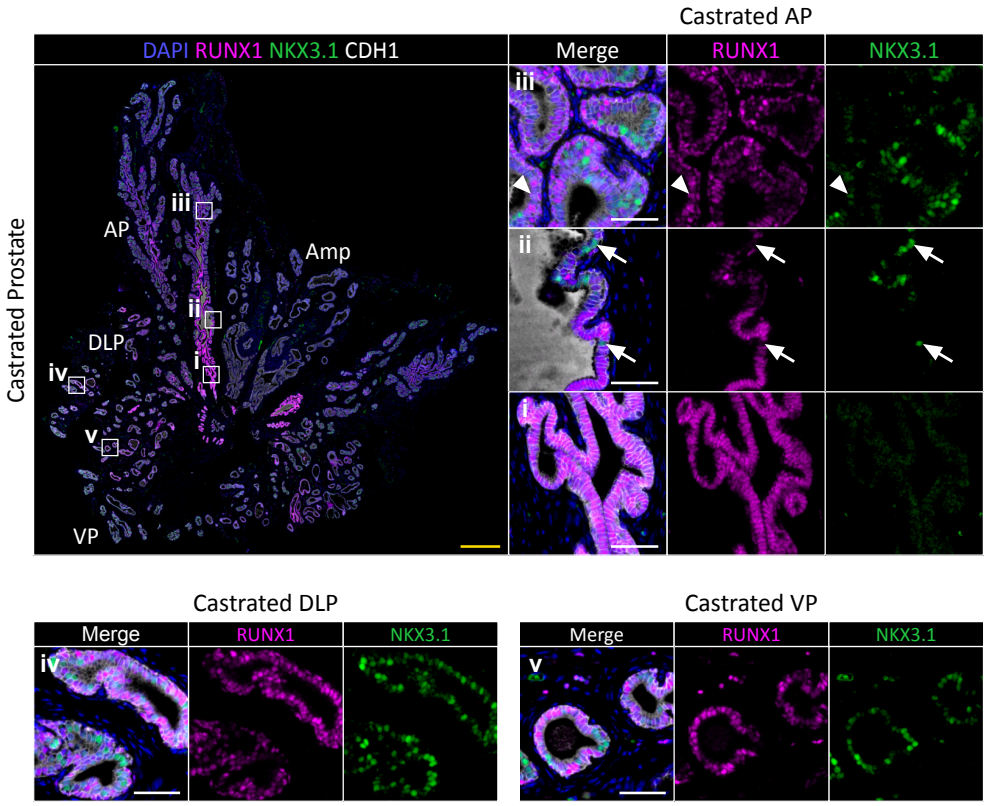

Figure 4.3. Phenotypic characterisation of RUNX1 and NKX3.1 expression in the castrated prostate epithelium.
Co-immunostaining of RUNX1, NKX3.1 and CDH1 in the castrated mouse prostate. Higher magnification images of (i) proximal AP, (ii) intermediate AP, (iii) distal AP, (iv) DLP and (v) VP are shown. Arrows indicate cells RUNX1- NKX3.1+ cells, arrowheads show luminal cell co-stained for RUNX1 and NKX3.1. Amp: ampullary gland. Scale bars: 500 µm (yellow) and 50 µm (white). Int: Intact, Cas: Castrated.

Castrated mouse prostates were then co-stained for TROP2 and SCA-1, previously found to be co-expressed with RUNX1 in intact prostates, especially in PLCs. Castrated RUNX1+ cells often co-expressed TROP2 (Figure 4.4 A), also known to be widely expressed in castrated prostate epithelium (Goldstein et al., 2008; Wang et al., 2007). However, the expression of SCA-1 was less evident, either due to the quality of the immunostainings, or reduced expression levels in the castrated prostate epithelium (Figure 4.4 B).

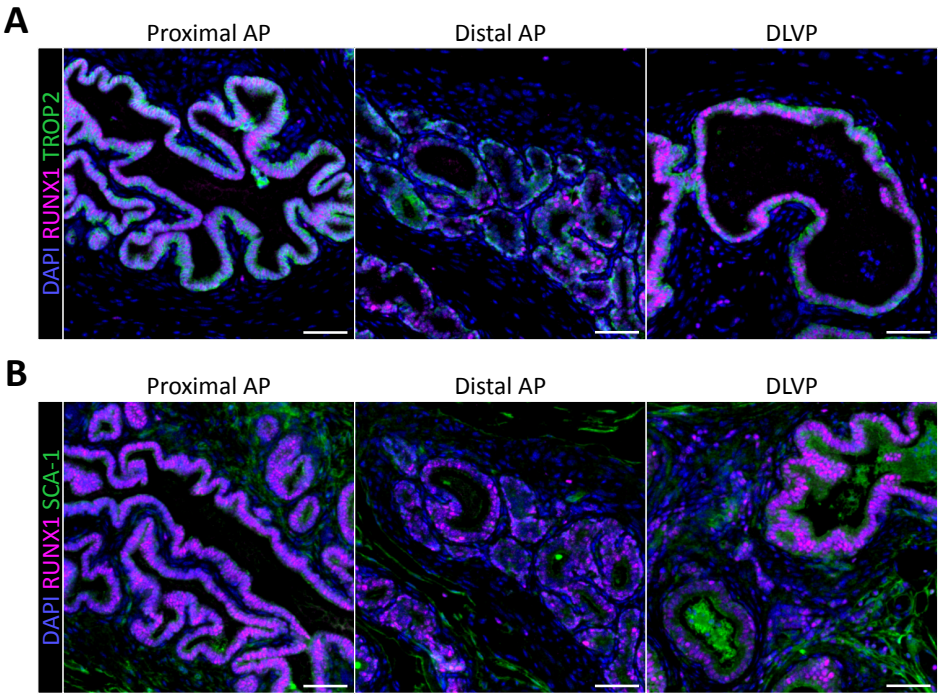

Figure 4.4. Phenotypic characterisation of RUNX1, TROP2 and SCA-1 expression in the castrated prostate epithelium.
(A) Co-immunostaining of RUNX1 and TROP2 in the proximal AP, distal AP and DLVP.
(B) Co-immunostaining of RUNX1 and SCA-1 in the proximal AP, distal AP and DLVP.
Scale bars: 50 µm (white).

Together, these results show that RUNX1 is expressed in the majority of castration-resistant cells. The RUNX1+ NKX3.1- subset identified in the proximal luminal epithelium of the intact prostate remained NKX3.1- following castration, supporting the notion that RUNX1+ PLCs constitute a distinct lineage from distal NKX3.1+ cells.

4.2.2. Lineage tracing of *Runx1* expressing cells using the *Runx1*^{CreER} *Rosa26*$^{LSL\text{-}RFP}$ model

The marked increased proportion of RUNX1+ cells following castration led to question whether this was a consequence of a broad induction of *Runx1* expression, or the existence of intrinsically castration-resistant RUNX1+ cells residing in the epithelium of intact mice. To address this question, a genetic lineage-tracing

approach was adopted using *Runx1*$^{mER\text{-}Cre\text{-}mER}$ *Rosa26*$^{loxP\text{-}STOP\text{-}loxP\text{-}RFP}$ mice (Luche et al., 2007; Samokhvalov et al., 2007), henceforth *Runx1*^{CreER} *Rosa26*$^{LSL\text{-}RFP}$.

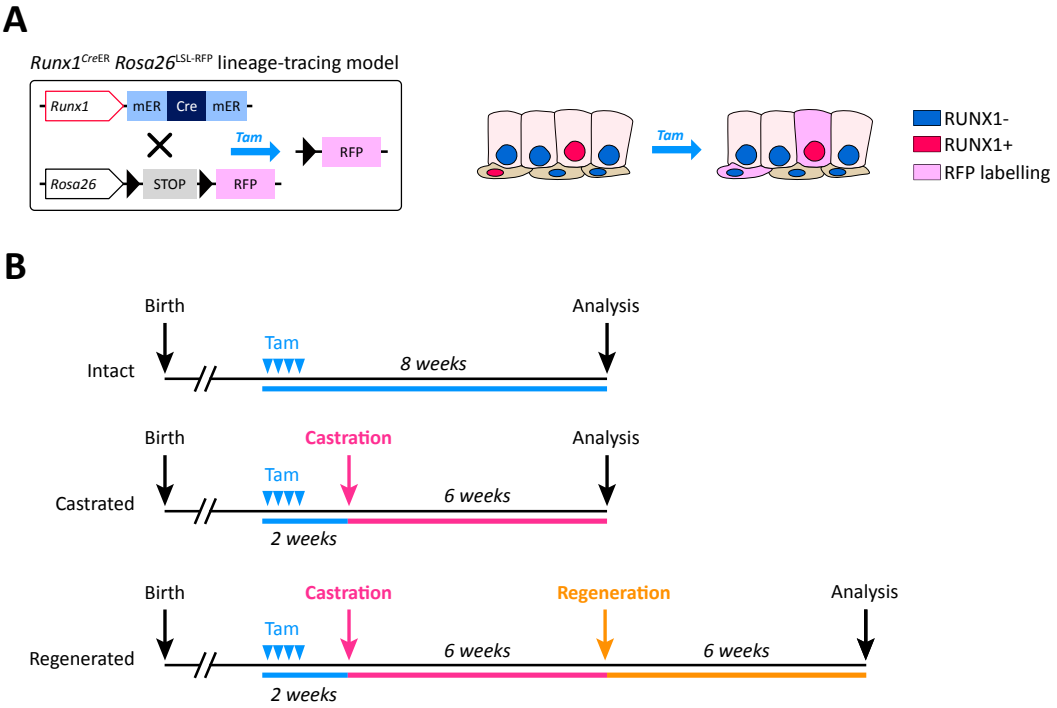

Figure 4.5. Experimental strategy to trace *Runx1* expressing cells.
(A) Schematic summary of the genetic lineage-tracing system employed. Tamoxifen triggers *Cre* nuclear translocation in *Runx1* expressing cells, which in turns activates constitutive RFP expression from the *Rosa26* locus.
(B) Experimental strategy for lineage-tracing in prostate castration-regeneration assays. One arm underwent only tamoxifen injections. One arm underwent tamoxifen injections followed by castration-induced regression. One arm underwent tamoxifen injections, castration-induced regression, and testosterone-induced regeneration.

In this model, the presence of tamoxifen results in the permanent labelling of *Runx1* expressing cells with RFP, thereby allowing long-term single-cell fate tracing *in vivo* (Figure 4.5 A). *Runx1* expressing cells were hypothesised to possess both intrinsic castration-resistant and regenerative properties. A 3-arm experiment was designed to test this hypothesis, consisting first in the labelling of *Runx1* expressing cells in intact mice of all 3 arms (Figure 4.5 B). One group of labelled mice served as control ("intact"), another group underwent surgical castration ("castrated"), and the remaining group underwent surgical castration followed by testosterone-induced regeneration ("regenerated").

To test whether this model would be suitable to track *Runx1* expressing cells in the mouse prostate, the labelling efficiency was initially evaluated in the prostate, bladder and thymus (Figure 4.6). daily intra-peritoneal injections of tamoxifen were carried out for 4 consecutive days and tissues were harvested 2 weeks after the first injection (Figure 4.6 A). Here, the labelling efficiency corresponds to the proportion of RFP labelled cells compared to the total number of RUNX1+ cells in the tissue. Using this approach, an average of 4.70 ± 2.8% prostate epithelial *Runx1* expressing cells could be genetically labelled with RFP upon tamoxifen injection, including RUNX1+ PLCs (Figure 4.6 B, C). RFP labelling efficiency was comparable in the thymus (2.8 ± 1.9%), but more efficient in the bladder (23.4 ± 15.5%). The higher labelling rates observed in the bladder could be explained by strong and homogeneous RUNX1 levels in this tissue (Figure 3.2), as well as possible increased diffusion rates of tamoxifen, normally excreted through urine. Overall, these results confirmed the suitability of the *Runx1*^{CreER} *Rosa26*$^{LSL\text{-}RFP}$ model to trace a representative cellular subset of mouse prostate *Runx1* expressing cells.

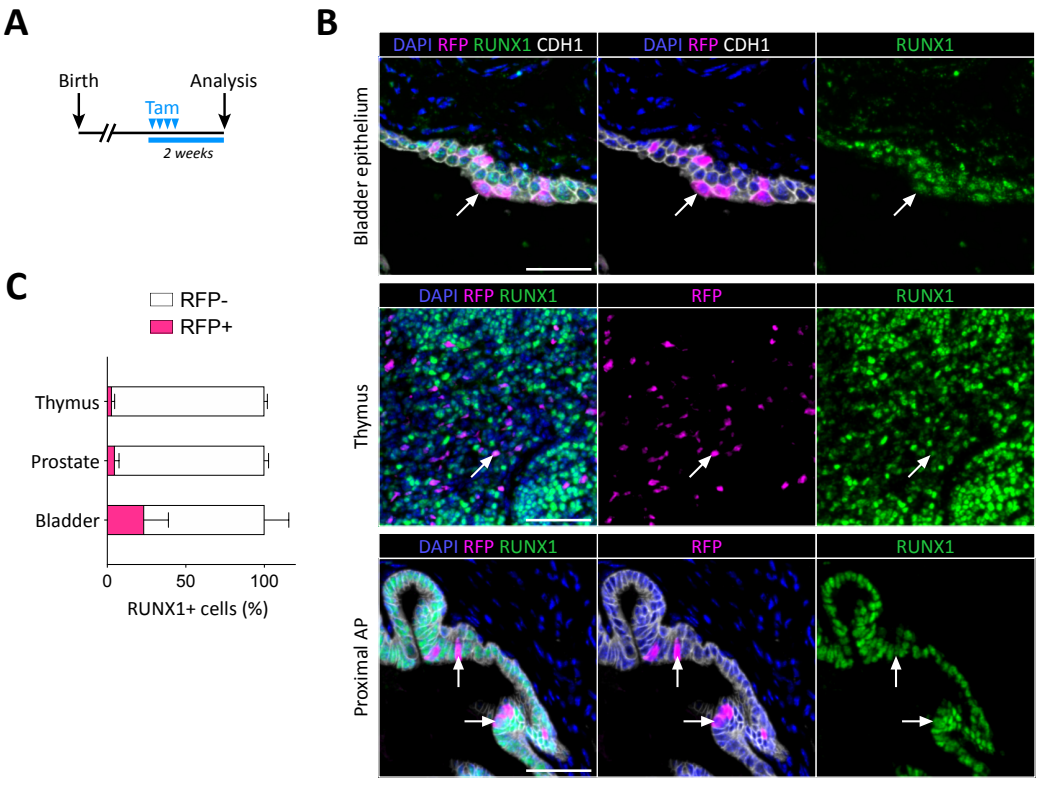

Figure 4.6. Labelling efficiency of *Runx1* expressing cells in the prostate, bladder and thymus of *Runx1*^{CreER} *Rosa26*$^{LSL\text{-}RFP}$ mice.
(A) Scheme of the tamoxifen labelling strategy to evaluate the labelling efficiency of *Runx1* expressing cells.
(B) Co-immunostaining of RUNX1 and RFP in the bladder epithelium (top), prostate (middle) and thymus (bottom). Scale bar: 50 µm.
(C) Quantification of the percentage of *Runx1* expressing cells labelled by RFP in the bladder, prostate, and thymus.

4.2.3. RUNX1+ cells are intrinsically castration-resistant but do not contribute to the regeneration of NKX3.1+ cells

Runx1^{CreER} *Rosa26*$^{LSL\text{-}RFP}$ mice were subjected to the experimental protocol described in Figure 4.5 B. In the total intact prostate epithelium, 0.54 ± 0.2‰ cells were initially found labelled with RFP by IHC (Figure 4.7 A, B). This was comparable to results obtained by flow-cytometry (0.3 ± 0.04‰ cells in intact mice) (Figure 4.7 C). Also, consistent with the expression pattern of *Runx1*, the majority of labelled cells were located in the proximal region of the prostate (Figure 4.7 D). Following surgical castration, the absolute number of RFP marked cells was stable (Figure 4.7 D) and remained mainly localised in proximal regions (Figure 4.7 D). However, the frequency of RFP+ cells in the epithelial compartment increased by ~4.3 fold (Figure 4.7 B) indicating that *Runx1* expressing cells have an enhanced capacity to survive castration compared to *Runx1* negative cells.

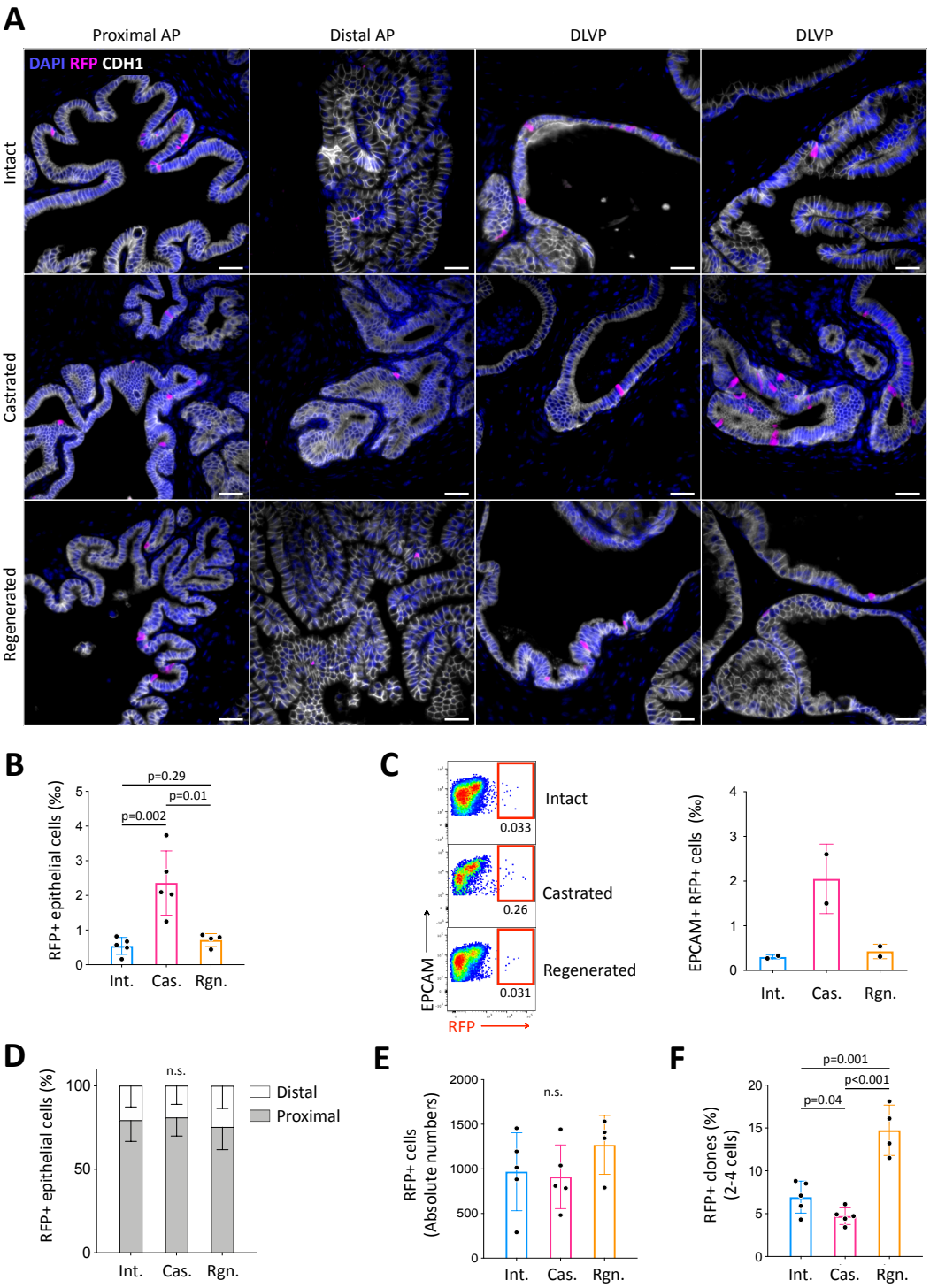

Figure 4.7. *Runx1* expressing cells are intrinsically castration-resistant but have limited regenerative potential in castration-regeneration assays.

(A) Co-immunostaining for RFP and CDH1 in the proximal AP, distal AP, and DLVP (DLP + VP) of intact, castrated, and regenerated prostates. Scale bar: 50 µm.

(B) Quantification of the ‰ of epithelial RFP+ cells in intact (n = 5), castrated (n = 5) and regenerated (n = 4) mice.

(C) Flow-cytometry analysis of intact (n = 2), castrated (n = 2), and regenerated (n = 2) *Runx1*^{CreER} *Rosa26*$^{LSL\text{-}RFP}$ mice, and corresponding quantification of the percentage of RFP+ cells in the epithelial EPCAM+ fraction.
(D) Quantification of the % of epithelial RFP+ cells in proximal and distal regions of the prostate in intact (n = 5), castrated (n = 4) and regenerated (n = 4) mice.
(E) Estimated absolute number of epithelial RFP+ cells in intact (n = 5) and castrated (n = 5) and regenerated (n = 4) prostates based on IHC quantifications.
(F) Quantification of the % of epithelial RFP+ clones comprising between 2 to 4 cells in intact (n = 5), castrated (n = 5) and regenerated (n = 4) mice.
Int: Intact, Cas: Castrated, Rgn: Regenerated.

After epithelial regeneration, only 0.71 ± 0.2‰ RFP+ epithelial cells were found in the regenerated prostate, which was comparable to the intact state (Figure 4.7 B). A similar trend was observed by flow-cytometry (Figure 4.7 C). Furthermore, no significant change in the absolute number of RFP+ cells was observed after regeneration (Figure 4.7 E). Although the majority of RFP+ clones consisted of single cells, there was in fact a minor ~2-fold increase in the frequency of larger clones (2-4 cells) in the regenerated prostate, highlighting a modest contribution of RFP labelled cells to the regenerative process (Figure 4.7 F).

The nature of RFP marked cells was then further investigated by staining for canonical basal (K5) and luminal (K8) markers. RFP marked cells were predominantly luminal K8+ in intact (86.3 ± 3.4% K8+), castrated (75.4 ± 7.4% K8+), and regenerated (79.2 ± 4.4% K8+) prostates (Figures 4.8 A, B). Thus, the ratio of basal-to-luminal cells remained largely unchanged from the intact state, both after castration and regeneration.

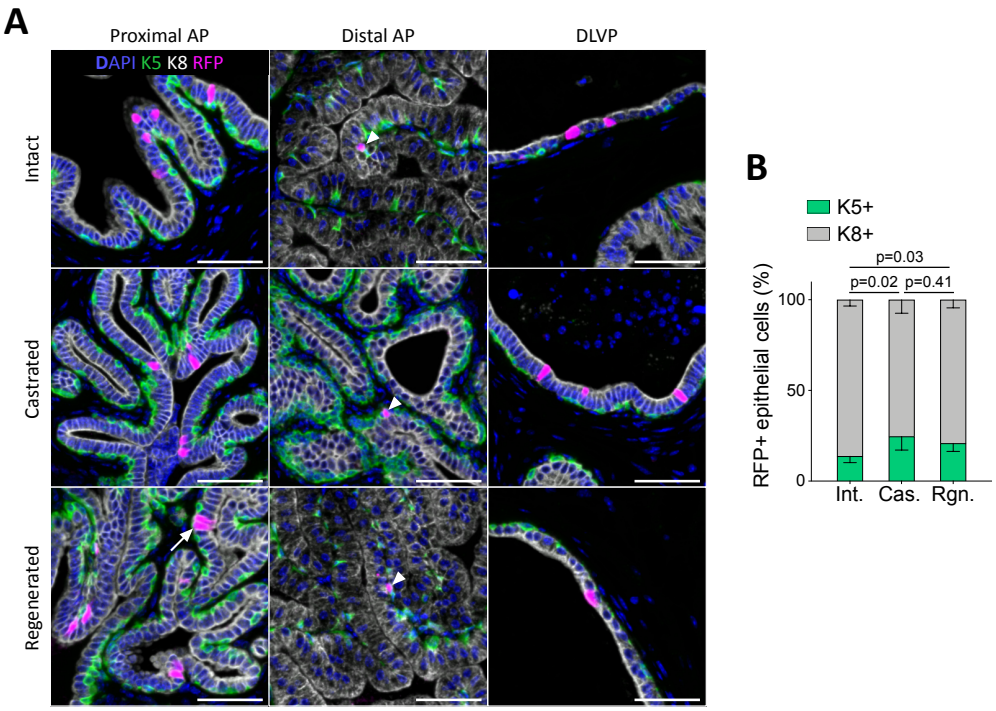

Figure 4.8. Characterisation of RFP labelled cells in castration-regeneration assays: K5/K8.
(A) Co-immunostaining of RFP, K5, K8 in the proximal AP, distal AP, and DLVP (DLP + VP). Arrowheads indicate RFP labelled basal cells (K5+) found in distal AP, the white arrow indicates a luminal (K8+) RFP+ clone made of 2 cells. Scale bar: 50 µm.
(B) Quantification of the % of RFP+ cells being K5+ or K8+ in intact (n = 5), castrated (n = 5) and regenerated (n = 4) mice. Int: Intact, Cas: Castrated, Rgn: Regenerated.

Finally, the expression of NKX3.1 was evaluated in RFP labelled cells. In line with the mutually exclusive expression pattern between RUNX1 and NKX3.1 in the mouse prostate, more than 90% of all RFP+ cells were negative for NKX3.1 in the intact prostate (Figures 4.9 A, B). Strikingly, the vast majority of RFP marked cells remained NKX3.1- in all experimental arms, both after castration and regeneration. This result suggested that *Runx1* expressing cells do not contribute to the regeneration of NKX3.1+ cells.

Taken together, these results confirmed the initial hypothesis that intact RUNX1+ are enriched in intrinsically castration-resistant cells, in particular within the proximal region. However, the observations that RUNX1+ cells remained largely proximal (Figure 4.7 D) luminal K8+ (Figure 4.8) NKX3.1- (Figure 4.9) and exhibited

only limited clonal expansion after regeneration (Figure 4.7 F) suggested that *Runx1* expressing cells constitute a self-contained lineage distinct from NKX3.1+ cells. Overall, RUNX1+ cells remain largely insensitive to fluctuations of AR signalling.

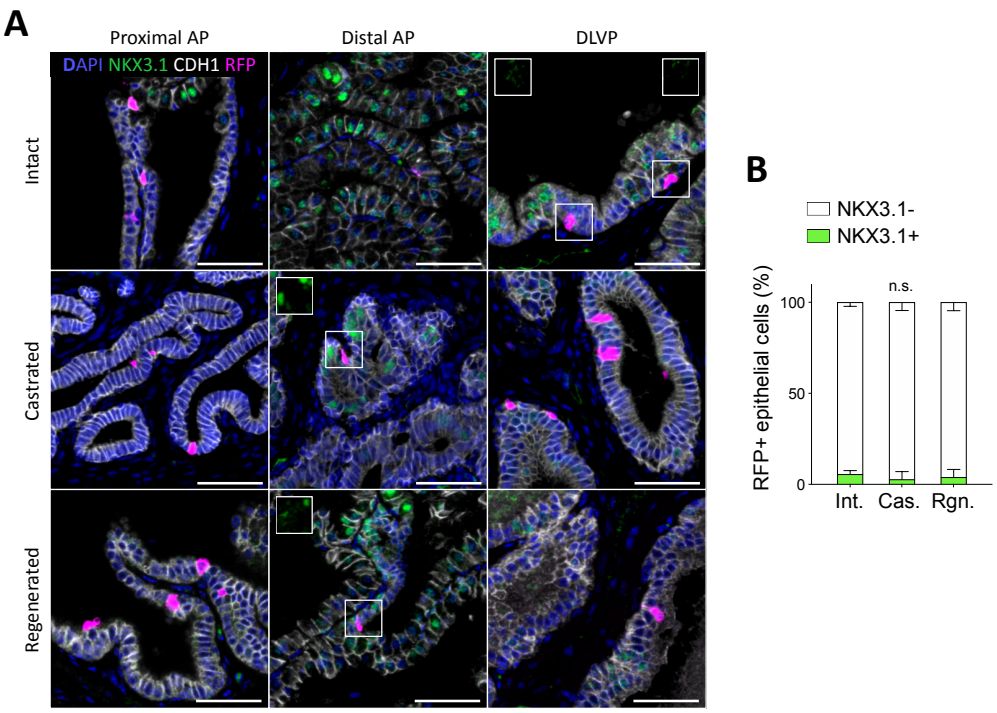

Figure 4.9. Characterisation of RFP labelled cells in castration-regeneration assays: NKX3.1.
(A) Co-immunostaining of RFP, NKX3.1, CDH1 in the proximal AP, distal AP, and DLVP (DLP + VP). Scale bar: 50 µm.
(B) Quantification of the percentage of RFP+ cells being NKX3.1+ or NKX3.1- in intact (n = 5), castrated (n = 5) and regenerated (n = 4) mice. Int: Intact, Cas: Castrated, Rgn: Regenerated.

4.2.4. Castration-resistant RUNX1+ cells have limited regenerative potential *in vivo*

The previous set of experiments demonstrated the intrinsic castration-resistant properties of *Runx1* expressing cells of the intact prostate, particularly in the proximal region. Also, these RUNX1+ cells did not contribute to the regeneration of distal NKX3.1+ cells. Since *Runx1* was also found broadly expressed in the castrated prostate, including the most distal cells (Figures 4.1, 4.2, and 4.3), it is possible that a large fraction of those distal cells also survived the castration

process, and activated *Runx1* expression after androgen withdrawal. Therefore, the lineage tracing labelling protocol was adapted to evaluate if cells that acquire *Runx1* expression upon castration were subsequently implicated in prostate regeneration. This protocol consisted in performing tamoxifen injections after the surgical castration in order to label castration-resistant *Runx1* expressing cells, before the addition of testosterone implants to stimulate regeneration (Figure 4.10 A). According to the broad expression of *Runx1* in the castrated prostate, 3.79 ± 2.4% castration-resistant cells were labelled with RFP in this experimental setup. Also, unlike in the previous experimental setup, RFP labelled cells were evenly distributed between proximal and distal regions (Figure 4.10 B). After regeneration, the proportion of RFP+ cells decreased to 0.47 ± 0.1%, suggesting that the majority of the expansion did not take place in the RFP+ compartment. Nevertheless, clone size quantification revealed the presence of larger RFP+ clones in the regenerated prostate, with up to 15% comprising 2 to 4 cells, and a tiny subset with more than 4 cells (Figure 4.10 D and Figure 4.11 A), which was not observed in the previous experimental setup (Figure 4.7 F). These results suggest that castration-resistant cells which acquire *Runx1* expression *de novo* upon castration have a greater *in vivo* regenerative potential upon testosterone add back compared with RUNX1+ cells pre-existing in the intact prostate.

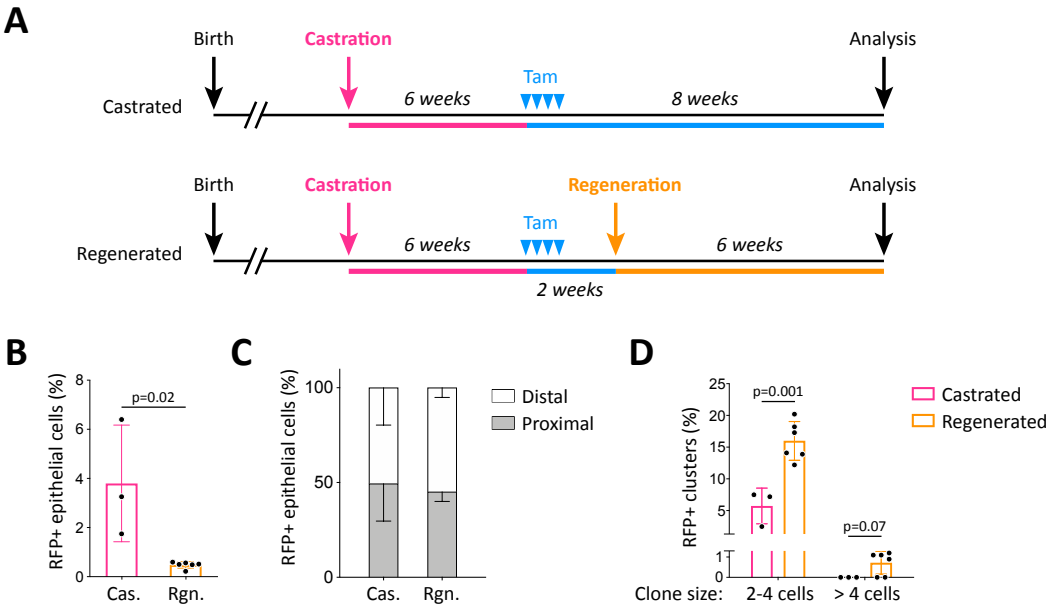

Figure 4.10. Lineage tracing of castration-resistant *Runx1* expressing cells.
(A) Experimental strategy for lineage-tracing of castration-resistant *Runx1* expressing cells during prostate regeneration.
(B) Quantification of the % of epithelial RFP+ cells in castrated (n = 3) and regenerated (n = 6) mice.
(C) Quantification of the % of epithelial RFP+ cells in proximal and distal regions of the prostate in castrated (n = 3) and regenerated (n = 6) mice.
(D) Quantification of the % of epithelial RFP+ clones comprising between 2 to 4 cells and more than 4 cells in castrated (n = 3) and regenerated (n = 6) mice.
Cas: Castrated, Rgn: Regenerated.

Additional characterisations showed that this labelling protocol resulted in an even proportion of basal K5+ and luminal K8+ cells labelled with RFP in the castrated prostate (Figure 4.11 A). Interestingly, the proportion of luminal RFP+ K8+ cells increased significantly after regeneration. This suggest that the expansion of the RFP+ compartment takes place predominantly in the luminal layer. This was further evidenced in Figure 4.11 A (asterisks) with the presence of multicellular luminal RFP+ clones. Also, distal luminal RFP+ cells were more frequently NKX3.1+ in the regenerated prostate than in the castrated one, with an increased proportion of RFP+ NKX3.1+ cells from 17.70 ± 2.3% to 77.8 ± 14.8% (Figure 4.11 C-D). Therefore, by marking cells which upregulate *Runx1* after castration, a significant fraction was found to contribute to the regeneration of the NKX3.1+ lineage.

To further evaluate the distribution of proliferation in RFP+ and RFP- subsets during the initial stages of tissue expansion, BrdU injections were performed during the first 3 days after testosterone addback (Figure 4.12 A). High rates of BrdU incorporation were observed in NKX3.1+ RFP- cells while only a small fraction of RFP+ cells were BrdU+ (Figure 4.12 B-D). Accordingly, proliferative cells were mainly observed in distal regions, and rarely found in the proximal NKX3.1- subset (Figure 4.12 C). Taken together, these results indicate that prostate regeneration is mainly driven by luminal castration-resistant NKX3.1+ cells, and that luminal castration-resistant RUNX1+ cells (RFP labelled) have a more restricted regenerative potential.

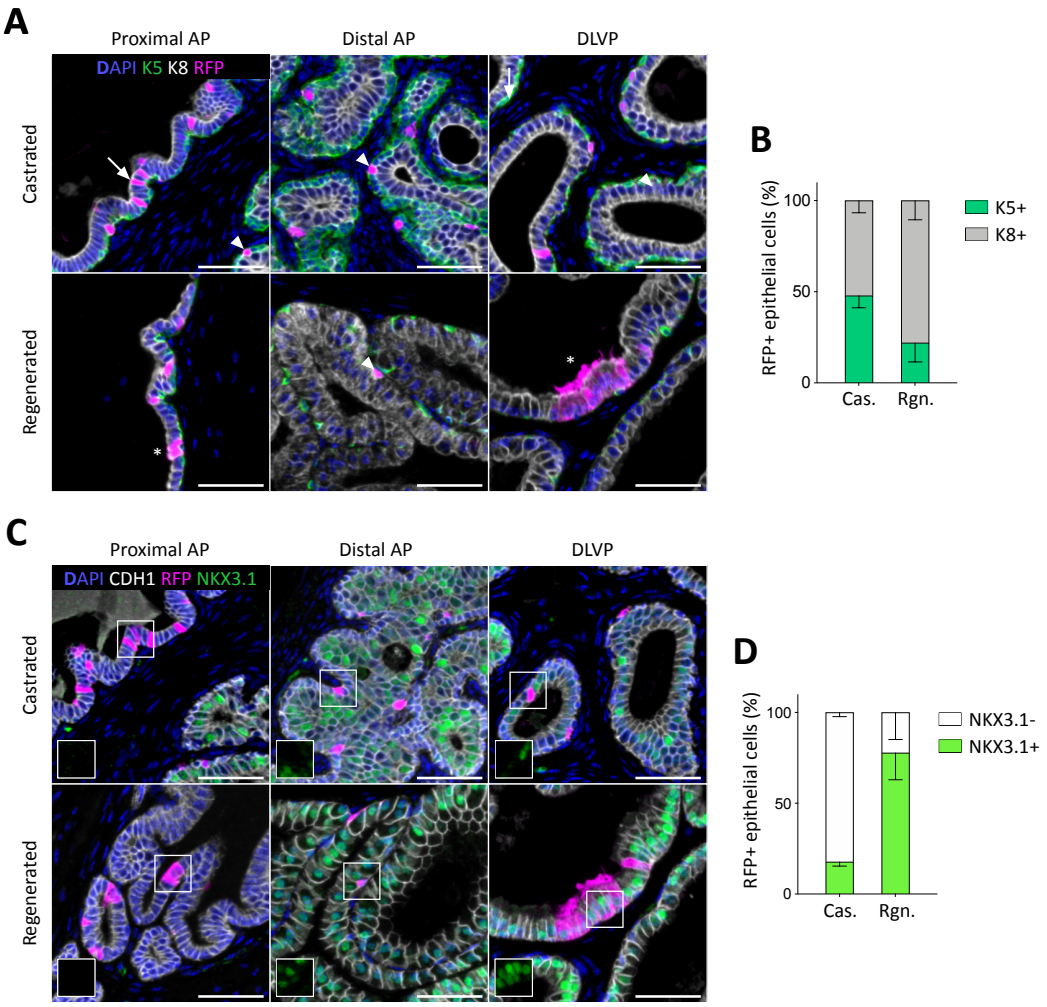

Figure 4.11. Characterisation of RFP labelled cells: K5/K8 and NKX3.1.

(A) Co-immunostaining of RFP, K5, K8 in the proximal AP, distal AP, and DLVP (DLP + VP). Arrowheads indicate RFP labelled basal cells (K5+) found in distal AP, white arrows indicate RFP labelled luminal cells (K8+), asterisks indicate multicellular RFP+ clones.

(B) Quantification of the % of RFP+ cells being K5+ or K8+ in castrated (n = 3) and regenerated (n = 6) mice.

(C) Co-immunostaining of RFP, CDH1, NKX3.1 in the proximal AP, distal AP, and DLVP (DLP + VP).

(D) Quantification of the % of RFP+ cells being NKX3.1+ or NKX3.1- in castrated (n = 3) and regenerated (n = 6) mice.

Scale bar: 50 µm. Cas: Castrated, Rgn: Regenerated.

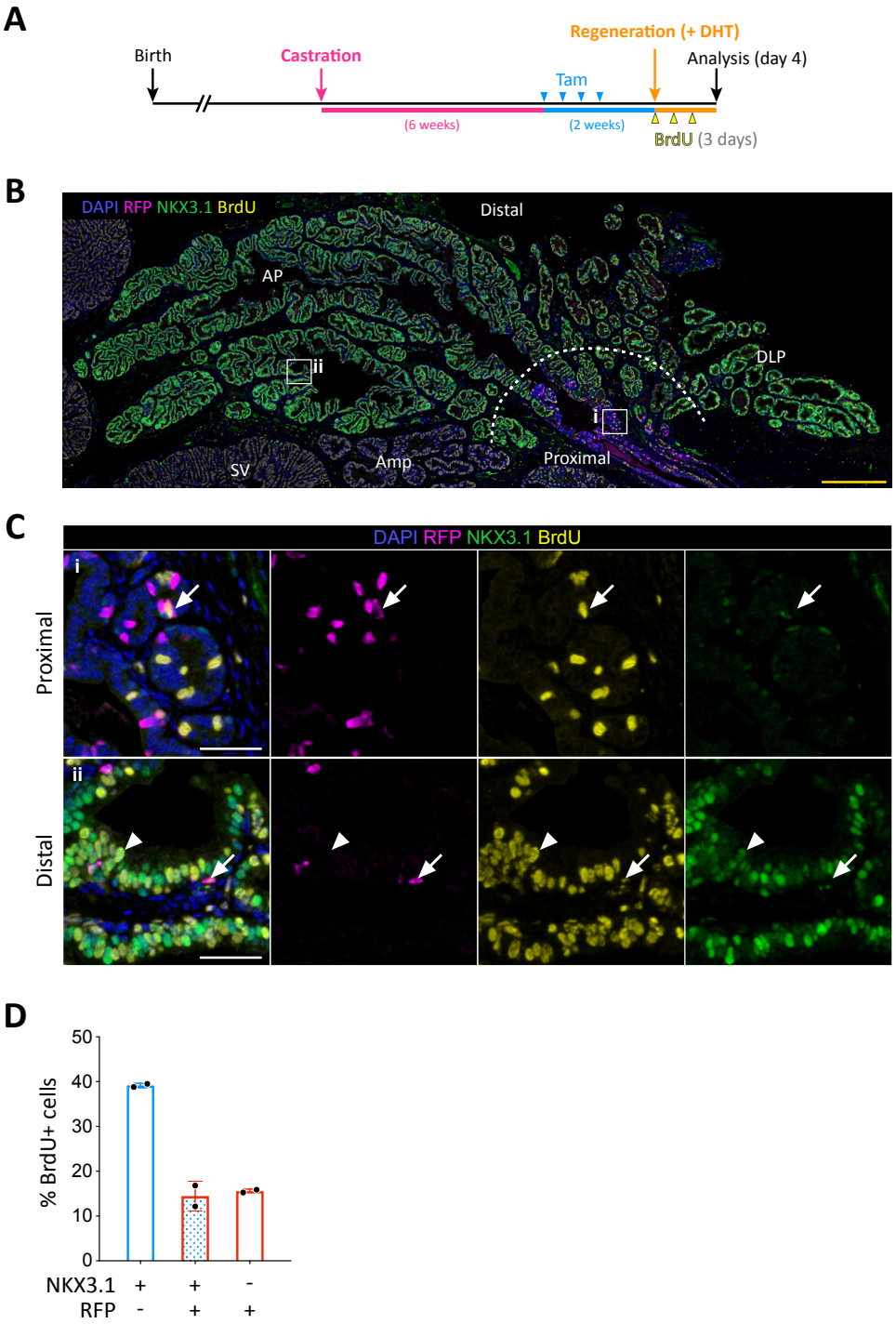

Figure 4.12. BrdU incorporation assay to track proliferative cells at the onset of prostate regeneration.

(A) Experimental strategy to evaluate the proliferative capacity of lineage marked castration-resistant *Runx1* expressing cells at the onset of prostate regeneration.

(B, C) Co-immunostaining of RFP, BrdU and NKX3.1 in proximal and distal prostates after 4 days of testosterone-induced prostate regeneration. Higher magnification images of proximal and distal regions are shown in (C). Yellow scale bar: 500 µm, white scale bar: 50 µm.

(D) Quantification of the % of BrdU+ cells being RFP- NKX3.1+, RFP+ NKX3.1+, RFP+ NKX3.1- (n = 2 mice).

4.3. Discussion

4.3.1. Summary

Despite initial therapeutic benefits of ADT, the emergence of castration-resistant disease is a major complication in the management of prostate cancer. Understanding the origin of these resistant cells and the mechanisms underlying the acquisition of castration-resistance are critical to improve PCa treatment. In light of the specific expression pattern of RUNX1 in the intact mouse prostate, the investigations presented in this Chapter aimed at evaluating the dynamics of *Runx1* expression as well as the response of RUNX1+ cells to androgen deprivation. This set of experiments were conducted using a combination of *Runx1* reporters and *Runx1* lineage tracing mouse models. Taking advantage of the possibilities offered by the *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ lineage tracing model, the *in vivo* regenerative properties of RUNX1+ cells were also evaluated in castration-regeneration assays. Collectively, the results presented here show that *Runx1* expressing cells, residing mainly in the proximal region of intact prostates and distinct from distal NKX3.1+ cells, are enriched in intrinsically castration-resistant cells (Figure 4.13). This distinct luminal lineage contributes marginally to the regeneration of the prostatic epithelium, likely due to its relative insensitivity to variations in androgen levels. Instead, among RUNX1+ cells, only those which upregulate *Runx1* in the distal region upon castration have the potential to contribute to the regeneration occurring in the NKX3.1+ compartment.

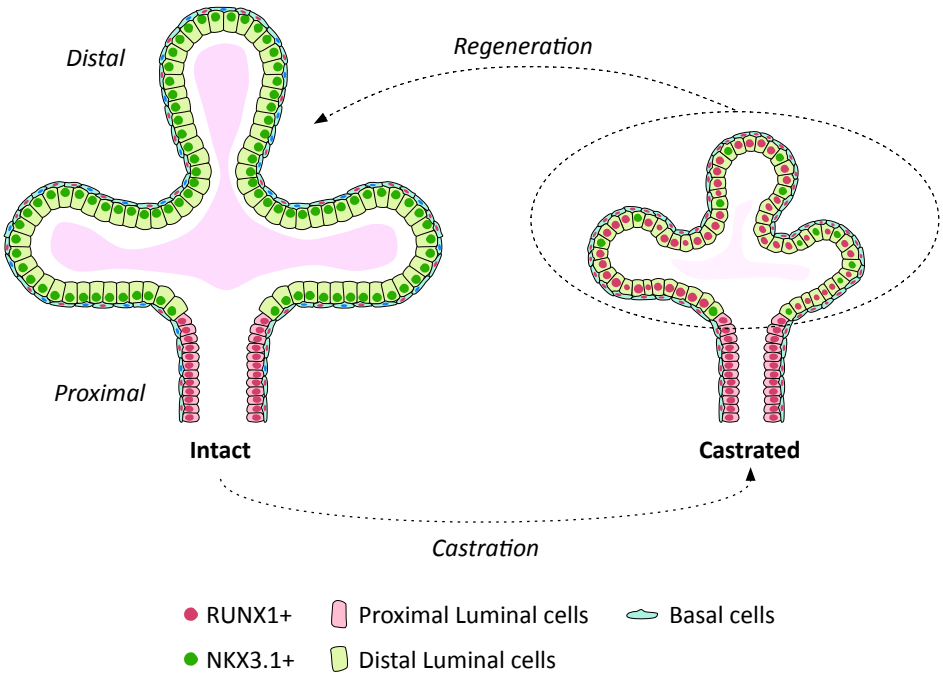

Figure 4.13. Working model of prostate castration-regeneration.
RUNX1+ proximal and NKX3.1+ distal luminal cells are distinct lineages of the intact prostate. Androgen deprivation causes the loss of a large fraction of distal NKX3.1+ cells, while a small subset survives and upregulates *Runx1*. The majority of the castrated prostate is RUNX1+, including basal cells. Testosterone driven regeneration is mainly driven by castration-resistant NKX3.1+ cells (CARNs) and a small proportion of castration-resistant RUNX1+ cells. In contrast, RUNX1+ proximal luminal cells are intrinsically castration-resistant, and do not regenerate distal NKX3.1+ lineages.

4.3.2. RUNX1 marks the majority of castration-resistant cells

Initially, the expression of *Runx1* was evaluated in the context of androgen deprivation following surgical castration of P2-*Runx1*:RFP reporters. Immunohistochemistry and FACS analyses of regressed adult prostates showed that RUNX1 was expressed in the majority of castration-resistant cells. While this result was obtained in the normal prostate and may not directly translate to a cancerous context, it suggests that RUNX1 could be a valuable marker of castration-resistant cells. Also, in line with its expression pattern in the intact normal prostate, castration-resistant RUNX1+ cells were largely distinct from previously described CARNs (Wang et al., 2009b). It should be noted that a restricted subset of RUNX1+

luminal cells co-expressed NKX3.1 after castration, indicating a small overlap between the populations expressing these two transcription factors. Beyond CARNs, other castration-resistant populations have been described, including CARBs (BMI1+) (Yoo et al., 2016), LY6D+ cells (Barros-Silva et al., 2018), SOX2+ cells (McAuley et al., 2019, 2), LGR5+ cells (Wang et al., 2015, 5), TROP2 (Goldstein et al., 2008), and SCA-1+ cells (Kwon et al., 2016). Of these, TROP2 was notably found to be often co-expressed with RUNX1 in an important fraction of the castrated epithelium. It is therefore expected that more of these markers would overlap at least partially with RUNX1. This will be addressed later at the transcriptomic level in Chapter 5.

4.3.3. Variations in AR signalling affects *Runx1* expression

The changes in *Runx1* expression observed upon modulation of AR signalling align with previous reports linking RUNX factors and hormones, including Estrogen Receptor (ER) and AR (Chimge et al., 2016; Lie-A-Ling et al., 2020; Little et al., 2014; Riggio and Blyth, 2017; Takayama et al., 2015; van Bragt et al., 2014). Along these lines, there is an intriguing parallel between the low or absent expression of *Runx1* in prostate secretory cells and the documented downregulation of *Runx1* in secretory alveolar cells of the mammary gland epithelium (van Bragt et al., 2014). Such parallel raises mechanistic questions of possible crosstalk between *Runx1* and AR. What are the signals mediated by AR signalling that trigger changes in RUNX1 expression? Do AR and RUNX1 cross-regulate each other? Do AR and RUNX1 regulate the expression of common target genes in a similar manner? Or do they have antagonistic activities? The work of Takayama and colleagues gave initial clues to these questions (Takayama et al., 2015). *RUNX1* was shown to be a transcriptional target of AR, and RUNX1 could be recruited to AR binding sites by interacting with AR, suggesting a cooperative role. Also, in AR-dependent PCa cell lines, deletion of *RUNX1* impaired AR-dependent transcription and proliferation, whereas downregulation of RUNX1 was proposed to be important for the proliferation of AR-independent cell lines. Although this study was carried out in human PCa cell lines, these results are somewhat contradictory with the

observation that RUNX1 is less expressed in intact compared with castrated mouse prostates. To clarify these findings, future work should aim at profiling the effect of androgen depletion on the DNA binding activity of both AR and RUNX1 in the mouse and human prostate.

4.3.4. RUNX1+ PLCs are intrinsically castration-resistant

The transition from proximal RUNX1+ to distal NKX3.1+ cells identified in the proximal luminal layer of intact mice appeared to be conserved after castration. However, since P2-*Runx1*:RFP reporters expressed RFP as a surrogate of *Runx1* expression, this mouse model could not be used to precisely estimate whether the increased proportion of RUNX1+ cells in the castrated prostate resulted from an induction of *Runx1* expression or the enhanced capacity of intact RUNX1+ cells to survive during prostate regression. To address this question, *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ were used to allow long term single-cell fate tracing via constitutive RFP expression in cells which expressed *Runx1* at the time of tamoxifen injection. After labelling *Runx1* expressing cells in intact mice, RFP+ cells were strongly enriched in the castrated prostate. Furthermore, most lineage traced cells were K8+ luminal cells expressing high RUNX1 levels in the proximal region of both intact and castrated prostates. This strongly indicated that RUNX1+ PLCs cells preserve their identity after castration induced regression and constitute a distinct intrinsically castration-resistant lineage. Furthermore, the absence of labelling in the distal compartment of intact and castrated *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ mice indirectly indicated that a fraction of the distal cells was also able to survive castration. While distal luminal cells did not express *Runx1* in the intact prostate, its broad expression in the castrated distal prostate indicates that at least a subset of the surviving distal fraction upregulated *Runx1* expression following castration. Therefore, surgical castration resulted not only in the preferential selection of intrinsically castration-resistant RUNX1+ cells of the proximal prostate, but also in the upregulation of *Runx1* in a large proportion of distal cells. Having established the intrinsic castration properties of RUNX1+ PLCs, it would next be of interest to study the specificities of this population, to understand what is intrinsically different at the molecular level

about this lineage, and how it relates to other prostate lineages. Also, these results raise the question of whether *Runx1* is functionally involved in the maintenance of RUNX1+ PLCs's identity, and whether it plays a functional role in the acquisition of a castration-resistant phenotype.

4.3.5. Technical limitations of lineage tracing mouse models

One important limitation of these experiments resides in the low recombination efficiencies of lineage tracing mouse models. In the prostate epithelium of *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ mice, only ~5% *Runx1* expressing cells were successfully labelled. This low recombination efficiency is not a specificity of the *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ mice. However, it is worth stating that labelling efficiencies are often referred to as "low" but rarely quantified, making the comparison difficult. Similar experiments performed in the mouse prostate reported ~1-2% efficiency in the case of a *Bmi1* lineage tracing mouse model (Yoo et al., 2016). In contrast, although this was measured during embryonic development, labelling efficiency was reported up to ~25% in the case of *Sox2* (McAuley et al., 2019). While the labelling efficiency of *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ mice was sufficient to establish the intrinsic castration-resistant properties of RUNX1+ PLCs, it is possible that the labelled fraction is not entirely representative of the initial RUNX1+ population. For example, FACS analysis of P2-*Runx1*:RFP reporters indicated broad but weak RUNX1 levels in basal cells (Figure 3.8), but RFP labelling in *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ mice seemed to be more efficient in RUNX1+ proximal luminal cells, characterised by higher RUNX1 expression. Also, the large fraction of distal RUNX1+ VP luminal cells seemed overall under-represented in *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ mice. Therefore, the results presented here apply mainly to RUNX1-high expressing cells of the proximal prostate.

4.3.6. Unlike distal compartments, RUNX1+ PLCs have limited regenerative potential

The adult mouse prostate encompasses extensive regenerative potential, which can be stimulated via ectopic testosterone delivery in castrated animals. This

leads to the full regeneration of the prostate epithelium, comparable to an intact prostate. The lineage tracing experiments offered a unique opportunity to evaluate the regenerative properties of RUNX1+ cells *in vivo*. Unexpectedly, RFP+ cells initially labelled in the intact prostate were found to have very limited regenerative potential, with only a small fraction of RFP+ clones having more than 1 cell. Also, these clones of 2 to 4 cells did not express NKX3.1 and were mainly found in the luminal layer of the proximal prostate. These observations showed that whilst capable of sustaining their own lineage, RUNX1+ PLCs did not contribute to the regeneration of distal NKX3.1+ cells. Basal and luminal lineages were previously shown to be largely self-sustained using generic basal and luminal *Cre* drivers (Choi et al., 2012; Ousset et al., 2012). However, whether distinct subpopulations of luminal cells contribute to the regeneration of the others remains poorly understood (Wang et al., 2009b; Yoo et al., 2016). Interestingly, the work presented here aligns with recent work by Guo et al proposed that a subpopulation of distal K4+ luminal cells can regenerate the distal prostate luminal lineage, whereas proximal K4+ luminal cells only contribute to regeneration of the proximal prostate luminal lineage (Guo et al., 2020). Thus, the results presented here suggest that the model of self-sustained basal and luminal populations might also extend to individual luminal subpopulations. This hypothesis should be tested in the future using additional *Cre* drivers, specific of other cell populations. The observation that RUNX1 and NKX3.1 mark distinct populations also suggests the existence of additional subpopulations. For example, distal luminal cells of each individual prostate lobes exhibit marked morphological differences whilst expressing NKX3.1. This additional degree of heterogeneity will be investigated in Chapter 5.

The goal of the initial labelling strategy used in lineage tracing experiments was to label intact RUNX1+ cells, to study their long-term fate in castration regeneration assays. However, a significant proportion of castration-resistant RUNX1+ cells only upregulated *Runx1* in the castrated prostate. This led to hypothesise that those distal castration-resistant RUNX1+ cells would contribute to the regeneration of distal NKX3.1+ luminal cells. Results from these experiments showed that there was indeed a small, but significant, fraction of distal castration-

resistant RUNX1+ cells which gave rise to NKX3.1+ lineage. Next, BrdU incorporation assays carried out at the onset of regeneration, revealed that most regeneration did not take place in the labelled fraction. These results are in line with a previous report showing that prostate regeneration is mainly driven by castration-resistant NKX3.1+ cells (Wang et al., 2009b). In fact, some castration-resistant RUNX1+ cells labelled with RFP already expressed NKX3.1 (~25%). Therefore, it is not possible to exclude that prostate regeneration was driven by CARNs expressing RUNX1, rather than castration-resistant RUNX1+ NKX3.1- cells.

4.3.7. Epithelial regenerative potential of individual lineages may be context-dependent

In a recent study, Karthaus et al used a generic *Krt8-Cre*ERT2 lineage tracing model in four colours "confetti mice" and proposed that prostate epithelial regeneration is driven by almost all luminal cells persisting in castrated prostates as opposed to a rare stem cell population (Karthaus et al., 2020). The results presented in this Chapter are compatible with this model, but further demonstrate that not all luminal subsets possess the same *in vivo* regenerative potential in response to androgen stimulation. Here, NKX3.1+ cells appear to be the most affected by androgen deprivation, but also to retain most regenerative potential upon testosterone addback. This observation contrasts with previous results shown in Chapter 3 showing that distal RUNX1- luminal cells, therefore mostly NKX3.1+ cells, have less clonogenic potential in organoids assays. These results suggest that castration-regeneration and *ex vivo* prostate organoid assays may not activate the same regenerative properties of a given cell. Likewise, while adult basal and luminal compartments have been shown to be self-sustained under homeostatic conditions (Ousset et al., 2012), luminal cells ablation results in the activation of basal multipotency (Centonze et al., 2020; Horton et al., 2019). Overall, epithelial regenerative potential may not solely depend on cell intrinsic properties, but also vary depending on the context in which it is stimulated.

Chapter 5 Characterisation of mouse prostate *Runx1* expressing cells by single-cell RNA-seq

[This section has been adapted from (Mevel et al., 2020)]

5.1. Introduction

The cell is the fundamental structural and functional unit of life. The generation of complex multicellular organisms is rendered possible by the variety of cell states that shape individual tissues despite sharing a nearly identical genotype. Understanding the relationship between genotypes and phenotypes represents a central challenge in biology. One of the most important level at which cell state and function are regulated occurs at the transcriptomic level. The ability to capture the abundances of individual mRNA molecules gives a direct read out of gene activity. At the scale of the entire transcriptome, these dynamical measurements provide a representative picture of functional cell states.

Historically, measuring the expression of a small number of transcripts was made possible using Northern blotting (Alwine et al., 1977) and RT-qPCR (Becker-André and Hahlbrock, 1989), but the emergence of high-throughput technologies has greatly facilitated quantitative analyses of whole transcriptomes. In the mid 90's, the first microarray study allowed the simultaneous quantification of 45 individual transcripts expressed at varying levels in *Arabidopsis thaliana* (Schena et al., 1995). The technology relied on the hybridisation of transcripts to an array of complementary probes. Microarrays technology then rapidly evolved and became a key approach to investigate gene expression of up to thousands of transcripts at once (Hoheisel, 2006). However, major limitations of these arrays were associated with the predefined set of probes and their variable specificities.

With the advent of next generation sequencing technologies, the first RNA-sequencing (RNA-seq) protocols allowing deep sequencing of entire transcriptomes were published in 2008 (Lister et al., 2008; Mortazavi et al., 2008; Nagalakshmi et

al., 2008). Briefly, RNA is isolated and converted into a library of short fragments of cDNA which are then loaded onto a sequencing platform. The latter generates millions of nucleotide sequences, known as reads, which are used for downstream computational analysis. Abundances are derived from the number of counts of each read that map onto specific locations of a reference genome or transcriptome. These are then often aggregated at the gene or transcript level. Unlike microarrays, RNA-seq does not required prior knowledge of existing sequences and therefore benefits from the ability to detect rare and novel transcripts. It also has a wider dynamic range, greater specificity and sensitivity (Roy et al., 2011; Wang et al., 2009a). Together with the decreasing cost of sequencing, RNA-seq has gradually become the method of choice for transcriptomic analyses, with applications in many aspects of biology, such as the characterisation of splicing dynamics (Mehmood et al., 2020), differential gene expression analyses between conditions (Love et al., 2014), as well as the discovery of predictive clinical biomarkers of disease progression (Yang et al., 2020). However, RNA-seq typically averages abundance measurement across "bulks" of hundreds to millions of cells per sample, which assumes a certain degree of homogeneity within the populations of cells being sequenced. While such assumption may be acceptable in specific experimental contexts, it does not apply to heterogeneous systems where population level averages do not reflect individual cell states (Loeffler and Schroeder, 2019; Raj and van Oudenaarden, 2008). Therefore, cell-to-cell variations cannot be reliably studied in bulk RNA-seq analyses, which are flawed by the unknown proportions of individual cell states present in the samples.

The observation that gene expression levels vary between cells of apparently homogenous populations has been known for a long time (Levsky and Singer, 2003). Novick and Weiner had previously described heterogeneous responses in beta-galactosidase production between individual bacteria of same strain (Novick and Weiner, 1957). Since, numerous techniques have helped describe and model the stochastic nature of gene expression, such as immunostaining for specific proteins, *in situ* hybridisation of RNA or even fluorescence microscopy (Raj and van

Oudenaarden, 2008). However, until recently, no technology could measure transcriptome wide gene expression levels at the single cell resolution. Almost three decades ago, pioneering work by Brady et al. and Eberwine et al. successfully reported the amplification of cDNA from single cells, but the expression levels of only a limited number of genes could be estimated (Brady et al., 1990; Eberwine et al., 1992). The method was later progressively scaled up to allow measurements of more genes and cells (Peixoto et al., 2004; Sheng et al., 1994), ultimately leading to its implementation on microarrays (Kurimoto et al., 2006; Tietjen et al., 2003; Warren et al., 2006). Multiplexed reverse transcription PCR combined with microfluidics allowed the interrogation of multiple targeted genes at the same time in a few hundred cells and became more widely accessible with commercially available solutions such as RT-qPCR based Fluidigm platforms. For instance, this enabled to investigate the expression of 48 genes on ~500 cells along the development of embryos from zygote to blastocyst stage (Guo et al., 2010). However, the resolution of these platforms remained relatively low throughput and biased by the predefined list of genes investigated.

The very first single cell (sc)RNA-seq study was published only one year after the initial bulk RNA-seq protocols. Tang et al., reported the gene expression profile of a single 4-cell stage blastomere (Tang et al., 2009). While this was a great technical achievement, the ability to perform scRNA-seq was initially reserved to a few specialised laboratories. Also, the cost of the first scRNA-seq protocols was prohibitive and involved several days of work to amplify the cDNA of individual cells. Therefore, in the early days of scRNA-seq, only a small number of single cells were sequenced at once, limiting the extent of biological interpretations. It was however evident that individual cells could be distinguished and grouped by homologies based of their individual transcriptomes (Pollen et al., 2014; Yan et al., 2013). In 2013, scRNA-seq was elected Method of the Year by Nature Methods (Method of the Year 2013, 2014), and a myriad of scRNA-seq protocols have been published, with exponential growth of the number of individual cells that can be sequenced in a single experiment (Svensson et al., 2018). Amongst popular

methods, plate-based methods such as Smart-seq2 (Picelli et al., 2013) or CEL-seq2 (Hashimshony et al., 2016) require single cell sort into individual wells. Smart-seq2 can produce full-length transcript coverage while CEL-seq2 only captures and sequences the 3' end. In contrast, bead based-methods were developed as a higher throughput alternative leveraging the need of upstream cell sorting, by capturing suspensions of cells into nanolitre droplet emulsions loaded with individually barcoded beads and amplification reagents. The most widely used 3' droplet-based methods include Drop-seq (Macosko et al., 2015), InDrop (Klein et al., 2015), and 10x Genomics Chromium (Zheng et al., 2017) platforms. Additional approaches to increase the number of cells captured were developed with the use of sub-nanolitre arrays containing barcoded beads such as Seq-Well (Gierahn et al., 2017) or Microwell-Seq (Han et al., 2018).

The field of scRNA-seq is still rapidly evolving (Svensson et al., 2018), and it has now become a gold-standard approach to characterise heterogeneous cell populations, study developmental and regenerative processes, as well as disease biology. Single cell transcriptomic data has allowed to redefine the cellular landscape of specific tissues (Treutlein et al., 2014; Zeisel et al., 2015), discover novel, rare and transient cell types (Grün et al., 2015), with the concomitant identification of new marker genes specific to cell populations, predict routes of differentiation (Saelens et al., 2019) as well as inferring gene regulatory networks defining specific cell states (Aibar et al., 2017; Chan et al., 2017; Matsumoto et al., 2017). The increased throughput of scRNA-seq has enabled researchers to capture enough cells to recapitulate the development of entire organisms such as zebrafish (Wagner et al., 2018) or planaria (Plass et al., 2018). Similarly, single cell transcriptomic landscapes of mouse embryos from blastocysts stage to gastrulation revealed insights into lineage relationships and specification at an unprecedented resolution (Nowotschin et al., 2019; Pijuan-Sala et al., 2019). Recently, scRNA-seq datasets of human lung and airways suggested that the expression pattern of mediators of SARS-CoV-2 viral entry in specific cell type may explain clinical and biological aspects of the disease (Muus et al., 2020). The widespread use of this

technology has led to the creation of large consortia, including the Human Cell Atlas (HCA) which aims at generating a comprehensive reference map of all human cells (Regev et al., 2017). In parallel to the development of scRNA-seq, other single cell technologies have emerged, enabling additional single cell 'omic' measurements such as cell surface proteins (Stoeckius et al., 2017), chromatin accessibility (Cusanovich et al., 2015; Cusanovich et al., 2018), DNA methylation (Hu et al., 2016), histone modifications (Kaya-Okur et al., 2019), transcription factor activity (Hainer et al., 2019). A telling example of the explosion of single cell technologies was the celebration of "single-cell multimodal omics" as 2019 Method of the Year (Method of the Year 2019: Single-cell multimodal omics, 2020). In addition to these advances, the field of spatial transcriptomics holds great promises to enable spatially resolved high dimensional transcriptomic measurements (Burgess, 2019) and is expected to reach single cell resolution in the near future.

In the prostate, Barros et al used multiplexed single-cell quantitative PCR to interrogate the expression of 313 cell surface markers and prostate related genes in 95 individual cells isolated from intact and castrated mice (Barros-Silva et al., 2018). This analysis revealed the existence of multiple luminal subsets with distinct gene expression patterns. Notably, they identified LY6D as a putative stem/progenitor marker also enriched in castration resistant cells, which may be a relevant marker in the context of prostate tumours. While a number of separate studies had already reported scRNA-seq atlases of the majority of the mouse and human organs, until recently, scRNA-seq datasets of the prostate were not available. In 2018, Henry et al. reported the first scRNA-seq study of human prostate samples and proposed the existence of SCGB1A1+ 'club' and KRT13+ 'hillock' cells in the peri-urethral zone (Henry et al., 2018). However, such resource remained unavailable for the mouse prostate, preventing comparisons between cell types identified. In 2020, a number of additional publications describing the cellular composition of the mouse and human prostate have been published (Crowley et al., 2020; Guo et al., 2020; Joseph et al., 2020; Karthaus et al., 2020), and will be discussed in section 5.3.

The data presented in Chapter 3 and 4 indicate that RUNX1 is expressed in different subpopulations of the intact mouse prostate. In particular, high RUNX1 expression was found to be highly expressed in PLCs, an intrinsically castration-resistant population capable of sustaining its own lineage *in vivo* and exhibiting facultative stem cell behaviour in organoid assays *ex vivo*. Except for NKX3.1 which displayed a largely mutually exclusive expression pattern with RUNX1, co-immunostainings with several markers previously shown to mark specific subpopulations of the mouse prostate were found to overlap, at least partially, with RUNX1. Also, RUNX1 was expressed in the majority of the castrated epithelium. Together, these results suggested that RUNX1 may be expressed in functionally and spatially distinct prostate epithelial lineages. However, the restricted number of markers that can be profiled on single tissue sections prevented from drawing a precise map of the complex prostate epithelial landscape. To leverage this limitation and characterise the different RUNX1+ and RUNX1- cell types present in the intact and castrated prostate epithelium, transcriptomic profiling of RFP+ and RFP- fractions was performed using droplet-based scRNA-seq of P2-*Runx1*:RFP reporter mice.

5.2. Results

5.2.1. Experimental design

To transcriptionally characterise RUNX1+ and RUNX1- fractions residing at different anatomical locations of the prostate, droplet-based single cell (sc)RNA-seq was performed using the 10x Genomics Chromium platform (Figure 5.1 A). EPCAM+ RFP+ and RFP- cells were sorted from individually dissected lobes of intact and castrated prostates isolated from P2-*Runx1*:RFP reporter mice (Figure 5.1 B). The AP, DLP and VP were kept separate in order to gain insights into the cellular composition of each lobe, shown to be particularly heterogeneous in Chapter 3. Basal and luminal cells of the same lobe and RFP fraction (for example AP RFP+) were sorted in a single tube under the assumption that basal and luminal cells would cluster apart in the analysis. A representative FACS profile of the populations

sorted is shown in Chapter 2, section 2.8.1.1. In order to minimize technical confounders such as multiplets and batch effects (McGinnis et al., 2019b), sorted populations were multiplexed using MULTI-seq lipid-tagged indices (Figure 5.1 C). In this approach, plasma membrane bound lipid-modified oligos (LMOs) act as anchors for specific DNA barcoded oligonucleotides. Following reverse transcription, library preparation and sequencing, cells can then be demultiplexed to assign their identity back to their population of origin. In total, three independent experiments (also referred to as "runs") were performed to obtain sufficient cell numbers from the different populations isolated.

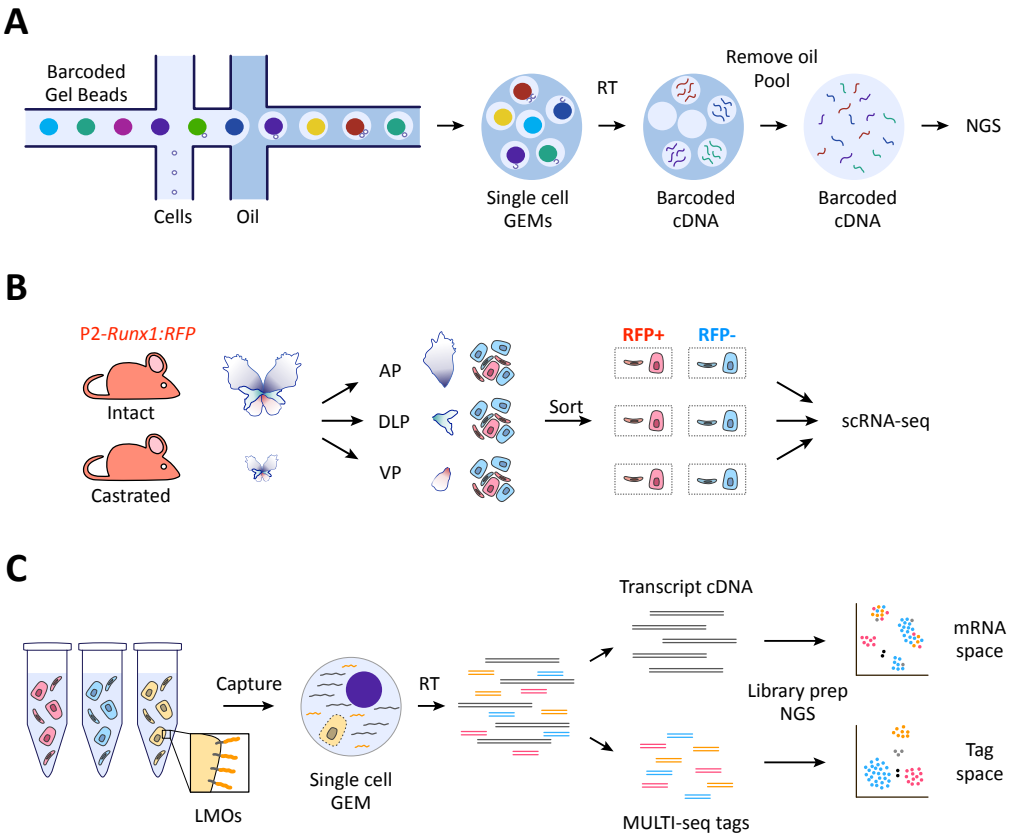

Figure 5.1. Experimental tools and design used for scRNA-seq of adult prostate epithelial populations.

(A) Diagram of the 10x Genomics Chromium workflow to capture and sequence single-cells transcriptomes. Single cells and barcoded gel beads are captured in aqueous droplets in an oil-surfactant solution to generate Gel Beads in EMulsion (GEMs). Cell lysis within the droplets allows reverse transcription of mRNA to produce barcoded cDNA. After breakage of the droplets, cDNA is amplified and sequenced.

(B) Experimental strategy for scRNA-seq of RFP+ and RFP- cells individually dissected lobes of intact and castrated prostates isolated from P2-*Runx1*:RFP reporter mice.

(C) Diagram of the MULTI-seq workflow. Samples are barcoded in individual tubes using MULTI-seq lipid-modified oligos (LMOs) bound to DNA barcodes containing a specific 8-bp sample specific sequence. Samples are pooled before cell encapsulation. After reverse transcription, cDNA and MULTI-seq libraries are separated by size selection and amplified independently. Sequencing of both libraries produces two count matrices corresponding to gene expression ('mRNA space') and sample-barcode abundances ('tag space'), allowing the assignment of each cell to its population of origin. This method allows the detection of GEM containing > 1 cell (black dots, multiplets) generated during cell encapsulation. Cells for which MULTI-seq labelling has failed are shown in grey.

5.2.2. Quality control and sample demultiplexing

Following sequencing and data pre-processing, droplets containing cells to be kept for downstream analysis were selected using the 10x's "Cell Ranger" software. The results are displayed using a commonly used "barcode rank plot" or "knee plot", representing decreasing Unique Molecular Identifiers (UMIs) counts in log scale for each cell barcode (Luecken and Theis, 2019). UMIs consist of random barcodes used to identify and remove duplicate reads generated during amplification. All three sequencing runs had a sharp drop-off between cell-associated barcodes and barcodes associated with empty droplets, indicating successful experiments (Figure 5.2 A). While this step is effective to remove empty droplets, it does not discriminate between high- and low-quality cells.

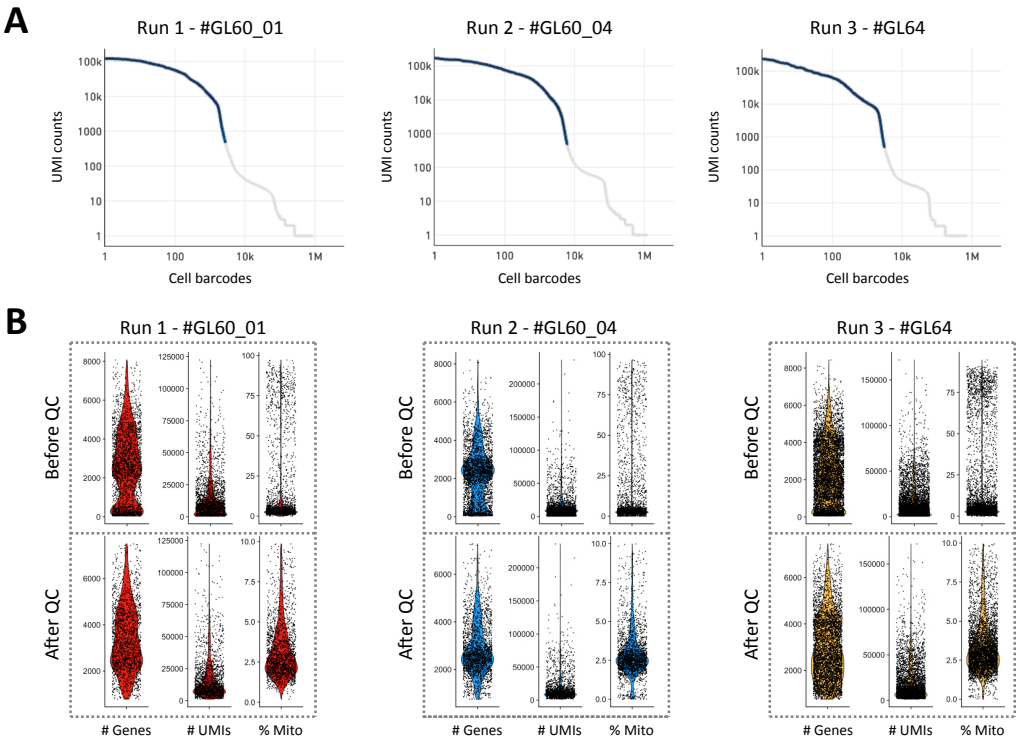

Figure 5.2. Quality control of the adult mouse prostate scRNA-seq dataset.
(A) Barcode rank plots produced by the 'Cell Ranger' software showing cell-associated barcodes (blue) and barcodes associated with empty droplets (grey).
(B) Violin plots of the number of genes detected per cell (# Genes), number of UMIs per cell (# UMIs) and percentage of mitochondrial transcripts (% Mito). Results are shown for each run, before and after quality control.

Next, high quality cells retained for analysis were demultiplexed using MULTI-seq tag data to assign them back to their population of origin, Overall, the MULTI-seq labelling showed varying levels of labelling efficiencies at the cellular, population, and experimental level (Figure 5.3 A). The labelling of "Bar3" (corresponding to the "Intact DLP RFP+" population) did not work in both Run 1 and Run 2 and was therefore replaced by a different MULTI-seq tag in Run 3. Despite the same protocol being used for all three runs, the MULTI-seq labelling of the last experiment performed better, with a greater signal-to-noise ratio and the presence of binomial distributions for each MULTI-seq tag. Uniform Manifold Approximation and Projection (UMAP) visualisation of the MULTI-seq tag count expression data was generated for each run and coloured by demultiplexed population (Figure 5.3 B). As expected from the suboptimal MULTI-seq labelling performance of Run 1 and

2, a significant proportion of the cells could not be matched to a specific population. These “negative” cells, as well as predicted “doublets” were excluded for downstream analyses. After these additional filtering steps, UMAP visualisation revealed the presence of distinct and homogenous clusters in MULTI-seq barcodes space, which confirmed that high quality labelled cells were retained in the final dataset.

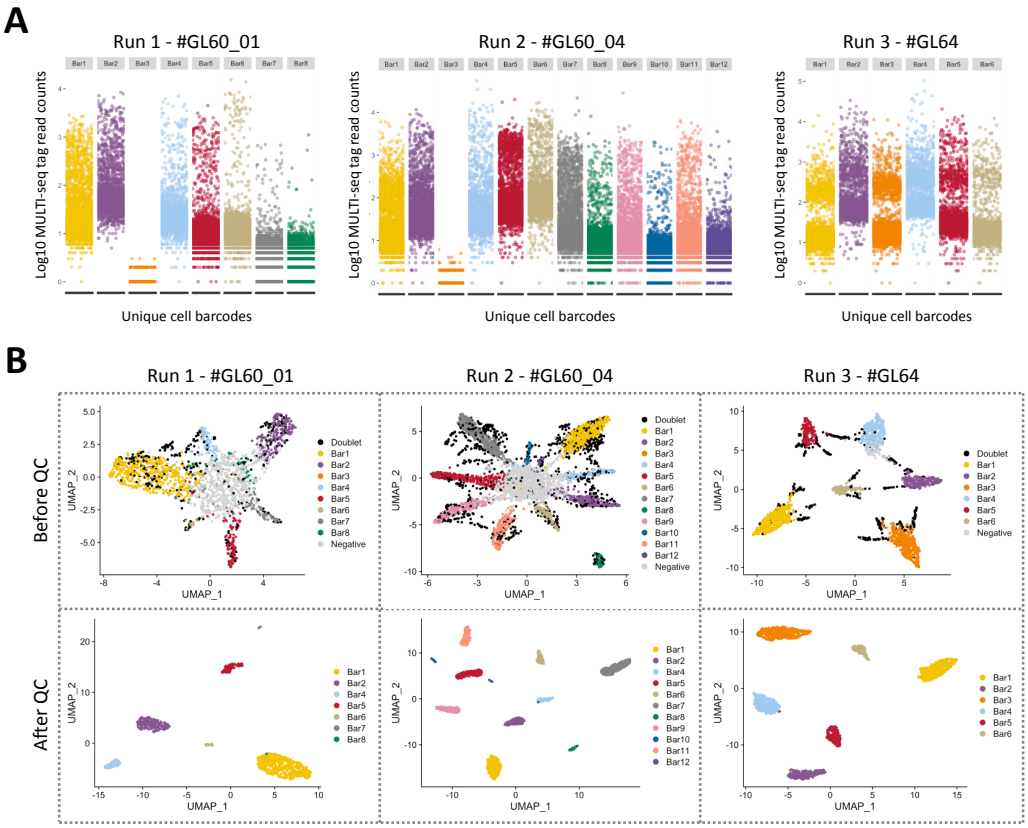

Figure 5.3. Quality control of the MULTI-seq labelling.
(A) Scatter plot showing the abundance of each MULTI-seq tag counts (log10) for each cell barcode.
(B) UMAP visualisation in MULTI-seq barcodes space for each independent experiment (Run1: left, Run2: middle, Run 3: right) before (top) and after (bottom) quality control and barcode filtration (see section 2.8.1.4. for details).

Table 5.1 shows the number of cells kept for analysis after the different quality control/filtering steps performed. In total, 4,516 cells were retained for downstream analyses.

Table 5.1. Number of cells after each quality control filtering step.

	Run 1: GL60_01	Run 2: GL60_04	Run 3: GL64
# of cells before QC	2612	6034	3028
# of cells after recovered QC	1891	4398	2243
# of after tags classification	777	2186	1673
# of cells after DoubletFinder	752	2091	-
Final # of cells	752	2091	1673

#: number

5.2.3. Integration and graph-based clustering of the whole dataset

The three independent experiments were combined into a single dataset using a standard batch-correction and integration procedure previously described for scRNA-seq datasets (Stuart et al., 2019). After linear log-normalisation, this method allows to effectively combine cells from different experiments or batches into a shared space through the identification of cell pairwise correspondences between cells across distinct datasets. After integration, UMAP visualisation in gene expression space showed the presence of multiple communities of cells, or clusters (Figure 5.4 A).

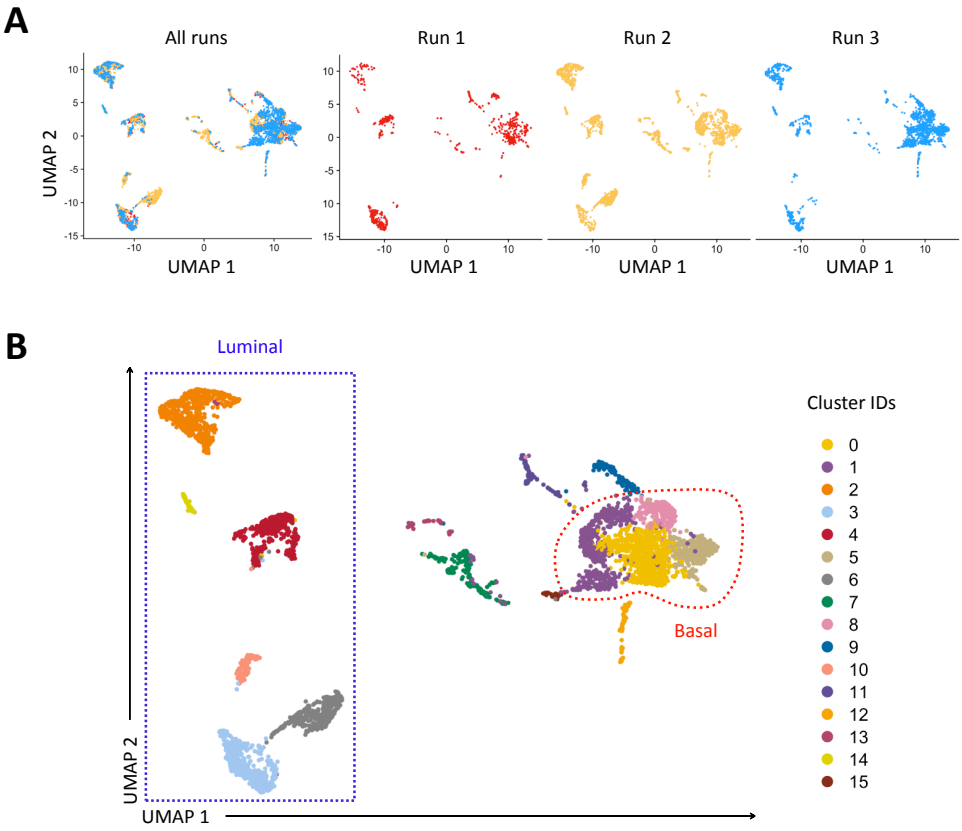

Figure 5.4. UMAP visualisation of the whole adult scRNA-seq dataset.
(A) UMAP visualisations of the integrated batch-corrected dataset (n = 4,516 cells from 3 independent experiments) combined and split by individual experiment (Run 1: red, Run 2: yellow, Run 3: blue).
(B) UMAP visualisation of the integrated batch-corrected dataset after unbiased graph-based clustering. Each cluster is represented by a unique colour.

Importantly, most clusters contained a combination of cells from all three runs, confirming the success of the integration procedure. Indeed, a poorly integrated dataset of cells known to have a similar biological origin would have clustered cells by experiment instead of the underlying biological differences. Here, unbiased graph-based clustering revealed the presence of 16 biologically distinct clusters (Figure 5.4 B).

Canonical marker genes of cell types residing within the prostate were iteratively explored to label and characterise each of the 16 clusters. This manual approach was combined with differential expression analyses between clusters to

confirm final labels (data not shown). Overall, 10 prostate epithelial clusters were labelled based on the expression of *Epcam*, *Krt8*, *Cd24a*, *Spink1*, *Krt19*, *Tacstd2*, *Psca*, *Krt4*, *Tgm4*, *Nkx3-1*, *Pbsn*, *Msmb*, *Piezo2*, *Trp63*, *Krt5* and *Krt14*. Based on the expression of basal and luminal markers, clusters 0, 1, 5, 8 were labelled as "Basal" while 2, 3, 4, 6, 10, 14 were labelled as "Luminal" (Figure 5.4 B and Figure 5.5). The remaining clusters were found to correspond to other populations of cells mistakenly isolated during cell sorting and carried over through to sequencing. Among these, 3 clusters (7, 13, 15) were found to contain haematopoietic cells, cluster 9 corresponded to endothelial cells, cluster 11 appeared to be fibroblasts, and cluster 12 had a gene expression profile resembling mesonephric derivatives such as vas deferens and/or seminal vesicles (Figure 5.5 A, B). Cells corresponding to the prostatic epithelial clusters were kept for further analyses.

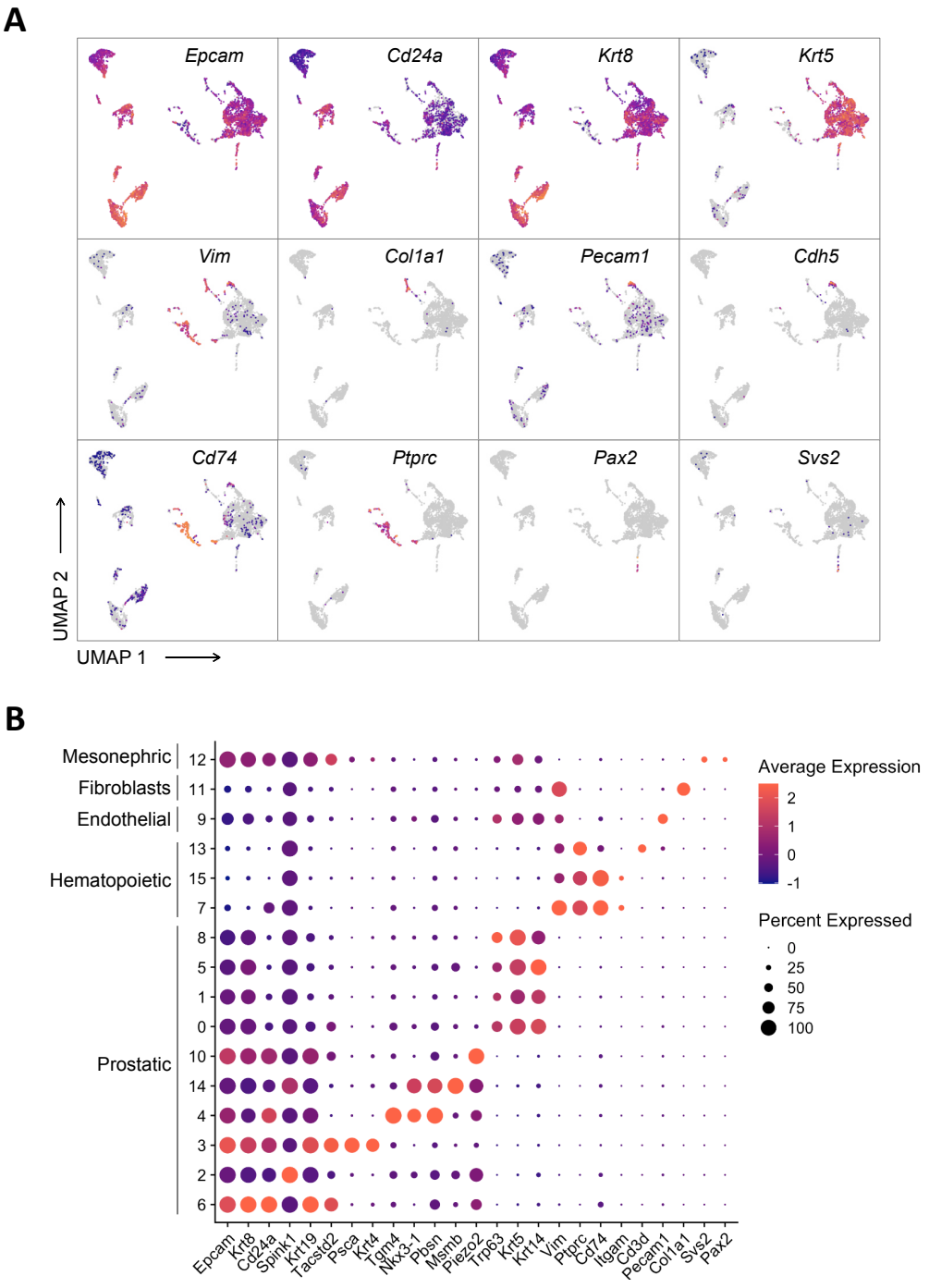

Figure 5.5. Examples of key marker genes used for cell type characterisation.
(A) UMAP visualisation of the integrated batch-corrected dataset. Cells are coloured by a gradient of log-normalised expression levels for each gene indicated.
(B) Dot plot showing the expression of selected marker genes associated with each cluster. Dot size corresponds to the proportion of cells having detectable gene expression levels in each cluster.

5.2.4. Graph-based clustering of the prostate epithelial subset

The adult mouse prostate epithelial subset was projected onto a UMAP layout (Figure 5.6). *Epcam* expression confirmed the presence of epithelial cells (Figure 5.6 A). Unsupervised graph-based clustering initially identified 9 clusters of cells, labelled 0 to 8 (Figure 5.6 B). One large population of basal cells was annotated by merging three tightly connected clusters (0, 1 and 5) broadly expressing *Krt5*, *Krt14*, and *Trp63* (Figure 5.6 B, C, E). Luminal populations expressed surprisingly heterogeneous levels of canonical luminal markers such as *Cd26*/*Dpp4*, *Cd24a*, *Krt8*, *Krt18* and *Krt19* (Figure 5.6 D). The different luminal clusters were annotated as Luminal-A (Lum-A), Lum-B, Lum-C, Lum-D, Lum-E and Lum-F (Figure 5.6 C).

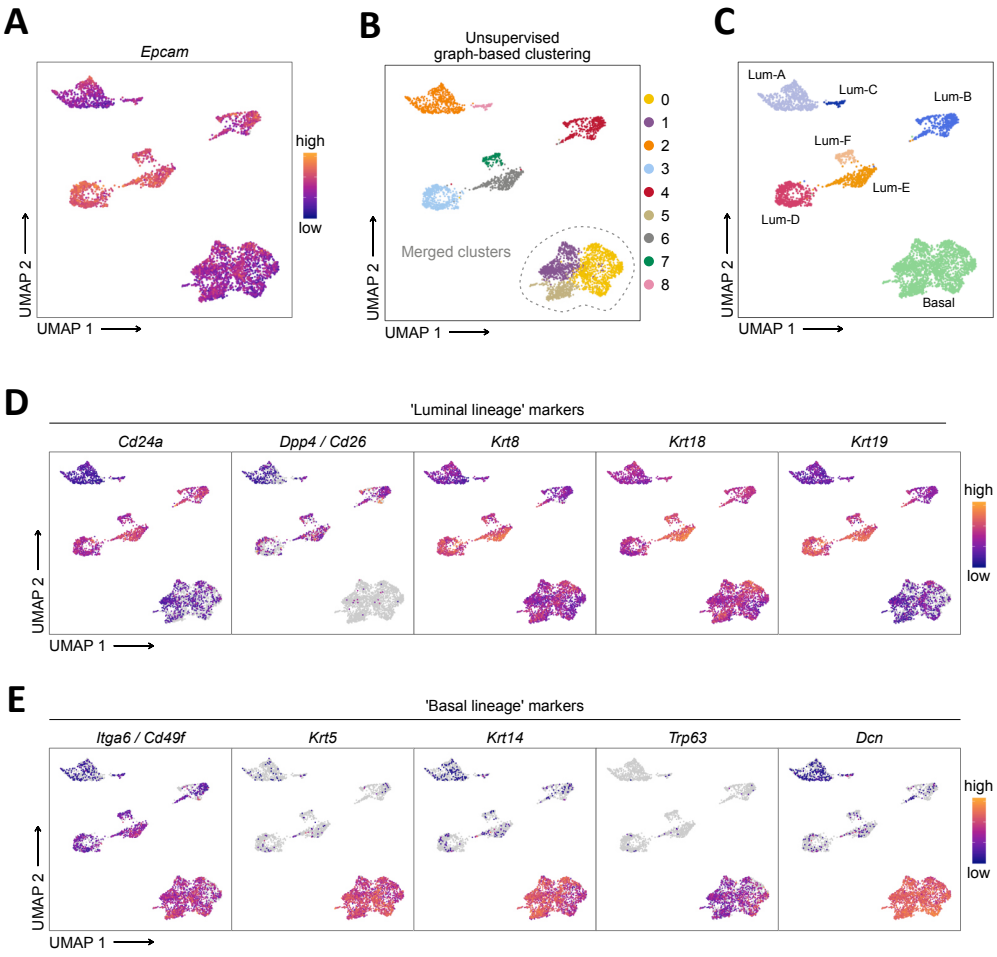

Figure 5.6. Characterisation of basal and luminal lineages in prostate epithelial cells. (A) UMAP visualisation of the prostate epithelial subset coloured by a gradient of log-normalised *Epcam* expression.

(B) Unsupervised graph-based clustering of the prostate epithelial subset coloured by clusters.
(C) UMAP visualisation of the prostate epithelial subset coloured by labelled clusters.
(D, E) UMAP visualisations of the prostate epithelial subset coloured by a gradient of log-normalised expression levels for canonical luminal (D) and basal (E) lineage markers.

To study the specific gene expression programs of those populations, differential expression analysis was performed by comparing each cluster to the remainder fraction. The top 10 most differentially upregulated genes per cluster are displayed as a heatmap in Figure 5.7 A to highlight cluster specific expression of certain genes. For example, Lum-A cells had very strong expression levels of *Sbp* (Spermine-binding protein), *Sbpl* (Spermine-binding protein-like) and *Spink1* (serine protease inhibitor Kazal-type 1). Lum-B expressed high levels of *Tgm4* (Transglutaminase 4), *Chn2* (Chimerin 2) and *C1rb* (Complement C1r-B subcomponent). Lum-C was characterised by higher expression of Msmb (Beta-microseminoprotein), *Defb50* (Beta-defensin 50 precursor) and *Itln1* (Intelectin 1). Lum-D was marked by the expression of *Psca* (Prostate Stem Cell Antigen), *Krt4* (Keratin 4) and *Pglyrp1* (Peptidoglycan Recognition Protein 1). Lum-E expressed elevated levels of *Clu* (Clusterin), *Basp1* (Brain Abundant Membrane Attached Signal Protein 1) and *Lpl* (lipoprotein lipase), and Lum-F had high expression of *Car2* (Carbonic Anhydrase 2), *Crym* (Crystallin Mu), *Ang5* (Angiogenin ribonuclease 5). Finally, basal cells strongly expressed *Krt5*, *Krt14* and *Trp63*. The cluster specific expression of some of these marker genes can also be appreciated in the form of a dot plot, which quantitatively describes the average expression of those markers as well as the proportion of cells expressing it (Figure 5.7 B) or overlaid onto the UMAP layout (Figure 5.7 C). Overall, cluster-based analysis revealed the existence of 6 distinct luminal populations, shared between intact and castrated prostates.

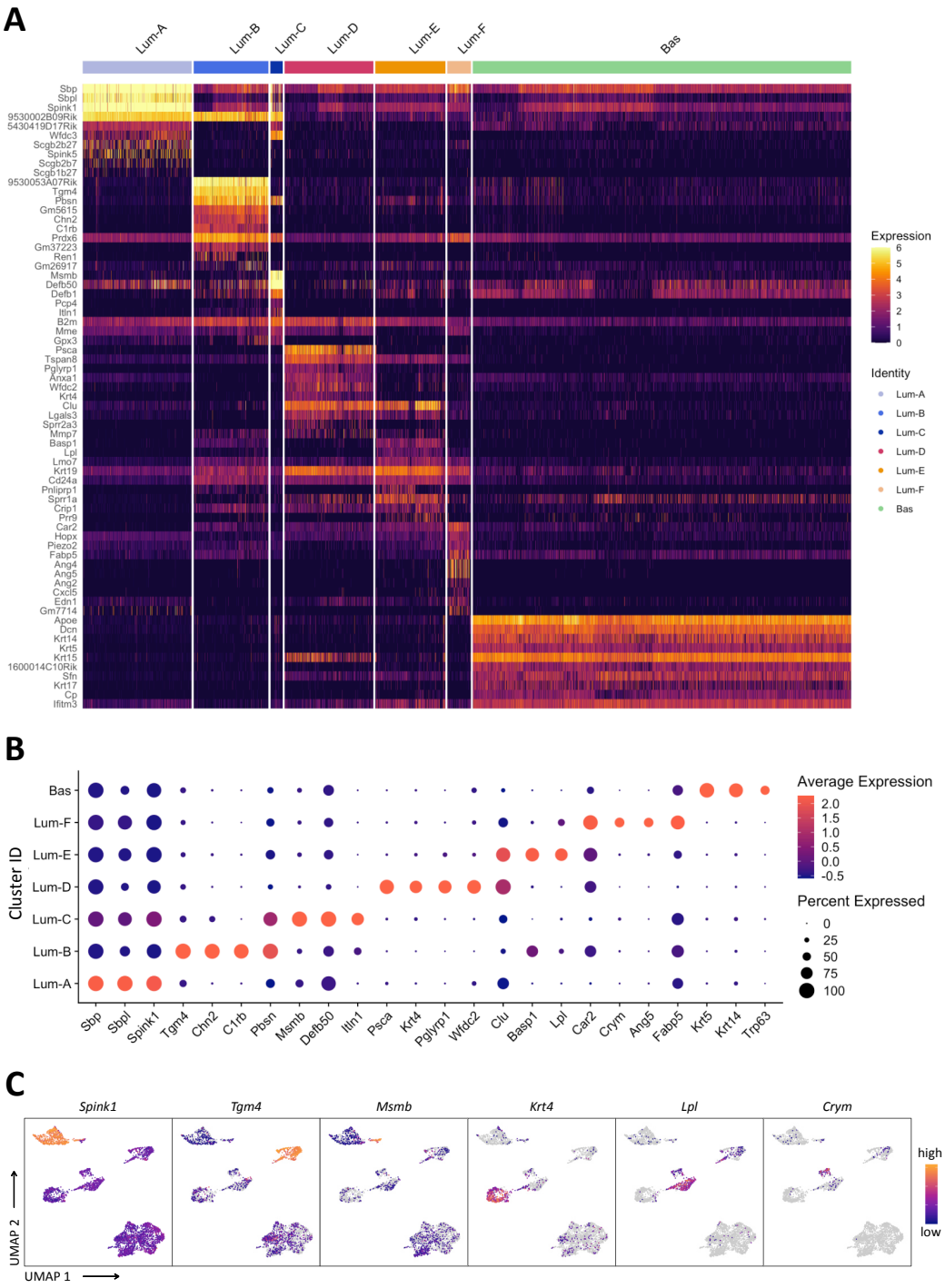

Figure 5.7. Characterisation of cluster-specific gene expression programs.
(A) Heatmap of the 10 most significantly differentially upregulated genes per cluster.
(B) Dot plot showing both the average expression and percentage of cells expressing the selected marker genes associated with each cluster.
(C) UMAP visualisation of the prostate epithelial subset coloured by a gradient of log-normalised expression levels for each gene indicated, showing specific expression patterns in distinct luminal clusters.

5.2.5. Composition of the prostate epithelial subset

The prostate epithelial subset comprised 3,825 prostate epithelial cells retrieved from the different sorted populations (Figure 5.8 B). While all sorted populations were present, castrated RFP- cells were largely under-represented due to their low frequency in the tissue. Visualisation of the dataset as a whole (Figure 5.8 A) or split by sorted population (Figure 5.8 C) gave a complex picture of the cellular heterogeneity residing in the mouse prostate epithelium, which translated into at least 7 distinct communities of cells.

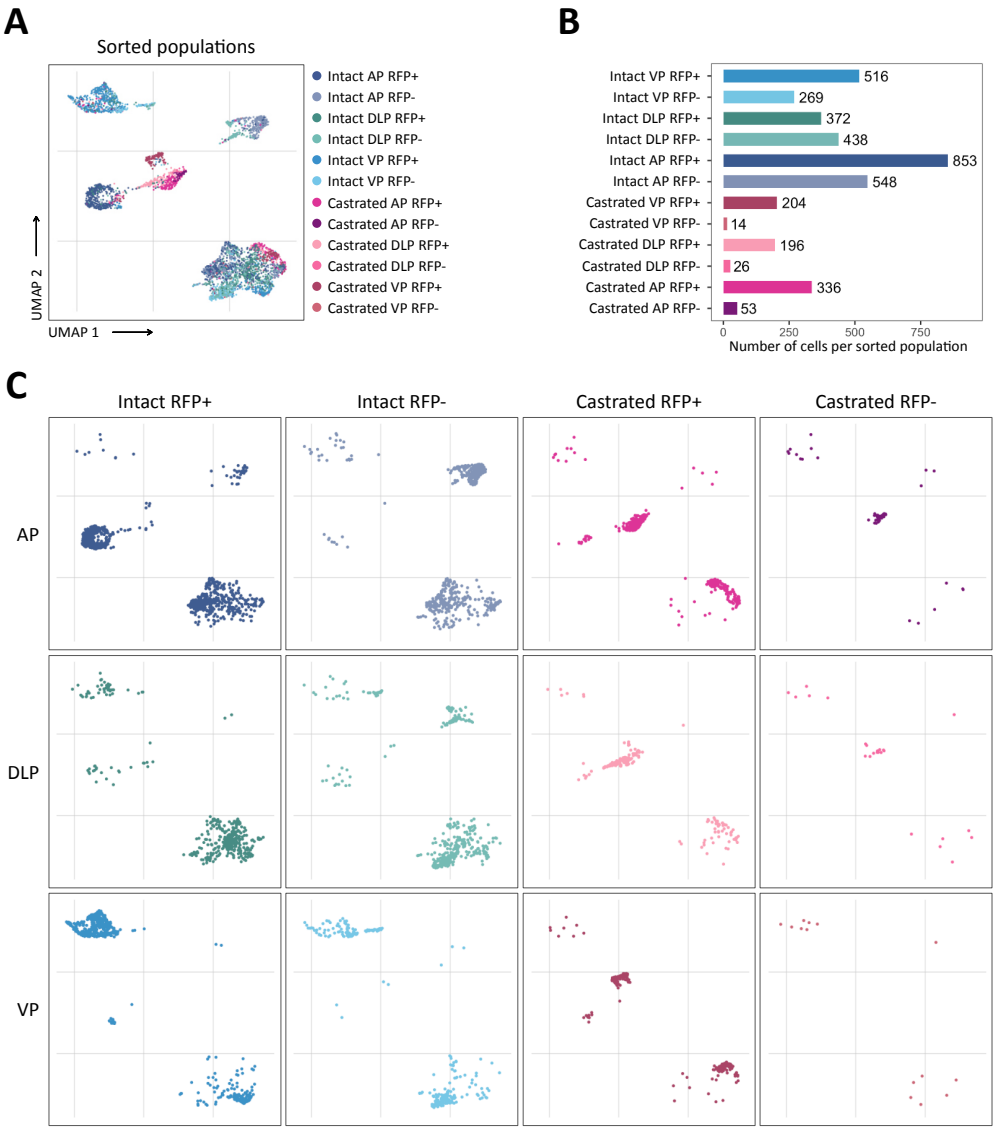

Figure 5.8. UMAP visualisations of the prostate epithelial subset by phenotypic sorted population.

(A) UMAP visualisation of the prostate epithelial subset coloured by phenotypic sorted populations. (n = 3,825 cells from 3 independent experiments).
(B) Number of cells retained by phenotypic sorted populations in the prostate epithelial subset.
(C) UMAP visualisation of the prostate epithelial subset as in A, split and coloured by individual phenotypic sorted populations.

Owing to the sample multiplexing approach and the complexity of interpreting such heterogeneous data, the composition of the dataset was visualised by individual phenotypic metadata, including the treatment of origin (intact, castrated, Figure 5.9 A-B), RFP FACS gate of origin (RFP+, RFP-, Figure 5.9 C-D) and lobe of origin (AP, DLP, VP, Figure 5.9 E-F). It is worth stating that the present dataset is not a faithful representation of the corresponding cellular frequency in the tissue, due to the upstream sorting strategy and sample multiplexing approaches. Notably, 2,996 cells were obtained from intact epithelia and 829 cells from castrated (Figure 5.9 B), corresponding to 1,790 AP cells, 1,032 DLP cells, and 1,003 VP cells (Figure 5.9 F). Due to the FACS enrichment strategy and high frequency in both basal cells and castrated epithelia, RFP+ cells were more abundant than the RFP- fraction (2,477 RFP+ cells compared to 1,348 RFP- cells, Figure 5.9 D).

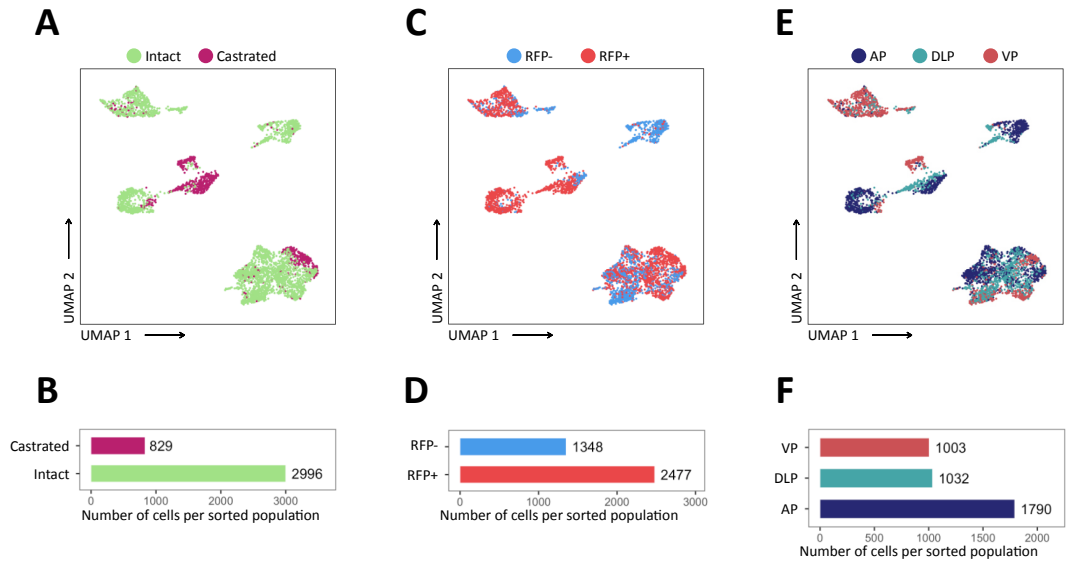

Figure 5.9. UMAP visualisations of the prostate epithelial subset by phenotypic sorted population.

(A, B) UMAP visualisation of the prostate epithelial subset coloured by treatment of origin (intact, castrated). Number of cells per group is shown in B.
(C, D) UMAP visualisation of the prostate epithelial subset coloured by RFP FACS gate of origin (RFP+, RFP-). Number of cells per group is shown in D.
(E, F) UMAP visualisation of the prostate epithelial subset coloured by lobes of origin (AP, DLP, VP). Number of cells per group is shown in F.

5.2.6. Effect of androgen withdrawal on lobe-specific cellular heterogeneity

This dataset offered the possibility to study the effect of androgen withdrawal on epithelial heterogeneity at the lobular level. As expected, basal cells of all three lobes appeared to be minimally affected by androgen-deprivation. Indeed, basal cells isolated from castrated prostates clustered together with basal cells of intact prostates (Figure 5.6 E and Figure 5.9 A). Among luminal populations, Lum-A/B/C/D were largely enriched in luminal cells originating from intact prostates, whereas Lum-E/F contained mainly castrated luminal cells (Figure 5.8 A and Figure 5.10 A). Of note, a small fraction of intact cells was also identified in the Lum-E/F clusters, while some castrated cells were found in each of the Lum-A/B/C/D clusters. In the predominantly intact clusters, Lum-A/C contained a large proportion of VP cells and some DLP cells, while Lum-B contained both AP and DLP cells (Figure 5.9 E and Figure 5.10 B). Lum-D had a majority of AP cells, as well as a small fraction of DLP and VP cells. In the castrated clusters, Lum-F was almost entirely made of VP cells, whereas Lum-E comprised both AP and DLP cells. These results suggest that a subset of intact Lum-A/C might undergo partial reprogramming during castration-induced regression and gives rise to the Lum-F cluster. Similarly, surviving Lum-B/D may predominantly reprogram into Lum-E cells upon castration. Alternatively, the small fraction of intact cells observed in Lum-E and Lum-F clusters might give rise to the expanded Lum-E/F clusters upon castration. Overall, these results highlight the dramatic changes occurring upon androgen deprivation in the representation of distinct luminal subpopulations. Also, lobe-specific identity of the luminal cells seems to be largely conserved after castration.

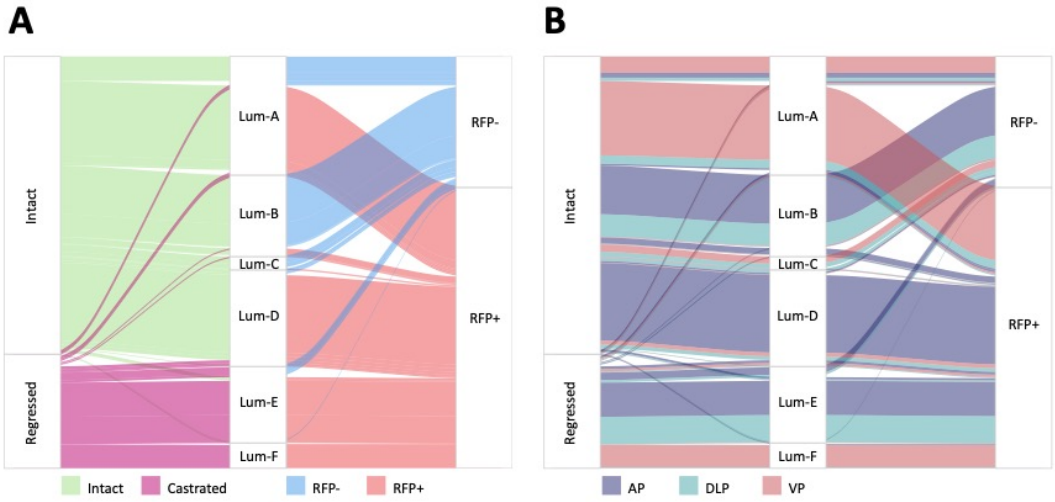

Figure 5.10. Characterisation of basal and luminal lineages in prostate epithelial cells. (A, B) Sankey diagrams summarising the proportion of cells between phenotypic groups in the prostate epithelial subsets, coloured by treatment of origin (intact, castrated) and RFP FACS gate of origin (RFP+, RFP-) in A, or lobes of origin (AP, DLP, VP) in B.

5.2.7. Expression profiles of stem/progenitor markers and transcriptional regulators

The scRNA-seq prostate epithelial dataset offered a unique opportunity to study the expression profile of selected genes of interest. First, a particular attention was given to genes previously associated with stem/progenitor properties, including *Cd44* (Collins et al., 2005), *Prom1*/*Cd133* (Richardson et al., 2004), *Tacstd2*/*Trop2* (Goldstein et al., 2008), and Ly6 family members such as *Sca-1* (*Ly6a*/*Ly6e*) (Kwon et al., 2016) and *Ly6d* (Barros-Silva et al., 2018). Overall, the expression of these markers was highly heterogeneous (Figure 5.11 A, B). *Tacstd2*/*Trop2* marked primarily Lum-D/E cells, as well as few Lum-A/F and basal cells. Both *Cd44* and *Ly6a* were found in about half of Lum-D/E clusters and a smaller fraction of basal cells. *Ly6e* expression was detected in all populations, with stronger levels in the Lum-E cluster. Conversely, *Ly6d* was only expressed in a small subset of Lum-E cells as well as in some basal cells. In contrast to the previous markers which generally marked Lum-D/E and basal subsets, *Prom1*/*Cd133* expression was largely restricted to the Lum-B population. Thus, these markers differentially label significantly heterogeneous cellular populations, suggesting that

neither of these used on their own would be a sufficient putative prostate stem cell marker.

Next, the expression profile of selected transcriptional regulators was explored throughout the different populations, including essential prostate specific transcription factors such as *Ar*, *Foxa1* and *Hoxb13*, as well as candidates thought to mark cells with castration-resistant and regenerative properties, including *Nkx3-1* (Wang et al., 2009b), *Sox2* (McAuley et al., 2019) and *Bmi1* (Yoo et al., 2016) (Figure 5.11 A, B). While *Ar*, *Foxa1* and *Hoxb13* were broadly expressed in the different clusters, *Nkx3-1* expression strongly marked cells of the Lum-B/C clusters, and a subset of Lum-A and basal cells. *Sox2* was mainly found in a small subset of Lum-D and basal cells, and *Bmi1* transcripts were weakly detected in a small number of cells of all populations. This suggests that *Ar*, *Foxa1*, *Hoxb13* and presumably *Bmi1* have a broader role in prostate homeostasis, compared to *Nkx3-1* and *Sox2* which may have a more restricted function in specific subsets.

Finally, the expression of the three RUNX transcription factors was evaluated together with their main binding partner *Cbfb* (Figure 5.11 A, B). Of all three RUNXs, *Runx1* was the most expressed in the prostate, present in the majority of the Lum-D/E/F clusters, and a large fraction of basal cells. Also, the distribution of *Runx1*-high expressing cells largely matched the distribution of RFP+ sorted cells (Figure 5.9 C).

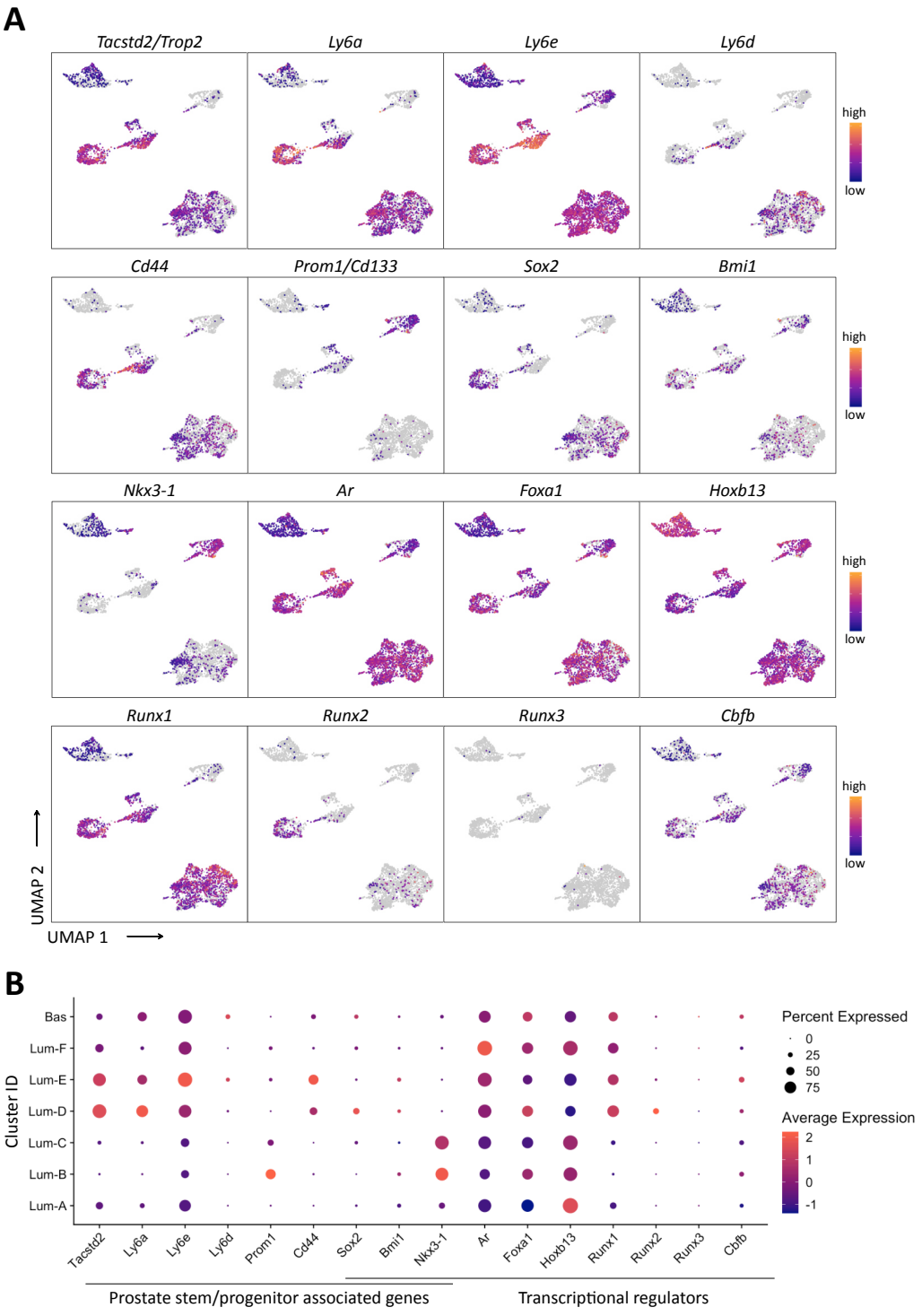

Figure 5.11. Gene expression profile of prostate stem/progenitor markers and transcriptional regulators.

(A) UMAP visualisation of the prostate epithelial subset coloured by a gradient of log-normalised expression levels for each gene indicated.

(B) Dot plot showing both the average expression and percentage of cells expressing each gene indicated.

Runx2 was most highly expressed in a subset of Lum-D and some basal cells, while *Runx3* was barely detected. According to its presumed ubiquitous expression profile, *Cbfb* was detected in all populations, albeit at relatively low frequency. These results suggest that *Runx1* is the dominant RUNX factor expressed in the prostate epithelium, and that its expression overlaps with *Runx2* in Lum-D and basal clusters.

5.2.8. The Lum-D cluster corresponds to Proximal Luminal Cells

Having described the main structure of the scRNA-seq prostate epithelial dataset, special attention was then given to RUNX1+ luminal cells. As previously mentioned, the Lum-D cluster predominantly consisted of AP derived RFP+ cells, as well as a small number of RFP+ DLP and VP cells (Figures 5.7, 5.8 and 5.11). Also, high *Runx1* expression in Lum-D correlated with higher levels of *Tacstd2*/*Trop2* and *Sca-1* (Figure 5.11), previously shown to be co-expressed with RUNX1 in PLCs (Figure 3.6). In contrast, *Runx1* was barely detected in Lum-B/C clusters which expressed high levels of *Nkx3-1,* while Lum-A cells expressed low levels of both *Runx1* and *Nkx3-1*. These observations suggested that RUNX1+ PLCs described in Chapter 3, enriched in the proximal area, and presenting a unique cell morphology, may correspond to the Lum-D cluster. To test this hypothesis, the expression of Keratin 4 (K4), one of the most specific marker gene expressed in the Lum-D population (Figures 5.9), was evaluated by immunostainings in combination with RUNX1. K4 staining largely overlapped with RUNX1 in PLCs (Figure 5.12). Taken together, these results indicate that the Lum-D cluster corresponds to the distinct RUNX1+ luminal subset identified in the proximal region of all three prostate lobes.

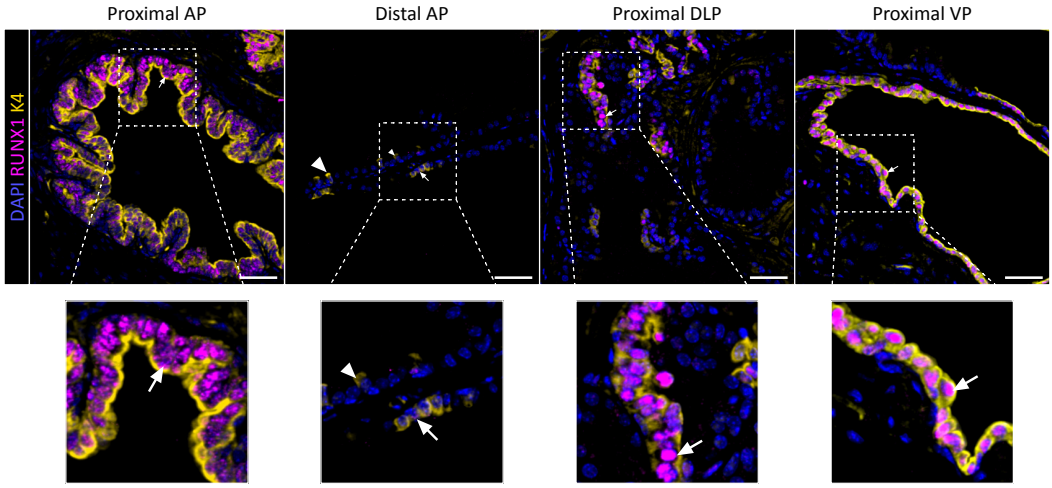

Figure 5.12. RUNX1 and K4 mark proximal luminal cells.
Co-immunostaining of RUNX1 and K4 in the proximal AP, distal AP, proximal DLP and proximal VP. Higher magnification images of selected regions are shown. Arrows indicate RUNX1+ K4+ cells, arrowheads indicate rare RUNX1- K4+ cells in the distal AP.

5.2.9. The Lum-D cluster is transcriptionally similar to castration resistant cells

To gain further insights into the nature of the Lum-D population in comparison to the other populations, gene ontology (GO) analysis was initially performed using the lists of genes significantly upregulated in each cluster (Figure 5.10 and supplementary data 6). The Lum-A cluster was enriched in GO terms referring to enzymatic activity such as “negative regulation of hydrolase activity” or “regulation of endopeptidase activity” (Figure 5.13 A). Lum-B and Lum-C clusters were enriched in GO terms related to protein synthesis function, including “retrograde protein transport, ER to cytosol” or “Endoplasmic-reticulum-associated protein degradation (ERAD) pathway” (Figure 5.13 B, C). Lum-D, Lum-E and Basal clusters had a drastically different GO enrichment profile, with terms such as “epithelium development”, “tissue development” and “developmental process” among the most enriched ones (Figure 5.13 D, E, G). Of note, Lum-F had too few specifically upregulated genes to give meaningful results in GO analysis (Figure 5.13 F). Thus, these results suggest that Lum-A/B/C clusters are highly distinct from Lum-D/E/F and Basal clusters. Importantly, the GO term enrichment of Lum-A/B/C

clusters suggests that these act as the secretory epithelial cells of the prostate, unlike cells of the Lum-D/E/F and Basal clusters.

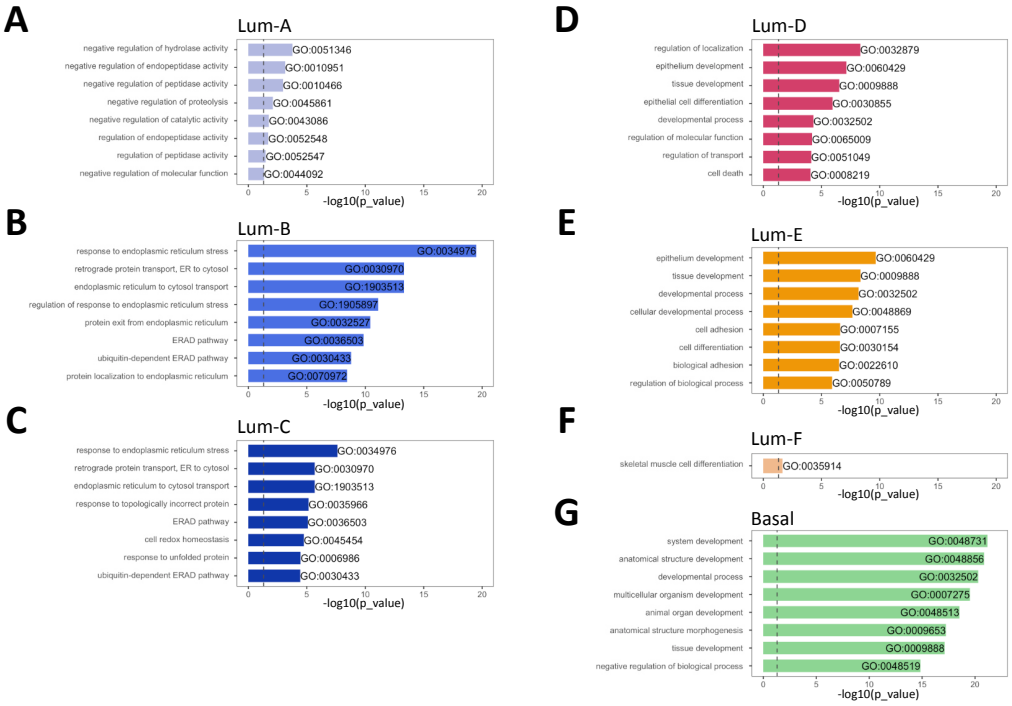

Figure 5.13. Gene Ontology and differential expression analysis within the scRNA-seq prostate epithelial dataset.
(A-G) Bar plots of the 8 most significantly (g:Profiler adjusted p-value < 0.05) enriched gene ontology (GO) terms (GO:BP, Biological Processes) using the list of differentially upregulated genes specific to (A) Lum-A, (B) Lum-B, (C) Lum-C, (D) Lum-D, (E) Lum-E, (F) Lum-F, and (G) Basal clusters.

To complement the characterisation of the Lum-D cluster mainly containing intact prostate cells, differential expression analysis was performed specifically between the Lum-D cluster and the remaining intact luminal fraction composed of Lum-A/B/C clusters (Figure 5.14 A). In line with the previous results, GO analysis showed that Lum-A/B/C were enriched in enzymatic activity and protein synthesis functions. In contrast, the Lum-D cluster was enriched in terms related to epithelial developmental processes, similar to the predominantly castrated Lum-E population. The transcriptional relationship between the distinct epithelial clusters was then evaluated using partition-based graph abstraction (PAGA) (Figure 5.14 B). In this representation, clusters are represented as nodes, and the weighted edges between the nodes represent a statistical measure of connectivity (Wolf et al.,

2019). Basal cells were expectedly the least connected to the other luminal subsets. Lum-A and Lum-C were tightly connected, indicating strong resemblance between the two populations. Similar observation was made for the mainly castrated Lum-E and Lum-F clusters. Interestingly, PAGA uncovered a strong degree of connectivity between the mainly intact Lum-D and castrated Lum-E population. This observation confirms the tight transcriptional relationship between Lum-D and Lum-E populations highlighted in GO analysis.

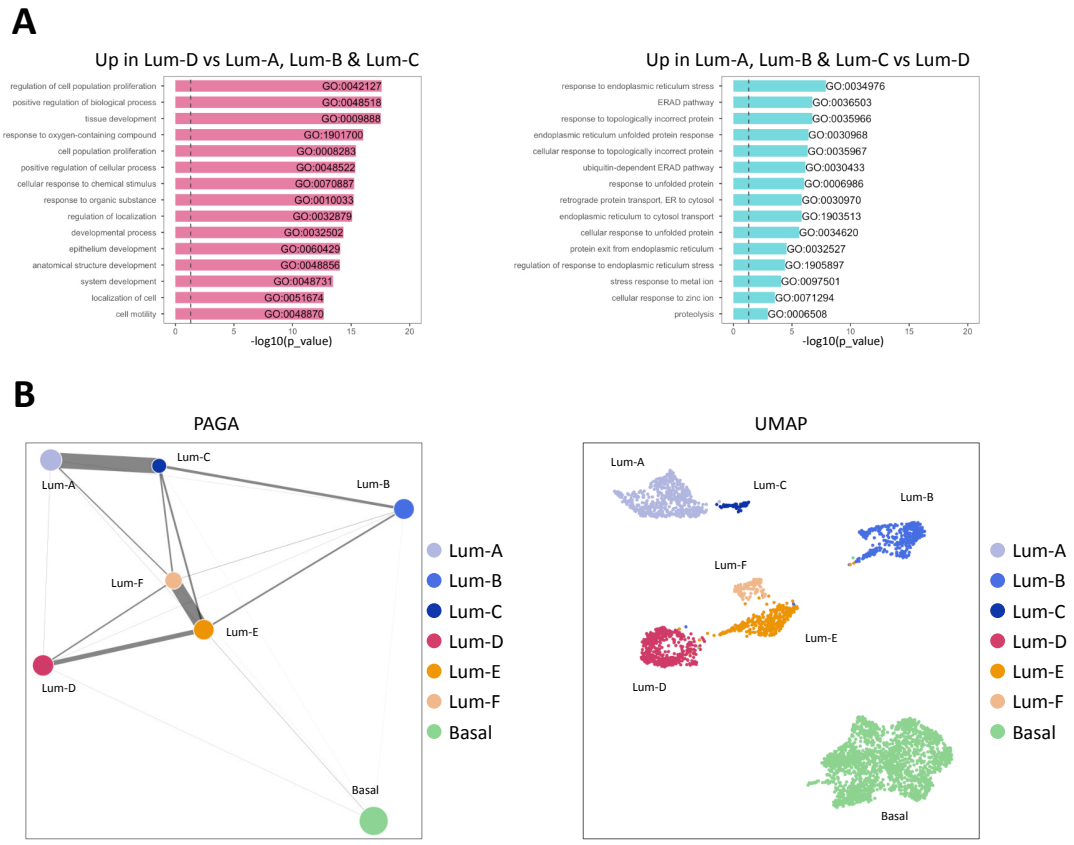

Figure 5.14. The Lum-D cluster is a non-secretory population transcriptionally similar to castration resistant cells.
(A) 15 most enriched GO terms (g:Profiler adjusted p-value < 0.05) for genes upregulated in the Lum-D cluster against combined Lum-A, Lum-B and Lum-C clusters (left), and the opposite comparison (right).
(B) Graph-abstracted representation (PAGA, left) and corresponding UMAP visualisation (right) of the prostate epithelial subset. Colours represent different clusters. In PAGA, clusters are linked by weighted edges that represent a statistical measure of connectivity.

Finally, the Lum-D cluster contained a small, but defined, subpopulation of castrated epithelial cells, suggesting the preservation of its identity upon androgen

deprivation (Figure 5.15 A). In this population, very few genes were found significantly differentially expressed between intact and castrated cells (n=103), of which only 18 varied by more than |1| log fold-change (Figure 5.15 B). AR-regulated genes such as *Psca* and *Tspan1* were downregulated in the castrated subset, while strong contributors of the Lum-D identity such as *Tacstd2*/*Trop2*, *Krt4* and *Runx1* maintained high expression levels (Figure 5.15 C). These results support that Lum-D/RUNX1+ PLCs maintain their identity following androgen-deprivation.

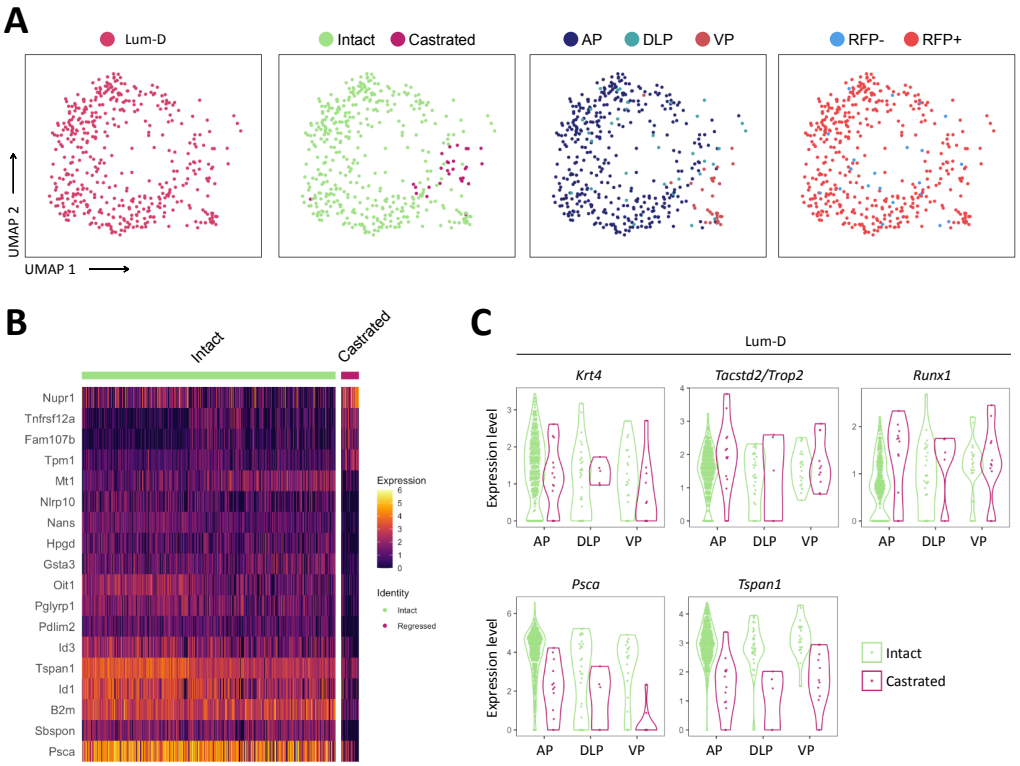

Figure 5.15. The Lum-D cluster preserves its identity upon androgen withdrawal.
(A) UMAP visualisation of the Lum-D cluster coloured by phenotypic groups, including cluster (Lum-D), treatment (intact, castrated), lobe (AP, DLP, VP), and RFP FACS gate (RFP+, RFP-).
(B) Heatmap of the most significantly differentially expressed genes between intact and castrated cells of the Lum-D cluster (p value < 0.05, average log-fold change > |1|).
(C) Violin plots showing the expression levels of specific genes within the Lum-D cluster. For each lobe, data are split and coloured by treatment (green: intact, pink: castrated).

5.2.10. Lineage tracing of *Runx1* expressing cells confirms that Lum-D cells are intrinsically castration resistant

The tight transcriptional relationship observed between the high *Runx1* expressing clusters Lum-D and Lum-E supported the notion that the Lum-D population, which corresponds to RUNX1+ PLCs, contains intrinsically castration-resistant luminal cells. To further validate this result, the expression of K4, which specifically marked Lum-D cells, was evaluated in the lineage tracing experiments combined with castration-regeneration assays previously described in Chapter 4. As expected, the majority of RFP labelled cells were K4+ and restricted to the proximal regions of intact, castrated, and regenerated prostates (Figure 5.16). This result confirmed that RUNX1+ PLCs maintain their identity in castration-regeneration assays.

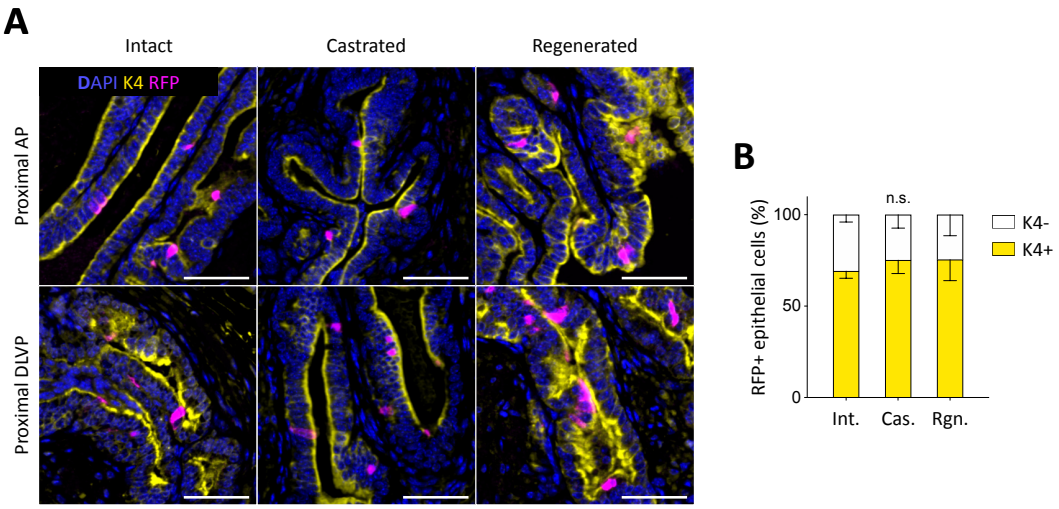

Figure 5.16. Characterisation of K4 expression in lineage tracing experiments.
(A) Co-immunostaining of RFP and K4 in the proximal AP and DLVP (DLP + VP) of intact, castrated, and regenerated prostates. Scale bar: 50 µm.
(B) Quantification of the percentage of RFP+ cells being K4+ and K4- in intact (n = 5), castrated (n = 5) and regenerated (n = 4) mice. Int: Intact, Cas: Castrated, Rgn: Regenerated.

5.3. Discussion

5.3.1. Summary

The prostate epithelial hierarchy is commonly described based on anatomical features of the basal and luminal layers, their histological characteristics and the expression of a small subset of markers (Shen and Abate-Shen, 2010; Toivanen and Shen, 2017). But beyond the classical basal and luminal lineages, the existence of distinct subpopulations of cells residing within the prostate epithelium has been suggested by multiple reports revealing the existence of population of cells with specific immunoprofiles and functional characteristics. However, until the advent in single-cell technologies, it remained particularly challenging to draw a complete picture of prostate epithelial heterogeneity. The aim of this Chapter was to take advantage of the recent advances in scRNA-seq to comprehensively characterise RUNX1+ and RUNX1- epithelial compartments at the single cell level, in intact and castrated lobes of the mouse prostate. Overall, the single-cell transcriptomic analysis uncovered a vast degree of heterogeneity within and between the luminal compartments of both intact and castrated mouse prostates (Figure 5.17). Among the different cell populations identified, the Lum-D cluster was found to correspond to the previously characterised RUNX1+ PLCs. The transcriptomic profile of this population confirmed the non-secretory phenotype of Lum-D/RUNX1+ PLCs. In line with previous lineage tracing experiments demonstrating the intrinsic castration resistant properties of these cells (Chapter 4), the Lum-D population strongly resembled castration-resistant cells.

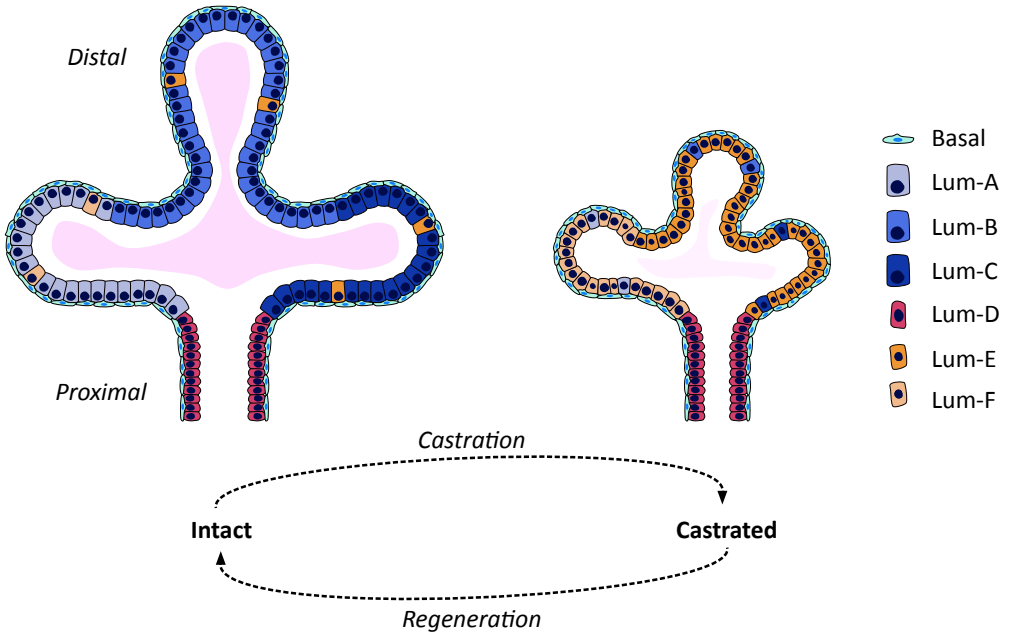

Figure 5.17. Working model of the cellular landscape of the prostate epithelium in intact and castrated conditions.
The proximal luminal lineage (Lum-D) is enriched in intrinsically castration resistant cells and remains barely affected by fluctuations of AR signalling. Distal lineages (Lum-A, Lum-B, Lum-C) undergo more dramatic changes in cellular representation, with the castrated distal prostate being predominantly composed of Lum-E/F clusters.

5.3.2. The luminal compartment of the mouse prostate contains distinct subpopulations of cells

The identification of multiple clusters of luminal cells in both intact and castrated mouse prostates highlights the cellular heterogeneity of the luminal compartment. Several scRNA-seq studies recently reported similar degree of heterogeneity (Crowley et al., 2020; Guo et al., 2020; Joseph et al., 2020; Karthaus et al., 2020). While these different studies are largely concordant, some discrepancies exist between the cell types identified. Table 5.2 summarises the labels given in each of these studies to the different cell clusters identified in intact prostates.

All studies identified a single basal population, indicating that basal cells represent a transcriptomically homogenous compartment compared to luminal cells. Karthaus et al profiled the AP and identified 3 populations of luminal cells, with the “L1” cluster resembling our “Lum-B” population and “L2” cluster which

appears to be the equivalent of our "Lum-D" population. However, no equivalent to the "L3" *Foxi1*-expressing cluster could be found in our dataset, which was instead labelled as "ductus/vas deferens" in another study (Crowley et al., 2020). Instead, the work presented here together with other studies (Crowley et al., 2020; Guo et al., 2020; Joseph et al., 2020) identified at least 2 additional major luminal subtypes were identified predominantly from DLP and VP lobes, namely "Lum-A" and "Lum-C". Joseph et al recently published a large dataset comprising > 30,000 single cells isolated from the whole dissociated intact prostate and urethra (Joseph et al., 2020). Although they did not profile individual lobes separately, the equivalent cell types of our Lum-A/Lum-B/Lum-C/Lum-D clusters were respectively labelled VP, DLP, AP and urethral luminal. This labelling assumed that the different clusters were entirely lobe specific, which does not seem entirely true from the studies which profiled individual lobes (this work; Crowley et al., 2020; Guo et al., 2020). Indeed, Crowley et al kept a greater level of spatial information associated with their dataset by micro-dissecting the proximal region of all prostate lobes, including the AP, DP, LP and VP. Guo et al profiled cells isolated from *Tmprss2-Cre*ERT2/+ *Rosa26*-EYFP/+ mouse models and found equivalent cell types to our Lum-A/B/C/D cells (Guo et al., 2020). Of note, Guo et al identified a restricted subset of "Luminal-C" ("Lum-D" in this work) located in the distal region, termed "Dist-Luminal-C" as opposed to "Prox-Luminal-C". Similarly, immunostainings for K4 carried out as part of this work also revealed sporadic K4+ cells (data not shown) and TROP2+ cells (Chapter 3) in distal areas, albeit with low or absent RUNX1 expression. However, these cells strongly resembled the proximal lineage, indicating that these may simply be an extension of the proximal luminal cells, perhaps carried away from the proximal region during prostate organogenesis, rather than a distinct lineage from proximal cells. Interestingly, while Joseph et al propose that their equivalent "Lum-D" population would be part of the urethral epithelium, the work presented here together with other studies (Crowley et al., 2020; Guo et al., 2020; Karthaus et al., 2020), indicates that this cell type is instead a prostate lineage, found in the peri-urethral region of the mouse prostate. Overall, while different experimental approaches were used, all these independently generated scRNA-seq datasets

reported the existence of a distinct luminal lineage enriched in the proximal or peri-urethral region of the mouse prostate, strongly confirming the distinct lineage relationship of the RUNX1+ PLCs/Lum-D cluster.

Table 5.2. Equivalence table between the different epithelial clusters identified in mouse prostate scRNA-seq studies.

This work	Karthaus et al., 2020	Joseph et al., 2020	Crowley et al., 2020	Guo et al., 2020
Lum-A	x	VP Luminal	LumV	Luminal-A
Lum-B	L1	AP Luminal	LumA + LumD	Luminal-B
Lum-C	x	DLP Luminal	LumL	x
Lum-D	L2	Urethral	LumP + PrU	Luminal-C
x	L3	x	ductus/vas deferens	x
Basal	Basal	Basal	Basal	Basal

5.3.3. Implications for the identification of rare prostate stem/progenitor populations

Unresolved questions in the field relate to the identification of prostate stem/progenitor cells, and if such population exist, what cellular and molecular mechanisms are involved in tissue maintenance. While several genes have been proposed to mark specific subset of cells with stem/progenitor properties and/or castration-resistant properties, the precise characterisation of these populations remains controversial (Toivanen and Shen, 2017; Zhang et al., 2018). The cellular landscape of the intact and castrated prostate epithelium depicted by scRNA-seq identifies 1 basal population and at least 6 distinct lineage-committed populations. Examination of the expression profile of previously identified stem/progenitor markers revealed significantly heterogeneous expression in several populations (section 5.2.7). In fact, the expression of these markers was often shared between basal and luminal cells or between distinct luminal subsets and varied in intensity between intact and castrated derived cells. Of note, several studies have proposed that basal and luminal lineages may be maintained by a pool of “stem cells” co-expressing canonical basal and luminal markers such as K5 and K8 (De Marzo et al., 1998; Hudson et al., 2001; Peehl et al., 1996; Taylor et al., 2010; Uzgare et al., 2004; Wang et al., 2001). To date, all published scRNA-seq studies tend to refute the existence of such rare population. However, it is evident that *Krt8* mRNA is

expressed throughout the basal compartment, albeit at lower levels than in luminal cells. Collectively, these observations tend to argue against a model whereby prostate lineages would be maintained by the presence of a common and rare population of “stem cells”. Instead, the lineage tracing results presented in Chapter 4, as well as the scRNA-seq dataset presented in this work and other studies (Crowley et al., 2020; Guo et al., 2020; Joseph et al., 2020; Karthaus et al., 2020), tend to favour the existence of self-sustained basal and luminal populations. Although current scRNA-seq studies failed to identify such population, this model does not entirely preclude the existence of smaller subpopulations within each of the main luminal populations. If tissue maintenance is not governed by a pool of stem cells, it is possible that individual lineage-committed cells retain the potential to contribute to epithelial maintenance to sustain their own lineage. In such model, it would be of interest to investigate what governs different cell fates, and whether it occurs in a stochastic or predefined manner. For example, since castration results in the loss of a large fraction of lineage committed luminal cells, this model would suggest that a fraction of each luminal lineage may have the capacity to acquire different fates, mainly consisting in cell death or castration resistance. Alternatively, since discrepancies in cell potential seem to exist between organoids and castration-regeneration assays (Chapter 4), different subpopulations of lineage-committed cells may also possess intrinsically distinct fate potentials. Overall, it would be important to investigate the plasticity of each distinct luminal lineage-committed subsets in the context of prostate development and tissue homeostasis. These models would also enable further functional characterisation of these populations, as well as their lineage relationship during prostate development. To this end, scRNA-seq provides a useful resource to identify promoters specific to each subpopulation to generate novel mouse models.

5.3.4. Comparison between mouse and human prostate epithelial landscapes

Mouse and human prostates exhibit important anatomical differences, and the lineage relationships between mouse and human prostate epithelia remains

poorly understood. In an attempt to compare cell types between species, some of the recent studies reporting scRNA-seq datasets of the mouse prostate also profiled human prostate samples by scRNA-seq. Using gene signatures derived from the mouse prostate samples, Karthaus et al identified “L1” and “L2” enriched subsets in the luminal compartment of human prostates (Karthaus et al., 2020). However, these cells appeared to form a continuum rather than discrete clusters. In another study, Joseph et al reported a large fraction of basal cells as well as the existence of “club” and “hillock” cells (Joseph et al., 2020), similar to a previous study from the same laboratory (Henry et al., 2018). While they did not directly compare all clusters identified between mouse and human samples, they propose that club and hillock cells are enriched in the peri-urethral region of the human prostate and may be the equivalent to what they defined “urethral luminal cells” in the mouse. Next, Crowley et al reported the presence of “ductal” and “acinar” luminal populations in human epithelia, as well as a peri-urethral cells (Crowley et al., 2020). They propose that these peri-urethral (“PrU”) cells are conserved between species, that “ductal” cells are similar to the mouse “LumP” population, and “acinar” closely related to the “LumL” and “LumV” clusters of the mouse prostate. Finally, Guo et al found a human cell type resembling their mouse “Luminal-C” cluster, and another broad population similar to “Luminal-A/B” of the mouse prostate (Guo et al., 2020). Overall, these studies seem to agree on the notion that an equivalent cell type to the mouse proximal luminal lineage exists in the human prostate. However, these studies also exhibit important differences in their interpretation and cellular contents which warrants further investigations. These discrepancies may notably be due to intrinsic differences between patient samples. Indeed, while they aimed at profiling “normal” human prostate tissue, these samples are usually obtained from aged patients who require surgical intervention due to the presence of BPH, bladder or prostate cancers.

5.3.5. Technical limitations associated with single cell isolation

Despite the many benefits of using scRNA-seq to study population heterogeneity, technical challenges were encountered throughout the experiments

presented in this Chapter. First, prostate lobes (AP, DLP, VP) were kept separate during prostate dissection in order to retain broad spatial information regarding the localisation of the cells sequenced. However, the boundaries between the different lobes are challenging to identify, in particular between the DLP and VP. Despite every effort to accurately split these lobes apart, there may have been a small proportion of mixed-up cells. Next, the quality of the cells retrieved for analysis was entirely dependent on the ability to isolate them from the tissue. Yet, the reference protocol used for mouse prostate dissociation (Drost et al., 2016) is accompanied with significant cell death, limiting the number of cells that can be successfully isolated from each mouse prostate. Therefore, if specific cell types were more sensitive to dissociation and downstream processing steps than others, it may have impacted the abundance of these cell populations in the final dataset. Also, while the dissociation protocol was specifically developed for single cell dissociation of mouse prostate epithelial cells, a large number of small cell clumps often remained in the sample prior cell sorting. In this regard, cell sorting was critical to exclude low quality cells, including dead or dying cells, debris, clumps of cells, and non-epithelial cells. *In fine*, cell sorting was beneficial to enrich for high quality live single cells, but the presence of some impurities was inevitable in the final cell preparation used for droplet encapsulation, including dead cells and small clumps. These should represent only a small fraction of the total cells isolated, but it is not possible to entirely exclude the presence of a small proportion of “cells” which are in fact small clumps of more than one cell, known as doublets or multiplets. In particular, if such events involved cells of distinct lineages, this may have caused biases in dimension reduction and cell clustering. Additionally, despite including EPCAM+ gates during cell sorting to isolate epithelial cells, a significant fraction of the cells appeared to be of non-epithelial nature (section 5.2.3). This indicates that sort impurity led to contaminating cell types being carried over through to cell encapsulation, which reduced the total number of epithelial cells of interest kept in the final prostate epithelial dataset.

5.3.6. Technical limitations associated with MULTI-seq sample multiplexing

Downstream of cell isolation, MULTI-seq was used to minimize technical confounders, including doublets generated during cell encapsulation, as well as the presence of batch effects (McGinnis et al., 2019b). The use of MULTI-seq was also cost-effective, since multiplexing reduced the number of individual samples (lanes) to load on the 10x Genomics Chromium platform. However, in the first 2 experimental runs, the MULTI-seq labelling completely failed for 1 population (Intact DLP RFP+). The unlabelled population was finally obtained in a third experimental run by changing the lipid-modified oligo used for labelling. Yet, the failed labelling of a single sample has resulted in a large number of captured cells being discarded from the analysis to preserve only those cells that could confidently be associated back to their population of origin. Further to this, the signal-to-noise ratio of MULTI-seq read counts observed between unlabelled and labelled fraction did not present the expected bimodal distribution for the first 2 experimental runs (Figure 5.3 A). This may be particularly important when studying the effect of castration on luminal cell heterogeneity, especially since a small fraction of intact cells were observed in predominantly castrated clusters, and vice versa. To minimise risks of misclassification, only cells with high read counts were kept for analysis. Manual examination of the MULTI-seq read counts in these rare cells did not reveal suspicious misclassification. However, one cannot rule out the possibility that a small fraction of cells have been assigned to the incorrect population. Overall, these technical challenges should be considered when drawing conclusions or generating biological hypotheses from this dataset. These limitations highlight the importance of experimental validation downstream *in silico* analyses.

5.3.7. Technical limitations of scRNA-seq

Several conceptual and technical aspects of scRNA-seq currently restrict the extent of biological conclusions that can be drawn. Here, the experimental design offered the possibility to study the effect of castration on the cellular content of the

mouse prostate at the single cell resolution. However, the destructive nature of scRNA-seq is one of its main limitation, giving only a "static snapshot" of cell state. Therefore, in the absence of information about cells history, it is not possible to directly conclude whether intact cells survive in the castrated prostate, or whether castrated cells pre-existed in the intact prostate and expanded upon castration. For example, the emergence of the Lum-E and Lum-F clusters could be due to the expansion of a small number of intact cells already present at the intact state or result from broad transcriptional reprogramming taking place in intact cells of Lum-A/B/C clusters which survived the process. In a recent study, Karthaus and colleagues profiled AP derived cells at sequential time points after castration (day 1, 7, 14 and 28) and regeneration (day 1, 2, 3, 7, 14, 28), and used the Phate algorithm (Moon et al., 2019) for dimension reduction to estimate cell state transitions (Karthaus et al., 2020). While the collection of sequential time points demonstrates the existence of gradual transcriptional changes during these processes, as well as possible transitional states, the destructive nature of the experiment prevents from drawing definitive conclusions about cell trajectories beyond computational inference. Plethora of computational methods have been developed to infer single-cell trajectories and transition states, including pseudotime inference (Saelens et al., 2019). However, these often assume the existence of continuous processes, and poorly model discontinuous molecular oscillations and discontinuous cell fate transitions (Leng et al., 2015; Loeffler and Schroeder, 2019). In fact, only time-lapse imaging coupled with scRNA-seq could currently overcome this obstacle. Unfortunately, it remains technically impossible to implement such *in vivo* tracking system for individual prostate cells. An alternative way of partially overcoming the challenges posed by the destructive nature of scRNA-seq would be to combine it with lineage tracing models specific to individual subpopulations, including Lum-A/B/C/D/E/F. Along these lines, barcoding systems enabling clonal labelling have recently emerged as a promising approach to combine scRNA-seq with lineage tracing, both *in vitro* (Weinreb et al., 2020) and *in vivo* (Bowling et al., 2020). These emerging technologies may prove useful to improve our understanding of the cellular mechanisms involved in castration resistance and regeneration.

Another important limitation of scRNA-seq is the need for tissue dissociation to obtain single cell suspensions, leading to the complete loss of spatial context. In light of the specific spatial expression pattern of *Runx1* in the mouse prostate, the different mouse prostate lobes where kept separate for the experiments carried out in this work. However, beyond the lobe of origin, it remains impossible to accurately infer the precise localisation of the cells sequenced. Since RUNX1+ cells are particularly enriched in the proximal region of the prostate, it would have been interesting to keep separate the proximal and distal regions of each lobe, similar to what was done in a recent study (Crowley et al., 2020). However, this would have complicated experimentations by doubling the number of individual samples to process. The high *Runx1* expressing Lum-D cluster was found to contain cells of all prostate lobes, and express markers specifically expressed in proximal luminal cells such as *Krt4*, *Tacstd2/Trop2*, *Ly6a/Sca-1*. After confirmatory immunostainings, Lum-D was proved to correspond to RUNX1+ PLCs. Nevertheless, the degree of spatial information remains very shallow. To overcome this issue and preserve cells in their spatial context, methods of spatial transcriptomics are starting to emerge and hold great promises to refine our understanding of tissue organisation. However, at the time of writing, these techniques are not yet capable to capture full transcriptomes at the single-cell resolution (Berglund et al., 2018; Burgess, 2019; Ståhl et al., 2016).

Further to these conceptual obstacles, scRNA-seq is limited in the number of single cells that can be sequenced in a single experiment at reasonable costs using existing technologies, including droplet-based methods (Klein et al., 2015; Macosko et al., 2015) or plate-based methods (Hashimshony et al., 2016; Picelli et al., 2013). Consequently, it is possible that the dataset presented here, comprising 3,825 prostate epithelial cells, does not fully recapitulate the heterogeneity of the mouse prostate epithelium made of several millions of cells. The recent publications of mouse prostate scRNA-seq datasets (Crowley et al., 2020; Guo et al., 2020; Joseph et al., 2020; Karthaus et al., 2020) provides an opportunity to increase the size of the dataset by integrating these in a single dataset. This may provide a more complete picture of the cellular landscape of the mouse prostate epithelium. Another aspect to consider is the number of genes expressed that can be detected

per cell. Even though tremendous progresses have been made in the recent years, the method used to generate this dataset is thought to capture only about a third of the actual transcriptome (Zheng et al., 2017), indicating that most of the genes expressed remain undetected. In line with this, the high "dropout" rates are another well characterised challenged posed by scRNA-seq data analysis (Hicks et al., 2018; Vallejos et al., 2017). This occurs when a gene is expressed in one cell but not detected in another cell of the same type, resulting in sparse data and high levels of technical noise. It is often attributed to the stochasticity of mRNA expression, the limiting amounts of mRNA available per cell or inefficient mRNA capture. For example, RFP+ and RFP- cells largely matched the distribution of *Runx1*+ and *Runx1*- cells, but some RFP+ cells had undetected *Runx1* expression (Figure 5.7 C and 5.11 A), indicating possible dropout events. In this case, this could also be explained by differences in *Runx1* mRNA and RFP protein stability in P2:*Runx1*-RFP reporters. Overall, while capturing a few thousand cells and a few thousand genes might be sufficient to generate a single-cell landscape which accurately reflects tissue heterogeneity, increased resolution at the cell and gene levels would certainly be beneficial. New technologies are starting to emerge to improve resolution, such as combinatorial indexing which recently enabled profiling 2 million cells representing sequential stages of embryonic development (Svensson et al., 2018). Considering the fast pace at which the field of scRNA-seq has evolved in the last decade, these new technologies will rapidly overtake current standards. Increasing the number of mouse prostate cells from thousands to hundreds of thousands may reveal the existence of additional small subpopulations. Similarly, more genes detected per cell may help reveal subtle differences within existing clusters. It could notably provide a greater understanding of the transcription factors expressed, which are known to be expressed at low levels while playing key roles in cell fate and cell identity (Iwafuchi-Doi and Zaret, 2016).

Chapter 6 Characterisation of *Runx1* expression in the developing mouse prostate

[This section has been adapted from (Mevel et al., 2020)*]*

6.1. Introduction

The mouse prostate starts developing from the endoderm-derived urogenital sinus (UGS). At embryonic day (E) 15.5, the UGS is composed of an outer layer of densely organised mesenchyme, which surrounds a stratified layer of epithelial cells (Georgas et al., 2015; Toivanen and Shen, 2017). The instruction to commit towards the prostatic fate is thought to occur between E15.5 and E16.5, before the emergence of the first prostate buds (Timms, 2008). At this stage, AR is expressed in both epithelial and mesenchymal compartments, and its activation is required for the induction of prostate budding. Indeed, previous work showed that mesenchymal AR signalling was essential for prostate specification (Cunha and Lung, 1978; Lasnitzki and Mizuno, 1977), whereas epithelial AR expression was required for the final stages of prostate morphogenesis as well as its secretory function (Donjacour and Cunha, 1993). By E17.5, the crosstalk between mesenchymal and epithelial compartments results in the induction of *Nkx3-1* expression in a restricted subset of specified epithelial cells (Bhatia-Gaur et al., 1999; Keil et al., 2012; Kruithof-de Julio et al., 2013). *Nkx3-1* is a master regulator of prostate development and considered the first prostate specific transcription factor expressed during embryogenesis (Bhatia-Gaur et al., 1999; Toivanen and Shen, 2017). NKX3.1+ cells form premature prostate buds that invaginate into the surrounding mesenchyme in a bilateral and symmetrical manner. From the initial budding events to the onset of puberty, the prostate undergoes and active phase of branching morphogenesis to generate a complex branched ductal network organised in four major pairs of lobes (Sugimura et al., 1986a; Tika et al., 2019).

During embryogenesis, the developing prostate epithelium comprises a heterogeneous fraction of K5/14+ K8/18+ cells, often co-expressing p63 and NKX3.1

(Kruithof-de Julio et al., 2013; Pignon et al., 2013). At early postnatal stages, the K5/14+ and K8/18+ compartments become progressively distinct. Genetic lineage tracing using either *Krt5*, *Krt14* or *Krt8* drivers suggest that the K5/14+ compartment contains a subset of unipotent basal progenitors as well as multipotent basal progenitors capable of giving rise to both basal and luminal lineages, while K8+ luminal progenitors are restricted to the luminal fate (Ousset et al., 2012). Spatiotemporal analyses of traced K5/14+ multipotent basal progenitors revealed their enrichment in the distal (or tip) region of the ducts, possibly to orchestrate branching morphogenesis (Tika et al., 2019). These multipotent basal progenitors progressively switch to unipotency by differentiating into unipotent basal cells and luminal progenitors prior to the onset of puberty (Tika et al., 2019). The majority of prostate expansion takes place in the luminal compartment, under the influence of hormonal changes occurring through puberty (Tika et al., 2019). While these studies have provided major advances in the understanding of prostate development, the precise epithelial hierarchy and the molecular events leading to the establishment of the differentiated prostate epithelial lineages remains elusive. In particular, it is currently unknown whether distinct adult prostate lineages derive from a common precursor, and at which point in development the definitive cell identities are established.

Previous Chapters demonstrated the specific expression pattern of RUNX1 in the intact, castrated and regenerated adult prostate. RUNX1 was shown to mark a distinct luminal lineage enriched in the proximal/peri-urethral region of the prostate, distinct from the previously identified NKX3.1+ luminal castration resistant cells. Using genetic lineage tracing, these RUNX1+ PLCs were found to be intrinsically castration resistant, and capable of sustaining their own lineage. However, RUNX1+ PLCs did not contribute to the regeneration of distal NKX3.1+ luminal cells. Single-cell transcriptomic analysis revealed the existence of multiple distinct luminal populations co-existing within the adult intact and castrated prostates. Importantly, the gene expression program of RUNX1+ PLCs/Lum-D cells indicated that these cells do not have a secretory function and resemble castration-

resistant cells. These results suggest that during prostate development, proximal and distal lineages may emerge from distinct progenitors. RUNX transcription factors are known to be widely implicated in the development of numerous tissues (Mevel et al., 2019). However, their expression in the development of the prostate has not yet been explored. Therefore, the expression of RUNX1 was first investigated in the developing mouse prostate, both during embryonic and postnatal stages. Next, to further characterise the early stages of prostate specification, an *ex vivo* culture model of embryonic prostate development was used in combination with scRNA-seq and genetic lineage tracing approaches.

6.2. Results

6.2.1. *Runx1* is expressed during prostate development

At E18.5, once the first prostate buds have emerged, RUNX1 was mainly found expressed in the K8-high inner layers of the stratified urogenital epithelium (UGE) (Figure 6.1 A). In contrast, RUNX1 expression was lower in p63+ and K5+ cells lining the outer UGE (Figure 6.1 A-B). The tips of the premature prostate buds were characterised by low or absent RUNX1 expression, and marked by high levels of NKX3.1, one of the first prostate-specific transcription factor expressed during prostate development (Figure 6.1 C). This observation mirrored the mutually exclusive expression pattern between RUNX1 and NKX3.1 found in the adult prostate. Interestingly, RUNX1+ cells of the UGE co-expressed K4 (Figure 6.1 D), previously found to mark RUNX1+ PLCs in the adult prostate, as well as LY6D, recently shown to mark a subset of adult luminal progenitors (Barros-Silva et al., 2018) (Figure 6.1 E). Also, cells marked by the expression of RUNX1, K4 and LY6D were completely negative for NKX3.1 expression, suggesting that the expression of these genes already identify distinct cellular populations.

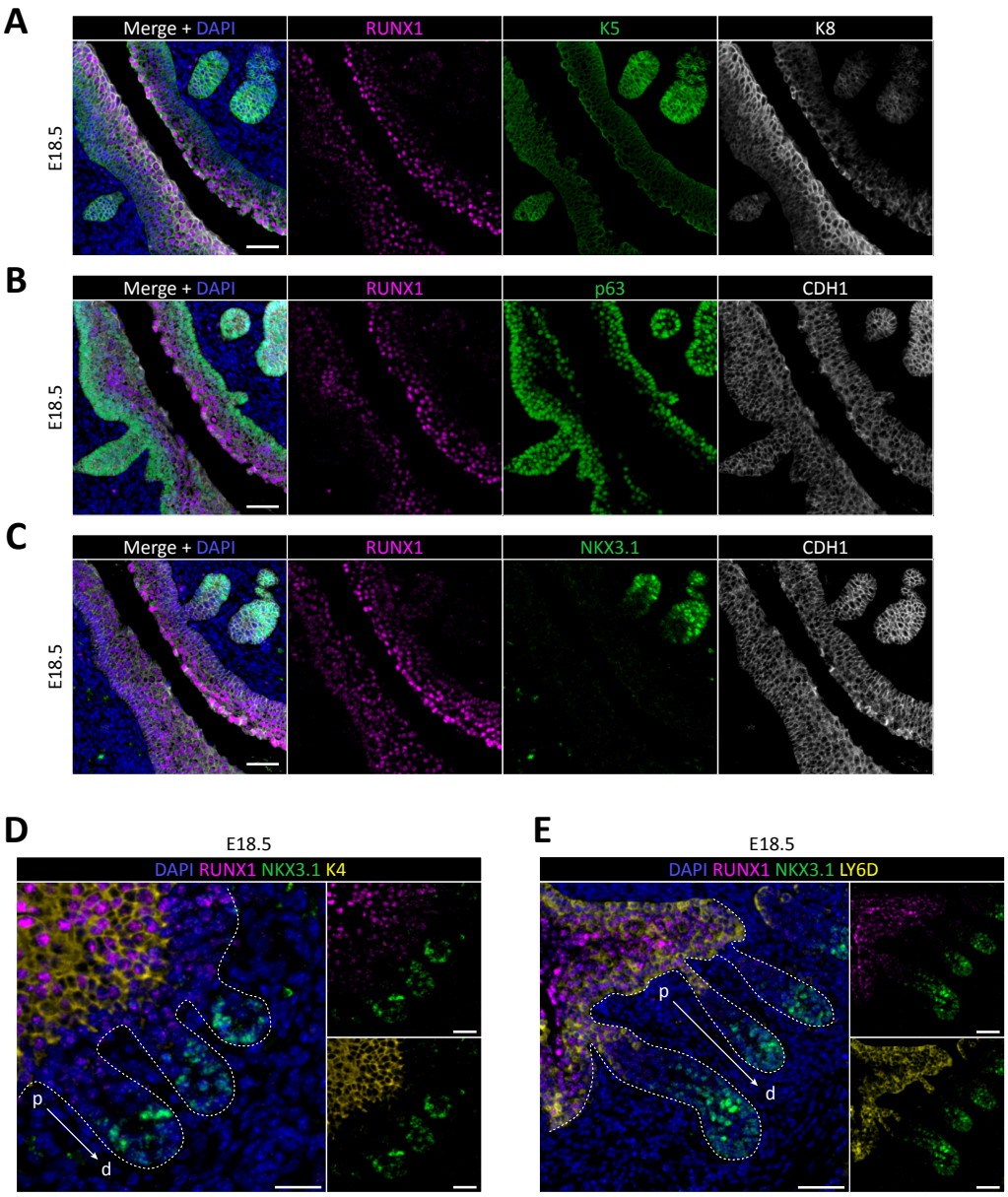

Figure 6.1. Characterisation of RUNX1 expression during embryonic prostate development at E18.5.
(A-E) Co-immunostainings of the mouse urogenital sinus at E18.5 for (A) RUNX1, K5, K8, (B) RUNX1, p63, CDH1, (C) RUNX1, NKX3.1, CDH1, (D) RUNX1, NKX3.1, K4, (E) RUNX1, NKX3.1, LY6D. Scale bar: 50 µm.

Later in development, postnatal day 14 (P14) corresponds to a prepubescent stage when most of the initial branching events have already occurred (Sugimura et al., 1986a; Tika et al., 2019). Consistent with its expression pattern observed at E18.5 and in the adult, RUNX1 was broadly and highly expressed in the proximal region of the developing prostate (Figure 6.2 A-B). RUNX1+ cells located in the

proximal region were largely distinct from distal NKX3.1+ cells (Figure 6.2 A). Also, at this developmental stage, RUNX1+ cells often co-expressed K4, mainly in the proximal region (Figure 6.2 B). Even though RUNX1 expression appeared to be higher in luminal cells, some RUNX1+ K5+ and RUNX1+ p63+ basal cells could also be observed (Figure 6.2 C-D).

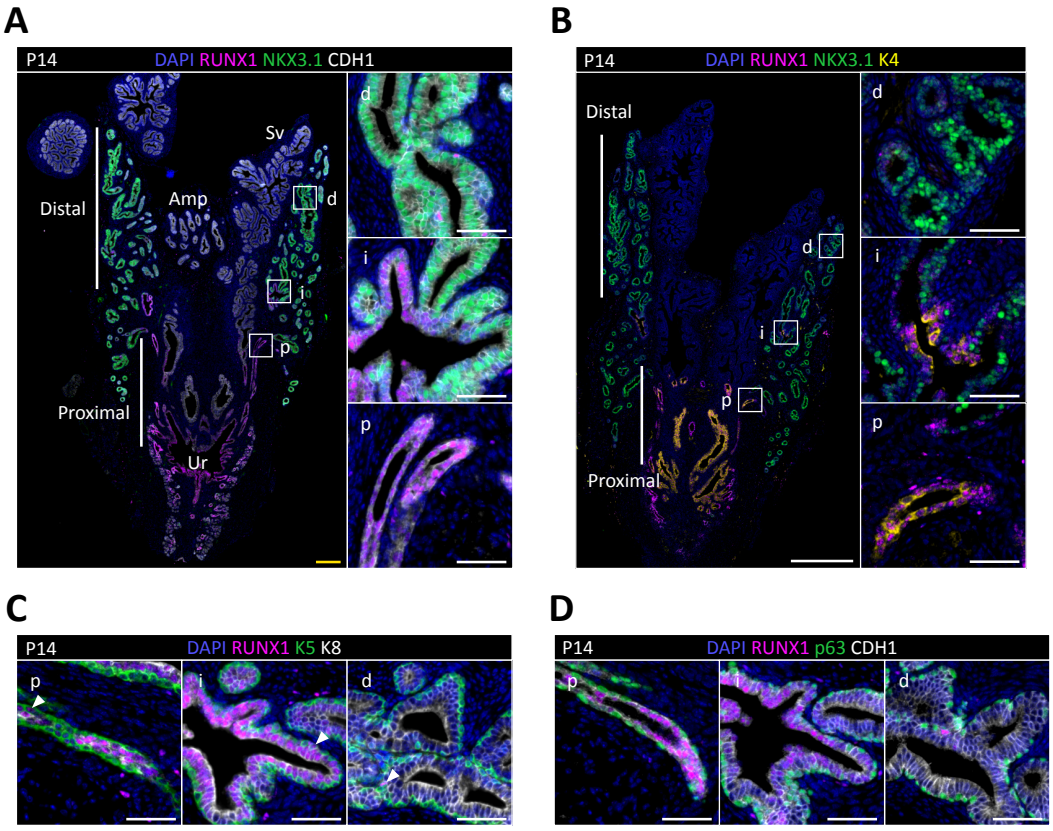

Figure 6.2. Characterisation of RUNX1 expression during postnatal prostate development. (A-E) Co-immunostainings of the mouse urogenital sinus at P14 for (A) RUNX1, NKX3.1, CDH1, (B) RUNX1, NKX3.1, K4, (C) RUNX1, K5, K8, (D) RUNX1, p63, CDH1. High magnification images of proximal (p), intermediate (i) and distal (d) regions are shown. Scale bar: 50 µm. Sv: Seminal vesicles, Amp: ampullary gland, Ur: urethra.

Overall, the specific spatial expression pattern of RUNX1 in proximal luminal cells, largely mutually exclusive with NKX3.1, suggests that these two transcription factors already mark distinct cellular lineages during embryonic prostate organogenesis. Distal cells broadly express NKX3.1, while proximal cells are marked by the co-expression of RUNX1, K4, as well as LY6D.

6.2.2. UGS explants: an *ex vivo* system to study the onset of prostate development

The early phases of prostate development were further investigated using an *ex vivo* explant culture system (Berman et al., 2004; Doles et al., 2005; Kruithof-de Julio et al., 2013; Lopes et al., 1996). Dissected E15.5 UGS were cultured for up to 7 days in the presence of dihydrotestosterone (Figure 6.3 A). This approach preserves both epithelial and mesenchymal compartments and has proven useful to study *ex vivo* the initial steps of prostate bud formation and branching morphogenesis. After 3 days, initial budding events could already be observed under the microscope, and then became more define in mature explants by day 7 (Figure 6.3 B). *In situ*, bud formation could be evidenced by day 2, and the epithelium was mainly composed of a double positive K5+ K8+ stratified epithelium (Figure 6.3 C-D). This epithelium then partially diversified by day 7 with the expansion of single positive K5+ and K8+ cells (Figure 6.3 C-D). Overall, the UGS explant model is a useful model to study embryonic prostate branching morphogenesis and cellular diversification processes *ex vivo*.

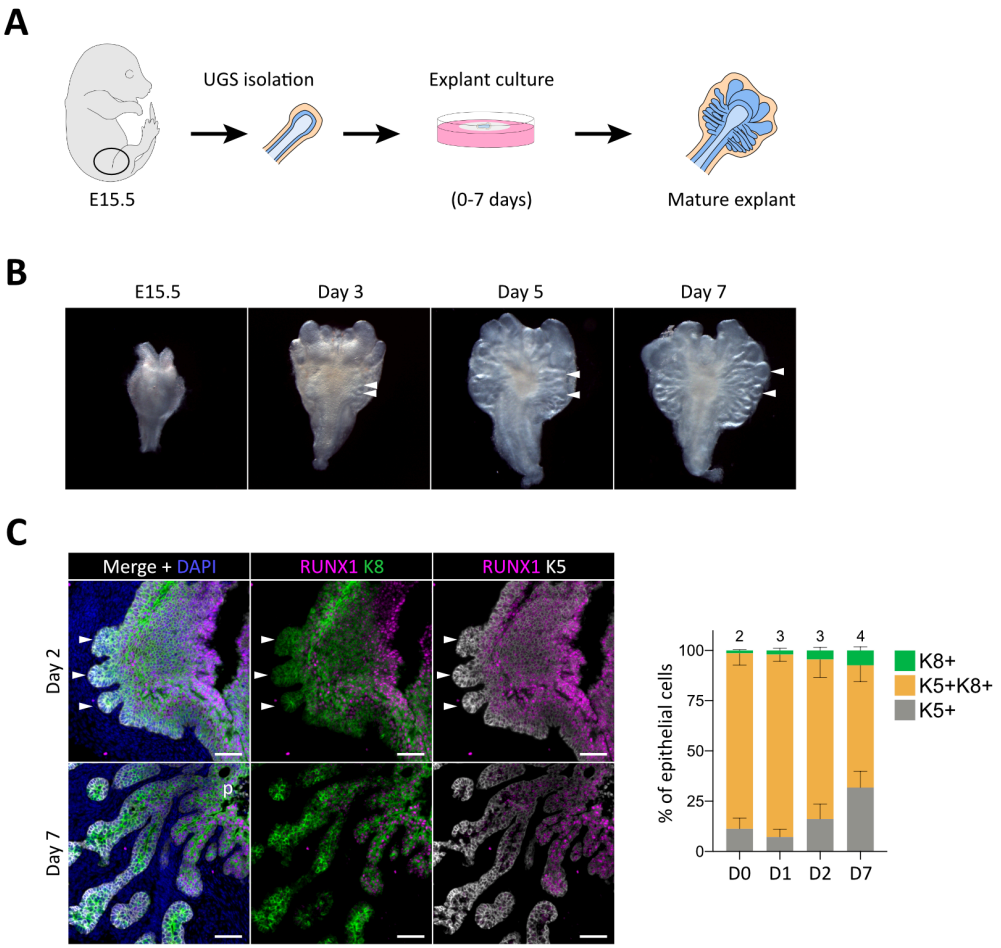

Figure 6.3. UGS explants model to study the onset of prostate development.
(A) Scheme of the protocol to culture *ex vivo* explants of mouse UGS harvested at E15.5.
(B) Representative images of UGS explants at E15.5 (day 0), day 3, day 5 and day 7 of culture showing the formation of premature prostate buds (arrowheads).
(C) Co-immunostaining of RUNX1, K5, K8 in UGS explants harvested at day 2 and day 7. Arrowheads indicate premature prostate buds emerging by day 2. p: proximal, d: distal. Scale bar: 50 µm.
(D) Quantification of the percentage of K5+, K8+, K5+ K8+ cells during UGS explant cultures at day 0 (n = 2 explants), day 1 (n = 3 explants), day 2 (n = 3 explants), day 7 (n = 4 explants), showing a progressive reduction in the proportion of K5+ K8+ double positive cells.

6.2.3. Expression pattern of *Runx1* in UGS explant cultures

Next, the dynamic of emergence of RUNX1+ cells was studied in UGS explants using quantitative image-based cytometry (QIBC) (Figure 6.4). On day 0 (E15.5), RUNX1 was detected at the rostral end of the UGE, particularly within the inner layers of the stratified epithelium (Figure 6.4 A). After 1 day in culture, NKX3.1

expression emerged in a subset of RUNX1+ cells located in the outer layers of the UGE, while defined budding was yet to be observed. On day 2, NKX3.1+ prostate buds were evident and had reduced or absent RUNX1 expression. This pattern was conserved in the mature explant, in which distal tips were mainly NKX3.1+, whereas the proximal area remained RUNX1+. These results suggest that during prostate specification, RUNX1 expression precedes NKX3.1 in a subset of cells located in the outer layer of the UGE. Once NKX3.1 is being expressed, RUNX1 is downregulated and these two transcription factors mark distinct cellular subsets along the proximal-to-distal axis.

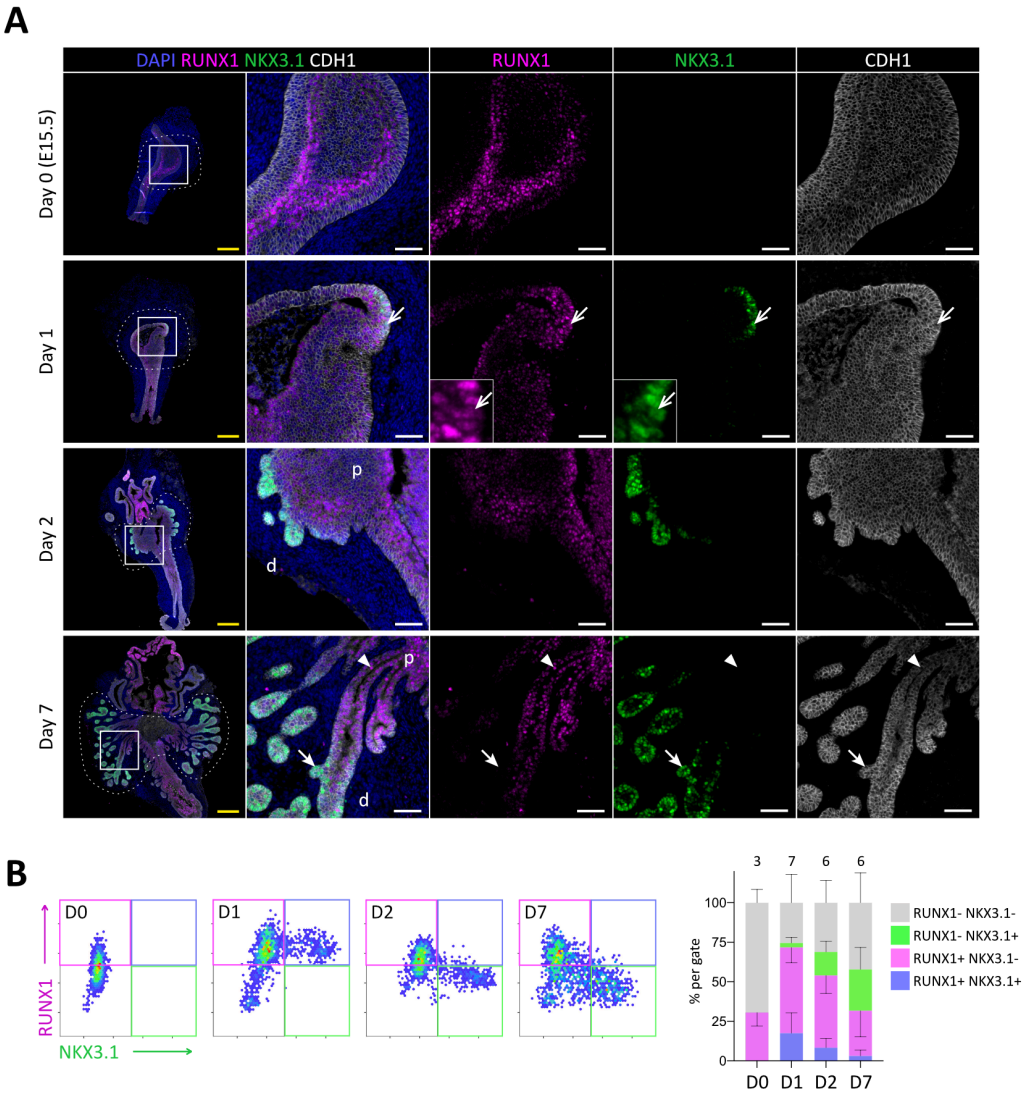

Figure 6.4. Characterisation of RUNX1 and NKX3.1 expression in UGS explant cultures. (A) Co-immunostaining of RUNX1, NKX3.1, CDH1 in UGS explants harvested at day 0, day 1, day 2 and day 7. Higher magnification images of each square (left) are shown for each time

point. Chevron arrows show RUNX1+ NKX3.1+ cells, closed arrows indicate RUNX1- NKX3.1+ cells, arrowheads show RUNX1+ NKX3.1- cells. Scale bars: 200 µm (yellow) and 50 µm (white).
(B) Quantification of RUNX1 and NKX3.1 nuclear intensity (log10) in CDH1+ epithelial cells of UGS explants by QIBC. Quantification was performed within the boundaries delimited in G by dotted lines, at day 0 (n = 3 explants), day 1 (n = 7 explants), day 2 (n = 6 explants) and day 7 (n = 6 explants).

Besides NKX3.1, the expression of LY6D and K4 was also investigated in UGS explants. As expected from E18.5 and P14 immunostainings (see section 6.2.1), RUNX1 expression largely overlapped with both K4 and LY6D, extending from the middle of the explant into the proximal region of the prostate buds (Figure 6.5 A-B). In the epithelial prostate buds, RUNX1 and K4 mainly localised in the inner part of the buds at both day 2 and 7, with lower expression levels in the outer layer (Figure 6.5 A). This contrasted with the expression of LY6D which marked proximal region of the buds, as well as the most outer layer of the stratified epithelium until the tips of NKX3.1+ distal buds (Figure 6.5 B). Of note, while K4 and LY6D have a mainly overlapping expression pattern, they also appear to differentially mark smaller cellular subset, in the middle of the explant for K4, and a small subset of cells of the distal prostate buds for LY6D.

Together, these results indicate that the expression pattern of RUNX1 in UGS explant cultures recapitulates its expression during embryonic prostate development *in vivo*, and its expression broadly overlaps with K4 and LY6D. Prostate budding originates from a subset of cells located in the outer layers of the stratified UGE, transiently marked by RUNX1 and NKX3.1. During embryonic prostate development, *Runx1* expression is already primarily confined to the proximal region of the prostatic ducts, in a distinct compartment from NKX3.1+ cells.

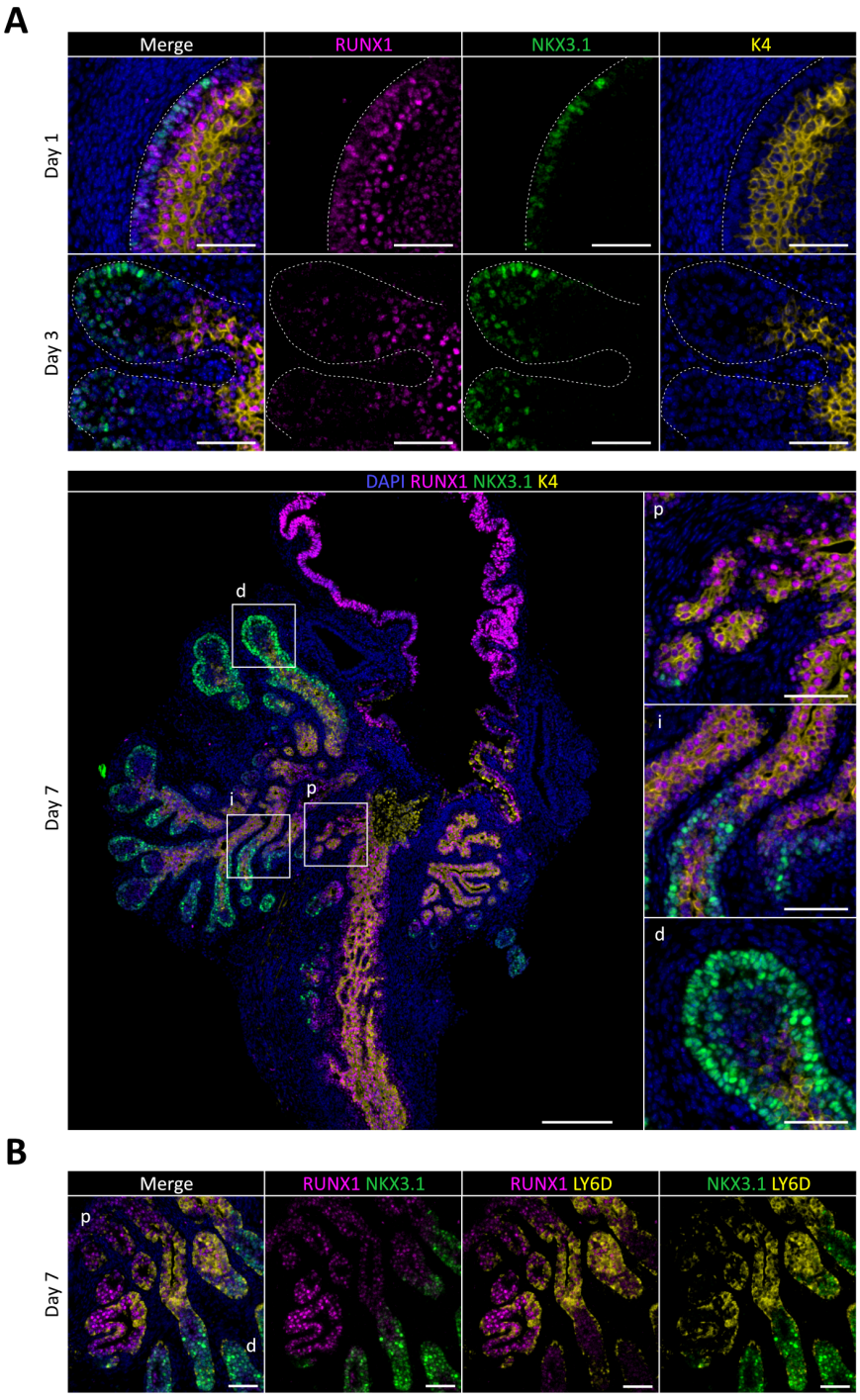

Figure 6.5. Characterisation of K4 and LY6D expression together with RUNX1 and NKX3.1 in UGS explant cultures.

(A) Co-immunostaining of RUNX1, NKX3.1, K4 in UGS explants harvested at day 1 (top), 3 (middle) and 7 (bottom). Higher magnification images of (p) proximal, (i) intermediate, and (d) distal regions are shown for day 7. Dotted lines indicate the urogenital epithelium.
(B) Co-immunostaining of RUNX1, NKX3.1, LY6D in UGS explants harvested at day 7. p: proximal, d: distal. Scale bars: 200 µm (yellow) and 50 µm (white).

6.2.4. scRNA-seq of UGS explant cultures

6.2.4.1. Experimental design

The characterisation of continuous developmental processes by immunostainings is generally constrained to a small number of markers at a time. To further study the specification of RUNX1+ and NKX3.1+ lineages, scRNA-seq was performed on UGS explant cultures collected at successive time points: E15.5 (D0), day 1 (D1), day 3 (D3), and day 6 (D6) (Figure 6.6). Epithelial EPCAM+ single cells sorted from different experimental time points were multiplexed using MULTI-seq lipid-tagged indices to minimise technical confounders (as for the adult scRNA-seq described in Chapter 5).

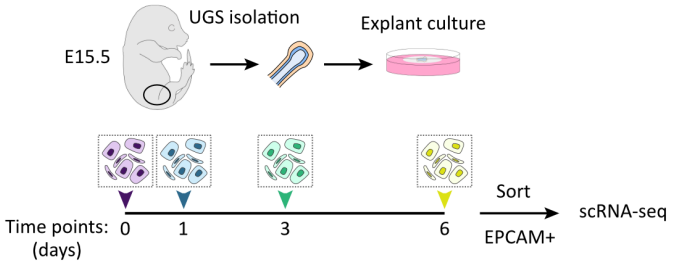

Figure 6.6. Experimental design used for scRNA-seq of UGS explant cultures.
Explants were harvested at day 0, day 1, day 3 and day 6.

6.2.4.2. Quality control and sample demultiplexing

After sequencing and data pre-processing, several quality control steps were performed to select high quality cells. The barcode rank plot had the expected "knee-shaped" profile, enabling the exclusion of empty droplets (Figure 6.7 A). Cells were then assigned to their population of origin by demultiplexing the MULTI-seq tag data. The spread of the MULTI-seq tag read counts displayed reasonable signal-to-noise (Figure 6.7 B), resulting in the classification of the majority of the cells to a specific barcode (Bar1 to Bar4 corresponding to D0 to D6) (Figure 6.7 C). Small fractions of "negative" and "doublet" cells were identified and excluded from the dataset. Finally, low quality cells were removed if they had less than 750 detected genes, more than 7500 UMIs, and more than 10% mitochondrial transcripts (Figure 6.7 D). In total, 5,122 cells were retained for analysis.

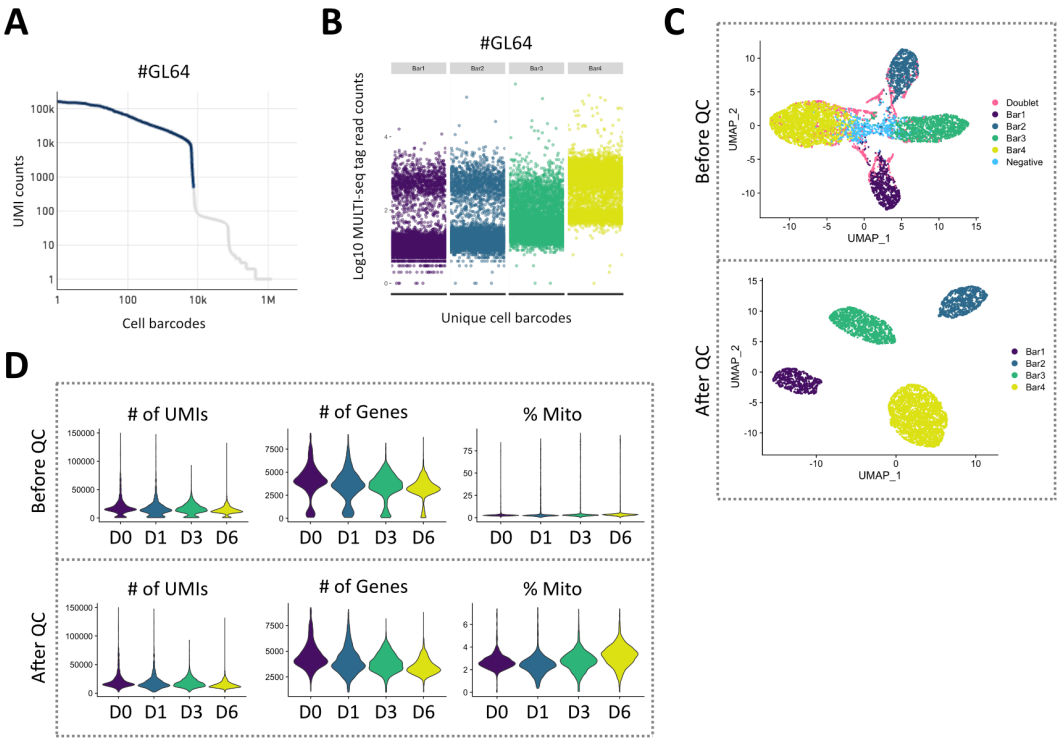

Figure 6.7. Quality control of the UGS explant cultures scRNA-seq dataset.
(A) Barcode rank plots produced by the 'Cell Ranger' software showing cell-associated barcodes (blue) and barcodes associated with empty droplets (grey).
(B) Scatter plot showing the abundance of each MULTI-seq tag counts (log10) for each cell barcode.
(B) UMAP visualisation in MULTI-seq barcodes space before (top) and after (bottom) barcode filtration (see section 2.8.1.5. for details).
(D) Violin plots of the number of UMIs per cell (# UMIs), number of genes detected per cell (# Genes), and percentage of mitochondrial transcripts (% Mito). Results are shown for each run, before and after quality control. Results are displayed by sample (D0, D1, D3, D6).

6.2.4.3. Graph-based clustering of the whole UGS explant cultures dataset

UMAP visualisation of the cells in gene expression space resulted in multiple cellular clusters. The effect of cell cycle has previously been shown to be one possible major confounder in scRNA-seq experiments (Luecken and Theis, 2019; Trapnell et al., 2014; Tsang et al., 2015). Since UGS explants are actively proliferating tissues, cell cycle phases were estimated at the single cell level using gene signatures of G1, S and G2/M phases (Figure 6.8 A). Here, cells appeared to

subcluster predominantly according to their predicted cell cycle phase, therefore the effect of cell cycle was regressed out for downstream analyses. Following cell cycle regression, graph-based clustering revealed the presence of 6 main clusters (Figure 6.8 B-C). The underlying cell types were characterised using a combination of manual and automated approaches. This resulted in the identification of several genes specifically marking each cluster (Figure 6.8 E). In particular, a "prostatic" cluster was defined by the expression of classical epithelial basal and luminal markers (*Epcam*, *Krt8*, *Krt5*, *Krt14*, *Krt15*), as well as the expression of well-characterised prostate specific transcription factors (*Ar*, *Hoxb13*, *Hoxd13*, *Nkx3-1*). Of note, one small cluster protruded from the main prostatic cluster and contained a majority of cells from the "Day 1" time point (Figure 6.8 B-C). This cluster was qualified as "hypoxic/stressed" due to the strong enrichment for gene ontology terms such as "response to hypoxia" or "cellular response to stress" (Figure 6.8 D). Additional clusters corresponded to mesonephric derivatives, bladder urothelium, as well as mesenchymal cells (see Figure 6.8 C, E and section 2.8.1.5 for further details). The prostatic cluster consisted of 3,937 developing prostatic cells and was retained for further analyses.

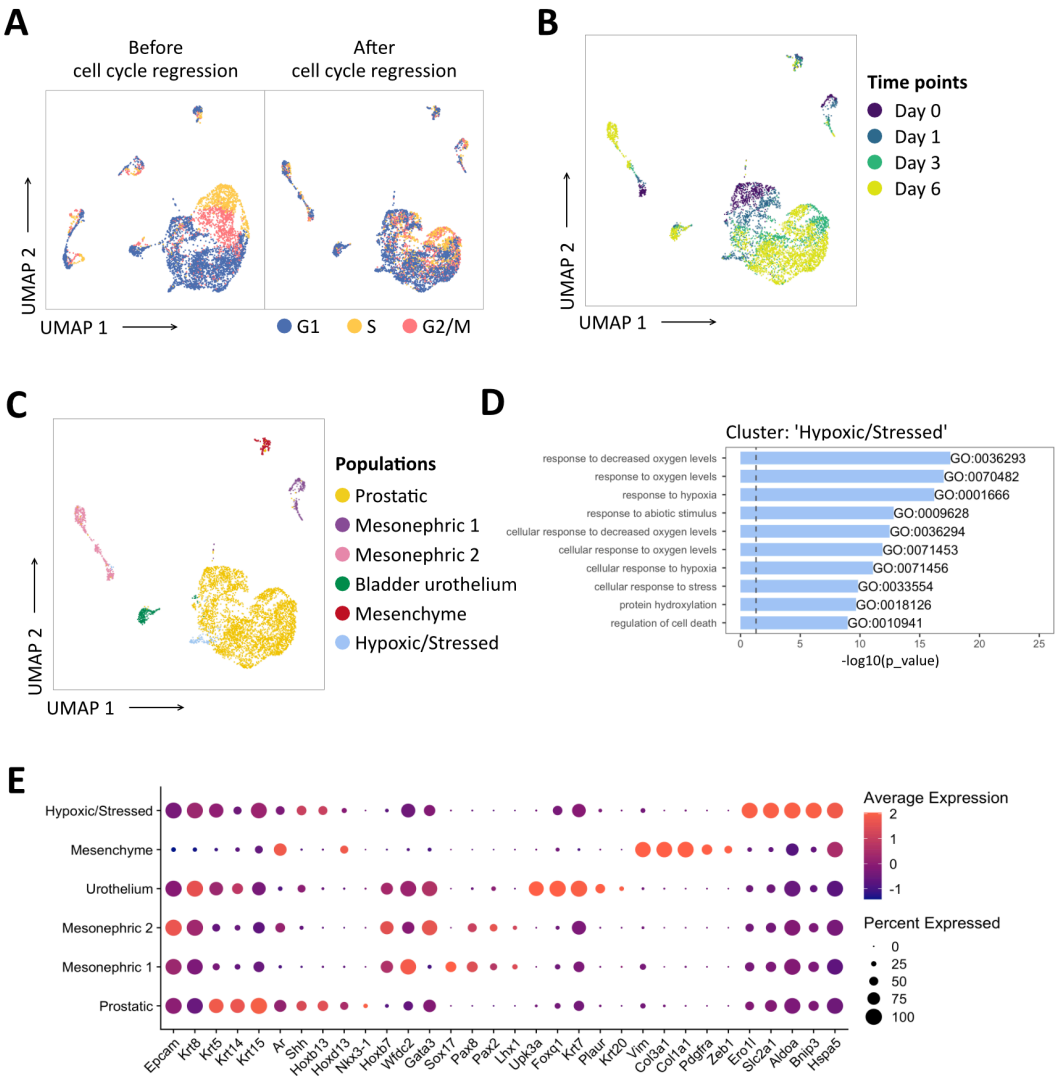

Figure 6.8. UMAP visualisation of the whole UGS explant cultures scRNA-seq dataset.
(A) UMAP visualisation of the dataset coloured by estimated cell cycle phase (top). Due to the strong underlying impact on clustering, the cell cycle effect was regressed out (bottom).
(B, C) UMAP visualisation of the dataset coloured in B by experimental time point and in C by labelled population.
(D) Bar plots of the 10 most significantly (g:Profiler adjusted p-value < 0.05) enriched gene ontology (GO) terms (GO:BP, Biological Processes) on the differentially upregulated genes specific to the 'Hypoxic/Stressed' cluster.
(E) Dot plot showing the expression of selected marker genes associated with each labelled population.

6.2.4.4. Graph-based clustering of the developing prostatic epithelium

UMAP visualisation of the developing prostatic subset revealed progressive changes in gene expression and cellular populations occurring between Day 0 and Day 6. This was particularly evident by the gradual shift in cell distribution along the "UMAP 1" axis (Figure 6.9 A). Graph-based clustering of the prostatic subset resulted in 10 distinct clusters, termed C0 to C9 (Figure 6.9 B). Since the effect of cell cycle had already been regressed, predicted cell cycle phases were spread out between the different clusters, except for C8 which contained only actively cycling cells in either S or G2/M phase (Figure 6.9 C-D).

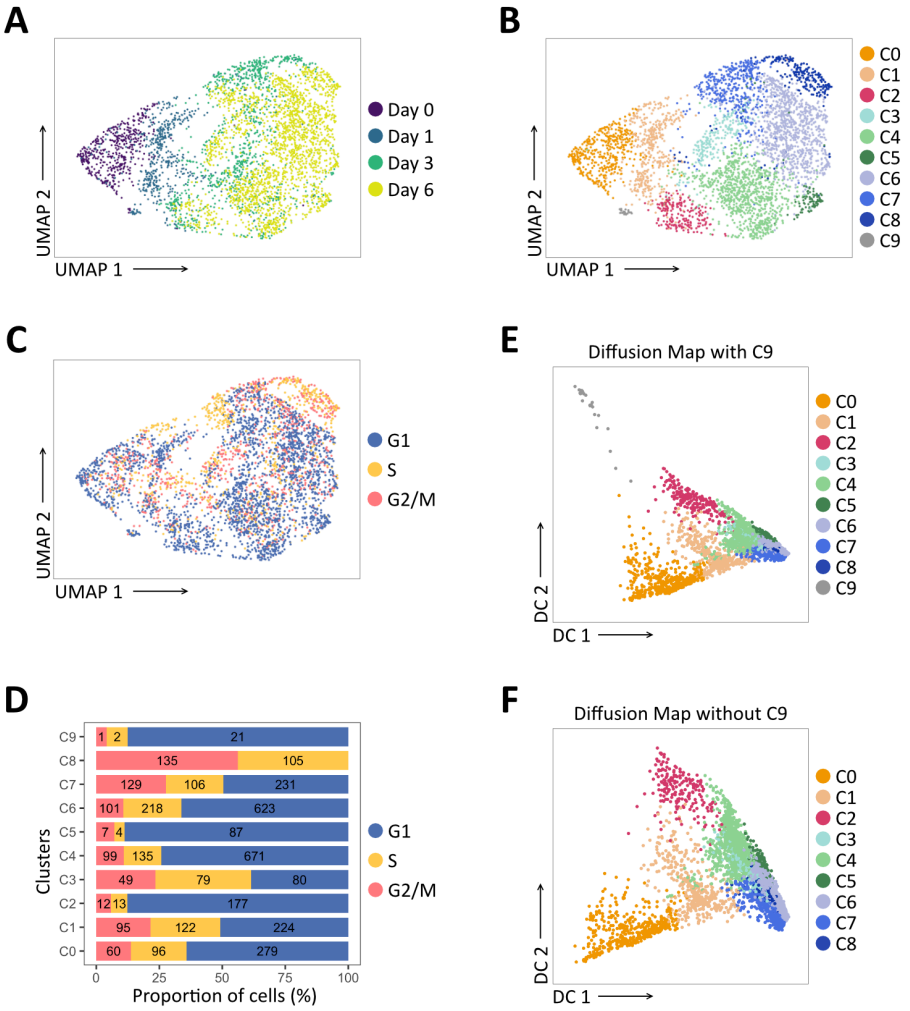

Figure 6.9. UMAP visualisation of the developing prostatic epithelium.
(A-C) UMAP visualisation of the prostatic subset (from Figure 6.8 C), labelled by time points in A, clusters in B, estimated cell cycle phase in C.
(D) Percentage of cells in the different estimated cell cycle phases in each cluster.

(E, F) Diffusion maps representation of the prostatic subset using the first 2 diffusion components with (E) and without C9 (F). C9 was excluded for downstream analyses due to its outlying profile in E.

Besides UMAP, diffusion maps represent another non-linear dimension reduction approach to study differentiation processes in a low-dimensional plot, allowing ordering of single cells in a pseudotemporal space (Haghverdi et al., 2015). Furthermore, diffusion maps have recently been used as an intermediate step to generate improved PAGA and force-directed visualisations, thereby allowing more faithful reconstruction of lineage trajectories (Schiebinger et al., 2017; van Dijk et al., 2018). After computing diffusion components and visualising the prostatic subset in diffusion space, the C9 cluster appeared to be outlying from the remainder fraction and was excluded for downstream analyses (Figure 6.9 E). The final diffusion representation gave a smoother and more connected cellular distribution (Figure 6.9 F), originating from the C0 cluster which mainly comprised cells from the initial day 0 time point (Figure 6.9 A, B, F).

Using diffusion components as an input dimension reduction, the dataset was next visualised using both a force-directed layout and partition-based graph abstraction (PAGA). Both approaches allow the representation of continuous gene expression topologies in a manner that preserves high-dimensional relationships (Wolf et al., 2019). These representations highlighted the progressive cellular diversification taking place from D0 to D6 (Figure 6.10 A-E). Interestingly, the developmental progression through experimental time points could be approximated via diffusion pseudotime (Figure 6.10 B). The C0/C1clusters contained a majority of D0 and D1 derived cells, while C2-C8 emerged and expanded at later time points (Figure 6.10 D). Differential expression analysis was then performed to characterise the specific gene expression profiles of the different clusters. The 10 most specific genes per cluster were visualised as a heatmap (Figure 6.10 E). While some genes were strongly associated with specific clusters, such as *Psca* in C2 or *Top2a* in C9, most of the genes exhibited a rather broad expression pattern across multiple clusters, suggesting transcriptomic proximity

between the clusters (Figure 6.10 E). This observation was further confirmed using PAGA, which connected most clusters by relatively strong edges (Figure 6.10 F). Overall, these results are compatible with progressive cell state changes occurring during continuous differentiation processes.

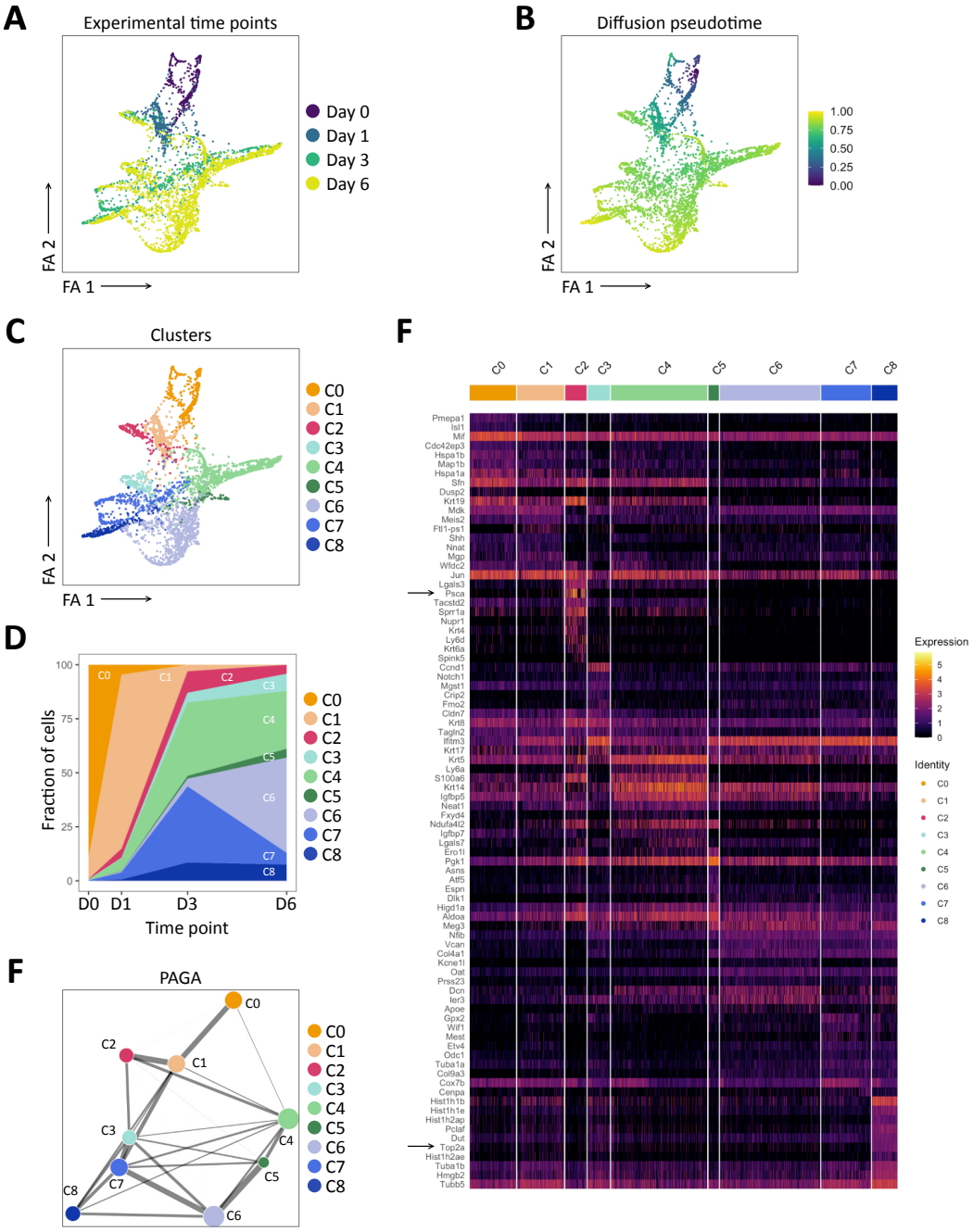

Figure 6.10. Visualisation of the developing prostatic epithelium.
(A-C) Force directed visualisation of the developing prostatic epithelium. In A, cells are coloured by experimental time points, in B cells are coloured by diffusion pseudotime, and in C cells are coloured by clusters.

(C) PAGA representation of the clusters as in C. Weighted edges between cluster nodes represent a statistical measure of connectivity.
(D) Fraction of cells per cluster at each experimental time point.
(E) Heatmap of the 10 most differentially upregulated genes per cluster.
(F) PAGA representations with cluster nodes coloured by a gradient representing the mean log-normalised expression levels of each gene.

The gene expression profile of selected genes of interest were then manually investigated using PAGA (Figure 6.11 A). Indeed, due the relatively small gene expression differences observed between the different clusters, PAGA offers a better representation to visualise minor changes in gene expression by displaying the average expression per cluster.

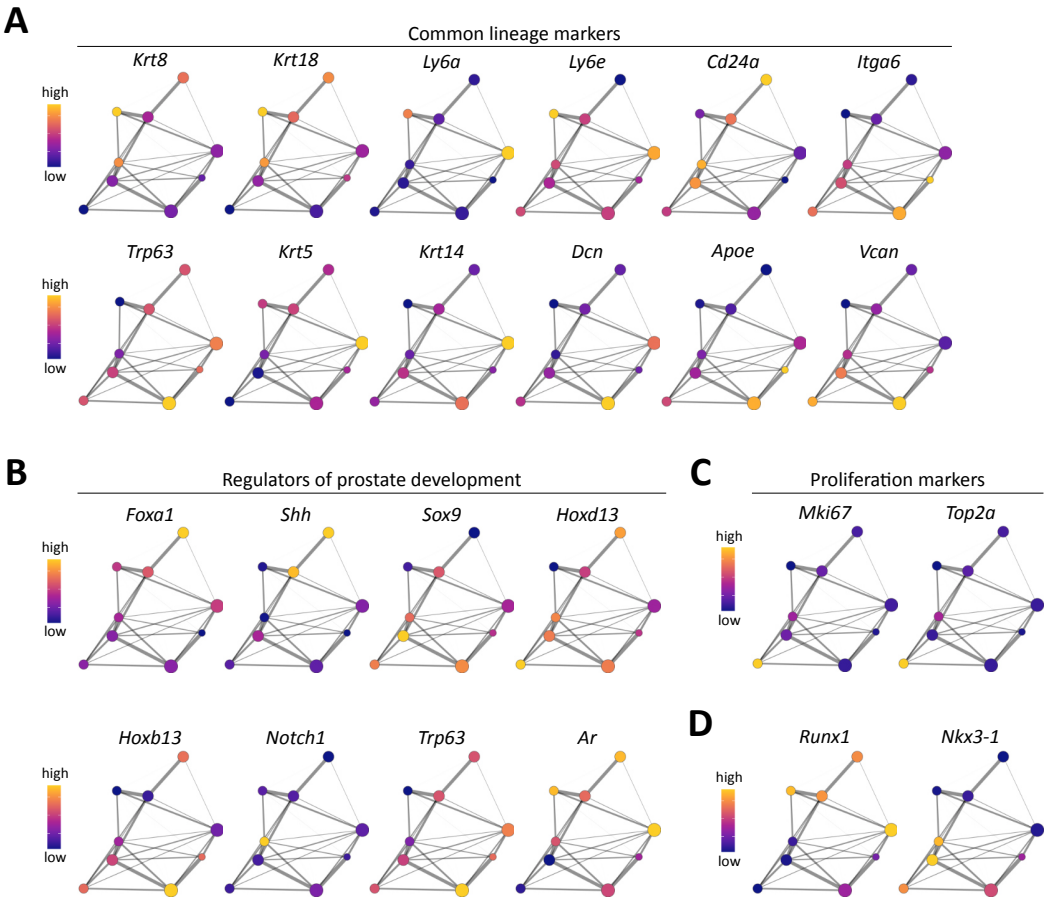

Figure 6.11. Investigation of the gene expression profile of the developing prostatic epithelium.
(A-E) PAGA representations with cluster nodes coloured by a gradient representing the mean log-normalised expression levels of each genes indicated. (A) shows the expression of

common prostate lineage markers, (B) shows known regulators of prostate development, (C) shows proliferation markers, (D) shows the expression of *Runx1* and *Nkx3-1*.

After querying canonical lineage markers, C4-C6 appeared to have a more pronounced 'basal' identity compared to the other clusters. *Krt5*/*Krt14* marked mainly C4, and additional basal markers including *Trp63*, *Dcn*, *Apoe*, or *Vcan* were higher in C5/C6. Comparatively, luminal markers such as *Krt8*/*Krt18* had higher expression levels in C0 to C3. Further to this, known regulators of prostate development (Toivanen and Shen, 2017) displayed a variable expression pattern across the different clusters. For example, *Foxa1* and *Shh* were strongly expressed in C0/C1, *Notch1* was higher in C3, and *Sox9* in C7 (Figure 6.11 B). Also, in line with the predicted cell cycle phases, C8 consisted in highly proliferative cells (Figure 6.11 C). Finally, consistent with previous results obtained by immunostainings, *Runx1* was highly expressed in clusters having lower *Nkx3-1* levels, including C0, C1, C2 and C4 (Figure 6.11 D). Due to the primitive nature of the UGE at these time points, these results suggest that the classical basal and luminal lineages are yet to be fully established. Nevertheless, *Runx1* and *Nkx3-1* already appear to mark broadly distinct cell populations. Furthermore, this dataset constitutes a novel resource to interrogate specific gene expression patterns during prostate development.

6.2.4.5. Investigation of the lineage relationships between embryonic and adult prostate clusters

To determine how the prostatic clusters related to differentiated prostate lineages found in fully developed prostates, population-specific gene signatures previously identified in the adult were interrogated in the UGS explant cultures dataset. The 'Basal' signature was enriched across all clusters, especially in C4/C6 (Figure 6.12 A-B). In comparison, the different luminal signatures tested displayed weaker enrichment (Figure 6.12 A, C). Nevertheless, the 'Lum-D' derived signature was highly enriched in C2 compared to all the other adult luminal population signatures. This observation suggested that the 'Lum-D' fate is determined early during prostate development. In fact, the singular identity of C2 was characterised by genes previously found highly expressed in the adult 'Lum-D' population,

including *Tacstd2*/*Trop2*, *Krt4*, *Psca*, as well as *Ly6d* and *Nupr1* (Figure 6.12 D). Collectively, scRNA-seq analysis show that adult 'Lum-D'/PLCs share strong similarities with the unique C2 population identified in embryonic explant cultures. This suggests that the distinct proximal luminal lineage is established at the very onset of prostate specification.

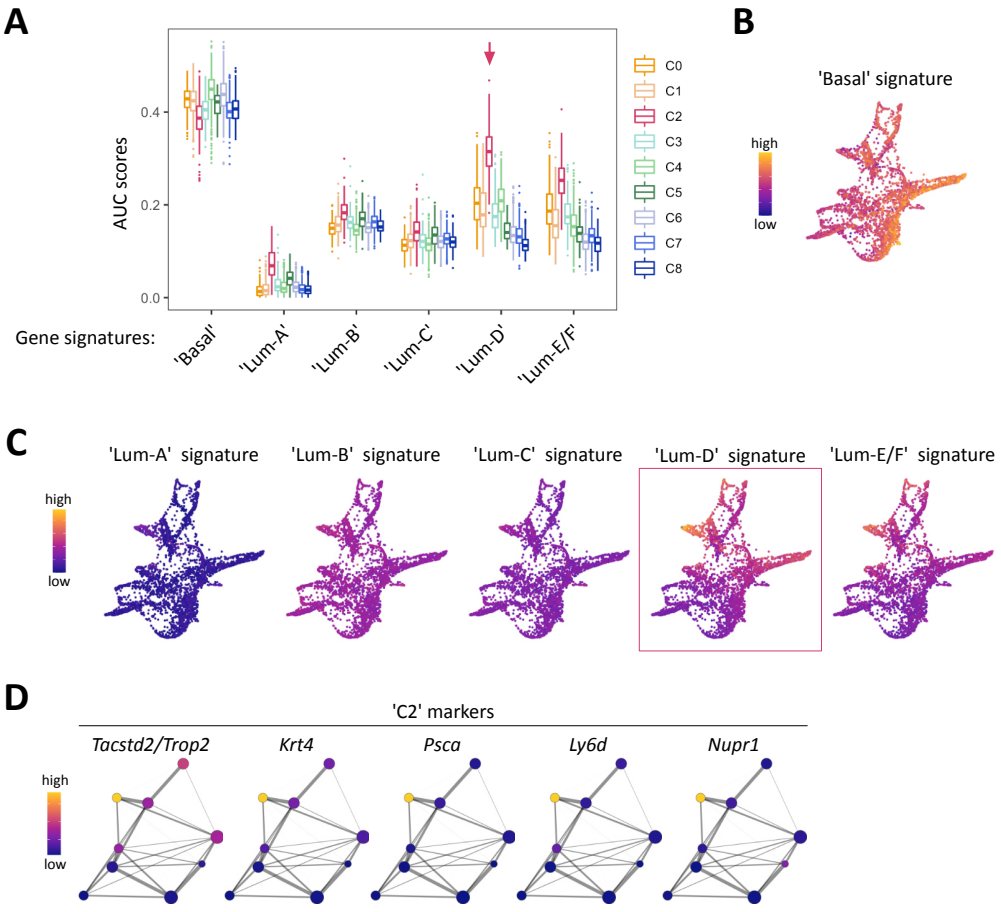

Figure 6.12. Interrogation of differentiated/adult prostate lineages gene signatures in UGS explant cultures.

(A) Box plots of per-cell AUC signature scores calculated using the 'AUCell' package for individual UGS explant clusters. Gene signatures for basal and luminal populations were generated using the list of differentially upregulated genes previously obtained from our adult mouse prostate clusters.

(B, C) Force directed visualisation of the developing prostatic epithelium in UGS explant cultures. Colour gradient represents AUC scores per cell. The basal gene signature is shown in B, and luminal signatures in C.

(D) PAGA representations with cluster nodes coloured by a gradient representing the mean log-normalised expression levels of genes highly expressed in the C2 cluster.

6.2.4.6. Inference of developmental lineage trajectories of prostate specification

The force-directed layout used to visualise the developing prostatic epithelium suggested that clusters were branching out along the diffusion pseudotime, in particular C2, C4, C6 and C7/C8 (Figure 6.10 A-C). While diffusion pseudotime ordered cells in a pseudotemporal manner that resembled the "real" time existing between experimental time points (Figure 6.10 B), it could not predict possible developmental trajectories. Therefore, the Slingshot R package (Street et al., 2018) was used to make *in silico* predictions of putative developmental trajectories. For trajectory estimation, this method can be used with or without the specification of known "priors" regarding the origin and end points of differentiation. Two example predictions are shown in Figure 6.13 A-B. When the starting point was specified (C0), Slingshot predicted 3 trajectories: C0-C1-C2, C0-C1-C4-C5-C6-C7-C3, C0-C1-C4-C5-C6-C7-C8. Alternatively, when the starting (C0) and end points (C2, C4, C6, C8) were specified, the algorithm predicted 5 trajectories: C0-C1-C2, C0-C1-C7-C3, C0-C1-C7-C8, C0-C1-C7-C5-C4, C0-C1-C7-C5-C6. Overall, despite several other attempts (data not shown), no prediction led to particularly credible scenarios. Nevertheless, in both PAGA and force-directed layouts, the C2 cluster appeared to be branching out from the C1 cluster, reinforcing the notion that this cluster corresponds to an emerging distinct lineage (Figures 6.10 B, D).

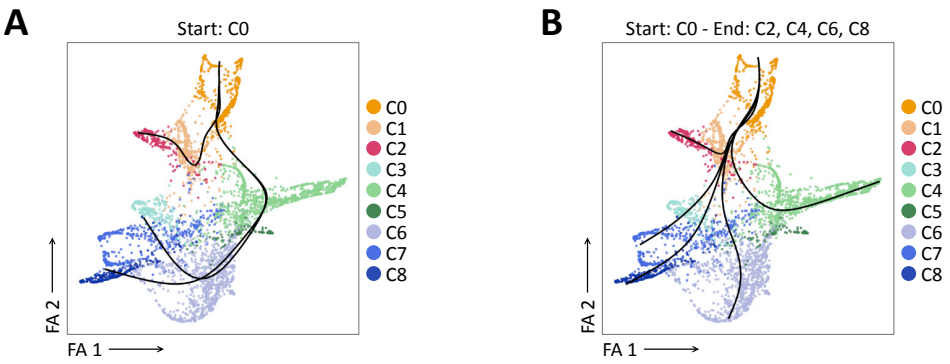

Figure 6.13. Putative developmental lineage trajectories of prostate specification using Slingshot.

(A-B) Predicted developmental trajectories overlaid onto force directed visualisation of the developing prostatic epithelium. Cells are coloured by clusters. In A, only the starting point

was specified (C0) for trajectory estimation. In B, the starting (C0) and end points (C2, C4, C6, C8) were specified.

6.2.5. Tracing the fate of *Runx1* expressing cells in UGS explants

Single-cell RNA-seq offered a snapshot view of the differentiation process occurring in UGS explant cultures and could be used to study intermediate cell states along putative lineage trajectories. However, because of the destructive nature of scRNA-seq experiments, this approach cannot recapitulate the history of a given cell. Taking advantage of the *Runx1*$^{Cre\text{ER}}$ *Rosa26*$^{\text{LSL-RFP}}$ lineage tracing model, the fate of RUNX1+ cells was investigated during embryonic prostate specification in UGS explant cultures. To maximise labelling efficiency, UGS explants were treated with tamoxifen on day 0 and day 1 of culture and analysed on day 2 and day 7 (Figure 6.14 A). At day 2, most RFP labelled cells co-expressed RUNX1 in the inner region of the stratified UGE, with rare RUNX1- RFP+ cells found on the outer layer (Figure 6.14 B). At day 7, the majority of the RFP labelled cells were in the most proximal RUNX1+ subset and rarely found in the distal area of the branches, where RUNX1- cells reside. Accordingly, the proportion of RFP+ RUNX1+ cells remained stable between day 2 and 7. The rare RFP+ RUNX1- cells detected in distal branches by day 7 often co-expressed NKX3.1, whereas very rare RFP+ NKX3.1+ could be detected on day 2 (Figure 6.14 C). Overall, this indicates that *Runx1* expressing cells marginally contribute to the expansion of the NKX3.1 compartment. Instead, the majority of *Runx1* expressing cells remain in the RUNX1+ compartment, located in the proximal region of the explants.

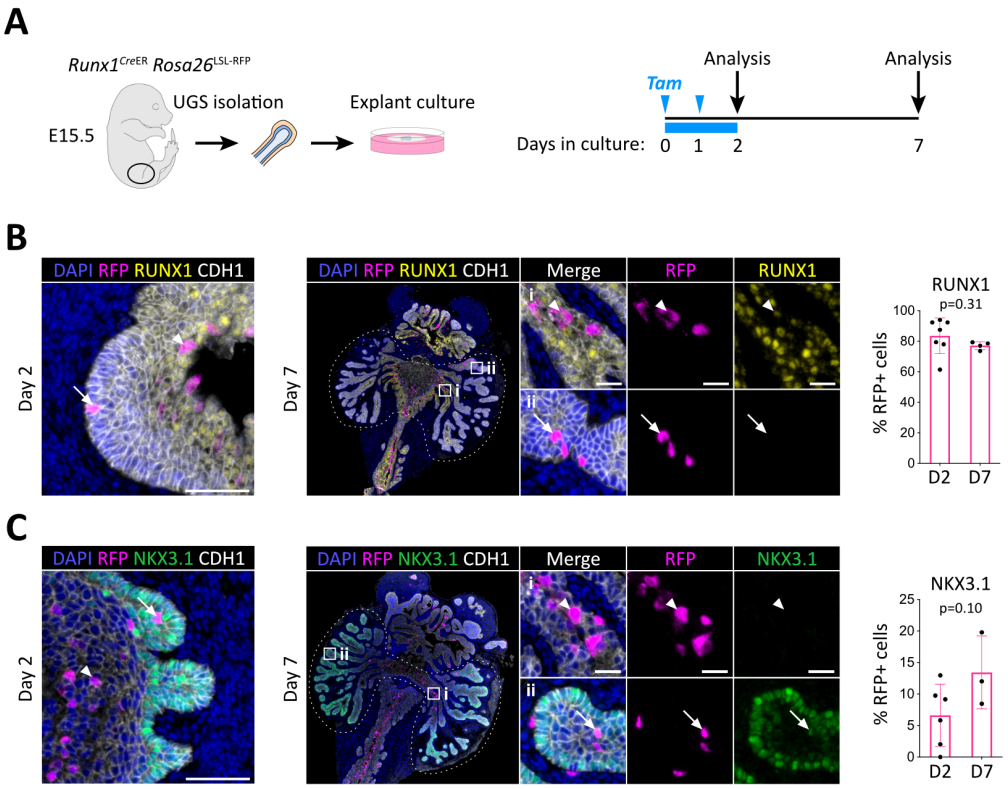

Figure 6.14. Lineage-tracing of RUNX1+ cells in UGS explant cultures.
(A) Strategy for lineage-tracing of RUNX1+ cells in UGS explant cultures. Tamoxifen was applied on day 0 and day 1 and washed out on day 2.
(B) Co-immunostaining of RFP, RUNX1, CDH1 in UGS explants harvested at day 2 (left) and day 7 (middle). Higher magnification images of proximal (i) and (ii) distal regions are shown for day 7. Arrows show RFP+ RUNX1-low cells, arrowheads show RFP+ RUNX1+ cells. Quantification of % of epithelial RUNX1+ cells in the RFP subset at day 2 (n = 7) and day 7 (n = 3) of UGS explant cultures (right). Quantification was performed within the boundaries delimited in B by dotted lines. Scale bars: 200 µm (yellow) and 50 µm (white).

Next, additional characterisation of the lineage traced cells was performed using canonical lineage markers, including p63, K5 and K8. These immunostainings showed that the fraction of RFP+ cells co-expressing p63 remained unchanged throughout the culture, corresponding to about half of the labelled cells (Figure 6.15 A). Probing K5 and K8 markers indicated that the majority of RFP labelled cells co-expressed both K5 and K8 at day 2 (Figure 6.15 B). Interestingly, the fraction of K5+/K8+ RFP+ cells decreased slightly by day 7, which was accompanied by a diversification into either K5+ or K8+ cells.

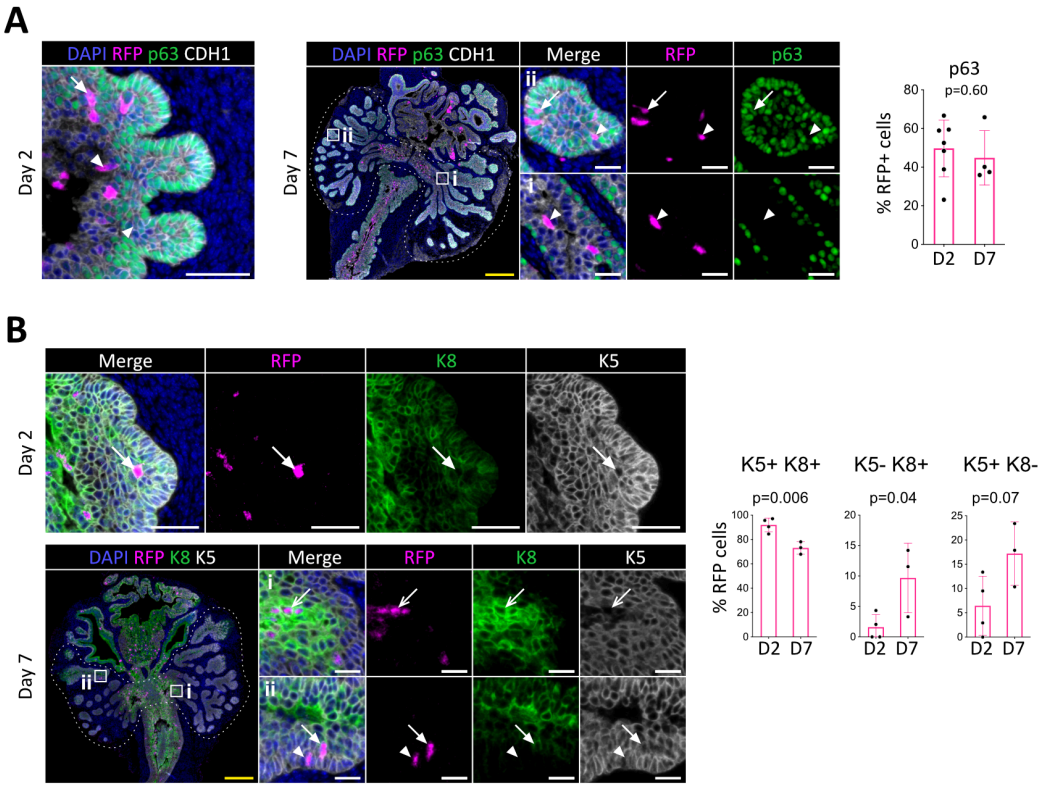

Figure 6.15. Characterisation of RUNX1+ lineage traced cells in UGS explant cultures.
(A) Co-immunostaining of RFP, p63, CDH1 in UGS explants harvested at day 2 (left) and day 7 (middle). Higher magnification images of (i) proximal and (ii) distal regions are shown for day 7. Arrows show RFP+ p63+ cells, arrowheads show RFP+ p63- cells. Quantification of the percentage of epithelial RFP+ p63+ cells at day 2 (n = 7) and day 7 (n = 4) of UGS explant cultures (right). Quantification was performed within the boundaries delimited by dotted lines.
(B) Co-immunostaining of RFP, K5, K8 in UGS explants harvested at day 2 (top) and day 7 (bottom). Higher magnification images of proximal (i) and (ii) distal regions are shown for day 7. Arrows show RFP+ K5+ K8+ cells, chevron arrows show RFP+ K5- K8+, arrowheads show RFP+ K5+ K8- cells. Quantification of the percentage of epithelial K5+ K8+, K5- K8+ cells and K5+ K8- cells in the RFP subset at day 2 (n = 4) and day 7 (n = 3) of UGS explant cultures (right). Quantifications were performed within the boundaries delimited in D by dotted lines.
Scale bars: 200 µm (yellow) and 50 µm (white).

Finally, a remaining question was whether RUNX1+ cells contributed to the establishment of the proximal luminal lineage previously identified by scRNA-seq. Since *Krt4* was previously identified as a marker of the developing C2 cluster as well as the adult Lum-D population, immunostainings were also performed in lineage tracing experiments of UGS explants. Interestingly, the fraction of RFP labelled K4+

cells increased from 56.9 ± 10.6% to 74.1 ± 3.0% between day 2 and 7 (Figure 6.16 A). Besides K4, *Nupr1* was also found to be specifically expressed in the C2 cluster, and its expression was evaluated in the UGS explants at the mRNA level by *in situ* hybridisation. As expected, the proportion of RFP+ cells expressing *Nupr1* increased significantly between day 2 and day 7 (Figure 6.16 B).

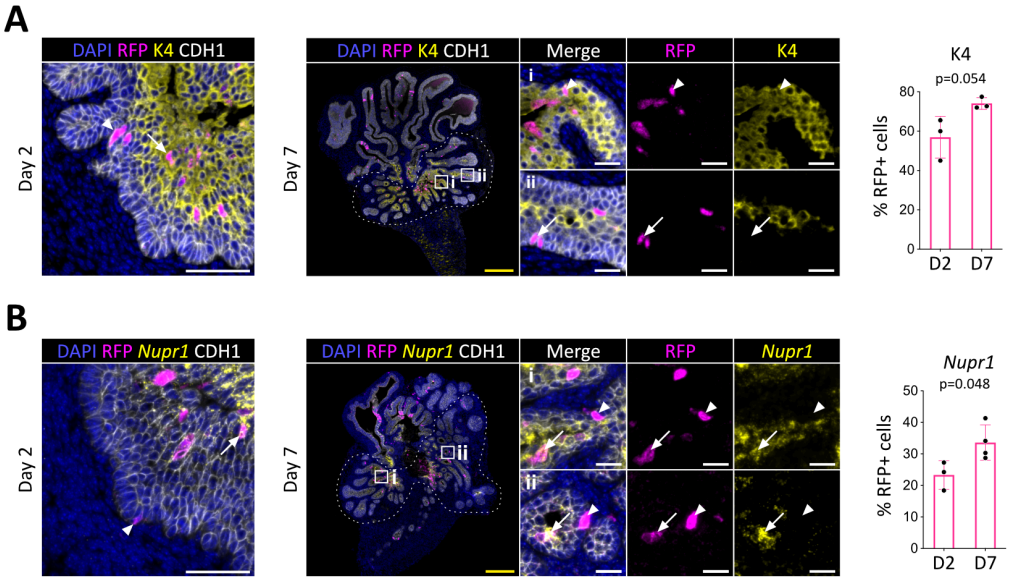

Figure 6.16. RUNX1+ cells contribute to the establishment of the proximal luminal lineage during embryonic prostate development.
(A) Co-immunostaining of RFP, K4, CDH1 in UGS explants harvested at day 2 (left) and day 7 (middle). Higher magnification images of (i) proximal and (ii) distal regions are shown for day 7. Arrows show RFP+ K4- cells, arrowheads show RFP+ K4+ cells. Quantification of the percentage of epithelial K4+ cells in the RFP subset at day 2 (n = 3) and day 7 (n = 3) of UGS explant cultures (right). Quantification was performed within the boundaries delimited in I by dotted lines.
(B) Co-immunostaining of RFP, *Nupr1* (mRNA), CDH1 in UGS explants harvested at day 2 (left) and day 7 (middle). Higher magnification images of (i) proximal and (ii) distal regions are shown for day 7. Arrows show RFP+ *Nupr1*+ cells, arrowheads show RFP+ *Nupr1*- cells. (right) Quantification of the percentage of epithelial *Nupr1*+ cells in the RFP subset at day 2 (n = 3) and day 7 (n = 4) of UGS explant cultures (right). Quantification was performed within the boundaries delimited in G by dotted lines.
Scale bars: 200 µm (yellow) and 50 µm (white).

Taken together, results of lineage tracing experiments performed in UGS explant cultures show that only a small subset of *Runx1* expressing cells contributes to the expansion of NKX3.1+ lineage found in the distal region of the developing

prostatic buds. Instead, the majority of *Runx1* expressing cells preferentially remain in the proximal region of the premature buds, where the proximal luminal lineage is established.

6.3. Discussion

6.3.1. Summary

In previous chapters, RUNX1 was shown to be strongly expressed in proximal luminal cells (PLCs), an intrinsic castration resistant luminal lineage enriched in the peri-urethral region of the adult mouse prostate. In light of the extensive contribution of RUNX transcription factors to developmental processes (Mevel et al., 2019), this raised the question of whether RUNX1 was expressed during prostate development, and if the RUNX1+ proximal luminal lineage was already emerging during prostate development. In embryos, UGS explant cultures, and during early postnatal period, *Runx1* marked a luminal subset located in the most proximal region of the prostate buds. Consistent with its expression pattern in the adult, RUNX1 was also largely mutually exclusive with NKX3.1, except at the very onset of prostate specification where the two transcription factors were found transiently co-expressed. Further characterisation of prostate development using scRNA-seq and genetic lineage tracing of *Runx1* expressing cells indicated that RUNX1+ PLCs emerge at the onset of embryonic prostate specification to populate the proximal region of the ducts (Figure 6.17).

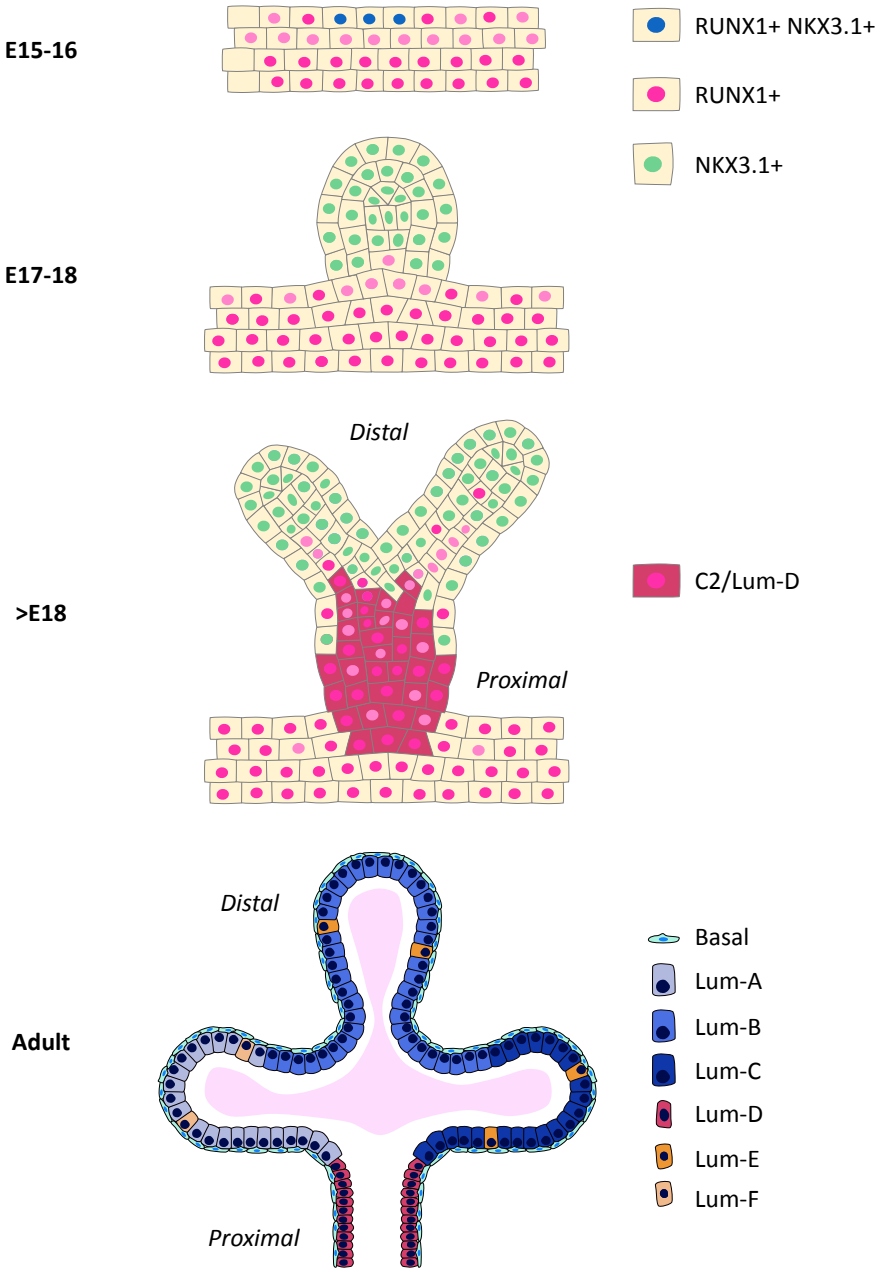

Figure 6.17. Working model showing the emergence of the proximal luminal lineage during embryonic prostate development.

At E15-E16, prostate budding initiates in a subset of cells marked by RUNX1 and NKX3.1. Then, RUNX1 and NKX3.1 mark 2 distinct cellular compartments, respectively in the proximal and distal section of the ducts. During embryonic prostate development, RUNX1+ proximal cells (corresponding to the C2 cluster by scRNA-seq) specify into the proximal luminal lineage, later found in the adult peri-urethral area (corresponding to the Lum-D cluster by scRNA-seq in Chapter 5).

6.3.2. The proximal region of the developing and adult prostate expresses high RUNX1 levels

At E15.5, before the onset of prostate development, RUNX1 was expressed in the stratified UGE, mainly in the inner layer marked by K8 expression. At E18.5, the early specified subset of prostatic epithelial cells expressed high levels of NKX3.1 and p63, but low levels of RUNX1. The dichotomy between RUNX1 and NKX3.1 was also evident at postnatal day 14, after most branching events had already occurred (Sugimura et al., 1986a; Tika et al., 2019). Similar observations were made in UGS explant cultures, a model which proved suitable to study *ex vivo* the early phases of prostate development. Interestingly, a transient state characterised by RUNX1 and NKX3.1 co-expression was evidenced at day 1 of UGS explant cultures. The significance of this observation would need to be further investigated. Indeed, this raises the question of whether both factors are cooperating, antagonising each other, or whether their expression and functional roles are linked to one another. It would be of interest to investigate whether such phenomenon also occurs *in vivo* around E16-E17, or whether this was specific to UGS explant cultures. Also, co-expression between *Runx1* and *Nkx3-1* mRNA could not be evidenced by scRNA-seq in the UGS explant cultures dataset. This could be due to differences between mRNA and protein stability, or because sample collection was carried out either too early or too late to identify this transient state. Alternatively, robust identification of RUNX1+ NKX3.1+ cells may require a greater number of cells captured, or scRNA-seq may be too noisy by nature to detect this feature. Overall, the abundance of RUNX1+ cells in the peri-urethral region and NKX3.1+ cells in the distal region of the ducts suggest that these transcription factors already mark distinct cellular compartments from early embryonic prostate development. However, it should be noted that prostate epithelial hierarchy cannot solely be explained by the expression of these factors. Indeed, as demonstrated by scRNA-seq, subpopulation of cells seems to exist within each *Runx1*-high and *Nkx3-1*-high subsets.

6.3.3. The proximal luminal lineage is established at the onset of prostate development

UGS explant cultures were used to model the initial stages of prostate development, and scRNA-seq of UGS explants evidenced the progressive cellular diversification taking place in these cultures. At such early stages, definitive cell types are yet to be established, and the overall majority of cells display a basal-like signature. The broad basal identity supports the presence of multipotent basal progenitors during embryonic development (Ousset et al., 2012), switching to unipotency postnatally (Tika et al., 2019). In contrast, a universal luminal progenitor could not be clearly identified. Instead, the identity of the C2 cluster correlated with the signature of the Lum-D population found in adult prostates, suggesting an early branching event towards the proximal luminal fate at the onset of prostate development. It is therefore tempting to speculate that adult Lum-A, Lum-B, and Lum-C derive later from multipotent-basal progenitors, while Lum-D already emerged during embryonic specification. These results would need to be validated *in vivo*, but they support strong intrinsic lineage differences existing as early as prostate development begins. Also, since the adult Lum-D population is not committed to prostate secretion and resembles the embryonic C2 cluster, Lum-D cells may retain a more embryonic-like program, which could relate to their intrinsic castration-resistant potential and have broader relevance to cancer treatment. Subsequent lineage tracing experiments indicated that *Runx1* expressing cells preferentially populate the emerging proximal luminal identity. However, beyond the restricted spatial localisation of RFP+ marked cells in mature explants, it is not possible to directly conclude about their long-term fate *in vivo*. For example, it is possible that these cells would eventually populate the medial or distal region of the mature prostate ducts. Therefore, it would be interesting to map the long-term fate of *Runx1* expressing cells. For this, pregnant females from the *Runx1*^{CreER} mouse line could be injected with tamoxifen, and newborn and adults harvested to analyse the spatial location of the lineage marked cells, as well as their contribution to the different prostate lineages. In the future, it would also be interesting to investigate the contribution of the proximal lineage during prostate development,

adult homeostasis and regeneration of the epithelium. This could notably be studied upon depletion of the proximal lineage, for example the *Runx1*^{CreER} mouse line could be crossed with the ROSA26-eGFP-DTA mouse line (Ivanova et al., 2005) to perform diphtheria toxin mediated cell depletion (Saito et al., 2001).

6.3.4. RUNX1 may be functionally involved in mouse prostate development

Since RUNX transcription factors are implicated in various developmental processes (Mevel et al., 2019), including mammary gland development (Browne et al., 2015; Riggio and Blyth, 2017), the detection of RUNX1 expression during prostate development suggests that it may be functionally involved in prostate development and lineage commitment. However, the embryonic lethality of *Runx1*-null embryos by E11.5-E12.5 prevents from directly investigating the requirement for RUNX1 during prostate development (Okuda et al., 1996; Wang et al., 1996). In the absence of phenotypic evidence, the data discussed below were not included in the results section.

Conditional deletion of *Runx1* using the *Pb-Cre4 Runx1*$^{Flox/Flox}$ mouse model failed to generate defects in the prostate epithelium (data not shown). One possible explanation for the absence of phenotype is that deletion occurs in a specific lineage that do not require nor expresses *Runx1*. Indeed, expression of the *Pb-Cre4* transgene has previously been shown to start during early postnatal period in the distal tips of the ducts, which are not expressing *Runx1* (Wu et al., 2001). Furthermore, *Pbsn* expression in the adult is elevated in Lum-B and Lum-C *Nkx3-1*-high clusters, whereas *Runx1* expression is low or absent in these cells (Chapter 5).

An alternative approach was tested using *Runx1*$^{flox/flox}$ Rosa26$^{Cre\text{ERT2/LSL-YFP}}$ mice, which consists in tamoxifen inducible *Cre* expression to knockout *Runx1* and activate YFP expression. UGS explants derived from these mice were grown in the presence of tamoxifen, but histological examination of these explants did not reveal any phenotypic alterations by day 7 of culture (data not shown). In this model, the low recombination efficiency due of the *Cre*ERT2 system could be an important limitation. For instance, the restricted number of cells in which the deletion of

Runx1 takes place may be too small to lead to detectable developmental defects. The possible disadvantage of these cells to contribute to prostate organogenesis may be compensated by the remaining untargeted *Runx1*-WT cellular fraction. Another possible limitation is the duration of the culture (7 days), which may be insufficient to exhibit strong lineage potential defects. This could potentially be overcome by transplanting mature explants into the subrenal capsule of recipient mice to allow further *in vivo* maturation of the explants. Alternatively, it would be interesting to perform scRNA-seq on YFP+ *Runx1*-KO versus YFP+ *Runx1*-WT cells in mature explants. This may reveal subtle differences in cell fate potential.

In an effort to circumvent these issues, mouse Embryonic Stem Cells (mESC) lines were generated with deletions of either *Runx1*, *Runx2*, or *Runx1* and *Runx2*. These cells were derived from the mTmG mouse model (Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J; The Jackson Laboratory) which constitutively express *tdTomato* on their membranes. These *tdTomato*+ mESC lines can then be used to generate chimaeras and analyse the contribution of *tdTomato*+ cells to the establishment of the different lineages of the prostate epithelium. With the help of the Transgenic Production Facility at CRUK MI, a pilot experiment was conducted with mTmG mESCs either *Runx1*-WT or *Runx1*-KO injected in *Runx1*-WT blastocysts. Chimaeras were harvested at postnatal day 14 to evaluate the percentage of *Runx1* KO mESCs *tdTomato*+ cells contributing to the different cellular compartments of the prostate in comparison to *Runx1* WT mESCs. Unfortunately, only 1 male was obtained in a pilot experiment, with less than 5% contribution from *tdTomato*+ cells, suggesting that additional optimisation would be required.

6.3.5. Limitations associated with scRNA-seq of UGS explant cultures

The UGS explants scRNA-seq dataset presented in this Chapter provides the first single-cell transcriptomic resource to study early prostate development. Yet, the general limitations associated with scRNA-seq approaches described in Chapter 5 (section 5.3) also apply to the dataset generated in this Chapter. In addition to these, the *ex vivo* nature of the UGS explant culture system limits the interpretations that can be formulated. While progressive changes in cell identity

could be evidenced at the transcriptomic level, it is virtually impossible to deconvolute the effect of the culture conditions, which should mainly be present between day 0 (E15.5) and day 1. Therefore, the transition between day 0 and day 1 may not accurately represent the evolution occurring between E15.5 to E16.5 *in vivo*. Also, as previously described, scRNA-seq consists of a "snapshot" view of cell states. Here, sequential time points were used in order to partially overcome the lack of continuity in the data and attempt to infer the transitional states occurring during the emergence of prostate lineages. To this aim, force-directed projection was preferred to UMAP as the latter gave a homogeneous representation of the data. Diffusion pseudotime could approximate the real time points (figure 6.10 A-B) but could not be used in itself to identify trajectories. Force-directed projection and PAGA representations suggested the existence of branching clusters, but it appeared particularly difficult to model possible scenarios using Slingshot, a package which often ranks among the best to infer cell lineages and pseudotime from single-cell gene expression data (Street et al., 2018). Overall, the analyses conducted using the scRNA-seq dataset did not permit the reconstruction of a full lineage hierarchy. The use of barcoding lineage tracing systems may be an alternative approach of interest to resolve these issues. It is also possible that the culture conditions do not allow extensive cellular differentiation, since most cells appeared to display a "basal-like" gene signature. The presence of more differentiated cell types should allow easier deconvolution of lineage trajectories. Attempts were made to integrate the adult dataset (Chapter 5) and the UGS explant cultures dataset (data not shown). However, the absence of common cell types between the 2 datasets, and the strong intrinsic difference between *in vivo* and *ex vivo* contexts rendered integration virtually impossible and prone to computational artefacts. In the future, it would be of particular interest to collect urogenital systems isolated from sequential developmental time points from embryos to adults. This should provide a more comprehensive view single cell landscape of prostate organogenesis and facilitate inference of cell lineage specification.

Chapter 7 Characterisation of RUNX1 expression in human patient samples

7.1. Introduction

Prostate Cancer (PCa) is the most commonly diagnosed cancer in men (Cancer Research UK, 2020). The increase in life expectancy and the widespread development of PCa diagnosis tests has been associated with a rise in PCa incidence, and it is estimated that 1 in 8 men will be diagnosed with PCa in their lifetime (Prostate Cancer UK, 2019). Overall, the survival rate of PCa is better than for other cancers, in part because a large proportion of prostate tumours remain latent or can be effectively managed via curative treatments (Litwin and Tan, 2017). Yet, the over-diagnosis of localised, low-grade, indolent tumours represents a central clinical challenge associated with early diagnosis of PCa leading to unnecessary and potentially harmful therapy (Donovan et al., 2016). Indeed, it remains particularly challenging to discriminate with accuracy between indolent and more aggressive forms of the disease leading to the development of lethal metastatic castration-resistant PCa. In PCa patients, ADT is often used as an adjunct to radiation therapy to target tumour cells which rely on androgens for their growth. However, despite the therapeutic benefits of targeting AR signalling, disease recurrence and resistance to ADT can develop after a few months or years. Understanding the origin of these resistant cells and the mechanisms underlying the acquisition of resistance are key to improve PCa treatment modalities. In this context, there is a need to identify better prognostic biomarkers to ameliorate patient stratification.

The human prostate is a walnut-sized gland located at the base of the bladder and surrounding the urethra. Unlike the mouse prostate, which is organised in anatomically distinct pairs of lobes, the human prostate has been divided into the central, transitional and peripheral zones, as well as the anterior fibromuscular stroma (see section 1.1.1). To date, there is no consensus on the relationship

between the zonal anatomy of the human prostate and the lobular anatomy of the mouse counterpart (Ittmann, 2018; Ittmann et al., 2013; Shappell et al., 2004). Histologically, the prostate of both species shares a similar ductal organisation and cell types, but the human prostate has a higher ratio of basal to luminal cells, and a denser architecture due notably to the presence of a thicker stromal layer (McNeal, 1981; Selman, 2011; Shappell et al., 2004). Despite similarities between the structure of the mouse and human prostate glands, the precise cellular composition of the human prostate remains poorly understood (Toivanen and Shen, 2017). More specifically, while basal and luminal cells have been described to be the main cell types of the prostate epithelium of the two species, it is still unclear whether the different subtypes of basal and luminal cells of the mouse prostate epithelium have a direct equivalent in the human setting. Recent advances in single cell transcriptomics have enabled to explore possible correspondence between cells of both species, but more work is required to reach a consensus (Crowley et al., 2020; Henry et al., 2018; Joseph et al., 2020; Karthaus et al., 2020). Overall, to facilitate translational research and deepen our understanding of prostate biology and pathology, it is important to evaluate whether markers and regulators of specific mouse prostate populations have a human equivalent, and vice versa. This is particularly challenging when characterising expression in normal homeostatic conditions since patient samples are typically obtained from diseased organs, and relatively old men in the case of prostate research.

In the mouse prostate, RUNX1 was found to be enriched in a luminal subset localised in the peri-urethral region. This distinct cell population encompasses self-sustaining properties and is not committed to the secretory function of the prostate. However, the relevance of this population to the human prostate and prostate cancer has not been evaluated. Interestingly, RUNX1 is expressed in foetal human prostatic ducts, suggesting that it may have similar expression dynamics as in mouse prostate development (Cunha et al., 2018). Another study reported RUNX1 expression in PCa cell lines and patient samples, and that it is a downstream target of AR signalling (Takayama et al., 2015). Yet, the precise expression pattern

of RUNX1 in the human prostate, and whether it mirrors its expression in the mouse prostate, remains unknown.

The objective of the work carried out in this chapter was to comprehensively characterise the expression pattern of RUNX1 in the human prostate, including normal and tumour samples. This was achieved by staining a wide range of human prostate tissue samples available for research purposes via the biobank of the Christie Hospital. Next, the biomarker potential of *RUNX1* was evaluated using two independent cohorts of samples available in TMA blocks previously generated by the GUCR group (University of Manchester). Finally, genomic alterations and mRNA expression levels of *RUNX1* were explored in publicly available datasets of PCa.

7.2. Results

7.2.1. Validation of a human RUNX1 antibody

In order to investigate the expression of RUNX1 in human patient samples, an anti-human RUNX1 (hRUNX1) antibody was validated using prostate cancer cell lines. First, the hRUNX1 antibody (#4336, Cell Signalling) was validated on 293T cells overexpressing the two main human transcriptionally active isoforms hRUNX1B and hRUNX1C. The two isoforms were detected by western blot at the expected size (around 50 kDa) with hRUNX1C migrating at a slightly higher molecular weight than RUNX1B due to the presence of 33 extra amino-acids at the N-terminus (Figure 7.1A).

To select an appropriate cell line for knockdown experiments and validate the specificity of the antibody, the expression of RUNX1 was investigated in human prostate cell lines by western blot (Figure 7.1B) and RT-qPCR (Figure 7.1C). *RUNX2* and *RUNX3* expression levels were also evaluated by RT-qPCR. RUNX1 expresion was higher in the normal immortalised prostate epithelial cell line RWPE-1 and in the castration-resistant PC-3, PC-3/m and DU145 cell lines compared to the androgen-dependent VCaP and LNCaP cell lines. RUNX2 was also expressed in RWPE-1, PC-3 and DU145 cells, while RUNX3 was mainly found in PC-3 and DU145.

Due to its high RUNX1 expression, the PC-3 cell line was selected for knockdown experiments.

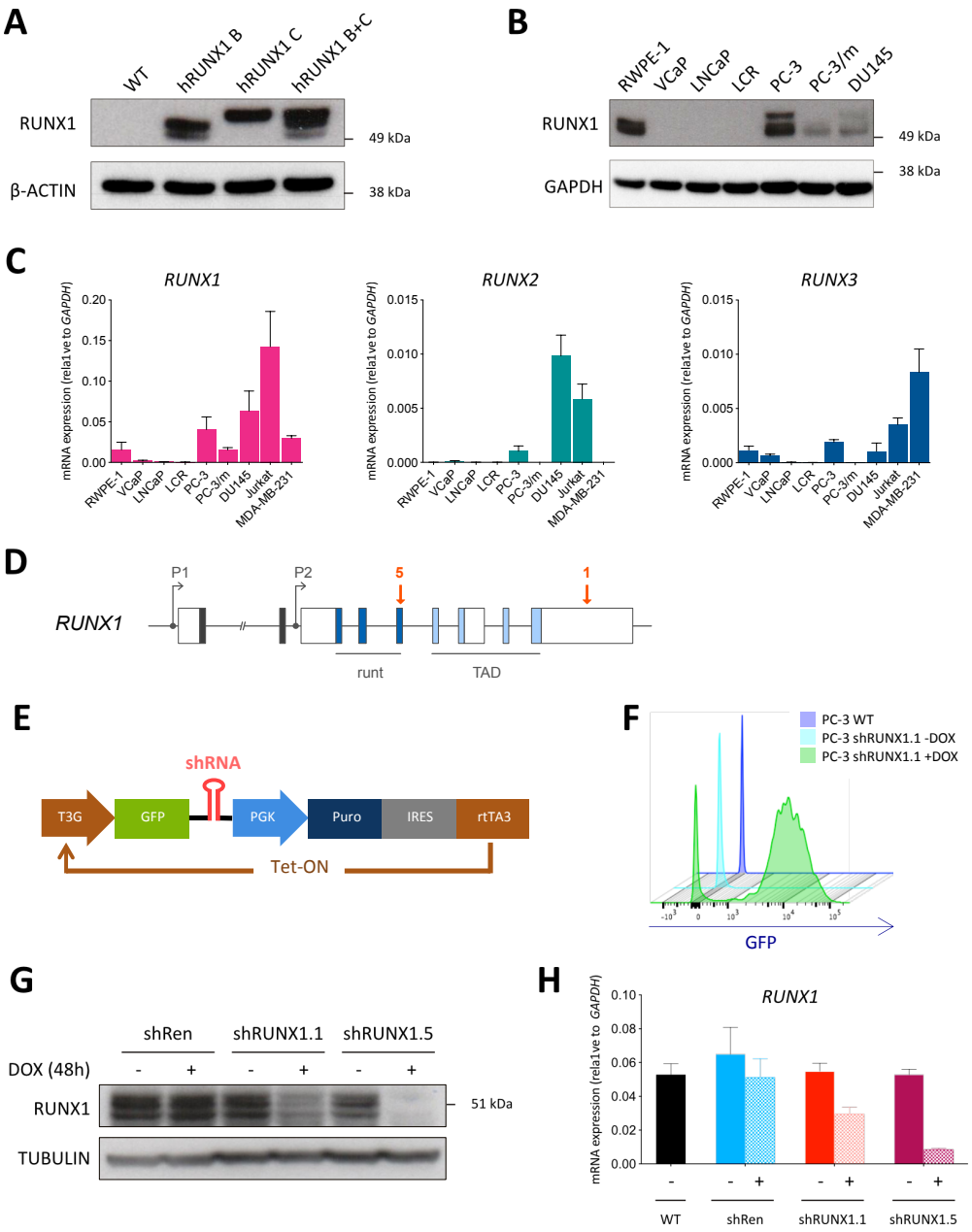

Figure 7.1. Validation of a human anti-RUNX1 antibody.
(A) Overexpression of human *RUNX1B* and *RUNX1C* cDNA in 293T cells. Cells were harvested 48h post-transfection.
(B) Western blot of RUNX1 expression in human PCa cell lines using the D33G6 clone (#4336, Cell Signalling).
(C) RT-qPCR of *RUNX1* (left), *RUNX2* (middle) and *RUNX3* (right) expression in human prostate cell lines. The T-cell derived Jurkat cell line was used as a positive control for *RUNX1* and *RUNX3* expression, and the breast cancer cell line MDA-MD-231 was a positive

control for *RUNX2* expression. Error bars indicate standard deviation (n=2 biological replicates of n=3 technical replicates).
(D) Two shRNAs were selected to target distinct regions of *RUNX1* mRNA (1: 3'UTR, 5: exon 5).
(E) Schematic representation of the Tet-regulated LT3GEPIR expression vector expressing GFP-coupled shRNA from a Tet-responsive element promoter (T3G) and rtTA3 expression upon dox treatment.
(F) Flow cytometry analysis of GFP-positive cells 24h after dox treatment in the induced PC-3 shRUNX1.1 line (green) in comparison with the non-induced (light blue) and parental (dark blue) lines.
(G) Western blot of RUNX1 after 48h of dox treatment in control shRenilla (shRen), shRUNX1.1 and shRUNX1.5 PC-3 cell lines.
(H) RT-qPCR of RUNX1 after 48h of dox treatment in control shRenilla (shRen), shRUNX1.1 and shRUNX1.5 PC-3 cell lines. The parental PC-3 (WT) cell line was included as control.

Two shRNA sequences were designed to target distinct regions of *RUNX1* mRNA (Figure 7.1D), in addition to a control shRNA targeting the *Renilla Luciferase* gene (shRen). The three shRNAs were cloned into doxycycline inducible lentiviral vectors that confer puromycin resistance and harbour a GFP cDNA co-expressed upon doxycycline induction (Figure 7.1E). PC-3 cells were infected with viruses produced with the different constructs and expanded under puromycin selection for two weeks. Transduced cell lines were induced with doxycycline which confirmed the expression of GFP after 24h (Figure 7.1F). RUNX1 knockdown was then evaluated by western blot 48h after doxycycline treatment (Figure 7.1G). As expected, no significant reduction in RUNX1 expression could be observed in the shRen control cell line, while shRUNX1.1 and shRUNX1.5 cell lines exhibited respectively moderate and strong RUNX1 knockdown. The extent of the knockdown was further corroborated by RT-qPCR analysis of *RUNX1* expression (Figure 7.1H). Taken together, overexpression and knockdown experiments of human *RUNX1* validated the specificity of the hRUNX1 antibody by western blot (#4336, Cell Signalling). This antibody will be used for all subsequent human based studies.

7.2.2. RUNX1 expression in Benign Prostatic Hyperplasia samples

Benign Prostatic Hyperplasia is a frequently diagnosed enlargement of the prostate, resulting from the hyper-proliferation of the stromal and epithelial

compartments of the transitional zone of the human prostate (McNeal et al., 1988). While BPH is non-malignant, it remains a major clinical complication, often associated with bladder outflow obstruction. It is commonly treated with 5-alpha reductase inhibitor and/or surgery using transurethral resection of the prostate (TURP).

Despite evident discrepancies between the anatomy of the mouse and human prostate (discussed in section 1.1.5), the location of the transitional zone near the urethra suggests that it may be analogous to the mouse proximal region, previously found to be enriched in RUNX1+ NKX3.1- cells. To test whether RUNX1+ NKX3.1- cells were associated with BPH, 16 BPH patient samples obtained after TURP were co-stained for RUNX1, NKX3.1, and a pan-cytokeratin (panCK) marker (Figure 7.2). The samples were selected to cover a wide range of BPH involvement, ranging from 3g to 56g (median: 28g) of prostate tissue resected via TURP.

NKX3.1 was found to be broadly expressed in luminal epithelial cells, whereas RUNX1 was detected at a lower frequency (Figure 7.2 A-C). Similar to previous observations in the mouse, QIBC analyses showed that RUNX1 and NKX3.1 were rarely co-expressed (Figure 7.2 A, C, D). However, neither RUNX1+ or NKX3.1+ compartments correlated with the amount of BPH tissue resected. These results suggest that RUNX1 do not represent a promising marker of BPH, and that BPH involvement does not correlate with the expansion of a RUNX1+ NKX3.1- cell type. Also, the absence of an abundant RUNX1+ NKX3.1- cell type in the transitional zone suggest either that it may not be the human counterpart of the mouse proximal region, or that these BPH specimens did not contain such region.

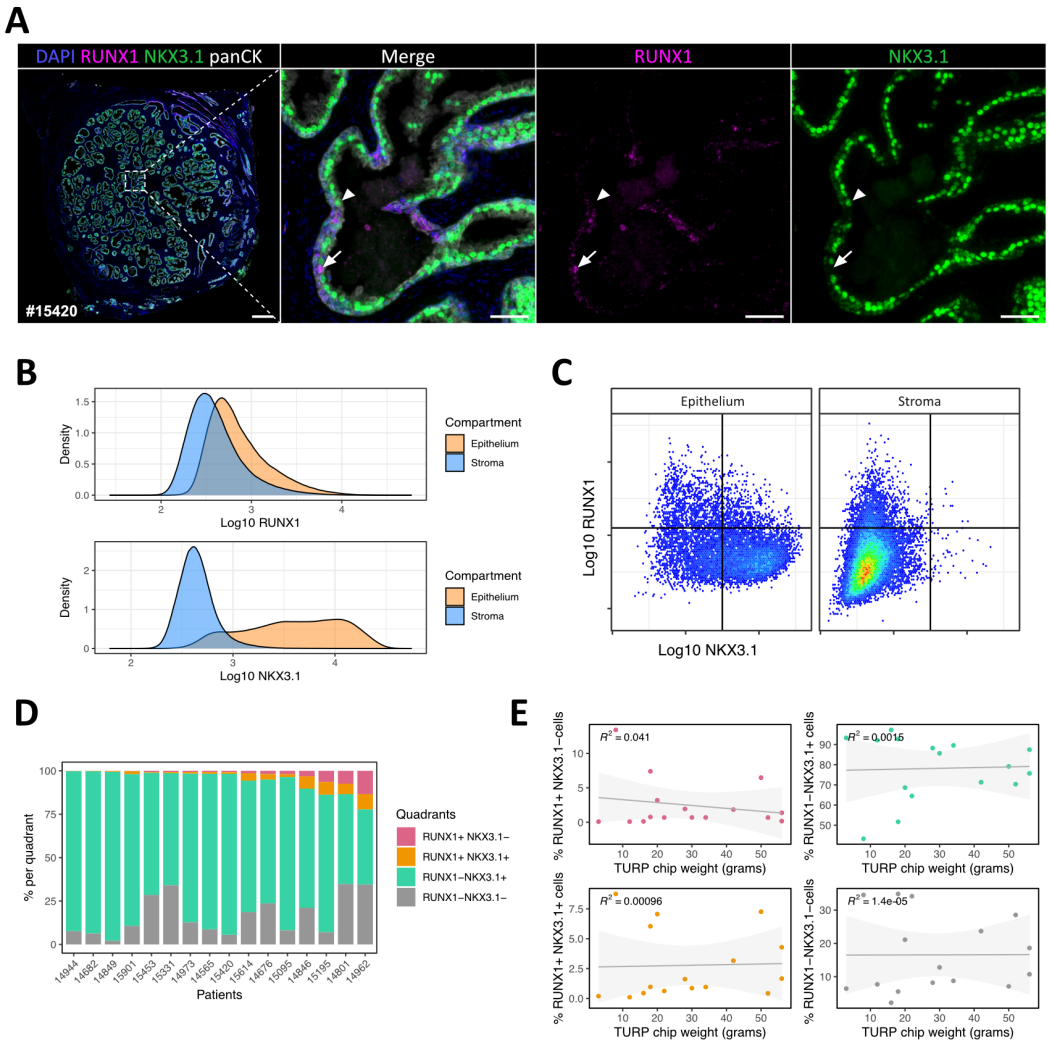

Figure 7.2. RUNX1 and NKX3.1 expression in Benign Prostatic Hyperplasia specimens.
(A) Co-immunostaining of RUNX1, NKX3.1 and panCK in a BPH nodule resected via TURP. Higher magnification images of the selected area are shown. The arrow indicates a RUNX1+ NKX3.1- cell, while the arrowhead indicates a NKX3.1+ RUNX1- cell. Scale bar: 50 µm.
(B) Density plots of RUNX1 and NKX3.1 nuclear intensity (log10 transformed) in the epithelial (panCK+) and stromal (panCK-) compartment of all BPH patient samples combined.
(C) Visualisation of RUNX1 and NKX3.1 nuclear intensity (log10 transformed) in epithelial (panCK+) and stromal (panCK-) cells by QIBC using all BPH patient samples combined.
(D) Quantification of the proportion of RUNX1+ NKX3.1-, RUNX1- NKX3.1+, RUNX1+ NKX3.1+ and RUNX1- NKX3.1- cells in individual BPH patient samples.
(E) Relationship between the proportion of the different cellular compartments quantified in (D) and the TURP chip weights (in grams), which corresponds to the weight of the resected BPH tissue.

7.2.3. RUNX1 expression in whole radical prostatectomy specimens

To further characterise the expression of RUNX1 in the normal human prostate, radical prostatectomy (RP) specimens were obtained from 7 patients diagnosed with localised PCa, low Gleason scores (3+3 and 3+4) and minimal tumour involvement (< 10%). Due to the complexity of qualitatively and quantitatively analyse the expression of RUNX1 in such complex and heterogeneous tissues, 4 distinct histopathological features ("Normal Secretory", "BPH", "Cystic Atrophy" and "Inflammatory Atrophy") were annotated on serial H&E sections of RP specimens by a pathologist (Dr. Pedro Oliveira, Consultant Urological Histopathologist at The Christie Hospital) (Figure 7.3). Of note, cancerous lesions of these patients were not analysed.

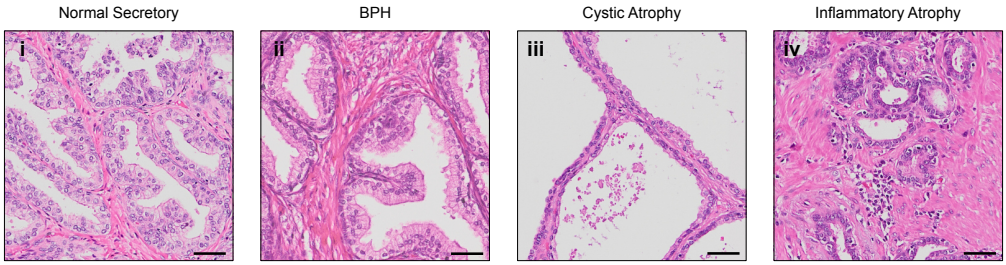

Figure 7.3. H&E staining of representative histopathological features.
The following 4 histopathological features annotated: (i) "Normal Secretory" epithelium, (ii) "BPH" nodules, and regions of (iii) "Cystic Atrophy" and (iv) "Inflammatory Atrophy". Scale bars: 50 µm.

Whole mount serial tissue sections were co-stained with the following 2 panels: (1) RUNX1, NKX3.1 and the high-molecular-weight cytokeratin (HMWCK) marker as well as (2) RUNX1, NKX3.1 and a pan-cytokeratin (panCK) marker. Representative example of staining panel 1 are shown in Figure 7.4. Similar to its expression pattern in mouse prostates, RUNX1 was found in the peri-urethral area (Figure 7.4 B-i), and often detected in the basal epithelium of normal glands (Figure 7.4 A-ii), including regions of basal hyperplasia (Figure 7.4 B-ii). Furthermore, similar to observations made in mice, the majority of luminal NKX3.1+ cells of normal secretory glands did not express RUNX1 (Figure 7.4 A-iii), which was also particularly rare in BPH nodules (Figure 7.4 B-iii). However, unlike RUNX1

expression pattern in specific mouse prostate lobes, RUNX1 expression was detected in all zones of the human organ. Interestingly, higher RUNX1 levels were often associated with prostatic atrophy, especially of inflammatory nature (Figure 7.4 A-iii, B-v). Also, these patches of atrophy with inflammation exhibited reduced NKX3.1 expression, indicating a generally non-overlapping pattern with RUNX1. This result is consistent with reports correlating low NKX3.1 expression to prostate atrophy and prostate inflammation (Bethel et al., 2006; Magnen et al., 2018). Prostatic atrophy is frequently found in human prostates and has been associated with chronic inflammation, variations of androgen levels, however its role in the development of PCa is controversial (Davidsson et al., 2011; De Marzo et al., 1999; Irani et al., 1999; McNeal, 1968; Murphy et al., 1991; Putzi and De Marzo, 2000; Reuter, 1997; Zhang et al., 2019).

Using staining panel 2, QIBC analyses of each annotated histopathological features further demonstrated the enrichment of RUNX1+ NKX3.1- cells in inflammatory atrophy, and very low levels of RUNX1 in BPH nodules (Figure 7.5 A, B). Patients B, C and G had more frequent RUNX1 expression in both normal secretory epithelium and cystic atrophy compared to the other patients, which highlights the existence of inter-patient heterogeneity of RUNX1 expression between similar histopathological features.

Taken together, these results indicate that RUNX1 is frequently detected in basal cells, and also strongly expressed in the luminal layer of atrophic glands of the human prostate. As in mouse prostates, RUNX1 and NKX3.1 were generally found in distinct cellular compartments. The similar expression pattern of RUNX1 in mouse and human prostates, especially in atrophic epithelia, suggests analogous roles in both species.

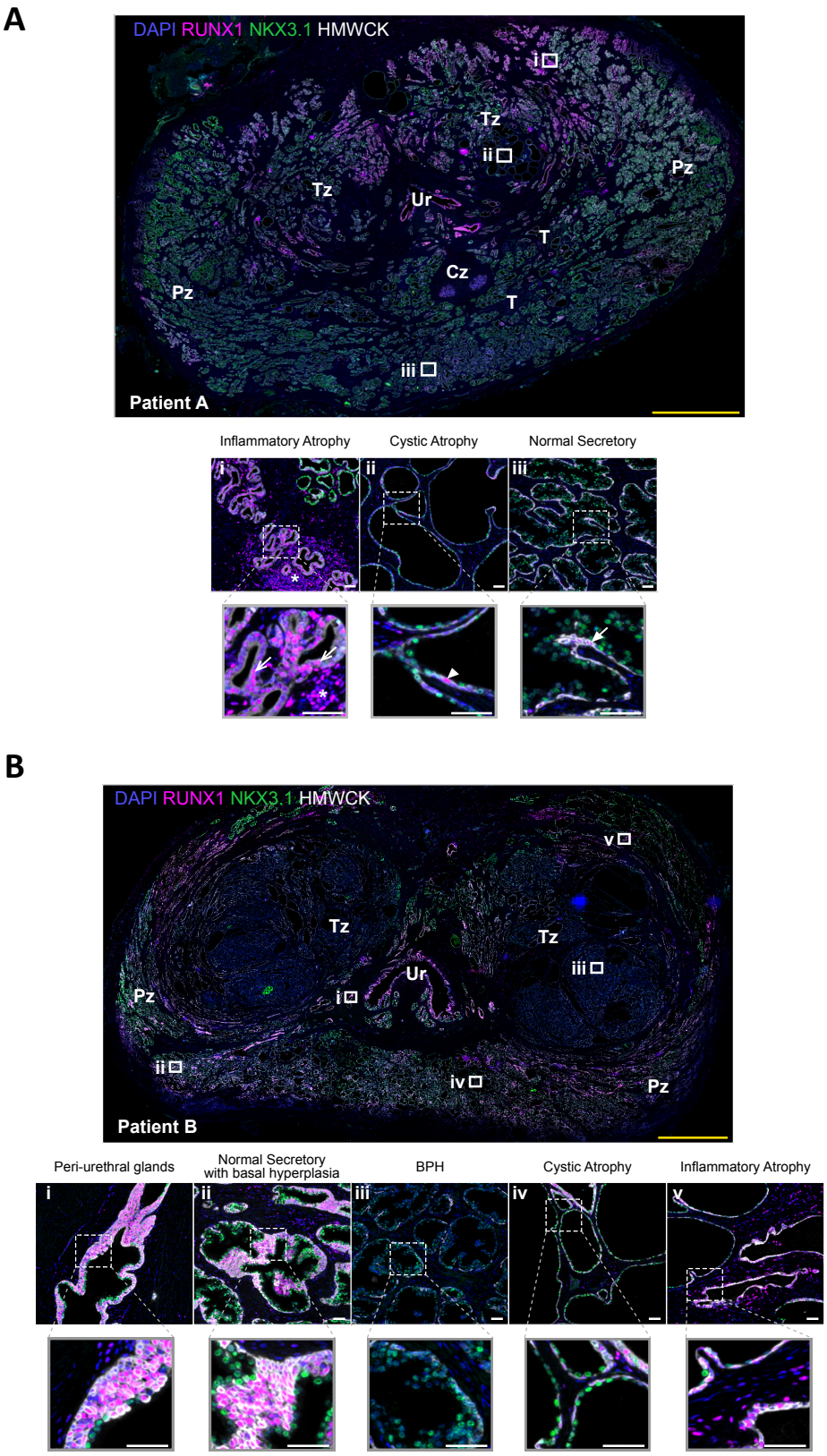

Figure 7.4. RUNX1 is enriched in human atrophic glands.

(A) Co-immunostaining of RUNX1, NKX3.1, HMWCK in a human radical prostatectomy sample (Patient A). Higher magnification images of (i) inflammatory atrophy, (ii) cystic atrophy and (iii) normal secretory epithelium are shown. Arrows indicate example RUNX1 expressing cells.
(B) Co-immunostaining of RUNX1, NKX3.1, HMWCK in a human radical prostatectomy sample (Patient B). Higher magnification images of (i) peri-urethral glands, (ii) normal secretory glands with basal hyperplasia, (iii) BPH, (iv) cystic atrophy and (v) inflammatory atrophy.
Ur: urethra, Cz: central zone, Tz: transition zone, Pz: peripheral zone, T: tumour.
Scale bars: 5 mm (yellow) and 50 µm (white).

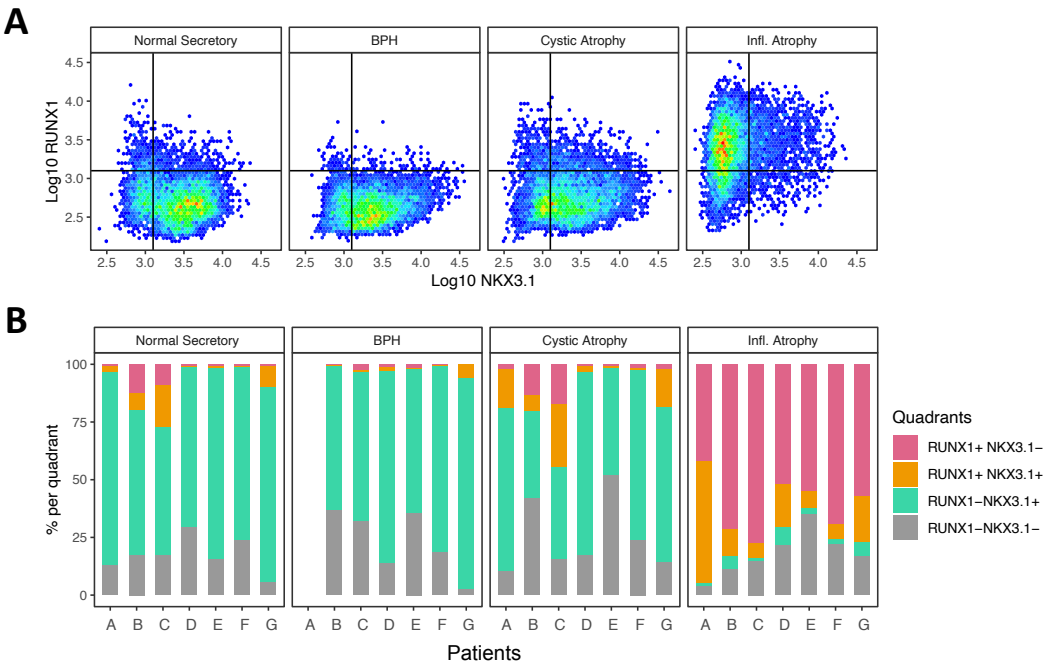

Figure 7.5. QIBC analyses of RUNX1 and NKX3.1 expression in prostate epithelial cells of radical prostatectomy specimens.
(A) Visualisation of RUNX1 and NKX3.1 nuclear intensity (log10 transformed) in epithelial cells (panCK+) by QIBC in each of the 4 histopathological features analysed. Data from all patients were combined and downsampled to achieve equal representation from each individual patient.
(B) Quantification of the percentage of RUNX1+ and NKX3.1+ cells in representative areas (as in Figure 7.3 and 7.4) for 7 human radical prostatectomy samples.

7.2.4. RUNX1 expression in matched biopsies pre- and post-ADT

In Chapter 4, RUNX1 was shown to be highly enriched in the castrated mouse prostate. To test whether RUNX1 was expressed in atrophic prostate glands after ADT, prostate needle biopsies of a PCa patient before and after 6 months of ADT

were co-stained for RUNX1 and NKX3.1 (Figure 7.6 A-B). While RUNX1 was barely detectable in benign untreated secretory glands, it was extensively expressed in benign atrophic glands post-ADT (Figure 7.6 B). The upregulation of RUNX1 in these atrophic glands tallied with reduced levels of NKX3.1. Of note, tumour cells remained NKX3.1+ RUNX1- following ADT. Although only a single patient with limited tissue material from needle biopsies could be evaluated, these results suggest that RUNX1 expression in normal/benign regions may be enriched in human atrophic glands post-ADT, similarly to its upregulation in the castrated mouse prostate. In contrast, tumour cells may respond differently to the treatment.

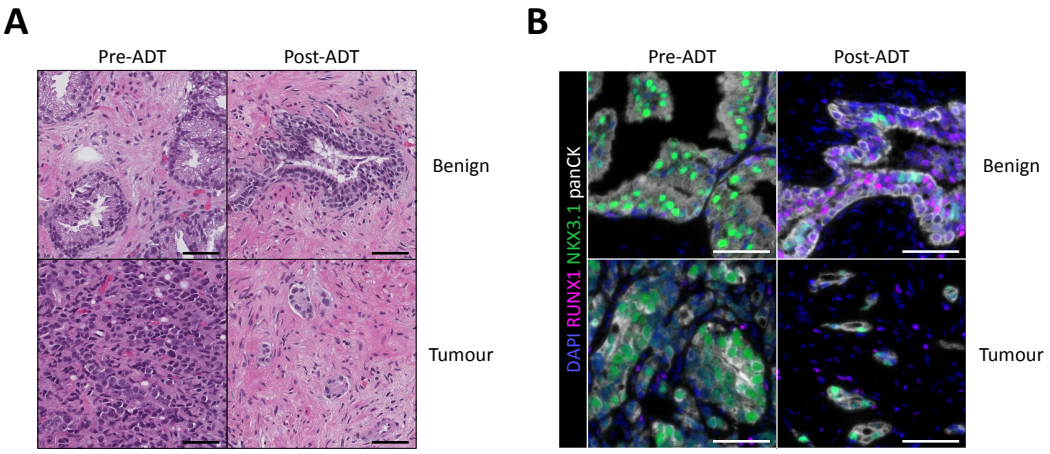

Figure 7.6. Expression of RUNX1 and NKX3.1 in needle biopsies before and after ADT.
(A) Representative H&E images of the benign and tumour areas stained in B, from needle biopsy samples of a patient pre- and post-ADT (treated for 6 months with leuprolide, a LHRH agonist).
(B) Co-immunostaining of RUNX1, NKX3.1 and pan-cytokeratin. Representative areas of normal and tumour glands are shown.
Scale bars: 50 µm.

7.2.5. TMA analysis of matched primary tumours with and without lymph node metastases

The previous set of analyses revealed a heterogeneous expression pattern of RUNX1 in non-malignant human prostate epithelia. To further characterise the extent of inter- and intra-patient heterogeneity of RUNX1 expression, a tissue micro-array (TMA) of normal-adjacent and tumour tissue was constructed from a cohort of radical prostatectomy (RP) patients diagnosed with localised and

metastatic prostate cancer (Table 2.2). Indeed, lymph nodes located near the surgical site are often resected and examined for the presence of metastatic disease during surgery. Patients with lymph node metastases (LNM, LN 1 status) tend to have a less favourable outcome (Boorjian et al., 2007), and their clinical management represents an important challenge that remains actively debated (Gupta, 2020). In keeping with its critical role in regulating the endothelial-to-haematopoietic transition (EHT) in the haematopoietic system, RUNX1 has recently been implicated in epithelial-to-mesenchymal transition (EMT) processes in breast cancer (Hong et al., 2017; Kulkarni et al., 2018; Pegg et al., 2019). However, while RUNX1 is important to initiate and complete the EHT, it appears to rather suppress EMT in breast cancer cells. For these reasons, the expression of RUNX1 was evaluated in patient samples diagnosed with and without LNM.

The pathologist selected multiple areas of normal adjacent and tumour tissue from resected RP specimens. At least 3 replicate cores were punched from each annotated region. To comprehensively characterise the expression pattern of RUNX1, the pathologist annotated histopathological features observed in each TMA core, at the glandular level (Figure 7.7 A). These annotations were then transferred onto the stained samples with RUNX1, NKX3.1 and pan-cytokeratin (Figure 7.7 B). Normal-adjacent tissue was annotated as either “Normal”, “Atrophy” or “Inflammatory (Infl.) Atrophy”, and tumour glands were scored using the Gleason grading system: Gleason (G) 3, 4 or 5 (Figure 7.7 A-C). Of note, due to the limited tissue material available for LNM, these samples were not included in the original TMA construction. Instead, LNM were stained separately on whole tissue sections. After removal of missing and poor-quality cores, 281 single tissue cores (spread out in 3 independent tissue blocks) and 13 LNM sections were kept for analysis (Figure 7.7 C). This dataset represents a heterogeneous and rich resource to study inter- and intra-patient heterogeneity in both normal-adjacent and tumour tissues of varying grades.

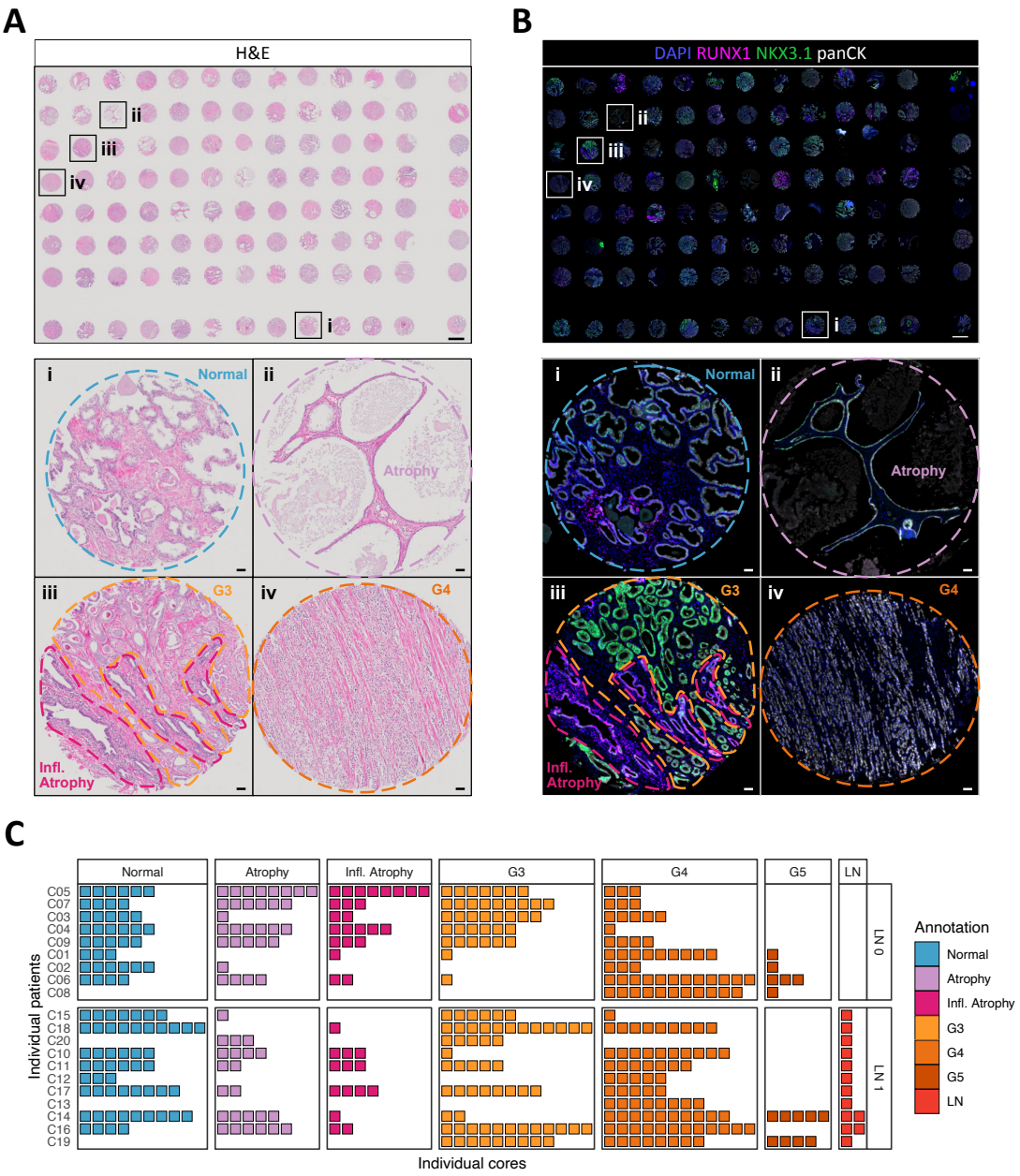

Figure 7.7. TMA of a RP cohort of matched primary tumours with and without lymph node metastases (LNM).

(A) Representative H&E image of a section of the RP TMA, with higher magnification images of selected cores (I to iv) showing example histopathological annotations.

(B) Example fluorescent immunostaining of the RP TMA, co-stained for RUNX1, NKX3.1 and pan-cytokeratin. Higher magnification images of the cores and annotations in A are shown (i to iv).

(C) Core-level tiled map representing the diversity of annotated histopathological features for each patient included in the construction of the RP TMA. Regions of "Normal" epithelium, "Atrophy", "Inflammatory (Infl.) Atrophy", Gleason (G) 3, 4, 5 were present in the TMA, while matched LNM for "LN 1" patients consisted of whole tissue sections. Individual squares represent distinct cores (TMA) or tissue sections (LNM).

LN 0: no LNM at surgery, LN 1: presence of LNM at surgery, G: Gleason. Scale bars: 1 mm (low magnification) and 50 μm (high magnification).

Quantitative image analysis was conducted on all stained samples and restricted to the epithelial (panCK+) compartment. Single-cell expression data was initially explored by histopathological feature after combining single-cells from all patients (Figure 7.8 A-B). Similar to previous analyses, RUNX1+ cells were rare in "Normal" epithelia, slightly more abundant in "Atrophy" and frequently found in "Inflammatory Atrophy". In malignant tissue, RUNX1 was very sparse in primary tumours, and almost absent in LNM. In contrast, NKX3.1 was broadly expressed in the different annotated regions of the prostate epithelium, with higher levels in LNM and lower expression in "Inflammatory Atrophy". These results reinforce the distinct expression pattern between NKX3.1 and RUNX1. Next, single-cell expression measurements were quantified at the patient level to evaluate the degree of inter-patient heterogeneity within the different histopathological annotations (Figure 7.8 C-D). Overall, the same conclusions were observed at the patient level. Both LN 0 and LN 1 patients displayed a similar expression pattern. In inflammatory atrophy, the frequency of RUNX1+ cells ranged from less than 25% to more than 75%. Instead, the proportion of NKX3.1+ cells was consistently low in these regions. In other histopathological annotations, the frequency of NKX3.1+ cells was highly variable. However, this quantification did not take into account possible heterogeneous expression levels within patients. Therefore, it is possible that selected areas, either in normal-adjacent or tumour tissue, display highly variable expression patterns that would be masked at the patient level.

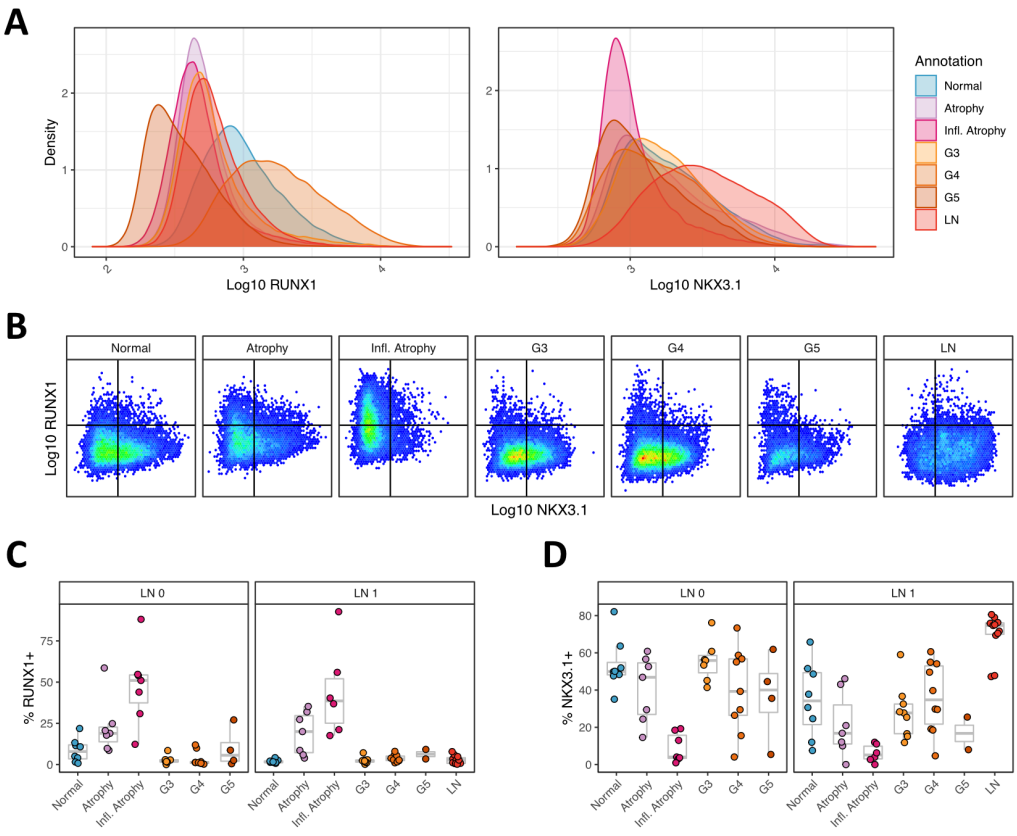

Figure 7.8. Quantitative analysis of RUNX1 and NKX3.1 expression in primary tumours and lymph node metastases.
(A) Density plots of RUNX1 and NKX3.1 nuclear intensity (log10 transformed) in the epithelial compartment of all patients combined, grouped and coloured by histopathological features.
(B) Visualisation of RUNX1 and NKX3.1 nuclear intensity (log10 transformed) by QIBC, in the epithelial compartment of all patients combined, grouped by histopathological features.
(C, D) Quantification of the proportion of (C) RUNX1+ and NKX3.1+ (D) cells in each histopathological annotation for LN 0 and LN 1 patients. Dots represent individual patients after aggregating cells of matching histopathological annotations.
LN 0: no lymph node metastasis at surgery, LN 1: presence of lymph node metastasis at surgery, G: Gleason.

The heterogeneity of RUNX1 and NKX3.1 expression could be better appreciated at the core-level after quantifying the proportion of cells being either RUNX1+ NKX3.1-, RUNX1+ NKX3.1+, RUNX1- NKX3.1+ or RUNX1- NKX3.1- (Figure 7.9). Core-level heatmap representation highlighted the high frequency of RUNX1+ cells in a restricted number of cores, which was previously masked by patient level averages. Along these lines, it is interesting to note that despite being

predominantly lowly expressed in cancerous human prostates, there was also tumour cores with higher RUNX1 levels. Also, a few normal-adjacent and tumour cores had a higher proportion of RUNX1+ NKX3.1+ cells. Of note, a significant proportion of cells had low levels of both RUNX1 and NKX3.1, indicating that these 2 transcription factors are not sufficient to fully describe the cellular heterogeneity of the prostate epithelia.

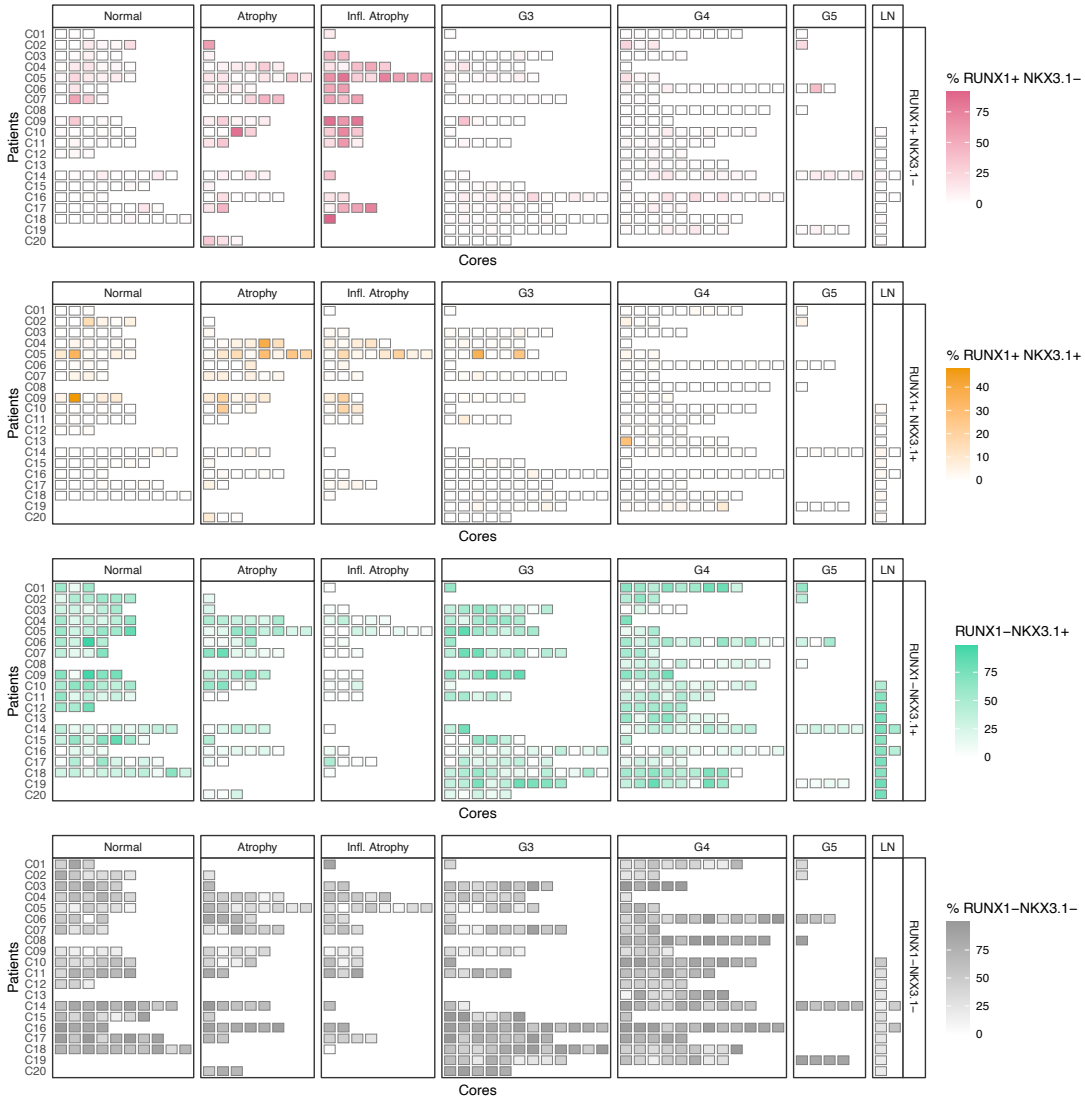

Figure 7.9. Inter- and intra-patient heterogeneity of RUNX1 and NKX3.1 compartments. Core-level heatmap representations of the proportion of RUNX1+ NKX3.1-, RUNX1+ NKX3.1+, RUNX1- NKX3.1+, RUNX1- NKX3.1- cells in primary tumours and lymph node metastases. Individual cores are represented by boxes. The proportion of each cell types in each core is represented by a colour gradient. LN 0: no lymph node metastasis at surgery, LN 1: presence of lymph node metastasis at surgery, G: Gleason.

Overall, these results indicate that neither RUNX1 or NKX3.1 expression can be used to describe intrinsic differences between localised and metastatic disease. All LNM analysed expressed high NKX3.1 and low RUNX1 levels. Instead, high RUNX1 and low NKX3.1 expression was associated with prostate atrophy, in particular when associated with inflammation. Finally, despite being predominantly lowly expressed in cancerous human prostates, there were also rare tumour foci with higher RUNX1 levels.

7.2.6. RUNX1 expression in large PCa TMA cohorts: TURP and needle biopsies

RUNX1 appeared to be rarely expressed in prostate tumours. However, only a restricted number of patient samples had been analysed, and there seem to be a small proportion of tumour foci with higher RUNX1 expression. To evaluate the clinical relevance of RUNX1 expression in PCa patient samples, a selection of tissue micro-arrays (TMAs) of PCa cases diagnosed between 1994 and 2004 in Salford (UK) were obtained from the GUCR (University of Manchester).

7.2.6.1. Clinical characteristics of the TMA cohorts

The material analysed comprised a total of 1189 individual cores from 220 patients, spread over 32 TMA blocks. Among these, 134 patients were sourced from transurethral resection of the prostate (TURP) samples, and 86 originated from transrectal ultrasound guided needle core biopsies (TRUSBx, referred to as biopsies thereafter). Patients of the TURP cohort had a shorter overall survival than the biopsy cohort, with a median of 56 months (95% CI 37–77) for the TURP against 76 months (95% CI 48-115) for the biopsy cohort (Figure 7.10 A). When combined together, patients diagnosed without metastases (M0) had a median overall survival of 90 months (95% CI 76-115) versus 32 months (95% CI 29-48) for patients diagnosed with metastases (M1) (Figure 7.10 B). As expected, shorter overall survival was observed for patients with a higher clinical Tumour (T) stage (Figure 7.10 C), higher Gleason Group (Figure 7.10 D), higher diagnostic PSA (Figure 7.10 E). Also, the majority of the cohort consisted of “High Risk” patients according to the

D'Amico (D'Amico et al., 1998) risk classification system (Figure 7.10 F), a model used to stratify patients into risk groups of biochemical recurrence after surgical treatment. Importantly, the whole cohort had a median overall survival follow-up time of > 16 years, making it suitable for long-term overall survival analyses.

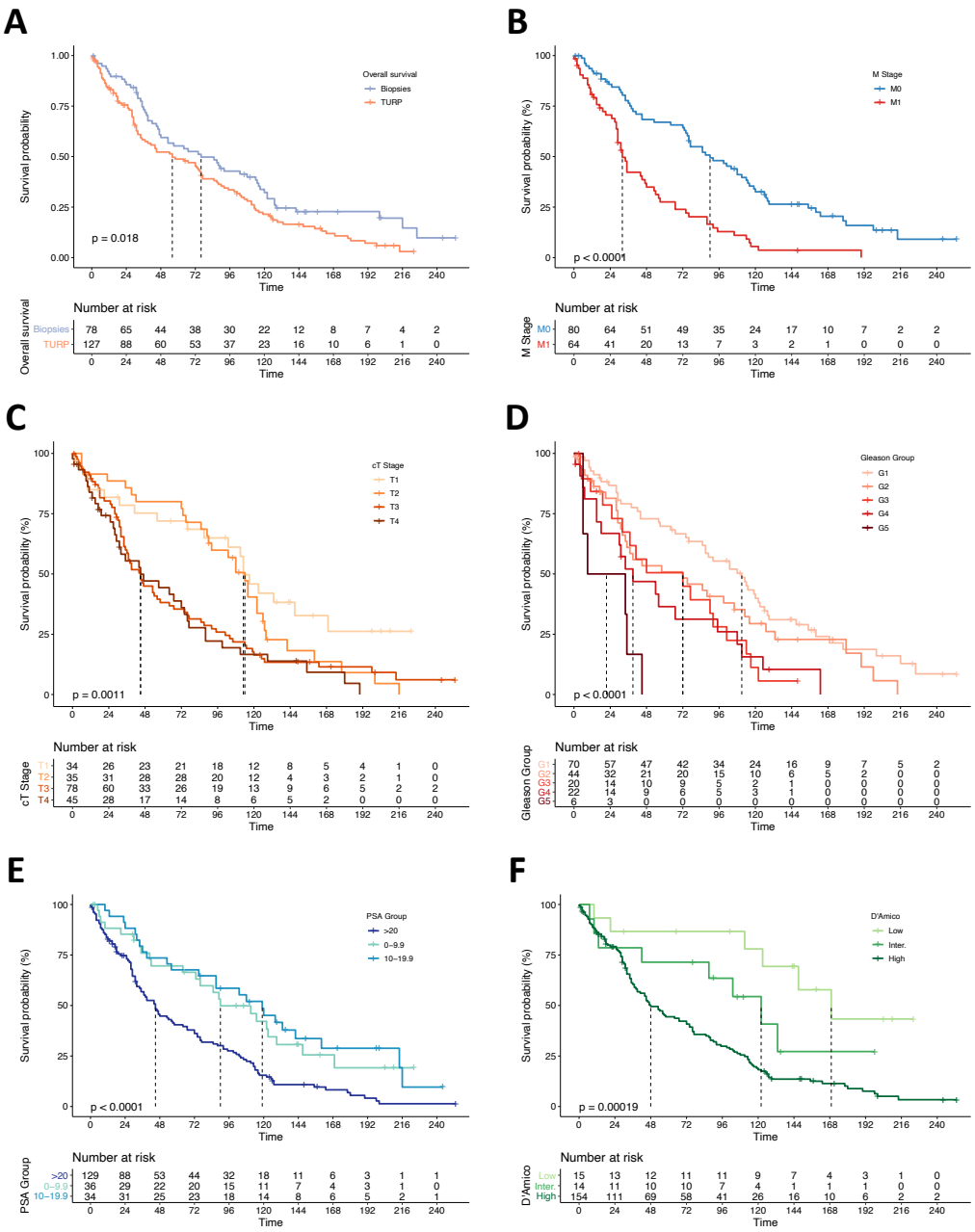

Figure 7.10. Survival analysis of the PCa TMA cohorts.
(A-F) Kaplan-Meier survival analysis of all patients present in the dataset by (A) tissue type, (B) metastasis stage, (C) clinical Tumour (cT) stage, (D) Gleason group, (E) PSA group and (F) D'Amico risk group. Log-rank test p values are indicated on the respective plots.

Clinical parameters available at the time of diagnosis can help improve the translational potential of biomarkers (Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma et al., 2008). To validate the quality of the TMA cohort to be used for prognostic purposes, established clinical co-variates available for this cohort were used in multivariate Cox regression analysis, including the age, log10-transformed PSA, Gleason group, clinical tumour stage, and metastasis stage. An additional variable was included to control for the source material (TURP or biopsy) (Figure 7.11). As expected, increasing age and PSA, clinical tumour stage T3 and T4, Gleason group 5, and presence of metastasis were independently associated with shorter overall survival in multivariate analysis. Of note, patients who underwent TURP were significantly and independently associated with a worse outcome compared with the biopsy cohort.

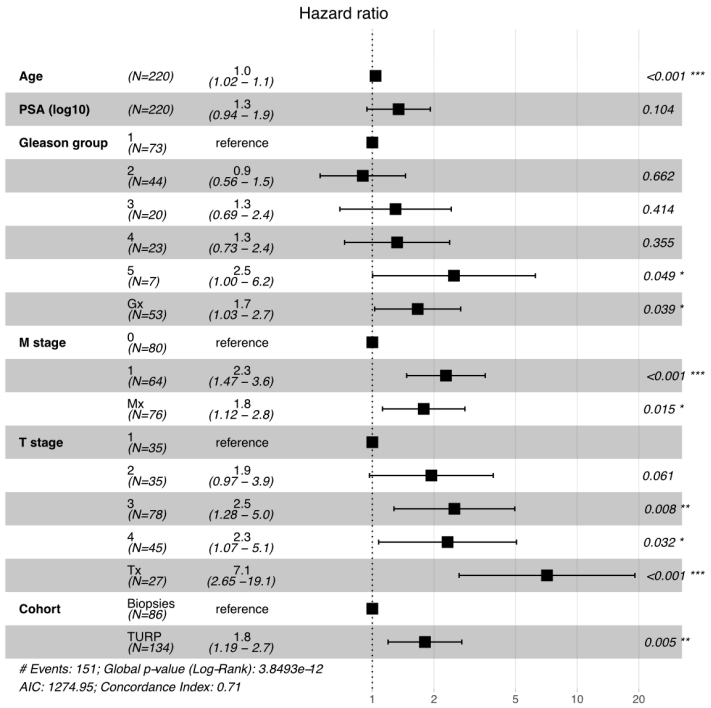

Figure 7.11. Forest plot of the multivariate Cox regression analysis.
The model was adjusted for age, PSA, Gleason group, clinical tumour stage, and presence of metastasis. Hazard ratios (HR) are reported with 95% confidence interval (CI). Gx: missing Gleason group; Tx: missing clinical tumour stage; Mx: missing metastasis stage.

7.2.6.2. Immunostainings of the TURP and biopsy TMA cohorts

In the mouse prostate (Chapters 3 to 6) and in human samples (sections 7.2.2 to 7.2.5), RUNX1 had a largely mutually exclusive expression pattern with NKX3.1,

suggesting that these 2 transcription factors are expressed in distinct epithelial lineages. Therefore, the expression of RUNX1 was evaluated in association with NKX3.1 and a pan-cytokeratin marker using multiplexed immunofluorescence (Figure 7.12). Example of stained TMA cores from TURP and biopsy cohorts with varying frequencies of RUNX1+ and NKX3.1+ cells are presented in Figure 7.12 A-B. Image analysis was performed on HALO (Indica Labs) to extract single-cell measurements. Briefly, the staining pattern of the pan-cytokeratin marker was used together with the nuclear marker DAPI to train a tissue segmentation algorithm to classify epithelial and stromal compartments. Then, single cells were detected using a nuclear detection module based on DAPI to locate each nucleus. Single-cell measurements were then exported for both epithelial and stromal compartments of all TMA cores.

The TURP cohort comprised 561 tumour annotated cores and 449 normal-adjacent cores. The biopsy cohort had less material available, with 167 tumour annotated cores and 12 normal adjacent cores. Patients with no clinical data, and/or having less than 100 epithelial tumour cells available for quantification were removed from the dataset. Also, normal-adjacent cores were excluded from the analysis to specifically focus on tumour-specific marker expression. In total, 718 tumour cores were retained, and stromal cells were excluded from all analyses. The amount of tissue available in TURP and biopsy cohorts was strikingly different (Figure 7.12 A-B). Needle biopsies were obtained using thin needles to extract prostate tissue used for diagnostic purposes, while whole chunks of tissues were resected out during TURP. Consequently, the number of epithelial tumour cells available per TMA cores ranged from 3 to 1,387 cells for the biopsies, and from 37 to 10,004 for the TURP (Figure 7.12 C). Once individual cores had been aggregated at the patient level, the amount of tumour epithelial cells analysed ranged from 101 to 2090 cells for the biopsies, and from 169 to 42,809 for the TURP (Figure 7.12 D).

These observations indicate that TURP and biopsy cohorts are not only heterogeneous due to the origin of tissue sampled, but also in the epithelial content available for analysis. Importantly, both cohorts contained TMA cores with both RUNX1-Hi NKX3.1-Lo and RUNX1-Lo NKX3.1-Hi tumours.

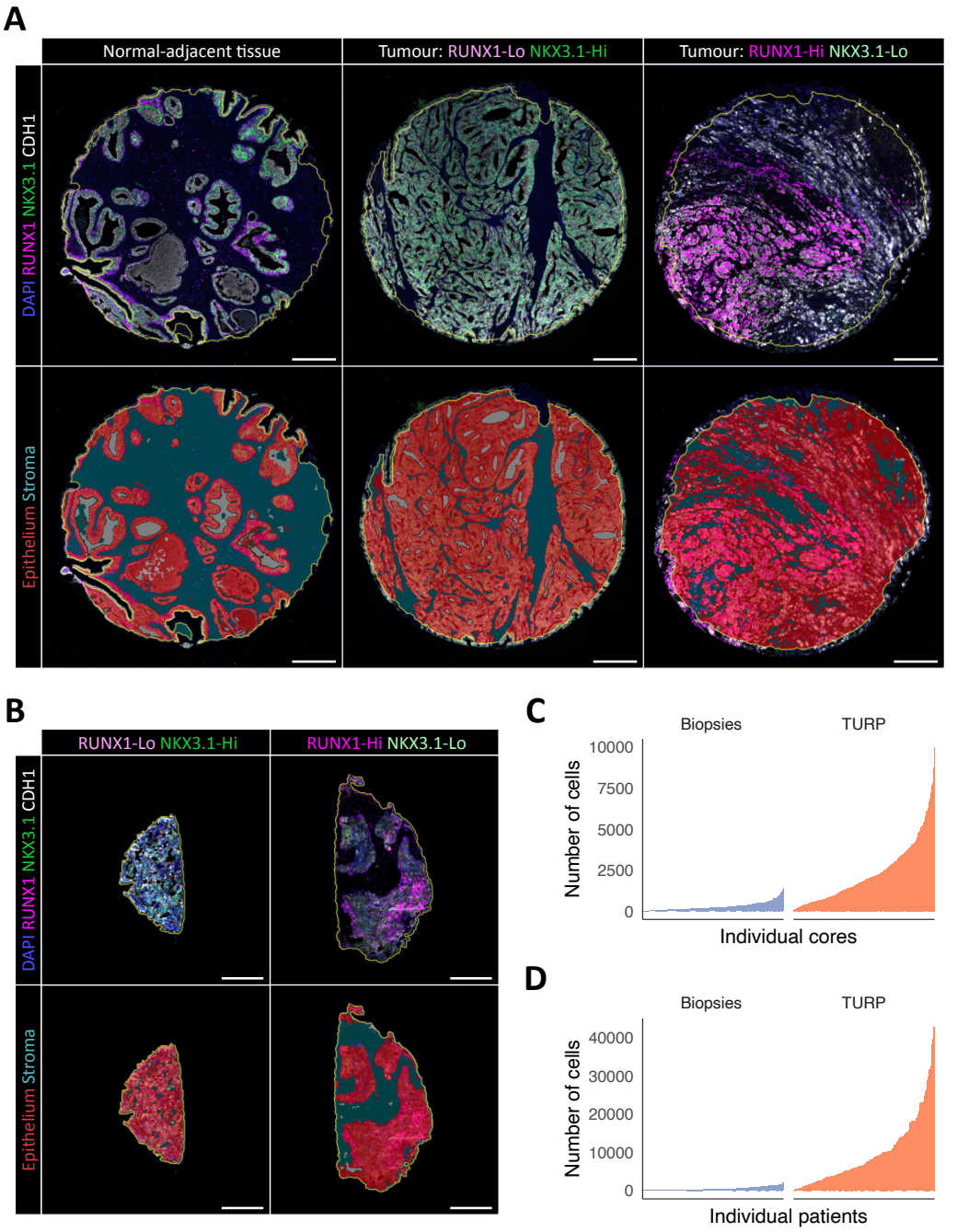

Figure 7.12. Image analysis of TURP and biopsy TMA cohorts.
(A) Example TURP TMA cores co-stained for RUNX1, NKX3.1 and pan-cytokeratin (top). A normal-adjacent core (left), and tumour cores with RUNX1-Lo NKX3.1-Hi (middle) and RUNX1-Hi NKX3.1-Lo (right) are shown. Corresponding tissue segmentation into "Epithelium" and "Stroma" compartments is shown (bottom).
(B) Example biopsy TMA cores co-stained for RUNX1, NKX3.1 and pan-cytokeratin (top). Tumour cores with RUNX1-Lo NKX3.1-Hi (left) and RUNX1-Hi NKX3.1-Lo (right) are shown. Corresponding tissue segmentation into "Epithelium" and "Stroma" compartments is shown (bottom).

(C) Waterfall plot showing the number of single epithelial tumour cells analysed per individual cores for the biopsy (left) and TURP (right) cohorts.
(D) Waterfall plot showing the number of single epithelial tumour cells analysed per individual patient for the biopsy (left) and TURP (right) cohorts.
Scale bars: 200 µm.

7.2.6.3. Correlation between RUNX1 abundance and clinical characteristics in human PCa TMAs

To evaluate the biomarker potential of RUNX1 in the two PCa TMA cohorts, the proportion of RUNX1+ cells was computed for each individual core. For each patient, the core with the highest proportion of RUNX1+ cells was used to rank patients by RUNX1 abundance. The upper quartile (75th percentile) was defined as the RUNX1-Hi group and compared with remaining RUNX1-Lo fraction. This approach was performed on individual cohorts either separately or after merging patients from both cohorts to improve statistical power. Core-level representation of the percentage of RUNX1+ cells in individual tumours of merged TURP and biopsy datasets confirmed that the RUNX1-Hi group adequately captured those patients with higher RUNX1 levels, while the majority of the tumour cores expressed low or no RUNX1 (Figure 7.13 A). When patients from both cohorts where pooled together, there was a trend suggesting shorter overall survival for RUNX1-Hi patients (Figure 7.13 B). In the TURP, but not in the biopsy cohort, Kaplan–Meier survival analysis showed that RUNX1-Hi patients had a significantly shorter overall survival than RUNX1-Lo patients (log-rank p value = 0.024) (Figure 7.13 C-D). RUNX1-Hi patients had a median overall survival of 34 months (95% CI 19-72) against 74 months (95% CI 45-99) for the RUNX1-Lo group.
The abundance of RUNX1+ cells was then evaluated against diagnostic clinical variables. Beyond a weak correlation with Gleason groups 2 and 3 (p = 0.028, Figure 7.13 E), no association was found between the frequency of RUNX1+ cells and clinical tumour stage (Figure 7.13 F), metastasis stage (Figure 7.13 G), D'Amico risk group (Figure 7.13 H), age (Figure 7.13 I) or PSA (Figure 7.13 J).

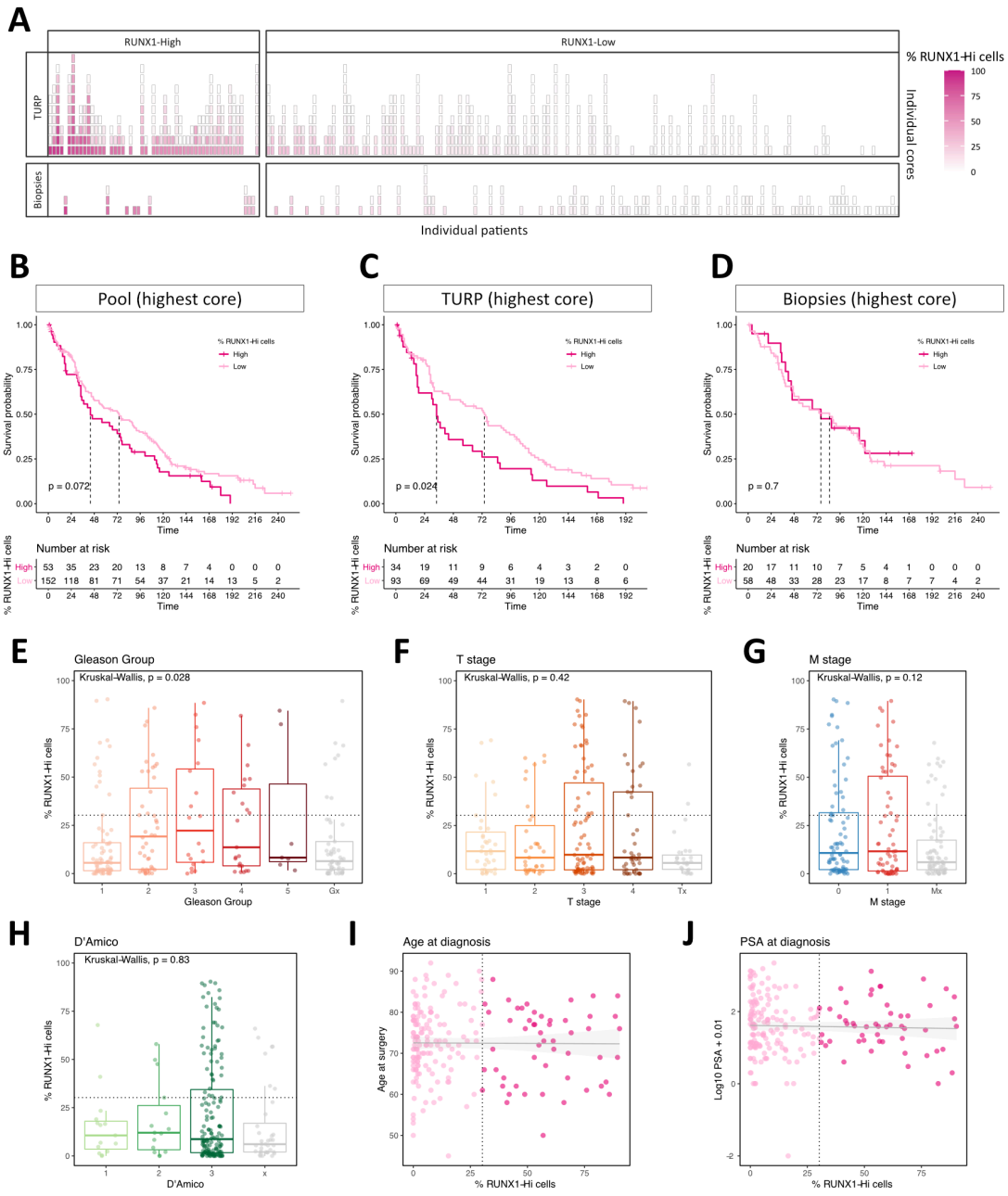

Figure 7.13. Correlation between the abundance of RUNX1 expression and clinical variables.

(A) Core-level heatmap representation of the proportion of RUNX1+ cells in TURP and biopsy tumour cores, split by RUNX1-Hi and RUNX1-Lo groups as defined in D.

(B-D) Kaplan-Meier survival analyses of RUNX1-Hi and RUNX1-Lo groups in the (B) pooled, (C) TURP and (D) biopsy cohorts. RUNX1-Hi groups were defined based on the upper quartile of the percentage of RUNX1+ cells. Log-rank test p values are indicated on the respective plots.

(E-J) Association between the frequency of RUNX1+ cells and diagnostic clinical parameters, including (E) Gleason group, (F) clinical tumour stage, (G) metastasis stage, (H) D'Amico risk

group, (I) age and (J) log10-transformed PSA. The dotted line indicates the cut-off used to split patients into RUNX1-Hi and RUNX1-Lo groups.

As previously mentioned, multivariate analyses represent a key approach to evaluate whether a given variable is independently associated with a specific prognostic outcome. RUNX1-Hi/Lo groups were therefore included in a multivariate Cox regression model with other co-variates. Similar to the univariate analysis (Figure 7.13 C), the RUNX1-Hi group was not associated with independent prognostic potential (p = 0.296) and was largely outperformed by the other co-variates (Figure 7.14 A).

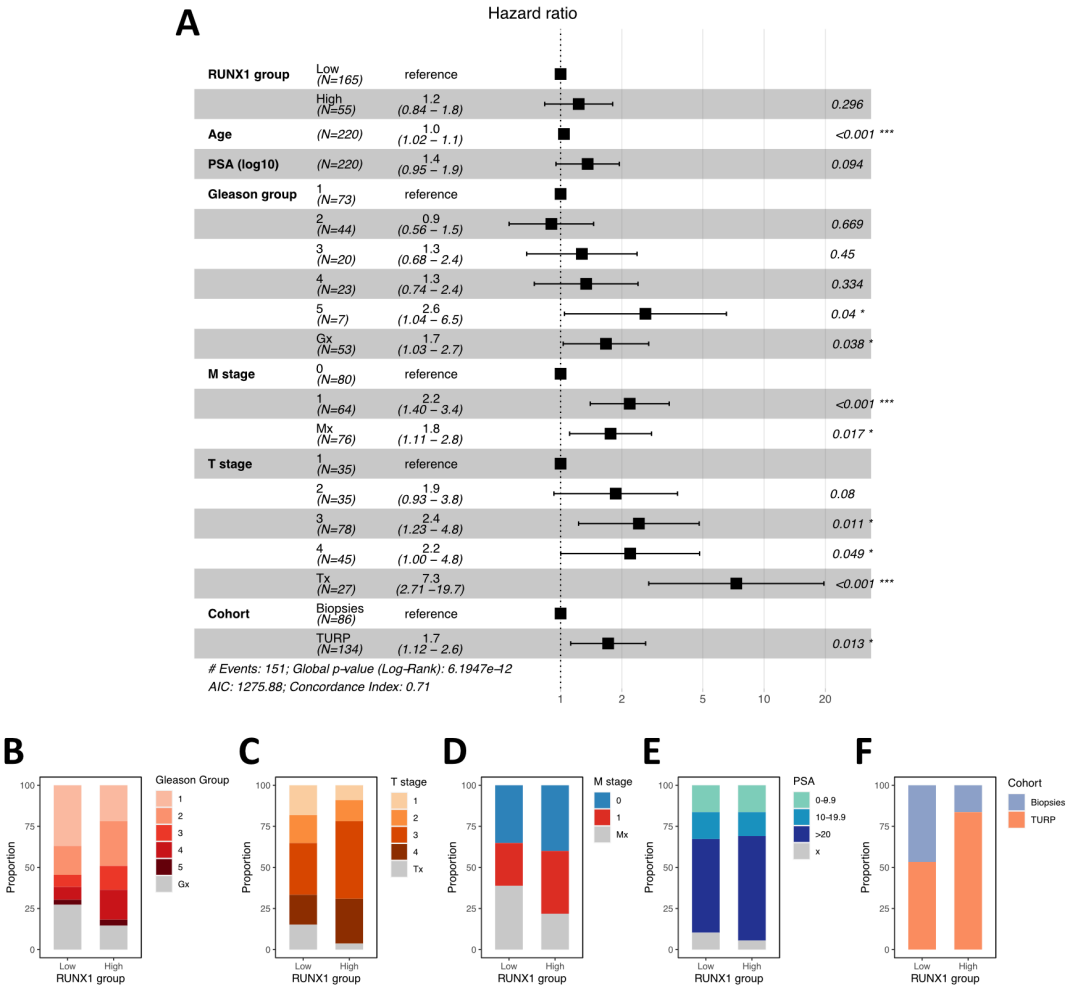

Figure 7.14. Multivariate Cox regression analysis of RUNX1-Hi/Lo groups with overall survival.

(A) Forest plot of the multivariate Cox regression analysis of RUNX1-Hi/Lo groups. The model was adjusted for age, PSA, Gleason group, clinical tumour stage, and presence of

metastasis. Hazard ratios (HR) are reported with 95% confidence interval (CI). Gx: missing Gleason group; Tx: missing clinical tumour stage; Mx: missing metastasis stage.
(B-E) Composition of RUNX1-Hi/Lo groups by (B) Gleason group, (C) clinical tumour stage, (D) metastasis stage (E) PSA groups and (F) TMA cohort.

Further investigation of the composition of RUNX1-Hi and RUNX1-Lo groups indicated that the RUNX1-Hi group contained less patients with Gleason group 1 and more patients with Gleason group 2, 3 and 4 than the RUNX1-Lo group (Figure 7.14 B). Similarly, there were more patients diagnosed with a clinical tumour stage T3 and T4 among RUNX1-Hi patients (Figure 7.14 C), as well as a slightly greater fraction of patients diagnosed with metastases (Figure 7.14 D), and diagnostic PSA above 20 ng/mL. Also, the majority of patients of the RUNX1-Hi group were part of the TURP cohort, itself associated with poorer outcome (Figure 7.13 D and Figure 7.14 A, F).

7.2.6.4. Correlation between NKX3.1 abundance and clinical characteristics in human PCa TMAs

Loss or decreased expression of NKX3.1 have previously been associated with increased tumour progression (Bowen et al., 2000). To verify whether these findings were applicable in the TURP and needle biopsies, the abundance of NKX3.1+ cells was calculated for each patient by combining cells from each individual core to obtain the average percentage of NKX3.1+ cells per patient. Since NKX3.1 is broadly expressed in the normal and cancerous human prostate, the lower quartile (25th percentile) was defined as NKX3.1-Lo group and compared with the remaining NKX3.1-Hi group. The analysis previously performed for the RUNX1-Hi/Lo groups was performed for NKX3.1. In the pooled patient cohort, there was a trend suggesting shorter overall survival for NKX3.1-Lo patients (Figure 7.15 B). In the TURP, but not in the biopsy cohort, NKX3.1-Lo patients had a significantly shorter overall survival than NKX3.1-Hi patients (log-rank p value = 0.015) (Figure 7.15 C-D).

In terms of clinical variables, there was a non-significant trend towards lower frequency of NKX3.1+ cells with increasing Gleason group and clinical tumour stage (Figure 7.15 E-F), which was significant in patients diagnosed with metastasis (p =

0.043, Figure 7.15 G) and higher D'Amico risk group (p = 0.038, Figure 7.15 H). However, there was no correlation between the percentage of NKX3.1+ cells and either age (Figure 7.15 I) or PSA (Figure 7.15 J).

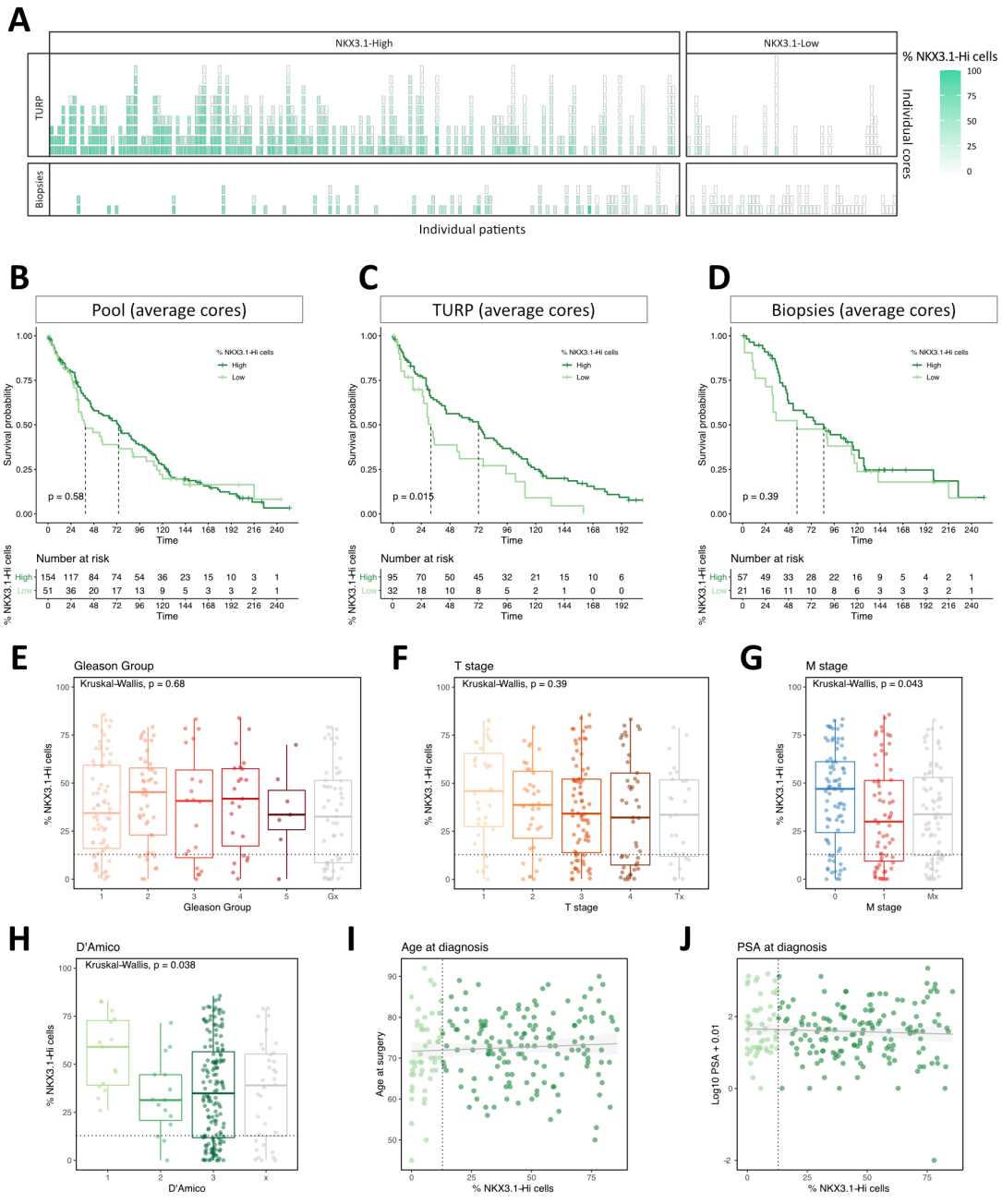

Figure 7.15. Correlation between the abundance of NKX3.1 expression and clinical variables.

(A) Core-level heatmap representation of the proportion of NKX3.1+ cells in TURP and biopsy tumour cores, split by NKX3.1-Hi and NKX3.1-Lo groups as defined in D.

(B-D) Kaplan-Meier survival analyses of NKX3.1-Hi and NKX3.1-Lo groups in the (B) pooled, (C) TURP and (D) biopsy cohorts. NKX3.1-Hi groups were defined based on the lower

quartile of the percentage of NKX3.1+ cells. Log-rank test p values are indicated on the respective plots.
(E-J) Association between the frequency of NKX3.1+ cells and diagnostic clinical parameters, including (E) Gleason group, (F) clinical tumour stage, (G) metastasis stage, (H) D'Amico risk group, (I) age and (J) log10-transformed PSA. The dotted line indicates the cut-off used to split patients into NKX3.1-Hi and NKX3.1-Lo groups.

In the context of a multivariate analysis, NKX3.1-Hi/Lo groups were not associated with independent prognosis for overall survival (Figure 7.16 A, p = 0.117). Indeed, the NKX3.1-Hi group comprised more patients with lower Gleason groups (Figure 7.16 B) and lower clinical tumour stage (Figure 7.16 C), lower diagnostic PSA (Figure 7.16 D), and less patients with metastasis at diagnosis (Figure 7.16 C). Finally, the NKX3.1-Hi group contained more patients of the TURP cohort.

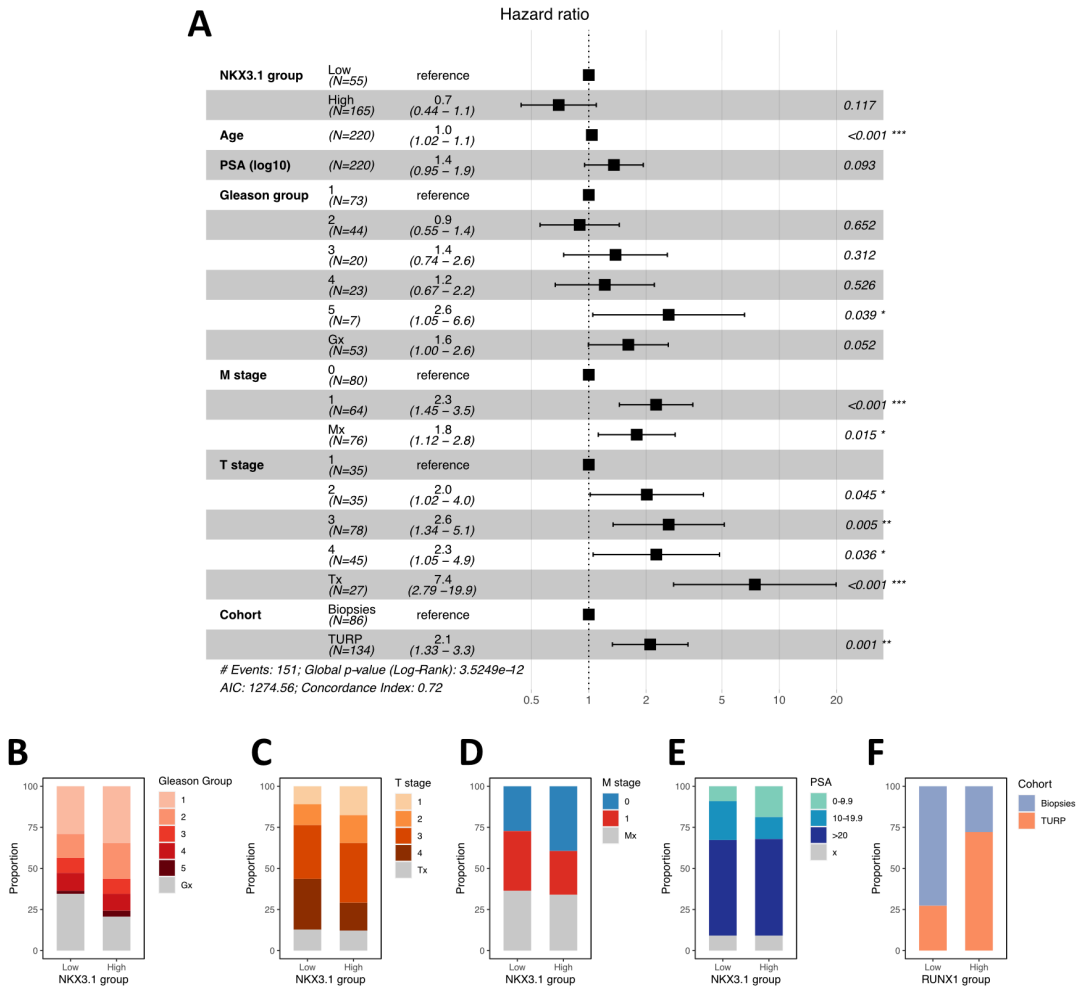

Figure 7.16. Multivariate Cox regression analysis of NKX3.1-Hi/Lo groups with overall survival.
(A) Forest plot of the multivariate Cox regression analysis of NKX3.1-Hi/Lo groups. The model was adjusted for age, PSA, Gleason group, clinical tumour stage, and presence of

metastasis. Hazard ratios (HR) are reported with 95% confidence interval (CI). Gx: missing Gleason group; Tx: missing clinical tumour stage; Mx: missing metastasis stage.
(B-E) Composition of NKX3.1-Hi/Lo groups by (B) Gleason group, (C) clinical tumour stage, (D) metastasis stage (E) PSA groups and (F) TMA cohort.

7.2.6.5. Combined analysis of RUNX1-Hi/Lo and NKX3.1-Hi/Lo groups in human PCa TMAs

RUNX1 and NKX3.1 used as single biomarkers had limited prognostic performance, except in the TURP cohort whereby RUNX1-Hi and NKX3.1-Lo groups were both associated with shorter overall survival in univariate analyses. The same strategy as in sections 7.2.6.3 and 7.2.6.4 was used to group patients into RUNX1-Hi/Lo and NKX.3.1-Hi/Lo groups. However, to maximise the number of patients per group, TURP and biopsy cohorts were pooled, and the median was chosen as cut-off (50th percentile) to categorise patients into Hi/Lo groups. Corresponding Kaplan-Meier analyses of the patient groups are shown in Figure 7.17 A-B. To test whether the combination of both markers could help improve patient stratification, patients were then classified into the following categories: RUNX1-Hi NKX3.1-Hi, RUNX1-Hi NKX3.1-Lo, RUNX1-Lo NKX3.1-Hi and RUNX1-Lo NKX3.1-Lo. Univariate and multivariate analyses of these patient groups did not improve the prognostic performance of RUNX1 and NKX3.1 used as single biomarkers (Figure 7.17 C-D). However, it is worth noting that there was a trend towards shorter overall survival (log-rank p value = 0.08, data not shown) for the RUNX1-Hi NKX3.1-Lo compared with the RUNX1-Lo NKX3.1-Hi (Figure 7.17 C). In particular, the RUNX1-Hi NKX3.1-Lo group was associated with a higher proportion of patients diagnosed with more advanced clinical tumour stage (Figure 7.17 E) and a greater proportion of patients with M1 disease (Figure 7.17 F).

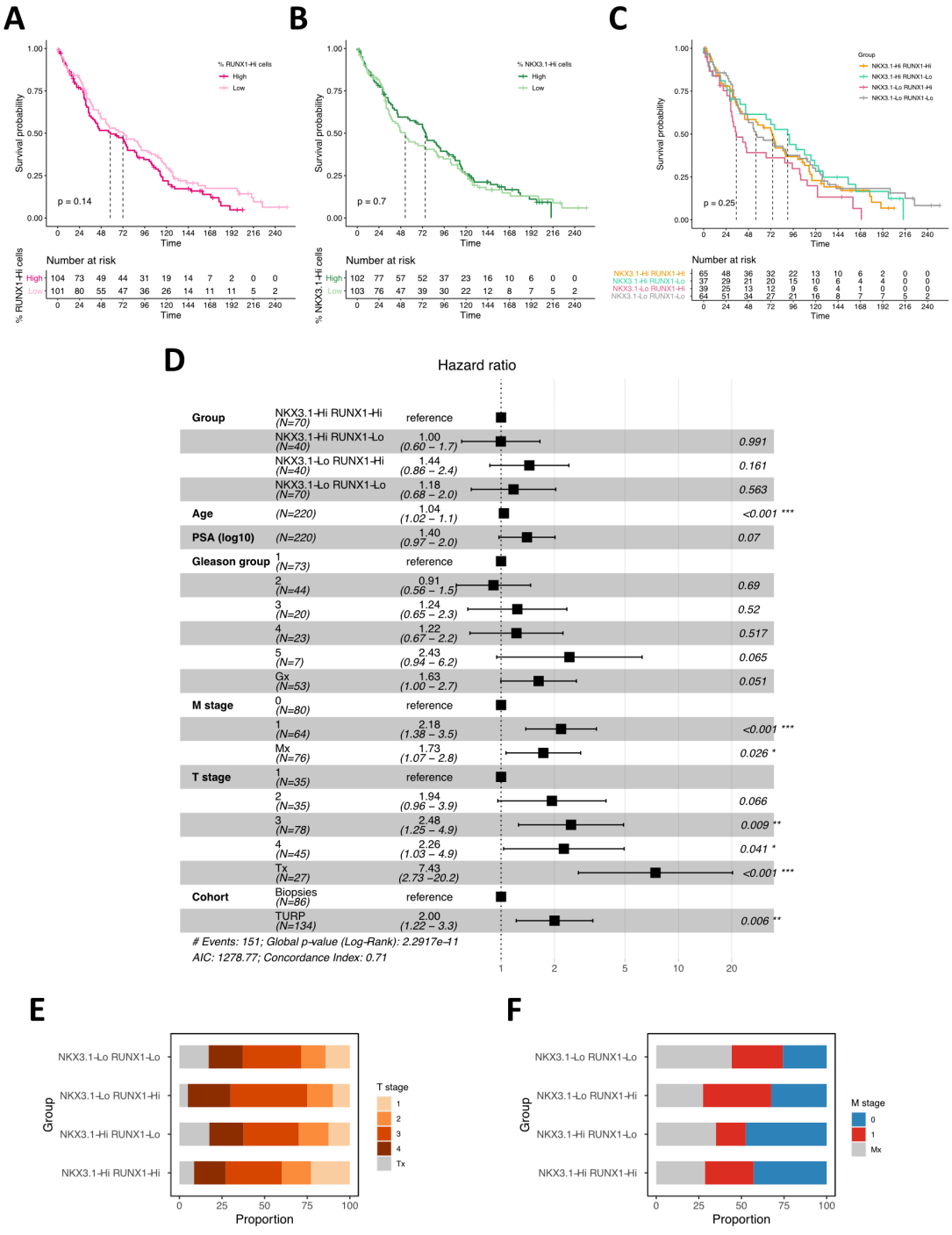

Figure 7.17. Univariate and multivariate analyses of RUNX1- and NKX3.1-Hi/Lo groups.
(A-B) Kaplan-Meier survival analyses of (A) RUNX1-Hi/Lo and (B) NKX3.1-Hi/Lo groups in both TURP and biopsy cohorts combined using the median as cut-off to split patients into Hi/Lo groups. Log-rank test p values are indicated on the respective plots.
(C) Kaplan-Meier survival analyses of RUNX1-Hi NKX3.1-Hi, RUNX1-Hi NKX3.1-Lo, RUNX1-Lo NKX3.1-Hi and RUNX1-Lo NKX3.1-Lo patient groups after combining RUNX1 and NKX3.1 status of each patient determined in A and B.

(D) Forest plot of the multivariate Cox regression analysis of RUNX1-Hi NKX3.1-Hi, RUNX1-Hi NKX3.1-Lo, RUNX1-Lo NKX3.1-Hi and RUNX1-Lo NKX3.1-Lo patient groups. The model was adjusted for age, PSA, Gleason group, clinical tumour stage, and presence of metastasis. Hazard ratios (HR) are reported with 95% confidence interval (CI). Gx: missing Gleason group; Tx: missing clinical tumour stage; Mx: missing metastasis stage.
(E-F) Composition of each RUNX1/NKX3.1 patient group by (E) clinical tumour stage and (F) metastasis stage.

7.2.7. Mining publicly available datasets of PCa

In the last decades, several micro-array and large-scale sequencing studies, have improved our understanding of the molecular alterations and gene expression programs found in prostate tumours. Publicly available datasets available on the cBioPortal platform were interrogated for genomic alterations of the RUNX genes (Figure 7.18 A-D) in a total of 6,037 patients from 19 studies. Overall, the proportion of patients with genomic alterations varied significantly between distinct studies (Figure 7.18 A-C) but was not associated with clinically distinct tumour subtypes (data not shown). *RUNX1* was altered in 7% of all patients tested, *RUNX2* in 11% and *RUNX3* in 13% (Figure 7.18 D). However, the majority of alterations were either labelled as “gain” or “shallow deletion”, suggesting they these may have a minor impact on the tumour phenotypes. When these were not considered, the frequency of severe genomic alterations dropped to 0.7% for *RUNX1*, 1% for *RUNX2* and 1.1% for *RUNX3* (Figure 7.18 E).

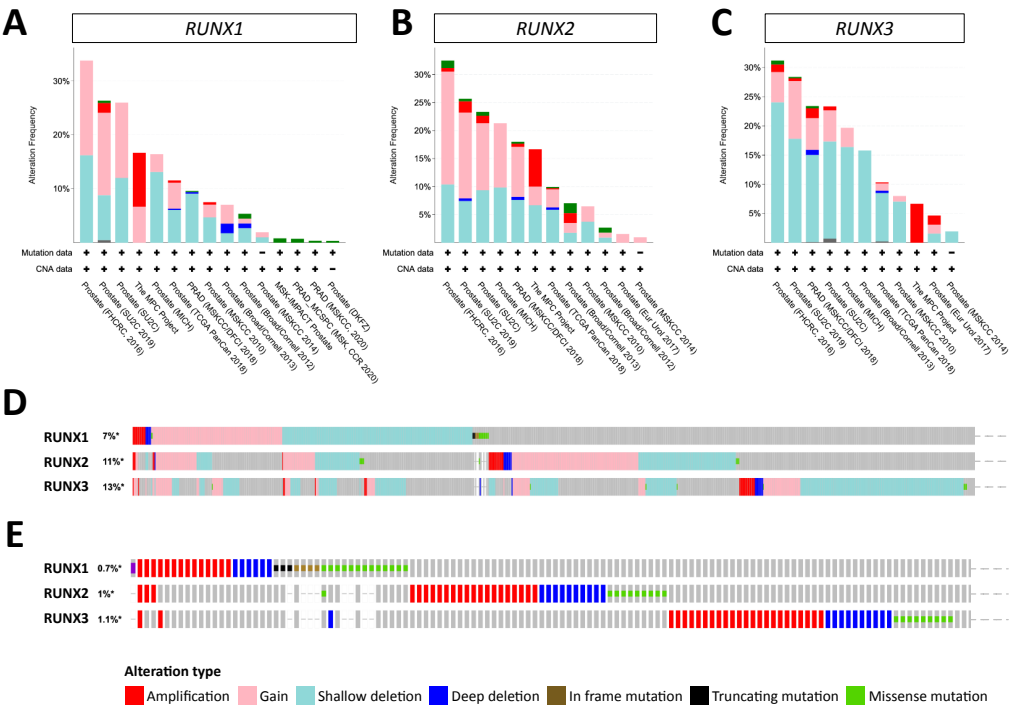

Figure 7.18. Genomic alterations of the RUNX genes in PCa (cBioPortal).
(A-C) Frequency of amplification (red), gain (light red), shallow deletion (light blue), deep deletion (blue) or mutations (green) in human PCa datasets available on cBioPortal with reported genomic alterations of (A) *RUNX1*, (B) *RUNX2* and (C) *RUNX3*. (D-E) Oncoprints for *RUNX1*, *RUNX2* and *RUNX3* generated on cBioPortal using all PCa datasets available and (D) all types of genomic alterations including amplification, gain, shallow deletion, deep deletion, and mutation or only (E) amplification, deep deletion and mutation. CNA: Copy Number Alterations.

Among the PCa datasets accessible via cBioPortal, the TCGA (Cancer Genome Atlas Research Network, 2015) and DKFZ (Gerhauser et al., 2018) studies had bulk RNA-seq and clinical data available for a subgroup of patients. Clinical endpoints of the TCGA study included disease-free survival and biochemical recurrence, while only the latter was available in the DKFZ study. Unlike the TMA analysis, overall survival data was not available for the DKFZ study, and the TCGA cohort had too few events (death) to be used reliably (Liu et al., 2018). Using a similar thresholding approach to the one previously used in TMAs (section 7.2.6), the upper quartile of *RUNX1* mRNA expression was chosen to classify patients into RUNX1-High versus RUNX1-Low groups. In the TCGA study, the RUNX1-High group did not show differences in outcome compared with the RUNX1-Low group, both in terms of

disease-free survival (p = 0.43, Figure 7.19 A) or biochemical recurrence (p = 0.83, Figure 7.19 B). The DKFZ study had less patients in the cohort, but there was a trend towards a shorter time to biochemical recurrence for the RUNX1-High group (p = 0.068, Figure 7.19 B). The expression of *RUNX1* was then evaluated against additional clinical variables available in these datasets. In the TCGA study, there was no correlation between RUNX1 expression and pathological tumour stage or the presence of lymph node metastases (Figure 7.19 C). In contrast, higher *RUNX1* expression was significantly associated with higher pathological tumour stage T4 (p = 0.02) and Gleason group 5 (p = 0.0094) in the DKFZ study (Figure 7.19 D). This observation mirrored the TURP TMA analysis showing that a fraction of high grade PCa with high *RUNX1* expression levels may be associated with a worse outcome.

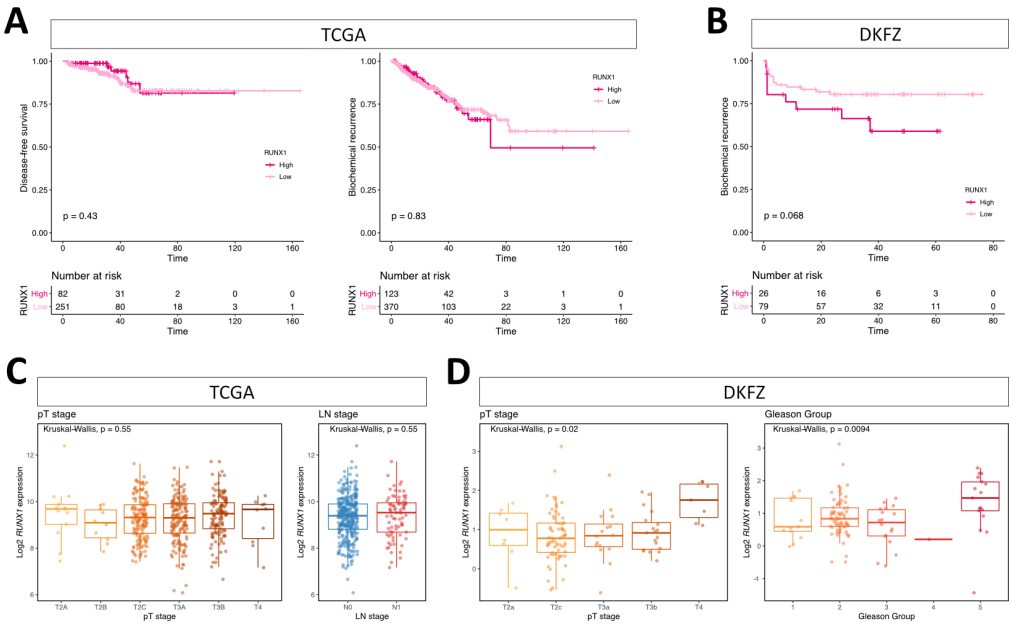

Figure 7.19. Genomic alterations of the RUNX genes in PCa (cBioPortal).
(A) Kaplan-Meier analysis of RUNX1-High and RUNX1-Low groups for disease-free survival (left) and biochemical recurrence (right) in the TCGA PRAD cohort. Clinical and mRNA expression data were obtained from cBioPortal.
(B) Kaplan-Meier analysis of RUNX1-High and RUNX1-Low groups for biochemical recurrence in the DKFZ cohort. Clinical and mRNA expression data were obtained from cBioPortal.
(C) Association between *RUNX1* mRNA expression (log2-transformed) and pathological tumour stage (left, n = 486) and presence of lymph node metastases (right, n = 420) in the TCGA cohort.

(D) Association between *RUNX1* mRNA expression (log2-transformed) and pathological tumour stage (left, n = 116) and Gleason group (right, n = 118) in the DKFZ cohort.

7.3. Conclusion

7.3.1. Summary

In the developing and adult normal mouse prostate, RUNX1 marks a specific subpopulation of luminal cells enriched in the peri-urethral/proximal region. However, the relevance of this finding to the normal and cancerous human prostate remained unexplored. The data presented in this Chapter aimed at characterising the expression pattern of RUNX1 in archived samples of prostate cancer, including normal-adjacent prostate tissue and heterogeneous cohorts of PCa samples. Another aim was to evaluate whether the expression of RUNX1 correlates with specific clinical outcomes to evaluate its biomarker potential in PCa. Analyses conducted on a broad range of patient samples revealed that RUNX1 expression was low in normal secretory glands. In contrast, high RUNX1 was found in basal cells and atrophic prostate epithelium, in particular in areas rich in inflammation associated with low NKX3.1 expression. In prostate tumours, RUNX1 was generally rare, with the exception of a limited number of samples which expressed very high RUNX1 levels. These RUNX1 high tumours often had concomitant low NKX3.1 expression and higher pathological grades. However, overall survival analyses revealed poor prognostic performance of RUNX1 used as a single marker or in combination with NKX3.1.

7.3.2. RUNX1 expression in benign human prostate glands

The expression pattern of RUNX1 was first investigated in benign human adult prostate glands. These samples were obtained from aged individuals who underwent surgery due to the presence of localised PCa or BPH in their prostate. It is possible that these samples do not accurately represent a completely “normal” adult prostate. In particular, the different histopathological features observed in

these samples such as inflammation, atrophy, BPH and basal cell hyperplasia are likely to be over-represented compared to entirely benign young prostates.

The intensity and spatial localisation of RUNX1 staining in human samples was highly patient-specific and showed an important degree of both inter- and intra-patient heterogeneity. Due to the complex and dense architecture of the human prostate, the spatial distribution of RUNX1+ cells was more difficult to appreciate based on 2D tissue sections compared with the mouse prostate. In mice, the expression of RUNX1 was strikingly restricted to a luminal subset enriched in the peri-urethral region of the prostate. In human samples, RUNX1 was also found to be broadly expressed in the peri-urethral region. However, RUNX1+ cells were also frequently identified in other areas of the prostate, including the transitional and peripheral zones (Figure 7.4). Although technically challenging because of the thickness and size of the tissue, it would be interesting to perform whole mount 3D tissue staining of entire prostatectomy specimens. Alternatively, a large number of stained serial sections could be used to virtually reconstruct human prostate samples in 3D in order to better visualise the spatial distribution of RUNX1+ cells in these samples (Moad et al., 2017).

Similar to the mouse prostate, RUNX1 was expressed in the basal layer of the human prostate but rarely in luminal secretory cells. Together with the downregulation of RUNX1 observed in secretory alveolar luminal cells of the mammary gland (van Bragt et al., 2014), these results suggest that there may be a general requirement for low or absent RUNX1 expression to maintain the secretory phenotype. It would be of interest to investigate this further by modulating the expression of RUNX1 in luminal secretory cells during development and under homeostatic conditions. Foci of high RUNX1 expression were often observed in regions of atrophy, especially when associated with inflammation and lower NKX3.1 levels. Several forms of atrophy have been described in the literature (De Marzo et al., 1999), but the clinical significance of these histopathological features remain largely debated. In patient samples, there is contrasting evidence suggesting that prostate atrophy is associated with either better (Zhang et al., 2019) or worse (Davidsson et al., 2011; De Marzo et al., 1999; Putzi and De Marzo, 2000) outcomes.

Angelo De Marzo proposed that proliferative inflammatory atrophy may be a hallmark of prostate carcinogenesis (De Marzo et al., 1999; De Marzo et al., 2003, 199). Along these lines, Le Magnen and colleagues showed that cancer initiation in aged *Nkx3-1* mutant mice correlated with inflammation (Magnen et al., 2018). Inflammation in mice was also found to correlate with aberrant basal to luminal differentiation (Kwon et al., 2014) and the expansion of CD38-Low tumour initiating cells (Liu et al., 2016). However, these studies focused on the presence of inflammation rather than the type or abundance of epithelial atrophy. Additional studies are warranted to investigate the functional and prognostic significance of prostatic atrophy in PCa samples, as well as to define the role played by RUNX1 in these regions. Finally, high RUNX1 expression was also observed in basal cell hyperplasia, though there is currently no evidence that basal cell hyperplasia correlates with PCa initiation and progression (Bennett and Gardner, 1985; Epstein and Armas, 1992; Freitas et al., 2019).

Immunostainings of normal human prostate glands did not allow the identification of a RUNX1+ NKX3.1- luminal subset that would be an equivalent to the Lum-D/PLCs found in mouse prostates. Whether such cell type exists in men, and whether RUNX1 can help identify these cells remain an open question. A number of studies involving scRNA-seq of patient samples have been published over the last year and represent a useful resource to study the nature of the multiple cell types present in patient samples. Notably, Crowley and colleagues proposed that the mouse proximal and distal regions respectively resemble human ductal and acinar regions (Crowley et al., 2020). The ductal phenotype shown in Crowley's study appears to be similar to the atrophic epithelium defined in this chapter. However, it is unclear what morphologically defines a ductal region as opposed to an acinar region, and where these regions are located. It would be interesting to explore these scRNA-seq datasets of human prostates to further evaluate whether a Lum-D/PLCs equivalent exists, and whether the expression of RUNX1 and NKX3.1 in combination with other markers can be used to identify these cells. If successful, such analyses could help define a key set of genes specific to this cell type and common to both human and mouse prostates to facilitate cross-

species studies. Indeed, it is clear that the cellular landscape of the normal prostate epithelium cannot be explained solely based on the expression of RUNX1 and NKX3.1.

7.3.3. RUNX1 has limited biomarker potential in PCa patient samples

Most of the human PCa samples stained in this chapter had low or absent RUNX1 expression in tumour cells. This observation is in line with a previous report indicating that RUNX1 expression tends to be decreased in tissues with a higher Gleason score (Takayama et al., 2015). However, among all the patients analysed, a subgroup displayed very high RUNX1 levels in their tumours. The abundance of RUNX1+ cells was used in two independent TMA cohorts of PCa patients to generate RUNX1-Hi and RUNX1-Lo groups. This resulted in poor prognostic performance of RUNX1 in terms of overall survival. Further stratification by NKX3.1 status did not improve the performance in both univariate and multivariate analyses. In other specific solid cancers, such as clear cell renal cell carcinoma and triple negative breast cancer, high RUNX1 levels have been associated with poorer outcomes (Ferrari et al., 2014; Rooney et al., 2020). It is therefore possible that RUNX1 would exhibit better performance in a specific molecular, phenotypic or clinical subgroup of PCa patients currently under-represented in the material analysed in this chapter. For example, basal cell carcinomas are extremely rare forms of PCa (Grignon, 2004), and together with the higher RUNX1 levels in basal cells and basal cell hyperplasia, it would be interesting to determine the expression and role of RUNX1 - if any - in this subtype of PCa. Here, RUNX1 showed better performance in the TURP cohort, which contained a greater proportion of high-risk patients, compared with the biopsy cohort. To confirm whether the expression of RUNX1 may be of greater clinical relevance in high-risk patients, it would be important to validate this result in an independent cohort of similar nature. Besides, it would be interesting to combine RUNX1 and NKX3.1 with other markers of specific prostate subpopulations in order to better stratify tumour phenotypes and evaluate whether this could help improve prognostic performance. Such analyses could be either performed on serial sections using a panel of markers or

using emerging multiplexing staining technologies such as CODEX (CO-Detection by indEXing, Akoya Biosciences) (Goltsev et al., 2018), imaging mass cytometry (Angelo et al., 2014; Giesen et al., 2014), or spatial transcriptomics (Burgess, 2019) to gather high-dimensional single cell data.

Since only immunostainings were performed on the TURP and biopsy TMA cohorts, it was not possible to determine whether high RUNX1 expression in specific tumours resulted from genomic alterations of *RUNX1*, if RUNX1 was upregulated as a consequence of a tumour-specific gene expression program, and whether high RUNX1 expression was required for tumour progression in these samples. However, *in silico* analyses carried out on publicly accessible databases (cBioPortal, Oncomine) did not show significant correlations between copy number alterations of *RUNX1* and expression levels (data not shown), suggesting that high RUNX1 expression may result from a tumour-specific gene expression program rather than genomic alterations. Also, the low alteration frequency of *RUNX1* in PCa patient samples suggests that *RUNX1* alterations is unlikely to represent a mechanism of PCa initiation. Nevertheless, to address whether RUNX1 is functionally required in specific tumour samples that express very high levels of RUNX1, functional analyses such as knockdown or knockout experiments should be performed using PCa models, including human cell lines and primary derived material.

Publicly available datasets accessed via cBioPortal were also explored to evaluate whether *RUNX1* mRNA levels could be used to stratify PCa patients. Of 19 studies available, the TCGA and DKFZ datasets were the only two comprising both RNA-seq and clinical outcome data (Cancer Genome Atlas Research Network, 2015; Gerhauser et al., 2018). Biochemical recurrence was also the only reliable clinical outcome data available for both cohorts, and therefore cannot be compared directly with the TMA analysis. Overall, RUNX1-Hi and RUNX1-Lo groups of the TCGA cohort did not exhibit marked differences in terms of biochemical recurrence. In the DKFZ study, there was a small subset of high-grade patients which displayed higher *RUNX1* expression and a non-significant trend towards shorter time to biochemical recurrence. This observation corroborated the TMA analysis in that

there is a small subset of PCa patients with high grade tumours and high RUNX1 expression with a possibly worse clinical outcome. However, both in the DKFZ study and the TURP TMA cohort, the restricted number of patients defining the RUNX1-Hi group limited the extent of the conclusions that could be drawn. It is also worth noting that these datasets correspond to bulk RNA-seq studies and may be biased by contaminations from non-epithelial cells, such as RUNX1+ immune cells infiltrating prostate tumours.

7.3.4. Limitations of the methodology used for TMA analyses

The analysis of archived TMA tissue cohorts has a number of limitations. TMA cores represent only a small fraction of large tumours. This is particularly important in the case of PCa, which is known to be highly multi-focal (Boutros et al., 2015; Carm et al., 2019, 201; Tomlins et al., 2015). Thus, it is difficult to determine whether sampling different regions in small cores is sufficient to capture tumour heterogeneity (Andreoiu and Cheng, 2010; Rubin et al., 2002; Tolkach and Kristiansen, 2018). For example, it is possible that tissues chosen during the sampling procedures did not always include a representative example of the most aggressive tumour foci driving the disease. TURP and biopsy samples are also intrinsically different types of tissue samples (Jhavar et al., 2005). The amount of tissue available for analysis was very limited in the case of biopsies compared with TURPs. Small biopsies are therefore likely to be less representative of the disease than TURPs, limiting their prognostic value for biomarker evaluation. Yet, needle biopsies are often used for diagnostic procedures and correspond to the material that pathologists evaluate in the first place. Further to this, TURP samples are normally enriched in tissue of the transitional zone while biopsies are expected to mainly originate from the peripheral zone. These discrepancies indicate a significant sampling bias which is important to consider when interpreting results. Therefore, the decision to pool samples from the 2 cohorts to improve statistical power may be questionable. Ideally, only samples from similar sources should be compared.

In addition to sampling biases, it is important to mention that the TURP and biopsy samples correspond to archive cohorts of patients diagnosed between 1994

and 2004. This constitutes a strength of those cohorts, with long term clinical follow up in terms of survival, but also a weakness since clinical management of PCa may not be entirely similar nowadays. Also, while overall survival is an important parameter when evaluating biomarker performance, it would be interesting to test other clinical parameters such as biochemical recurrence to evaluate the success rate of ADT according to RUNX1 status. Here, it is difficult to determine if patients remained disease-free after receiving a curative form of treatment such as surgery or radiotherapy, and which patients received ADT. In fact, anecdotical annotations present in the clinical database suggest that patients received very heterogeneous treatments. Heterogeneity in subsequent treatments is also likely to introduce additional biases in a fraction of patients. Unfortunately, it was not possible to include these layers of information in multivariate analyses due to the amount of missing data.

Despite the benefits of analysing large number of patient samples in TMAs, the technical approach chosen here is associated with additional limitations. TMA sections were co-stained with 3 markers by immunofluorescence. This approach provides a greater dynamic range compared with classical chromogenic immunohistochemistry thereby improving quantitative analyses (Camp et al., 2002; Levenson et al., 2015; McCabe et al., 2005). Another important advantage was the use of a pan-cytokeratin to restrict the analysis to the epithelial compartment. However, despite the use of this parameter, it is possible that cellular classification into epithelial and stromal compartments was not always accurate. This may be problematic in cases where RUNX1+ tumour infiltrating immune cells would be misclassified as epithelial during image analysis. Overall, the use of 3 markers remains a limitation when studying population dynamics in complex and heterogeneous tumour samples. Increasing the number of proteins probed would allow more accurate classification of stromal and epithelial compartments, but also enable to group patient tumours into more similar subtypes which may in turn improve prognostic performance.

Besides, antibody-based stainings of archived samples are associated with technical artefacts difficult to account for in semi-automated quantitative analyses,

such as differences in tissue fixation post-surgery and variations in antigen preservation. These aspects can introduce biases between samples originating from different patient cohorts. For example, TURP samples were likely to be affected by diathermy artefacts due to the surgical procedure. Conversely, biopsies suffered from more pronounced edge effects and tissue loss due to the smaller sample sizes. Tissue sections also present a certain degree of autofluorescence, which varies depending on tissue composition. In principle, autofluorescence should have a minimal impact on intensity measurements due to the high signal-to-noise ratio obtained by the use of TSA-conjugated fluorophores. But despite taking into account these aspects before extracting single cell measurements to minimise quantitative artefacts, it is not possible to rule out the presence of unidentified technical biases. A direct consequence of this quality control step resulted in some cases in significant loss of material, which may have an impact on downstream analyses. We estimated that for both the TURP and biopsy cohorts, about half of the cores were not available for analysis due to poor staining quality or artefacts, as well as missing, rolled and crushed TMA cores.

Finally, an important point to discuss relates to the strategy used to analyse the TMA cohorts. Single cell measurements were extracted from scanned images and used to classify patients into high and low RUNX1 or NKX3.1 expression based on the percentage of positive cells. The first critical parameter to set was the intensity threshold used to classify single cells as positive or negative. This value was determined empirically for both RUNX1 and NKX3.1 by manually evaluating a large number of cells in different section and tissue cores. Next, an arbitrary cut-off was chosen using quantiles in order to separate patients with the highest or lowest RUNX1 and NKX3.1 levels respectively. This was done by considering only the core with the highest proportion of RUNX1+ cells in the case of RUNX1 and using the pooled material from all cores for NKX3.1. The rationale for this was that a single tumour focus with high RUNX1 expression or overall low expression levels of NKX3.1 may be important and unusual features to consider (Bethel and Bieberich, 2007; Bowen et al., 2000). Varying the thresholds and approaches used to conduct the analysis could have important consequences in terms of prognostic

performance. Therefore, it would be important to evaluate the robustness of the approach chosen in this chapter on independent and larger patient cohorts. It is possible that the thresholds chosen would not apply to different cohorts, or even that biopsies should be treated differently than TURP. Staining larger patient cohorts would likely help to optimise the determination of relevant thresholds. Of note, other quantification approaches were tested, including the calculation of a weighted histological H-score which takes into account the proportion of cells of increasing intensities for a given marker (Fedchenko and Reifenrath, 2014). However, this did not improve the prognostic performance of RUNX1 and NKX3.1 (data not shown). Despite all these limitations, RUNX1 seems to have limited clinical utility as a single marker or in combination with NKX3.1 in PCa patients.

Chapter 8 Conclusion and perspectives

Cancer is a highly heterogeneous disease, capable of using multiple strategies to transform a normal cell into a malignant one (Hanahan and Weinberg, 2011). While the mechanisms involved in tumorigenesis are increasingly studied and understood, cancer remains amongst the principal causes of death worldwide (Ferlay et al., 2015). It is largely admitted that early diagnosis is the most effective way to improve patient outcomes (Crosby et al., 2020). In the case of PCa, the development and widespread availability of PSA testing has improved early detection, but it is not specific and sensitive enough to distinguish between low- and high-risk patients. This is of particular importance since the vast majority of PCa cases are diagnosed with indolent or clinically insignificant tumours that will never progress to more advanced forms such as mCRPC (Shen and Abate-Shen, 2010; Welch and Black, 2010; Wilt et al., 2008). Yet, the development of mCRPC in a subgroup of patients is associated with high mortality, due to the limited effective treatment options for metastatic and hormone-refractory tumours (Teo et al., 2019). In this context, there is a strong need for better biomarkers of disease progression and recurrence to improve the clinical management of PCa. In the recent years, the master regulator of haematopoiesis RUNX1 has been increasingly implicated in epithelial tissues (Blyth et al., 2005; Ito et al., 2015; Mevel et al., 2019; Scheitz et al., 2012), including hormone-regulated ones (Lie-A-Ling et al., 2020; Riggio and Blyth, 2017). This led to the hypothesis of my PhD that RUNX1 is involved in normal prostate biology and prostate cancer, which I investigated using a combination of experimental approaches involving mouse models and human patient samples.

8.1. Key findings

In mice, RUNX1 was found to be expressed during embryonic prostate development in the urogenital epithelium, prior to the emergence of the first prostatic buds. Throughout mouse prostate development, RUNX1 expression was

largely restricted to the peri-urethral or proximal region, while the most distal parts of the epithelial ducts were marked by NKX3.1 expression. RUNX1 was also expressed in a large proportion of basal cells, though this aspect was not further investigated. The pattern of expression of RUNX1 was conserved in the adult mouse prostate, whereby RUNX1 marked a subset of proximal luminal cells (PLCs). Together with extensive *in situ* phenotypic characterisation using multiple immunofluorescent panels, the use of single-cell transcriptomics of developmental and adult stages suggested that RUNX1+ PLCs constitutes a distinct luminal lineage of the prostate. In fact, this luminal lineage appears to be established at the very onset of prostate development, and predominantly confined until adulthood to the most proximal region of all mouse prostate lobes. Beyond the study of RUNX1 biology, scRNA-seq of the normal and castrated adult mouse prostate highlighted an unprecedented level of heterogeneity within the prostatic epithelium, and in particular within the luminal compartment with the detection of at least six transcriptomically divergent mature populations.

In addition to a distinct gene expression program and specific anatomical localisation, RUNX1+ PLCs have unique functional characteristics. First, the histological phenotypes of these cells suggest that they are not committed to the secretory function of the prostate. This was further reinforced by analyses of their gene expression program by scRNA-seq showing the absence of a secretory program. Then, RUNX1+ PLCs displayed stem cell behaviour in organoid assays, with an enhanced clonogenic potential compared with other luminal subsets, and a lineage bias towards the luminal fate. Finally, genetic lineage tracing experiments revealed that RUNX1+ PLCs are intrinsically castration-resistant, but do not contribute to the regeneration of distinct lineages, including NKX3.1+ cells. Together, these results provide new insights into the lineage relationship of the prostate epithelium and warrant further studies of RUNX1+ PLCs in the context of tumorigenesis and treatment resistance.

Inherent anatomical and histological differences between species remain an obstacle to directly translate these findings to the human prostate. Nevertheless, in non-transformed human prostate tissue, RUNX1 was also expressed in the peri-

urethral region, basal cells, and strongly enriched in areas of epithelial atrophy and inflammation. Overall, while the expression of RUNX1 was low in human tumours, a subset of patients displayed high RUNX1 levels, which was often associated with concomitant low NKX3.1 expression and higher pathological grades. However, RUNX1 expression was not significantly associated with differences in overall survival using archived TMA cohorts and publicly available databases, indicating limited clinical utility as a single marker or in combination with NKX3.1.

8.2. Towards a greater understanding of RUNX1 biology in the prostate epithelium

8.2.1. Regulation of *Runx1* expression in prostate epithelial cells

Runx1 expression can be driven by two alternative promoters, distal (P1) and proximal (P2). The tight regulation between P1 and P2 has been best studied in the context of haematopoiesis, both in development and adult (described in section 1.4.1.3). However, little is known about the activity of RUNX1 promoters in other tissues. In the mouse prostate epithelium, the expression of *Runx1* appeared to be exclusively mediated by the P2 promoter, thought to be the primordial promoter (Levanon and Groner, 2004), and responsible for the expression of the RUNX1b isoform. Additional work carried out as part of a collaboration, but not shown in this thesis, indicates that RUNX1b is also the predominant isoform expressed in mammary epithelial cells. More work is therefore required to assess whether any other epithelial population expresses RUNX1c, but the data presented here would suggest that RUNX1c is not expressed in epithelial lineages. This observation contrasts with the haematopoietic system, in which RUNX1b and RUNX1c are expressed during embryonic and adult haematopoiesis, either sequentially or in distinct cell types. One possible explanation for this differential isoform usage could relate to the specific requirements for RUNX1 in haematopoietic and epithelial cells. The use of two isoforms might be more relevant to the haematopoietic setting where RUNX1 is a master regulator and requires fine tuning of its expression (Lie-A-Ling et al., 2020). It is also striking that during haematopoietic development,

RUNX1b is expressed in endothelial cells, before any lineage committed blood cell emerge, reinforcing the notion that RUNX1c expression may be restricted to the haematopoietic system (Lie-A-Ling et al., 2018; Sroczynska et al., 2009). Promoter activity may also depend on upstream regulators involved, leading to differential enhancer accessibility in prostate cells, or a requirement for subtle functional differences between RUNX1b and RUNX1c in distinct tissues. Overall, it would be interesting to investigate further the regulatory mechanisms of *Runx1* expression in epithelial cells, and test whether replacing RUNX1b by RUNX1c in these cells would alter its downstream function and the phenotype of RUNX1+ cells. Finally, it would be interesting to investigate whether similar regulation occurs in the human prostate, and in PCa.

8.2.2. Functional role of *Runx1* in the prostate

The main role of RUNX1 is to act as a DNA-binding transcription factor to regulate gene expression. The expression pattern of RUNX1 in the prostate suggests a specific role. However, the present work focused on the description of RUNX1 expression, but it did not uncover the transcriptional program regulated by RUNX1 in the prostate. *Runx1* knockout experiments in prostate organoid cultures failed to exhibit a phenotype. Similarly, experiments not shown in this thesis but performed in UGS explant cultures did not result in noticeable changes upon *Runx1* deletion. Conditional deletion of *Runx1* using *Pb-Cre4 Runx1*$^{Flox/Flox}$ mouse model also failed to generate defects in the prostate epithelium, even in aged mice or in castration-regeneration assays (data not shown). The absence of phenotype may indicate that RUNX1 is not essential for prostate cells, or that loss of RUNX1 is compensated by another RUNX family member or an alternative pathway. Thus, additional work is required to investigate the role and requirement of RUNX1 in prostate cells. Since genetic deletion experiments did not result in evident phenotypic changes, it would be interesting, in the future, to focus on the DNA-binding activity of RUNX1 rather than the phenotypic consequences of its deletion. In the haematopoietic system, RUNX1 has been extensively studied in normal development and adult homeostasis, as well as in blood malignancies including leukaemia (Lichtinger et al.,

2012; Ptasinska et al., 2012; Tijssen et al., 2011; Wilson et al., 2010). However, the precise gene expression program regulated by RUNX1 in other tissues remains poorly characterised. In particular, the high cell numbers required for Chromatin Immunoprecipitation Sequencing (ChIP-Seq) experiments constitutes one of the main limitations to perform genome-wide DNA-binding profile of RUNX1 *in vivo*. A novel method, termed Cleavage Under Targets and Release Using Nuclease (CUT&RUN), has recently been developed to profile transcription factors binding using low cell numbers (Skene and Henikoff, 2017; Skene et al., 2018), and may therefore leverage part of the limitations associated with ChIP-Seq. While it still relies on the use of an anti-RUNX1 antibody, experimental conditions differ from ChIP-Seq, and the cost of individual samples has been dramatically reduced. It would be interesting to optimise and apply CUT&RUN on RUNX1+ prostate epithelial cells in order to characterise the transcriptional program regulated by RUNX1 in distinct populations of the prostate epithelium. Furthermore, recent advances in single cell technologies hold promises to profile transcription factors at the single cell resolution, such as single-cell CUT&RUN (Hainer et al., 2019). Alternatively, single-cell Assay for Transposase-Accessible Chromatin (ATAC)-seq could be used to profile open chromatin regions and infer gene regulatory networks from transcription factor binding sites (Cusanovich et al., 2015; Cusanovich et al., 2018).

8.2.3. Contribution of RUNX1 to hormone regulated tissues

Beyond the haematopoietic system, RUNX1 is increasingly implicated in the homeostasis and tumorigenesis of hormone regulated tissues, suggesting a possible mechanistic link between hormones and RUNX1. Previous work focused mainly on the study of RUNX1 in mammary glands and breast cancer, but our study adds further evidence that RUNX1 may also be involved in the prostate, predominantly regulated by male hormones.

In the mammary gland, *Runx1* is expressed in both basal and luminal cells, with fluctuations during pregnancy and lactation (Blyth et al., 2010; Browne et al., 2015; McDonald et al., 2014; van Bragt et al., 2014). Specifically, RUNX1 is totally

absent in secretory alveolar cells, a luminal lineage involved in milk production. Also, deletion of *Runx1* led to a reduction of mature luminal cells, and RUNX1 was shown to repress the master regulator of alveolar cells *Elf5* and promote the expression of a mature ductal luminal transcriptional program (van Bragt et al., 2014). Here, a parallel seem to exist between the mammary and prostate glands, with higher RUNX1 expression in non-secretory prostate luminal cells of the proximal region, basal cells, and castrated cells. Protein secretion by luminal cells is dynamically regulated by hormone signalling, and in particular AR in the prostate. Since AR exerts also a DNA-binding function and can bind to RUNX1 promoter (Takayama et al., 2015), it would be interesting to profile both RUNX1 and AR activity in RUNX1+ prostate epithelial cells at homeostasis, and under hormone-deprived conditions. This would help understand how hormonal fluctuations influence the gene expression program regulated by RUNX1, and whether RUNX1 and AR cooperate to regulate gene expression. Also, since RUNX1 was highly expressed after surgical castration, understanding RUNX1 activity in castrate conditions may have relevance to hormone refractory tumours in which RUNX1 is expressed at high levels.

Despite apparent similarities between the mammary and prostate glands, the requirement for RUNX1 may also differ between the two tissues. In organoid cultures of primary prostate epithelial cells, no defects could be observed upon deletion of *Runx1*. This contrasts with the mammary setting, where RUNX1 was shown to prevent EMT in mammary epithelial cell lines, with its deletion leading to abnormal morphogenesis in 3D cultures (Hong et al., 2017; Pegg et al., 2019; Sokol et al., 2015; Wang et al., 2011b). Nevertheless, these divergent phenotypic observations could also be explained by differences in experimental conditions and do not entirely rule out a role for RUNX1 in regulating 3D morphogenesis of prostate organoids. Since RUNX1 is expressed in both contexts, genome-wide profiling of RUNX1 transcriptional activity during mammary and prostate organoid formation could help resolve whether it regulates a similar program in both tissues.

In breast cancer patient samples, RUNX1 expression was found to be higher in the triple-negative subgroup (Karn et al., 2011; Rody et al., 2011), and patients with

higher RUNX1 levels in triple-negative and ER-negative breast cancers were associated with poorer outcomes (Ferrari et al., 2014). Conversely, the present work suggests that RUNX1 expression levels are not predictive of PCa patient outcomes. Yet, a small number of patients diagnosed with high grade PCa expressed higher levels of RUNX1. Additional work should be performed to understand the reason behind this phenomenon. Indeed, while PCa has been stratified in a few histologically and molecularly distinct subtypes, it may be beneficial to refine this characterisation, and assess if the observation that a subgroup of tumours expresses high RUNX1 levels would have any clinical utility on larger and independent patient cohorts.

8.2.4. Role of RUNX1 in PCa

The mouse work presented in this thesis focused on the characterisation of RUNX1 expression in the normal mouse prostate. Indeed, it is essential to deepen our understanding of normal biology and physiology to better comprehend disease. However, this study undoubtedly lacks a functional analysis of RUNX1 in the context of PCa. The expression of RUNX1 was only assessed in archived human patient samples of PCa. Thus, it would now be important to complement this descriptive work in benign prostate cells by studying RUNX1 in models of PCa. First, existing mouse models of PCa could be screened for changes in RUNX1 expression in normal and transformed cells. Next, genetic manipulation of RUNX1 expression could be modelled with the generation of prostate-specific conditional knockout or overexpression mouse models, in combination with additional oncogenic events. These models could then be surgically castrated in order to mimic ADT and test whether RUNX1 plays a role in the development of castration resistant tumours. Since several populations of mature luminal cells were identified by scRNA-seq, it would also be interesting to direct those genetic manipulations using cell type specific promoters, including RUNX1+ PLCs, to assess their behaviour during prostate tumorigenesis and castration resistance in the presence or absence of RUNX1.

8.2.5. Relevance to the human prostate

The mouse was chosen as a model to study multiple aspects of RUNX1 biology in the prostate, but it is associated with inevitable limitations preventing direct translation of the main findings to the human prostate. In humans, the expression pattern of RUNX1 was more difficult to appreciate, notably due to the nature of the samples obtained from diseased individuals. Yet, some similarities in the expression pattern of RUNX1 have emerged between the mouse and human prostate. RUNX1 was notably enriched in the peri-urethral region and basal cells of both species. However, the relevance of high RUNX1 expression in regions of epithelial atrophy in human samples is unclear. Similarly, it is difficult to assess based on a single patient analysed after-ADT whether increased RUNX1 expression would be systematically observed in atrophic human prostates post-ADT. While this would mimic the increased RUNX1 expression observed after surgical castration in mice, it remains unclear if increased RUNX1 expression is functionally linked to the acquisition of a castration resistant phenotype, or if it is only circumstantial due to other concomitant stimuli during drastic tissue remodelling induced by castration. Additional differences in anatomy and cell type composition between mice and men warrant additional work to reconcile the biology of the mouse and human prostate. It would be particularly interesting to further investigate the relevance - if any - of RUNX1+ PLCs to the human prostate gland, in human prostate development, adult homeostasis, and disease.

8.3. Investigating the lineage hierarchy of the normal prostate epithelium

The relevance of the work presented in this thesis extends beyond RUNX1 biology. Studying the heterogeneity of the adult mouse prostate epithelium by scRNA-seq revealed the existence of several mature population of cells, in particular within the luminal compartment. In 2020, a number of concomitant studies have reported similar degree of cellular heterogeneity in the adult mouse prostate (Crowley et al., 2020; Guo et al., 2020; Joseph et al., 2020; Karthaus et al.,

2020). However, the origin of this heterogeneity and how it is established during development is largely unknown. The work presented in this thesis attempted to decipher the developmental origin of mature prostate lineages by combining single cell transcriptomic analyses of embryonic and adult stages. However, despite a correlation between an embryonic (C2) and adult (Lum-D) cluster of cells suggesting a parental relationship, the precise differentiation trajectories of individual lineages remain undetermined. Lineage tracing using canonical basal and luminal *Cre* drivers is considered the gold-standard to study long-term cellular differentiation in developing organisms, but it also lacks the possibility to track individual cells along the full hierarchical tree of differentiation. Although alternative methods have been used in specific systems, such as continuous time-lapse imaging of haematopoietic cells *in vitro* (Hoppe et al., 2016; Loeffler and Schroeder, 2019), these approaches are virtually impossible to implement in solid organs *in vivo*. An emerging scRNA-seq technology consists in the use of expressed barcodes to track single cell history (Bowling et al., 2020; He et al., 2020; Kong et al., 2020; Umkehrer et al., 2020; VanHorn and Morris, 2020; Weinreb et al., 2020). For example, the CRISPR array repair lineage tracing (CARLIN) mouse model can be used to induce “scaring” events in the expressed lineage barcodes arrays allowing partial reconstruction of cell lineage hierarchy (Bowling et al., 2020). While this technology has only emerged recently, it holds promises to refine our understanding of lineage differentiation. Similar approaches could be applied to study whole organ development, as well as lineage plasticity in organoid formation and castration-regeneration assays.

8.4. Mapping the phenotypic landscape of prostate tumours to inform clinical management

Independent studies of RUNX1 (this work) and NKX3.1 (Wang et al., 2009b) showed that these transcription factors were expressed in distinct cell types of the mouse prostate, with specific castration resistant and regenerative properties. However, neither had been investigated simultaneously in human PCa patient samples. This study suggests that RUNX1 has limited biomarker potential used either as a single marker or in combination with NKX3.1. Indeed, quantitative

analyses of the expression levels of these markers in PCa patient samples could not retrospectively serve to predict patient outcomes. Other candidates have previously been described to mark specific cell types, such as *Bmi1* (Yoo et al., 2016), *Sox2* (McAuley et al., 2019), *Ly6d* (Barros-Silva et al., 2018), *Tacstd2*/*Trop2* (Goldstein et al., 2008). But to date, neither of these has shown potential to be used on their own as clinical biomarkers for PCa progression or castration resistant disease. Thus, the discovery of a reliable biomarker to predict the emergence of castration resistant disease is still lacking. One of the main limitations for the discovery of individual markers reside in the extensive molecular and phenotypic heterogeneity of PCa. To circumvent this problem, the combinatorial analysis of several markers to define phenotypic subtypes of tumours has been proposed to have clinical predictive relevance (Federer-Gsponer et al., 2020). Along these lines, the emergence of high dimensional proteomic platforms such as CODEX (CO-Detection by indEXing, Akoya Biosciences) (Goltsev et al., 2018) or imaging mass cytometry (Giesen et al., 2014) open new avenues. Rather than investigating single candidates, large panels of antibodies can be designed to characterise their expression simultaneously in single cells and quantify the presence of phenotypically similar groups of cells irrespective of their mutational status. This approach should enable greater resolution to identify clusters of clinically relevant cell types responsible for the development of aggressive disease. The use of large cohorts of patient samples archived in TMAs could be exploited to test whether such approach would leverage current limitations associated with the use of single markers. Together with established diagnostic parameters, these data could be combined, with the ultimate aim to improve the clinical management of PCa.

Bibliography

Abate-Shen, C., Shen, M. M. and Gelmann, E. (2008). Integrating differentiation and cancer: The Nkx3.1 homeobox gene in prostate organogenesis and carcinogenesis. *Differentiation* **76**, 717–727.

Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., Cieslik, M., Benelli, M., Robinson, D., Van Allen, E. M., et al. (2019). Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci USA* **116**, 11428–11436.

Abrahamsson, P. A. (1999). Neuroendocrine cells in tumour growth of the prostate. *Endocr Relat Cancer* **6**, 503–519.

Abramowitz, M. C., Li, T., Buyyounouski, M. K., Ross, E., Uzzo, R. G., Pollack, A. and Horwitz, E. M. (2008). The Phoenix definition of biochemical failure predicts for overall survival in patients with prostate cancer. *Cancer* **112**, 55–60.

Agarwal, S., Hynes, P. G., Tillman, H. S., Lake, R., Abou-Kheir, W. G., Fang, L., Casey, O. M., Ameri, A. H., Martin, P. L., Yin, J. J., et al. (2015). Identification of Different Classes of Luminal Progenitor Cells within Prostate Tumors. *Cell Rep* **13**, 2147–2158.

Ahmed, H. U., El-Shater Bosaily, A., Brown, L. C., Gabe, R., Kaplan, R., Parmar, M. K., Collaco-Moraes, Y., Ward, K., Hindley, R. G., Freeman, A., et al. (2017). Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* **389**, 815–822.

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086.

Aikawa, Y., Nguyen, L. A., Isono, K., Takakura, N., Tagata, Y., Schmitz, M. L., Koseki, H. and Kitabayashi, I. (2006). Roles of HIPK1 and HIPK2 in AML1- and p300-dependent transcription, hematopoiesis and blood vessel formation. *EMBO J.* **25**, 3955–3965.

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *PNAS* **74**, 5350–5354.

Andreoiu, M. and Cheng, L. (2010). Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. *Hum Pathol* **41**, 781–793.

Angelo, M., Bendall, S. C., Finck, R., Hale, M. B., Hitzman, C., Borowsky, A. D., Levenson, R. M., Lowe, J. B., Liu, S. D., Zhao, S., et al. (2014). Multiplexed ion beam imaging (MIBI) of human breast tumors. *Nat Med* **20**, 436–442.

Antonarakis, E. S. and Armstrong, A. J. (2011). Evolving standards in the treatment of docetaxel-refractory castration-resistant prostate cancer. *Prostate Cancer Prostatic Dis.* **14**, 192–205.

Arora, K. and Barbieri, C. E. (2018). Molecular Subtypes of Prostate Cancer. *Curr Oncol Rep* **20**, 58.

Ayala, A. G., Ro, J. Y., Babaian, R., Troncoso, P. and Grignon, D. J. (1989). The prostatic capsule: does it exist? Its importance in the staging and treatment of prostatic carcinoma. *Am J Surg Pathol* **13**, 21–27.

Bae, S. C., Ogawa, E., Maruyama, M., Oka, H., Satake, M., Shigesada, K., Jenkins, N. A., Gilbert, D. J., Copeland, N. G. and Ito, Y. (1994). PEBP2 alpha B/mouse AML1 consists of multiple isoforms that possess differential transactivation potentials. *Mol Cell Biol* **14**, 3242–3252.

Baena, E., Shao, Z., Linn, D. E., Glass, K., Hamblen, M. J., Fujiwara, Y., Kim, J., Nguyen, M., Zhang, X., Godinho, F. J., et al. (2013). ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev.* **27**, 683–698.

Balk, S. P., Ko, Y.-J. and Bubley, G. J. (2003). Biology of prostate-specific antigen. *J Clin Oncol* **21**, 383–391.

Banach-Petrosky, W., Jessen, W. J., Ouyang, X., Gao, H., Rao, J., Quinn, J., Aronow, B. J. and Abate-Shen, C. (2007). Prolonged exposure to reduced levels of androgen accelerates prostate cancer progression in Nkx3.1; Pten mutant mice. *Cancer Res.* **67**, 9089–9096.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K. K., Carter, S. L., Frederick, A. M., Lawrence, M. S., Sivachenko, A. Y., Sougnez, C., Zou, L., et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409.

Bangsow, C., Rubins, N., Glusman, G., Bernstein, Y., Negreanu, V., Goldenberg, D., Lotem, J., Ben-Asher, E., Lancet, D., Levanon, D., et al. (2001). The RUNX3 gene--sequence, structure and regulated expression. *Gene* **279**, 221–232.

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., et al. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports* **7**, 16878.

Barbieri, C. E., Baca, S. C., Lawrence, M. S., Demichelis, F., Blattner, M., Theurillat, J.-P., White, T. A., Stojanov, P., Van Allen, E., Stransky, N., et al. (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689.

Barros-Silva, J. D., Linn, D. E., Steiner, I., Guo, G., Ali, A., Pakula, H., Ashton, G., Peset, I., Brown, M., Clarke, N. W., et al. (2018). Single-Cell Analysis Identifies LY6D as a Marker Linking Castration-Resistant Prostate Luminal Cells to Prostate Progenitors and Cancer. *Cell Rep* **25**, 3504-3518.e6.

Becker-André, M. and Hahlbrock, K. (1989). Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res* **17**, 9437–9446.

Bee, T., Liddiard, K., Swiers, G., Bickley, S. R. B., Vink, C. S., Jarratt, A., Hughes, J. R., Medvinsky, A. and de Bruijn, M. F. T. R. (2009). Alternative Runx1 promoter usage in mouse developmental hematopoiesis. *Blood Cells Mol. Dis.* **43**, 35–42.

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A. C., Li, P. S., Taniuchi, I. and de Bruijn, M. F. T. R. (2010). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. *Blood* **115**, 3042–3050.

Behrens, K., Triviai, I., Schwieger, M., Tekin, N., Alawi, M., Spohn, M., Indenbirken, D., Ziegler, M., Müller, U., Alexander, W. S., et al. (2016). Runx1 downregulates stem cell and megakaryocytic transcription programs that support niche interactions. *Blood* **127**, 3369–3381.

Bellissimo, D. C. and Speck, N. A. (2017). RUNX1 Mutations in Inherited and Sporadic Leukemia. *Front Cell Dev Biol* **5**, 111.

Bello, D., Webber, M. M., Kleinman, H. K., Wartinger, D. D. and Rhim, J. S. (1997). Androgen responsive adult human prostatic epithelial cell lines immortalized by human papillomavirus 18. *Carcinogenesis* **18**, 1215–1223.

Beltran, H., Prandi, D., Mosquera, J. M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B. V. S. K., Varambally, S., et al. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* **22**, 298–305.

Ben-Ami, O., Pencovich, N., Lotem, J., Levanon, D. and Groner, Y. (2009). A regulatory interplay between miR-27a and Runx1 during megakaryopoiesis. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 238–243.

Ben-Ami, O., Friedman, D., Leshkowitz, D., Goldenberg, D., Orlovsky, K., Pencovich, N., Lotem, J., Tanay, A. and Groner, Y. (2013). Addiction of

t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep* **4**, 1131–1143.

Bennett, B. D. and Gardner, W. A. (1985). Embryonal hyperplasia of the prostate. *The Prostate* **7**, 411–417.

Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstråhle, J., Tarish, F., Tanoglidi, A., Vickovic, S., Larsson, L., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* **9**, 2419.

Berman, D. M., Desai, N., Wang, X., Karhadkar, S. S., Reynon, M., Abate-Shen, C., Beachy, P. A. and Shen, M. M. (2004). Roles for Hedgehog signaling in androgen production and prostate ductal morphogenesis. *Dev. Biol.* **267**, 387–398.

Berquin, I. M., Min, Y., Wu, R., Wu, H. and Chen, Y. Q. (2005). Expression Signature of the Mouse Prostate. *J. Biol. Chem.* **280**, 36442–36451.

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S., Stainier, D. Y. R. and Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108–111.

Bethel, C. R. and Bieberich, C. J. (2007). Loss of Nkx3.1 expression in the transgenic adenocarcinoma of mouse prostate model. *The Prostate* **67**, 1740–1750.

Bethel, C. R., Faith, D., Li, X., Guan, B., Hicks, J. L., Lan, F., Jenkins, R. B., Bieberich, C. J. and Marzo, A. M. D. (2006). Decreased NKX3.1 Protein Expression in Focal Prostatic Atrophy, Prostatic Intraepithelial Neoplasia, and Adenocarcinoma: Association with Gleason Score and Chromosome 8p Deletion. *Cancer Res* **66**, 10683–10690.

Bhatia-Gaur, R., Donjacour, A. A., Sciavolino, P. J., Kim, M., Desai, N., Young, P., Norton, C. R., Gridley, T., Cardiff, R. D., Cunha, G. R., et al. (1999). Roles for Nkx3.1 in prostate development and cancer. *Genes Dev* **13**, 966–977.

Biggs, J. R., Peterson, L. F., Zhang, Y., Kraft, A. S. and Zhang, D.-E. (2006). AML1/RUNX1 Phosphorylation by Cyclin-Dependent Kinases Regulates the Degradation of AML1/RUNX1 by the Anaphase-Promoting Complex. *Mol Cell Biol* **26**, 7420–7429.

Binnerts, M. E., Kim, K.-A., Bright, J. M., Patel, S. M., Tran, K., Zhou, M., Leung, J. M., Liu, Y., Lomas, W. E., Dixon, M., et al. (2007). R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. *Proc Natl Acad Sci U S A* **104**, 14700–14705.

Blyth, K., Cameron, E. R. and Neil, J. C. (2005). The runx genes: gain or loss of function in cancer. *Nat Rev Cancer* **5**, 376–387.

Blyth, K., Vaillant, F., Jenkins, A., McDonald, L., Pringle, M. A., Huser, C., Stein, T., Neil, J. and Cameron, E. R. (2010). Runx2 in normal tissues and cancer cells: A developing story. *Blood Cells Mol. Dis.* **45**, 117–123.

Boisset, J.-C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E. and Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120.

Boorjian, S. A., Thompson, R. H., Siddiqui, S., Bagniewski, S., Bergstralh, E. J., Karnes, R. J., Frank, I. and Blute, M. L. (2007). Long-term outcome after radical prostatectomy for patients with lymph node positive prostate cancer in the prostate specific antigen era. *J Urol* **178**, 864–870; discussion 870-871.

Bostwick, D. G. (1989). Prostatic intraepithelial neoplasia (PIN). *Urology* **34**, 16–22.

Boutros, P. C., Fraser, M., Harding, N. J., de Borja, R., Trudel, D., Lalonde, E., Meng, A., Hennings-Yeomans, P. H., McPherson, A., Sabelnykova, V. Y., et al. (2015). Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* **47**, 736–745.

Bowen, C., Bubendorf, L., Voeller, H. J., Slack, R., Willi, N., Sauter, G., Gasser, T. C., Koivisto, P., Lack, E. E., Kononen, J., et al. (2000). Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. *Cancer Res* **60**, 6111–6115.

Bowling, S., Sritharan, D., Osorio, F. G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.-C., Fujiwara, Y., Li, B. E., et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* **181**, 1410-1422.e27.

Brady, G., Whiteman, H. J., Spender, L. C. and Farrell, P. J. (2009). Downregulation of RUNX1 by RUNX3 Requires the RUNX3 VWRPY Sequence and Is Essential for Epstein-Barr Virus-Driven B-Cell Proliferation. *Journal of Virology* **83**, 6909–6916.

Brady, G., Elgueta Karstegl, C. and Farrell, P. J. (2013). Novel function of the unique N-terminal region of RUNX1c in B cell growth regulation. *Nucleic Acids Res.* **41**, 1555–1568.

Brady, G., Barbara, M. and Iscove, N. N. Representative in Vitro cDNA Amplification From Individual Hemopoietic Cells and Colonies. 9.

Bratt, O., Drevin, L., Akre, O., Garmo, H. and Stattin, P. (2016). Family History and Probability of Prostate Cancer, Differentiated by Risk Category: A Nationwide Population-Based Study. *JNCI: Journal of the National Cancer Institute* **108**,.

Bravo, J., Li, Z., Speck, N. A. and Warren, A. J. (2001). The leukemia-associated AML1 (Runx1)--CBF beta complex functions as a DNA-induced molecular clamp. *Nat. Struct. Biol.* **8**, 371–378.

Bresciani, E., Carrington, B., Wincovitch, S., Jones, M., Gore, A. V., Weinstein, B. M., Sood, R. and Liu, P. P. (2014). CBFβ and RUNX1 are required at 2 different steps during the development of hematopoietic stem cells in zebrafish. *Blood* **124**, 70–78.

Browne, G., Taipaleenmäki, H., Bishop, N. M., Madasu, S. C., Shaw, L. M., van Wijnen, A. J., Stein, J. L., Stein, G. S. and Lian, J. B. (2015). Runx1 is associated with breast cancer progression in MMTV-PyMT transgenic mice and its depletion in vitro inhibits migration and invasion. *J. Cell. Physiol.* **230**, 2522–2532.

Burger, P. E., Xiong, X., Coetzee, S., Salm, S. N., Moscatelli, D., Goto, K. and Wilson, E. L. (2005). Sca-1 expression identifies stem cells in the proximal region of prostatic ducts with high capacity to reconstitute prostatic tissue. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7180–7185.

Burgess, D. J. (2019). Spatial transcriptomics coming of age. *Nature Reviews Genetics* **20**, 317–317.

Buyyounouski, M. K., Choyke, P. L., McKenney, J. K., Sartor, O., Sandler, H. M., Amin, M. B., Kattan, M. W. and Lin, D. W. (2017). Prostate cancer – major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA: A Cancer Journal for Clinicians* **67**, 245–253.

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, R. J. and Dzierzak, E. (2000). Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity* **13**, 423–431.

Cai, X., Gaudet, J. J., Mangan, J. K., Chen, M. J., De Obaldia, M. E., Oo, Z., Ernst, P. and Speck, N. A. (2011). Runx1 loss minimally impacts long-term hematopoietic stem cells. *PLoS ONE* **6**, e28430.

Cai, X., Gao, L., Teng, L., Ge, J., Oo, Z. M., Kumar, A. R., Gilliland, D. G., Mason, P. J., Tan, K. and Speck, N. A. (2015). Runx1 Deficiency Decreases Ribosome Biogenesis and Confers Stress Resistance to Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell* **17**, 165–177.

Camp, R. L., Chung, G. G. and Rimm, D. L. (2002). Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med* **8**, 1323–1327.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70.

Cancer Genome Atlas Research Network (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025.

Carceles-Cordon, M., Kelly, W. K., Gomella, L., Knudsen, K. E., Rodriguez-Bravo, V. and Domingo-Domenech, J. (2020). Cellular rewiring in lethal prostate cancer: the architect of drug resistance. *Nature Reviews Urology* **17**, 292–307.

Carm, K. T., Hoff, A. M., Bakken, A. C., Axcrona, U., Axcrona, K., Lothe, R. A., Skotheim, R. I. and Løvf, M. (2019). Interfocal heterogeneity challenges the clinical usefulness of molecular classification of primary prostate cancer. *Scientific Reports* **9**, 13579.

Carter, H. B., Helfand, B., Mamawala, M., Wu, Y., Landis, P., Yu, H., Wiley, K., Na, R., Shi, Z., Petkewicz, J., et al. (2019). Germline Mutations in ATM and BRCA1/2 Are Associated with Grade Reclassification in Men on Active Surveillance for Prostate Cancer. *European Urology* **75**, 743–749.

Castro, E., Goh, C., Olmos, D., Saunders, E., Leongamornlert, D., Tymrakiewicz, M., Mahmud, N., Dadaev, T., Govindasami, K., Guy, M., et al. (2013). Germline BRCA Mutations Are Associated With Higher Risk of Nodal Involvement, Distant Metastasis, and Poor Survival Outcomes in Prostate Cancer. *J Clin Oncol* **31**, 1748–1757.

Ceder, J. A., Aalders, T. W. and Schalken, J. A. (2017). Label retention and stem cell marker expression in the developing and adult prostate identifies basal and luminal epithelial stem cell subpopulations. *Stem Cell Res Ther* **8**, 95.

Centonze, A., Lin, S., Tika, E., Sifrim, A., Fioramonti, M., Malfait, M., Song, Y., Wuidart, A., Van Herck, J., Dannau, A., et al. (2020). Heterotypic cell–cell communication regulates glandular stem cell multipotency. *Nature* **584**, 608–613.

Challen, G. A. and Goodell, M. A. (2010). Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells. *Exp. Hematol.* **38**, 403–416.

Chan, T. E., Stumpf, M. P. H. and Babtie, A. C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Systems* **5**, 251-267.e3.

Chen, C. D., Welsbie, D. S., Tran, C., Baek, S. H., Chen, R., Vessella, R., Rosenfeld, M. G. and Sawyers, C. L. (2004). Molecular determinants of resistance to antiandrogen therapy. *Nat Med* **10**, 33–39.

Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. and Speck, N. A. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891.

Chimge, N.-O., Little, G. H., Baniwal, S. K., Adisetiyo, H., Xie, Y., Zhang, T., O'Laughlin, A., Liu, Z. Y., Ulrich, P., Martin, A., et al. (2016). RUNX1 prevents oestrogen-mediated AXIN1 suppression and β-catenin activation in ER-positive breast cancer. *Nat Commun* **7**, 10751.

Choi, N., Zhang, B., Zhang, L., Ittmann, M. and Xin, L. (2012). Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. *Cancer Cell* **21**, 253–265.

Chua, C. W., Shibata, M., Lei, M., Toivanen, R., Barlow, L. J., Bergren, S. K., Badani, K. K., McKiernan, J. M., Benson, M. C., Hibshoosh, H., et al. (2014). Single luminal epithelial progenitors can generate prostate organoids in culture. *Nat. Cell Biol.* **16**, 951–961, 1–4.

Chuang, L. S. H., Ito, K. and Ito, Y. (2013). RUNX family: Regulation and diversification of roles through interacting proteins. *Int. J. Cancer* **132**, 1260–1271.

Chung, D. D., Honda, K., Cafuir, L., McDuffie, M. and Wotton, D. (2007). The Runx3 distal transcript encodes an additional transcriptional activation domain. *FEBS J.* **274**, 3429–3439.

Clark, J. P. and Cooper, C. S. (2009). ETS gene fusions in prostate cancer. *Nat Rev Urol* **6**, 429–439.

Clevers, H. (2016). Modeling Development and Disease with Organoids. *Cell* **165**, 1586–1597.

Clevers, H. and Watt, F. M. (2018). Defining Adult Stem Cells by Function, not by Phenotype. *Annual Review of Biochemistry* **87**, 1015–1027.

Cohen, R. J., Shannon, B. A., Phillips, M., Moorin, R. E., Wheeler, T. M. and Garrett, K. L. (2008). Central zone carcinoma of the prostate gland: a distinct tumor type with poor prognostic features. *J Urol* **179**, 1762–1767; discussion 1767.

Collins, A. T., Berry, P. A., Hyde, C., Stower, M. J. and Maitland, N. J. (2005). Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res* **65**, 10946–10951.

Cook, C., Vezina, C. M., Allgeier, S. H., Shaw, A., Yu, M., Peterson, R. E. and Bushman, W. (2007). Noggin is required for normal lobe patterning and ductal budding in the mouse prostate. *Dev Biol* **312**, 217–230.

Corrò, C., Novellasdemunt, L. and Li, V. S. W. (2020). A brief history of organoids. *American Journal of Physiology-Cell Physiology* **319**, C151–C165.

Costa, G., Kouskoff, V. and Lacaud, G. (2012). Origin of blood cells and HSC production in the embryo. *Trends Immunol.* **33**, 215–223.

Crosby, D., Lyons, N., Greenwood, E., Harrison, S., Hiom, S., Moffat, J., Quallo, T., Samuel, E. and Walker, I. (2020). A roadmap for the early detection and diagnosis of cancer. *The Lancet Oncology* **21**, 1397–1399.

Crowell, P. D., Fox, J. J., Hashimoto, T., Diaz, J. A., Navarro, H. I., Henry, G. H., Feldmar, B. A., Lowe, M. G., Garcia, A. J., Wu, Y. E., et al. (2019). Expansion of Luminal Progenitor Cells in the Aging Mouse and Human Prostate. *Cell Reports* **28**, 1499-1510.e6.

Crowley, L., Cambuli, F., Aparicio, L., Shibata, M., Robinson, B. D., Xuan, S., Li, W., Hibshoosh, H., Loda, M., Rabadan, R., et al. (2020). A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *Elife* **9**,.

Culp, M. B., Soerjomataram, I., Efstathiou, J. A., Bray, F. and Jemal, A. (2020). Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur Urol* **77**, 38–52.

Cunha, G. R. (1975). Age-dependent loss of sensitivity of female urogenital sinus to androgenic conditions as a function of the epithelia-stromal interaction in mice. *Endocrinology* **97**, 665–673.

Cunha, G. R. and Lung, B. (1978). The possible influence of temporal factors in androgenic responsiveness of urogenital tissue recombinants from wild-type and androgen-insensitive (Tfm) mice. *J Exp Zool* **205**, 181–193.

Cunha, G. R., Donjacour, A. A., Cooke, P. S., Mee, S., Bigsby, R. M., Higgins, S. J. and Sugimura, Y. (1987). The endocrinology and developmental biology of the prostate. *Endocr. Rev.* **8**, 338–362.

Cunha, G. R., Vezina, C. M., Isaacson, D., Ricke, W. A., Timms, B. G., Cao, M., Franco, O. and Baskin, L. S. (2018). Development of the human prostate. *Differentiation* **103**, 24–45.

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914.

Cusanovich, D. A., Reddington, J. P., Garfield, D. A., Daza, R. M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H. A., Christiansen, L., Qiu, X., Steemers, F. J., et

al. (2018). The cis -regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542.

D'Amico, A. V., Whittington, R., Malkowicz, S. B., Schultz, D., Blank, K., Broderick, G. A., Tomaszewski, J. E., Renshaw, A. A., Kaplan, I., Beard, C. J., et al. (1998). Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* **280**, 969–974.

Daskivich, T. J., Chamie, K., Kwan, L., Labo, J., Palvolgyi, R., Dash, A., Greenfield, S. and Litwin, M. S. (2011). Overtreatment of men with low-risk prostate cancer and significant comorbidity. *Cancer* **117**, 2058–2066.

Dauge, M. C., Delmas, V. and Mandarim de Lacerda, C. A. (1986). [Development of the human prostate during the first stages of fetal life. Morphometric study]. *Bull Assoc Anat (Nancy)* **70**, 5–11.

Davidsson, S., Fiorentino, M., Andrén, O., Fang, F., Mucci, L. A., Varenhorst, E., Fall, K. and Rider, J. R. (2011). Inflammation, focal atrophic lesions, and prostatic intraepithelial neoplasia with respect to risk of lethal prostate cancer. *Cancer Epidemiol. Biomarkers Prev.* **20**, 2280–2287.

Davis, I. D., Martin, A. J., Stockler, M. R., Begbie, S., Chi, K. N., Chowdhury, S., Coskinas, X., Frydenberg, M., Hague, W. E., Horvath, L. G., et al. (2019). Enzalutamide with Standard First-Line Therapy in Metastatic Prostate Cancer. *New England Journal of Medicine*.

de Lau, W., Barker, N., Low, T. Y., Koo, B.-K., Li, V. S. W., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P. J., van de Wetering, M., et al. (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476**, 293–297.

De Marzo, A. M., Meeker, A. K., Epstein, J. I. and Coffey, D. S. (1998). Prostate stem cell compartments: expression of the cell cycle inhibitor p27Kip1 in normal, hyperplastic, and neoplastic cells. *Am J Pathol* **153**, 911–919.

De Marzo, A. M., Marchi, V. L., Epstein, J. I. and Nelson, W. G. (1999). Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. *Am J Pathol* **155**, 1985–1992.

De Marzo, A. M., Meeker, A. K., Zha, S., Luo, J., Nakayama, M., Platz, E. A., Isaacs, W. B. and Nelson, W. G. (2003). Human prostate cancer precursors and pathobiology. *Urology* **62**, 55–62.

De Marzo, A. M., Platz, E. A., Sutcliffe, S., Xu, J., Grönberg, H., Drake, C. G., Nakai, Y., Isaacs, W. B. and Nelson, W. G. (2007). Inflammation in prostate carcinogenesis. *Nat Rev Cancer* **7**, 256–269.

Ding, Z., Wu, C.-J., Chu, G. C., Xiao, Y., Ho, D., Zhang, J., Perry, S. R., Labrot, E. S., Wu, X., Lis, R., et al. (2011). SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature* **470**, 269–273.

Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden, K., Taylor, J. M. G., Enkemann, S. A., Tsao, M.-S., Yeatman, T. J., Gerald, W. L., Eschrich, S., Jurisica, I., Giordano, T. J., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822–827.

Doles, J. D., Vezina, C. M., Lipinski, R. J., Peterson, R. E. and Bushman, W. (2005). Growth, morphogenesis, and differentiation during mouse prostate development in situ, in renal grafts, and in vitro. *Prostate* **65**, 390–399.

Doll, A., Gonzalez, M., Abal, M., Llaurado, M., Rigau, M., Colas, E., Monge, M., Xercavins, J., Capella, G., Diaz, B., et al. (2009). An orthotopic endometrial cancer mouse model demonstrates a role for RUNX1 in distant metastasis. *Int. J. Cancer* **125**, 257–263.

Donjacour, A. A. and Cunha, G. R. (1993). Assessment of prostatic protein secretion in tissue recombinants made of urogenital sinus mesenchyme and urothelium from normal or androgen-insensitive mice. *Endocrinology* **132**, 2342–2350.

Donovan, J. L., Hamdy, F. C., Lane, J. A., Mason, M., Metcalfe, C., Walsh, E., Blazeby, J. M., Peters, T. J., Holding, P., Bonnington, S., et al. (2016). Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *New England Journal of Medicine* **375**, 1425–1437.

Draper, J. E., Sroczynska, P., Tsoulaki, O., Leong, H. S., Fadlullah, M. Z. H., Miller, C., Kouskoff, V. and Lacaud, G. (2016). RUNX1B Expression Is Highly Heterogeneous and Distinguishes Megakaryocytic and Erythroid Lineage Fate in Adult Mouse Hematopoiesis. *PLoS Genet.* **12**, e1005814.

Draper, J. E., Sroczynska, P., Leong, H. S., Fadlullah, M. Z. H., Miller, C., Kouskoff, V. and Lacaud, G. (2017). Mouse RUNX1C regulates premegakaryocytic/erythroid output and maintains survival of megakaryocyte progenitors. *Blood* **130**, 271–284.

Draper, J. E., Sroczynska, P., Fadlullah, M. Z. H., Patel, R., Newton, G., Breitwieser, W., Kouskoff, V. and Lacaud, G. (2018). A novel prospective isolation of murine fetal liver progenitors to study in utero hematopoietic defects. *PLoS Genet.* **14**, e1007127.

Drissi, H., Luc, Q., Shakoori, R., Chuva De Sousa Lopes, S., Choi, J. Y., Terry, A., Hu, M., Jones, S., Neil, J. C., Lian, J. B., et al. (2000). Transcriptional

autoregulation of the bone related CBFA1/RUNX2 gene. *J. Cell. Physiol.* **184**, 341–350.

Drost, J., Karthaus, W. R., Gao, D., Driehuis, E., Sawyers, C. L., Chen, Y. and Clevers, H. (2016). Organoid culture systems for prostate epithelial and cancer tissue. *Nat. Protocols* **11**, 347–358.

Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. and Karsenty, G. (1997). Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754.

Ducy, P., Starbuck, M., Priemel, M., Shen, J., Pinero, G., Geoffroy, V., Amling, M. and Karsenty, G. (1999). A Cbfa1-dependent genetic pathway controls bone formation beyond embryonic development. *Genes & Development* **13**, 1025–1036.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. and Coleman, P. (1992). Analysis of gene expression in single live neurons. *PNAS* **89**, 3010–3014.

Eilken, H. M., Nishikawa, S.-I. and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature* **457**, 896–900.

Elagib, K. E. (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* **101**, 4333–4341.

Elbadawy, M., Abugomaa, A., Yamawaki, H., Usui, T. and Sasaki, K. (2020). Development of Prostate Cancer Organoid Culture Models in Basic Medicine and Translational Research. *Cancers (Basel)* **12**,.

Eliades, A., Wareing, S., Marinopoulou, E., Fadlullah, M. Z. H., Patel, R., Grabarek, J. B., Plusa, B., Lacaud, G. and Kouskoff, V. (2016). The Hemogenic Competence of Endothelial Progenitors Is Restricted by Runx1 Silencing during Embryonic Development. *Cell Rep* **15**, 2185–2199.

Ellis, M. J., Ding, L., Shen, D., Luo, J., Suman, V. J., Wallis, J. W., Van Tine, B. A., Hoog, J., Goiffon, R. J., Goldstein, T. C., et al. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360.

English, H. F., Santen, R. J. and Isaacs, J. T. (1987). Response of glandular versus basal rat ventral prostatic epithelial cells to androgen withdrawal and replacement. *Prostate* **11**, 229–242.

Epstein, J. I. and Armas, O. A. (1992). Atypical Basal Cell Hyperplasia of the Prostate. *The American Journal of Surgical Pathology* **16**, 1205–1214.

Epstein, J. I., Allsbrook, W. C., Amin, M. B., Egevad, L. L., and ISUP Grading Committee (2005). The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am. J. Surg. Pathol.* **29**, 1228–1242.

Epstein, J. I., Egevad, L., Amin, M. B., Delahunt, B., Srigley, J. R., Humphrey, P. A., and Grading Committee (2016a). The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* **40**, 244–252.

Epstein, J. I., Zelefsky, M. J., Sjoberg, D. D., Nelson, J. B., Egevad, L., Magi-Galluzzi, C., Vickers, A. J., Parwani, A. V., Reuter, V. E., Fine, S. W., et al. (2016b). A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol* **69**, 428–435.

Ewing, C. M., Ray, A. M., Lange, E. M., Zuhlke, K. A., Robbins, C. M., Tembe, W. D., Wiley, K. E., Isaacs, S. D., Johng, D., Wang, Y., et al. (2012). Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* **366**, 141–149.

Fatehullah, A., Tan, S. H. and Barker, N. (2016). Organoids as an in vitro model of human development and disease. *Nature Cell Biology* **18**, 246–254.

Fedchenko, N. and Reifenrath, J. (2014). Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. *Diagn Pathol* **9**, 221.

Federer-Gsponer, J. R., Müller, D. C., Zellweger, T., Eggimann, M., Marston, K., Ruiz, C., Seifert, H.-H., Rentsch, C. A., Bubendorf, L. and Magnen, C. L. (2020). Patterns of stemness-associated markers in the development of castration-resistant prostate cancer. *The Prostate* **80**, 1108–1117.

Feil, R., Wagner, J., Metzger, D. and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* **237**, 752–757.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D. Y., Barbosa, I. A. M., Kwon, J. S., Guan, Y., et al. (2013). An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep* **5**, 1704–1713.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359-386.

Ferrari, N., Mohammed, Z. M. A., Nixon, C., Mason, S. M., Mallon, E., McMillan, D. C., Morris, J. S., Cameron, E. R., Edwards, J. and Blyth, K. (2014). Expression

of RUNX1 Correlates with Poor Patient Prognosis in Triple Negative Breast Cancer. *PLoS ONE* **9**, e100759.

Fijneman, R. J. A., Anderson, R. A., Richards, E., Liu, J., Tijssen, M., Meijer, G. A., Anderson, J., Rod, A., O'Sullivan, M. G., Scott, P. M., et al. (2012). Runx1 is a tumor suppressor gene in the mouse gastrointestinal tract. *Cancer Sci.* **103**, 593–599.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16**, 278.

Frame, J. M., Fegan, K. H., Conway, S. J., McGrath, K. E. and Palis, J. (2016). Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity. *Stem Cells* **34**, 431–444.

Freedland, S. J., Sutter, M. E., Dorey, F. and Aronson, W. J. (2003). Defining the ideal cutpoint for determining PSA recurrence after radical prostatectomy. Prostate-specific antigen. *Urology* **61**, 365–369.

Freitas, D. M. O., Andriole, G. L., Freedland, S. J., Neto, B. S. and Moreira, D. M. (2019). Baseline Basal Cell Hyperplasia Is not Associated With Baseline Lower Urinary Tract Symptoms, Baseline Clinical Prostatitis or Prostate Cancer in Repeat Biopsies. *Urology* **129**, 160–164.

Fu, Y., Sun, S., Man, X. and Kong, C. (2019). Increased expression of RUNX1 in clear cell renal cell carcinoma predicts poor prognosis. *PeerJ* **7**,.

Fujita, Y., Nishimura, M., Taniwaki, M., Abe, T. and Okuda, T. (2001). Identification of an alternatively spliced form of the mouse AML1/RUNX1 gene transcript AML1c and its expression in early hematopoietic development. *Biochem. Biophys. Res. Commun.* **281**, 1248–1255.

Fukushima-Nakase, Y. (2005). Shared and distinct roles mediated through C-terminal subdomains of acute myeloid leukemia/Runt-related transcription factor molecules in murine development. *Blood* **105**, 4298–4307.

Gao, D., Vela, I., Sboner, A., Iaquinta, P. J., Karthaus, W. R., Gopalan, A., Dowling, C., Wanjala, J. N., Undvall, E. A., Arora, V. K., et al. (2014). Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176–187.

Geoffroy, V., Corral, D. A., Zhou, L., Lee, B. and Karsenty, G. (1998). Genomic organization, expression of the human CBFA1 gene, and evidence for an

alternative splicing event affecting protein function. *Mamm. Genome* **9**, 54–57.

Georgas, K. M., Armstrong, J., Keast, J. R., Larkins, C. E., McHugh, K. M., Southard-Smith, E. M., Cohn, M. J., Batourina, E., Dan, H., Schneider, K., et al. (2015). An illustrated anatomical ontology of the developing mouse lower urogenital tract. *Development* **142**, 1893–1908.

Gergen, J. P. and Butler, B. A. (1988). Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. *Genes Dev.* **2**, 1179–1193.

Gerhauser, C., Favero, F., Risch, T., Simon, R., Feuerbach, L., Assenov, Y., Heckmann, D., Sidiropoulos, N., Waszak, S. M., Hübschmann, D., et al. (2018). Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell* **34**, 996-1011.e8.

Ghozi, M. C., Bernstein, Y., Negreanu, V., Levanon, D. and Groner, Y. (1996). Expression of the human acute myeloid leukemia gene AMLI is regulated by two promoter regions. *Proc. Natl. Acad. Sci. USA* 6.

Gierahn, T. M., Wadsworth, M. H., Hughes, T. K., Bryson, B. D., Butler, A., Satija, R., Fortune, S., Love, J. C. and Shalek, A. K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature Methods* **14**, 395–398.

Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolimund, D., Buhmann, J. M., Brandt, S., et al. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* **11**, 417–422.

Gillessen, S., Omlin, A., Attard, G., de Bono, J. S., Efstathiou, E., Fizazi, K., Halabi, S., Nelson, P. S., Sartor, O., Smith, M. R., et al. (2015). Management of patients with advanced prostate cancer: recommendations of the St Gallen Advanced Prostate Cancer Consensus Conference (APCCC) 2015. *Ann Oncol* **26**, 1589–1604.

Gillessen, S., Attard, G., Beer, T. M., Beltran, H., Bjartell, A., Bossi, A., Briganti, A., Bristow, R. G., Chi, K. N., Clarke, N., et al. (2020). Management of Patients with Advanced Prostate Cancer: Report of the Advanced Prostate Cancer Consensus Conference 2019. *Eur Urol* **77**, 508–547.

Gleason, D. F. and Mellinger, G. T. (1974). Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J. Urol.* **111**, 58–64.

Goh, Y.-M., Cinghu, S., Hong, E. T. H., Lee, Y.-S., Kim, J.-H., Jang, J.-W., Li, Y.-H., Chi, X.-Z., Lee, K.-S., Wee, H., et al. (2010). Src kinase phosphorylates RUNX3 at tyrosine residues and localizes the protein in the cytoplasm. *J. Biol. Chem.* **285**, 10122–10129.

Goldstein, A. S., Lawson, D. A., Cheng, D., Sun, W., Garraway, I. P. and Witte, O. N. (2008). Trop2 identifies a subpopulation of murine and human prostate basal cells with stem cell characteristics. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20882–20887.

Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S. and Nolan, G. P. (2018). Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15.

Goto, K., Salm, S. N., Coetzee, S., Xiong, X., Burger, P. E., Shapiro, E., Lepor, H., Moscatelli, D. and Wilson, E. L. (2006). Proximal Prostatic Stem Cells Are Programmed to Regenerate a Proximal-Distal Ductal Axis. *STEM CELLS* **24**, 1859–1868.

Goyama, S., Yamaguchi, Y., Imai, Y., Kawazu, M., Nakagawa, M., Asai, T., Kumano, K., Mitani, K., Ogawa, S., Chiba, S., et al. (2004). The transcriptionally active form of AML1 is required for hematopoietic rescue of the AML1-deficient embryonic para-aortic splanchnopleural (P-Sp) region. *Blood* **104**, 3558–3564.

Goyama, S., Schibler, J., Cunningham, L., Zhang, Y., Rao, Y., Nishimoto, N., Nakagawa, M., Olsson, A., Wunderlich, M., Link, K. A., et al. (2013). Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest* **123**, 3876–3888.

Goyama, S., Huang, G., Kurokawa, M. and Mulloy, J. C. (2015). Posttranslational modifications of RUNX1 as potential anticancer targets. *Oncogene* **34**, 3483–3492.

Grasso, C. S., Wu, Y.-M., Robinson, D. R., Cao, X., Dhanasekaran, S. M., Khan, A. P., Quist, M. J., Jing, X., Lonigro, R. J., Brenner, J. C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243.

Greenberg, N. M., DeMayo, F., Finegold, M. J., Medina, D., Tilley, W. D., Aspinall, J. O., Cunha, G. R., Donjacour, A. A., Matusik, R. J. and Rosen, J. M. (1995). Prostate cancer in a transgenic mouse. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3439–3443.

Grignon, D. J. (2004). Unusual subtypes of prostate cancer. *Mod Pathol* **17**, 316–327.

Groner, Y. (2017). *RUNX proteins in development and cancer*. New York, NY: Springer Berlin Heidelberg.

Growney, J. D., Shigematsu, H., Li, Z., Lee, B. H., Adelsperger, J., Rowan, R., Curley, D. P., Kutok, J. L., Akashi, K., Williams, I. R., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* **106**, 494–504.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H. and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255.

Guo, H. and Friedman, A. D. (2011). Phosphorylation of RUNX1 by cyclin-dependent kinase reduces direct interaction with HDAC1 and HDAC3. *J. Biol. Chem.* **286**, 208–215.

Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D. and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18**, 675–685.

Guo, W., Li, L., He, J., Liu, Z., Han, M., Li, F., Xia, X., Zhang, X., Zhu, Y., Wei, Y., et al. (2020). Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nature Genetics* **52**, 908–918.

Gupta, M. (2020). Management of Men With Lymph Node Metastases Following Radical Prostatectomy: What Is the Optimal Treatment Strategy? *Rev Urol* **22**, 37–39.

Haghverdi, L., Buettner, F. and Theis, F. J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998.

Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. and Fazzio, T. G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.e11.

Hamidi, S. and Sheng, G. (2018). Epithelial-mesenchymal transition in haematopoietic stem cell development and homeostasis. *J. Biochem.* **164**, 265–275.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107.e17.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* **144**, 646–674.

Harada, H., Tagashira, S., Fujiwara, M., Ogawa, S., Katsumata, T., Yamaguchi, A., Komori, T. and Nakatsuka, M. (1999). Cbfa1 isoforms exert functional differences in osteoblast differentiation. *J. Biol. Chem.* **274**, 6972–6978.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K. J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology* **17**, 77.

He, Z., Gerber, T., Maynard, A., Jain, A., Petri, R., Santel, M., Ly, K., Sidow, L., Callejal, F. S., Riesenberg, S., et al. (2020). Lineage recording reveals dynamics of cerebral organoid regionalization. *bioRxiv* 2020.06.19.162032.

Henry, G. H., Malewska, A., Joseph, D. B., Malladi, V. S., Lee, J., Torrealba, J., Mauck, R. J., Gahan, J. C., Raj, G. V., Roehrborn, C. G., et al. (2018). A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Reports* **25**, 3530-3542.e5.

Herglotz, J., Kuvardina, O. N., Kolodziej, S., Kumar, A., Hussong, H., Grez, M. and Lausen, J. (2013). Histone arginine methylation keeps RUNX1 target genes in an intermediate state. *Oncogene* **32**, 2565–2575.

Hicks, S. C., Townes, F. W., Teng, M. and Irizarry, R. A. (2018). Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578.

Hidalgo, M., Amant, F., Biankin, A. V., Budinská, E., Byrne, A. T., Caldas, C., Clarke, R. B., de Jong, S., Jonkers, J., Mælandsmo, G. M., et al. (2014). Patient Derived Xenograft Models: An Emerging Platform for Translational Cancer Research. *Cancer Discov* **4**, 998–1013.

Higuchi, M., O'Brien, D., Kumaravelu, P., Lenny, N., Yeoh, E.-J. and Downing, J. R. (2002). Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell* **1**, 63–74.

Hofman, M. S., Lawrentschuk, N., Francis, R. J., Tang, C., Vela, I., Thomas, P., Rutherford, N., Martin, J. M., Frydenberg, M., Shakher, R., et al. (2020). Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *The Lancet* **395**, 1208–1216.

Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics* **7**, 200–210.

Hoi, C. S. L., Lee, S. E., Lu, S.-Y., McDermitt, D. J., Osorio, K. M., Piskun, C. M., Peters, R. M., Paus, R. and Tumbar, T. (2010). Runx1 Directly Promotes

Proliferation of Hair Follicle Stem Cells and Epithelial Tumor Formation in Mouse Skin. *Mol Cell Biol* **30**, 2518–2536.

Hong, D., Messier, T. L., Tye, C. E., Dobson, J. R., Fritz, A. J., Sikora, K. R., Browne, G., Stein, J. L., Lian, J. B. and Stein, G. S. (2017). Runx1 stabilizes the mammary epithelial cell phenotype and prevents epithelial to mesenchymal transition. *Oncotarget* **8**, 17610–17627.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A., et al. (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood* **114**, 299–309.

Hoppe, P. S., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K. D., Hilsenbeck, O., Moritz, N., Endele, M., Filipczyk, A., Gambardella, A., Ahmed, N., et al. (2016). Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* **535**, 299–302.

Horoszewicz, J. S., Leong, S. S., Kawinski, E., Karr, J. P., Rosenthal, H., Chu, T. M., Mirand, E. A. and Murphy, G. P. (1983). LNCaP model of human prostatic carcinoma. *Cancer Res.* **43**, 1809–1818.

Horton, C., Liu, Y., Yu, C., Xie, Q. and Wang, Z. A. (2019). Luminal-contact-inhibition of epithelial basal stem cell multipotency in prostate organogenesis and homeostasis. *Biol Open* **8**,.

Hu, Y., Huang, K., An, Q., Du, G., Hu, G., Xue, J., Zhu, X., Wang, C.-Y., Xue, Z. and Fan, G. (2016). Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biology* **17**, 88.

Huang, G., Shigesada, K., Ito, K., Wee, H.-J., Yokomizo, T. and Ito, Y. (2001). Dimerization with PEBP2β protects RUNX1/AML1 from ubiquitin–proteasome-mediated degradation. *EMBO J* **20**, 723–733.

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J. D., et al. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat. Genet.* **40**, 51–60.

Huang, S.-P., Lan, Y.-H., Lu, T.-L., Pao, J.-B., Chang, T.-Y., Lee, H.-Z., Yang, W.-H., Hsieh, C.-J., Chen, L.-M., Huang, L.-C., et al. (2011). Clinical significance of runt-related transcription factor 1 polymorphism in prostate cancer. *BJU International* **107**, 486–492.

Huang, H., Woo, A. J., Waldon, Z., Schindler, Y., Moran, T. B., Zhu, H. H., Feng, G.-S., Steen, H. and Cantor, A. B. (2012). A Src family kinase-Shp2 axis controls

RUNX1 activity in megakaryocyte and T-lymphocyte differentiation. *Genes Dev.* **26**, 1587–1601.

Hudson, D. L., Guy, A. T., Fry, P., O'Hare, M. J., Watt, F. M. and Masters, J. R. W. (2001). Epithelial Cell Differentiation Pathways in the Human Prostate: Identification of Intermediate Phenotypes by Keratin Expression. *J Histochem Cytochem.* **49**, 271–278.

Huggins C, H. C. (1941). The effect of castration, of estrogens, and of androgen injection on serum phosphatase in metastatic carcinoma of prostate. **Cancer Res**, 293–297.

Humphrey, P. A. (2007). Diagnosis of adenocarcinoma in prostate needle biopsy tissue. *J. Clin. Pathol.* **60**, 35–42.

Hyde, R. K., Zhao, L., Alemu, L. and Liu, P. P. (2015). Runx1 is required for hematopoietic defects and leukemogenesis in Cbfb-MYH11 knock-in mice. *Leukemia* **29**, 1771–1778.

Ichikawa, M., Asai, T., Saito, T., Yamamoto, G., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Hirai, H., et al. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nature Medicine* **10**, 299–304.

Ilic, D., Djulbegovic, M., Jung, J. H., Hwang, E. C., Zhou, Q., Cleves, A., Agoritsas, T. and Dahm, P. (2018). Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ* **362**, k3519.

Illendula, A., Gilmour, J., Grembecka, J., Tirumala, V. S. S., Boulton, A., Kuntimaddi, A., Schmidt, C., Wang, L., Pulikkan, J. A., Zong, H., et al. (2016). Small Molecule Inhibitor of CBFβ-RUNX Binding for RUNX Transcription Factor Driven Cancers. *EBioMedicine* **8**, 117–131.

Imai, Y., Kurokawa, M., Tanaka, K., Friedman, A. D., Ogawa, S., Mitani, K., Yazaki, Y. and Hirai, H. (1998). TLE, the human homolog of groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. *Biochem. Biophys. Res. Commun.* **252**, 582–589.

Imai, Y., Kurokawa, M., Yamaguchi, Y., Izutsu, K., Nitta, E., Mitani, K., Satake, M., Noda, T., Ito, Y. and Hirai, H. (2004). The corepressor mSin3A regulates phosphorylation-induced activation, intranuclear location, and stability of AML1. *Mol. Cell. Biol.* **24**, 1033–1043.

Inman, J. L., Robertson, C., Mott, J. D. and Bissell, M. J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, 1028–1042.

Inoue, K., Ozaki, S., Shiga, T., Ito, K., Masuda, T., Okado, N., Iseda, T., Kawaguchi, S., Ogawa, M., Bae, S.-C., et al. (2002). Runx3 controls the axonal projection of proprioceptive dorsal root ganglion neurons. *Nat. Neurosci.* **5**, 946–954.

Irani, J., Goujon, J. M., Ragni, E., Peyrat, L., Hubert, J., Saint, F. and Mottet, N. (1999). High-grade inflammation in prostate cancer as a prognostic factor for biochemical recurrence after radical prostatectomy. Pathologist Multi Center Study Group. *Urology* **54**, 467–472.

Ito, Y., Bae, S.-C. and Chuang, L. S. H. (2015). The RUNX family: developmental regulators in cancer. *Nat. Rev. Cancer* **15**, 81–95.

Ittmann, M. (2018). Anatomy and Histology of the Human and Murine Prostate. *Cold Spring Harb Perspect Med* **8**,.

Ittmann, M., Huang, J., Radaelli, E., Martin, P., Signoretti, S., Sullivan, R., Simons, B. W., Ward, J. M., Robinson, B. D., Chu, G. C., et al. (2013). Animal models of human prostate cancer: The Consensus Report of the New York Meeting of the Mouse Models of Human Cancers Consortium Prostate Pathology Committee. *Cancer Res* **73**, 2718–2736.

Ivanova, A., Signore, M., Caro, N., Greene, N. D. E., Copp, A. J. and Martinez-Barbera, J. P. (2005). In Vivo Genetic Ablation by Cre-Mediated Expression of Diphtheria Toxin Fragment A. *Genesis* **43**, 129–135.

Iwafuchi-Doi, M. and Zaret, K. S. (2016). Cell fate control by pioneer transcription factors. *Development* **143**, 1833–1837.

Jacob, B., Osato, M., Yamashita, N., Wang, C. Q., Taniuchi, I., Littman, D. R., Asou, N. and Ito, Y. (2010). Stem cell exhaustion due to Runx1 deficiency is prevented by Evi5 activation in leukemogenesis. *Blood* **115**, 1610–1620.

Jaffredo, T., Gautier, R., Eichmann, A. and Dieterlen-Lièvre, F. (1998). Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development* **125**, 4575–4583.

Jain, P., Nattakom, M., Holowka, D., Wang, D. H., Thomas Brenna, J., Ku, A. T., Nguyen, H., Ibrahim, S. F. and Tumbar, T. (2018). Runx1 Role in Epithelial and Cancer Cell Proliferation Implicates Lipid Metabolism and Scd1 and Soat1 Activity: Runx1 Mediates Proliferation Via Changes in Lipid Metabolism. *STEM CELLS* **36**, 1603–1616.

Jamaspishvili, T., Berman, D. M., Ross, A. E., Scher, H. I., De Marzo, A. M., Squire, J. A. and Lotan, T. L. (2018). Clinical implications of PTEN loss in prostate cancer. *Nat Rev Urol* **15**, 222–234.

Jeon, E.-J., Lee, K.-Y., Choi, N.-S., Lee, M.-H., Kim, H.-N., Jin, Y.-H., Ryoo, H.-M., Choi, J.-Y., Yoshida, M., Nishino, N., et al. (2006). Bone morphogenetic protein-2 stimulates Runx2 acetylation. *J. Biol. Chem.* **281**, 16502–16511.

Jhavar, S., Corbishley, C. M., Dearnaley, D., Fisher, C., Falconer, A., Parker, C., Eeles, R. and Cooper, C. S. (2005). Construction of tissue microarrays from prostate needle biopsy specimens. *Br J Cancer* **93**, 478–482.

Jin, Y.-H., Jeon, E.-J., Li, Q.-L., Lee, Y. H., Choi, J.-K., Kim, W.-J., Lee, K.-Y. and Bae, S.-C. (2004). Transforming growth factor-beta stimulates p300-dependent RUNX3 acetylation, which inhibits ubiquitination-mediated degradation. *J. Biol. Chem.* **279**, 29409–29417.

Johnson, D. C., Raman, S. S., Mirak, S. A., Kwan, L., Bajgiran, A. M., Hsu, W., Maehara, C. K., Ahuja, P., Faiena, I., Pooli, A., et al. (2019). Detection of Individual Prostate Cancer Foci via Multiparametric Magnetic Resonance Imaging. *Eur Urol* **75**, 712–720.

Jones, D., Friend, C., Dreher, A., Allgar, V. and Macleod, U. (2018). The diagnostic test accuracy of rectal examination for prostate cancer diagnosis in symptomatic patients: a systematic review. *BMC Fam Pract* **19**, 79.

Joseph, D. B., Henry, G. H., Malewska, A., Iqbal, N. S., Ruetten, H. M., Turco, A. E., Abler, L. L., Sandhu, S. K., Cadena, M. T., Malladi, V. S., et al. (2020). Urethral luminal epithelia are castration-insensitive cells of the proximal prostate. *Prostate*.

Joshi, A., Roberts, M. J., Alinezhad, S., Williams, E. D. and Vela, I. (2020). Challenges, applications and future directions of precision medicine in prostate cancer - the role of organoids and patient-derived xenografts. *BJU Int* **126**, 65–72.

Kadota, M., Yang, H. H., Gomez, B., Sato, M., Clifford, R. J., Meerzaman, D., Dunn, B. K., Wakefield, L. M. and Lee, M. P. (2010). Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. *PLoS ONE* **5**, e9201.

Kaighn, M. E., Narayan, K. S., Ohnuki, Y., Lechner, J. F. and Jones, L. W. (1979). Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest Urol* **17**, 16–23.

Kamachi, Y., Ogawa, E., Asano, M., Ishida, S., Murakami, Y., Satake, M., Ito, Y. and Shigesada, K. (1990). Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J. Virol.* **64**, 4808–4819.

Kamikubo, Y. (2018). Genetic compensation of RUNX family transcription factors in leukemia. *Cancer Science* **109**, 2358–2363.

Karn, T., Pusztai, L., Holtrich, U., Iwamoto, T., Shiang, C. Y., Schmidt, M., Müller, V., Solbach, C., Gaetje, R., Hanker, L., et al. (2011). Homogeneous Datasets of Triple Negative Breast Cancers Enable the Identification of Novel Prognostic and Predictive Signatures. *PLOS ONE* **6**, e28403.

Karthaus, W. R., Iaquinta, P. J., Drost, J., Gracanin, A., van Boxtel, R., Wongvipat, J., Dowling, C. M., Gao, D., Begthel, H., Sachs, N., et al. (2014). Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175.

Karthaus, W. R., Hofree, M., Choi, D., Linton, E. L., Turkekul, M., Bejnood, A., Carver, B., Gopalan, A., Abida, W., Laudone, V., et al. (2020). Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* **368**, 497–505.

Kas, S. M., de Ruiter, J. R., Schipper, K., Annunziato, S., Schut, E., Klarenbeek, S., Drenth, A. P., van der Burg, E., Klijn, C., Ten Hoeve, J. J., et al. (2017). Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nat. Genet.* **49**, 1219–1230.

Kavanagh, J. P. (1985). Sodium, potassium, calcium, magnesium, zinc, citrate and chloride content of human prostatic and seminal fluid. *J Reprod Fertil* **75**, 35–41.

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K. and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications* **10**, 1930.

Keil, K. P., Mehta, V., Abler, L. L., Joshi, P. S., Schmitz, C. T. and Vezina, C. M. (2012). Visualization and quantification of mouse prostate development by in situ hybridization. *Differentiation* **84**, 232–239.

Keita, M., Bachvarova, M., Morin, C., Plante, M., Gregoire, J., Renaud, M.-C., Sebastianelli, A., Trinh, X. B. and Bachvarov, D. (2013). The RUNX1 transcription factor is expressed in serous epithelial ovarian carcinoma and contributes to cell proliferation, migration and invasion. *Cell Cycle* **12**, 972–986.

Kendrick, H., Regan, J. L., Magnay, F.-A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M. and Smalley, M. J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* **9**, 591.

Khani, F., Mosquera, J. M., Park, K., Blattner, M., O'Reilly, C., MacDonald, T. Y., Chen, Z., Srivastava, A., Tewari, A. K., Barbieri, C. E., et al. (2014). Evidence for molecular differences in prostate cancer between African American and Caucasian men. *Clin Cancer Res* **20**, 4925–4934.

Kiciński, M., Vangronsveld, J. and Nawrot, T. S. (2011). An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PLoS One* **6**, e27130.

Kilbey, A., Blyth, K., Wotton, S., Terry, A., Jenkins, A., Bell, M., Hanlon, L., Cameron, E. R. and Neil, J. C. (2007). Disruption Promotes Immortalization and Confers Resistance to Oncogene-Induced Senescence in Primary Murine Fibroblasts. *Cancer Research* **67**, 11263–11271.

Kim, M. J., Cardiff, R. D., Desai, N., Banach-Petrosky, W. A., Parsons, R., Shen, M. M. and Abate-Shen, C. (2002). Cooperativity of Nkx3.1 and Pten loss of function in a mouse model of prostate carcinogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2884–2889.

Kim, H.-R., Oh, B.-C., Choi, J.-K. and Bae, S.-C. (2008). Pim-1 kinase phosphorylates and stabilizes RUNX3 and alters its subcellular localization. *J. Cell. Biochem.* **105**, 1048–1058.

Kim, J.-H., Jang, J.-W., Lee, Y.-S., Lee, J.-W., Chi, X.-Z., Li, Y.-H., Kim, M.-K., Kim, D.-M., Choi, B.-S., Kim, J., et al. (2014a). RUNX family members are covalently modified and regulated by PIAS1-mediated sumoylation. *Oncogenesis* **3**, e101.

Kim, W., Barron, D. A., Martin, R. S., Chan, K. S., Tran, L. L., Yang, F., Ressler, S. J. and Rowley, D. R. (2014b). RUNX1 is essential for mesenchymal stem cell proliferation and myofibroblast differentiation. *PNAS* **111**, 16389–16394.

Kissa, K. and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–115.

Kitabayashi, I., Aikawa, Y., Nguyen, L. A., Yokoyama, A. and Ohki, M. (2001). Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO J.* **20**, 7184–7196.

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. and Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201.

Knouf, E. C., Metzger, M. J., Mitchell, P. S., Arroyo, J. D., Chevillet, J. R., Tewari, M. and Miller, A. D. (2009). Multiple integrated copies and high-level production of the human retrovirus XMRV (xenotropic murine leukemia virus-related virus) from 22Rv1 prostate carcinoma cells. *J. Virol.* **83**, 7353–7356.

Komeno, Y., Yan, M., Matsuura, S., Lam, K., Lo, M.-C., Huang, Y.-J., Tenen, D. G., Downing, J. R. and Zhang, D.-E. (2014). Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood* **123**, 3760–3769.

Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R. T., Gao, Y. H., Inada, M., et al. (1997). Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89**, 755–764.

Kong, W., Biddy, B. A., Kamimoto, K., Amrute, J. M., Butka, E. G. and Morris, S. A. (2020). CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nature Protocols* **15**, 750–772.

Korenchuk, S., Lehr, J. E., MClean, L., Lee, Y. G., Whitney, S., Vessella, R., Lin, D. L. and Pienta, K. J. (2001). VCaP, a cell-based model system of human prostate cancer. *In Vivo* **15**, 163–168.

Kruithof-de Julio, M., Shibata, M., Desai, N., Reynon, M., Halili, M. V., Hu, Y.-P., Price, S. M., Abate-Shen, C. and Shen, M. M. (2013). Canonical Wnt signaling regulates Nkx3.1 expression and luminal epithelial differentiation during prostate organogenesis. *Dev Dyn* **242**, 1160–1171.

Kulkarni, M., Tan, T. Z., Syed Sulaiman, N. B., Lamar, J. M., Bansal, P., Cui, J., Qiao, Y. and Ito, Y. (2018). RUNX1 and RUNX3 protect against YAP-mediated EMT, stem-ness and shorter survival outcomes in breast cancer. *Oncotarget* **9**, 14175–14192.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K. D., Yamada, R. G., Ueda, H. R. and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Research* **34**, e42–e42.

Kuvardina, O. N., Herglotz, J., Kolodziej, S., Kohrs, N., Herkt, S., Wojcik, B., Oellerich, T., Corso, J., Behrens, K., Kumar, A., et al. (2015). RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. *Blood* **125**, 3570–3579.

Kwon, O.-J., Zhang, L., Ittmann, M. M. and Xin, L. (2014). Prostatic inflammation enhances basal-to-luminal differentiation and accelerates initiation of prostate cancer with a basal cell origin. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E592-600.

Kwon, O.-J., Zhang, L. and Xin, L. (2016). Stem Cell Antigen-1 Identifies a Distinct Androgen-Independent Murine Prostatic Luminal Cell Lineage with Bipotent Potential. *Stem Cells* **34**, 191–202.

Kwon, O.-J., Zhang, Y., Li, Y., Wei, X., Zhang, L., Chen, R., Creighton, C. J. and Xin, L. (2019). Functional Heterogeneity of Mouse Prostate Stromal Cells Revealed by Single-Cell RNA-Seq. *iScience* **13**, 328–338.

Kwon, O.-J., Choi, J. M., Zhang, L., Jia, D., Wei, X., Li, Z., Zhang, Y., Jung, S. Y., Creighton, C. J. and Xin, L. (2020). The Sca-1+ and Sca-1− mouse prostatic

luminal cell lineages are independently sustained. *STEM CELLS* **38**, 1479–1491.

Lacaud, G., Gore, L., Kennedy, M., Kouskoff, V., Kingsley, P., Hogan, C., Carlsson, L., Speck, N., Palis, J. and Keller, G. (2002). Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. *Blood* **100**, 458–466.

Lacaud, G., Kouskoff, V., Trumble, A., Schwantz, S. and Keller, G. (2004). Haploinsufficiency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells. *Blood* **103**, 886–889.

Lam, E. Y. N., Chau, J. Y. M., Kalev-Zylinska, M. L., Fountaine, T. M., Mead, R. S., Hall, C. J., Crosier, P. S., Crosier, K. E. and Flores, M. V. (2009). Zebrafish runx1 promoter-EGFP transgenics mark discrete sites of definitive blood progenitors. *Blood* **113**, 1241–1249.

Lam, E. Y. N., Hall, C. J., Crosier, P. S., Crosier, K. E. and Flores, M. V. (2010). Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. *Blood* **116**, 909–914.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V. and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* **457**, 892–895.

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, O., Wilson, N. K., Möröy, T., Bonifer, C., et al. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **120**, 314–322.

Lasnitzki, I. and Mizuno, T. (1977). Induction of the rat prostate gland by androgens in organ culture. *J Endocrinol* **74**, 47–55.

Lawson, D. A., Xin, L., Lukacs, R. U., Cheng, D. and Witte, O. N. (2007). Isolation and functional characterization of murine prostate stem cells. *Proc Natl Acad Sci U S A* **104**, 181–186.

Lee, J., Hoi, C. S. L., Lilja, K. C., White, B. S., Lee, S. E., Shalloway, D. and Tumbar, T. (2013). Runx1 and p21 synergistically limit the extent of hair follicle stem cell quiescence in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4634–4639.

Lee, S. E., Sada, A., Zhang, M., McDermitt, D. J., Lu, S. Y., Kemphues, K. J. and Tumbar, T. (2014). High Runx1 Levels Promote a Reversible, More-Differentiated Cell State in Hair-Follicle Stem Cells during Quiescence. *Cell Reports* **6**, 499–513.

Lee, J. J., Thomas, I.-C., Nolley, R., Ferrari, M., Brooks, J. D. and Leppert, J. T. (2015). Biologic Differences Between Peripheral and Transition Zone Prostate Cancer. *Prostate* **75**, 183–190.

Lee, J.-W., Kim, D.-M., Jang, J.-W., Park, T.-G., Song, S.-H., Lee, Y.-S., Chi, X.-Z., Park, I. Y., Hyun, J.-W., Ito, Y., et al. (2019). RUNX3 regulates cell cycle-dependent chromatin dynamics by functioning as a pioneer factor of the restriction-point. *Nature Communications* **10**, 1897.

Leissner, K. H. and Tisell, L. E. (1979). The weight of the human prostate. *Scand J Urol Nephrol* **13**, 137–142.

Leng, N., Chu, L.-F., Barry, C., Li, Y., Choi, J., Li, X., Jiang, P., Stewart, R. M., Thomson, J. A. and Kendziorski, C. (2015). Oscope identifies oscillatory genes in unsynchronized single cell RNA-seq experiments. *Nat Methods* **12**, 947–950.

Leong, K. G., Wang, B.-E., Johnson, L. and Gao, W.-Q. (2008). Generation of a prostate from a single adult stem cell. *Nature* **456**, 804–808.

Lesche, R., Groszer, M., Gao, J., Wang, Y., Messing, A., Sun, H., Liu, X. and Wu, H. (2002). Cre/loxP-mediated inactivation of the murine Pten tumor suppressor gene. *Genesis* **32**, 148–149.

Levanon, D. and Groner, Y. (2004). Structure and regulated expression of mammalian RUNX genes. *Oncogene* **23**, 4211–4219.

Levanon, D., Goldstein, R. E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z. and Groner, Y. (1998). Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11590–11595.

Levanon, D., Glusman, G., Bangsow, T., Ben-Asher, E., Male, D. A., Avidan, N., Bangsow, C., Hattori, M., Taylor, T. D., Taudien, S., et al. (2001). Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. *Gene* **262**, 23–33.

Levanon, D., Bettoun, D., Harris-Cerruti, C., Woolf, E., Negreanu, V., Eilam, R., Bernstein, Y., Goldenberg, D., Xiao, C., Fliegauf, M., et al. (2002). The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. *EMBO J.* **21**, 3454–3463.

Levenson, R. M., Borowsky, A. D. and Angelo, M. (2015). Immunohistochemistry and mass spectrometry for highly multiplexed cellular molecular imaging. *Lab Invest* **95**, 397–405.

Levsky, J. M. and Singer, R. H. (2003). Gene expression and the myth of the average cell. *Trends in Cell Biology* **13**, 4–6.

Li, Q., Lai, Q., He, C., Fang, Y., Yan, Q., Zhang, Y., Wang, X., Gu, C., Wang, Y., Ye, L., et al. (2019). RUNX1 promotes tumour metastasis by activating the Wnt/β-catenin signalling pathway and EMT in colorectal cancer. *Journal of Experimental & Clinical Cancer Research* **38**, 334.

Liakhovitskaia, A., Rybtsov, S., Smith, T., Batsivari, A., Rybtsova, N., Rode, C., de Bruijn, M., Buchholz, F., Gordon-Keylock, S., Zhao, S., et al. (2014). Runx1 is required for progression of CD41+ embryonic precursors into HSCs but not prior to this. *Development* **141**, 3319–3323.

Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S. A., Lie-A-Ling, M., Noailles, L., Vijayabaskar, M. S., Wu, M., et al. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis: RUNX1 shifts transcription factor binding patterns. *The EMBO Journal* **31**, 4318–4333.

Lie-A-Ling, M., Marinopoulou, E., Li, Y., Patel, R., Stefanska, M., Bonifer, C., Miller, C., Kouskoff, V. and Lacaud, G. (2014). RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. *Blood* **124**, e11-20.

Lie-A-Ling, M., Marinopoulou, E., Lilly, A. J., Challinor, M., Patel, R., Lancrin, C., Kouskoff, V. and Lacaud, G. (2018). Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development* **145**,.

Lie-A-Ling, M., Mevel, R., Patel, R., Blyth, K., Baena, E., Kouskoff, V. and Lacaud, G. (2020). RUNX1 Dosage in Development and Cancer. *Mol. Cells* **43**, 126–138.

Lilja, H., Oldbring, J., Rannevik, G. and Laurell, C. B. (1987). Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. *J Clin Invest* **80**, 281–285.

Lilly, A. J., Costa, G., Largeot, A., Fadlullah, M. Z. H., Lie-A-Ling, M., Lacaud, G. and Kouskoff, V. (2016). Interplay between SOX7 and RUNX1 regulates hemogenic endothelial fate in the yolk sac. *Development* **143**, 4341–4351.

Lin, D., Xue, H., Wang, Y., Wu, R., Watahiki, A., Dong, X., Cheng, H., Wyatt, A. W., Collins, C. C., Gout, P. W., et al. (2014). Next generation patient-derived prostate cancer xenograft models. *Asian J Androl* **16**, 407–412.

Linja, M. J., Savinainen, K. J., Saramäki, O. R., Tammela, T. L., Vessella, R. L. and Visakorpi, T. (2001). Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. *Cancer Res* **61**, 3550–3555.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. and Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523–536.

Little, G. H., Baniwal, S. K., Adisetiyo, H., Groshen, S., Chimge, N.-O., Kim, S. Y., Khalid, O., Hawes, D., Jones, J. O., Pinski, J., et al. (2014). Differential effects of RUNX2 on the androgen receptor in prostate cancer: synergistic stimulation of a gene set exemplified by SNAI2 and subsequent invasiveness. *Cancer Res.* **74**, 2857–2868.

Litwin, M. S. and Tan, H.-J. (2017). The Diagnosis and Treatment of Prostate Cancer: A Review. *JAMA* **317**, 2532–2542.

Liu, J. C., Lengner, C. J., Gaur, T., Lou, Y., Hussain, S., Jones, M. D., Borodic, B., Colby, J. L., Steinman, H. A., van Wijnen, A. J., et al. (2011). Runx2 protein expression utilizes the Runx2 P1 promoter to establish osteoprogenitor cell number for normal bone formation. *J. Biol. Chem.* **286**, 30057–30070.

Liu, X., Grogan, T. R., Hieronymus, H., Hashimoto, T., Mottahedeh, J., Cheng, D., Zhang, L., Huang, K., Stoyanova, T., Park, J. W., et al. (2016). Low CD38 Identifies Progenitor-like Inflammation-Associated Luminal Cells that Can Initiate Human Prostate Cancer and Predict Poor Outcome. *Cell Rep* **17**, 2596–2606.

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., Kovatich, A. J., Benz, C. C., Levine, D. A., Lee, A. V., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11.

Loeb, S., Bjurlin, M. A., Nicholson, J., Tammela, T. L., Penson, D. F., Carter, H. B., Carroll, P. and Etzioni, R. (2014). Overdiagnosis and overtreatment of prostate cancer. *Eur. Urol.* **65**, 1046–1055.

Loeffler, D. and Schroeder, T. (2019). Understanding cell fate control by continuous single-cell quantification. *Blood* **133**, 1406–1414.

Lomas, D. J. and Ahmed, H. U. (2020). All change in the prostate cancer diagnostic pathway. *Nature Reviews Clinical Oncology* **17**, 372–381.

Lopes, E. S., Foster, B. A., Donjacour, A. A. and Cunha, G. R. (1996). Initiation of secretory activity of rat prostatic epithelium in organ culture. *Endocrinology* **137**, 4225–4234.

Lorsbach, R. B., Moore, J., Ang, S. O., Sun, W., Lenny, N. and Downing, J. R. (2004). Role of RUNX1 in adult hematopoiesis: analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression. *Blood* **103**, 2522–2529.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.

Lu, T.-L., Huang, Y.-F., You, L.-R., Chao, N.-C., Su, F.-Y., Chang, J.-L. and Chen, C.-M. (2013). Conditionally ablated Pten in prostate basal cells promotes basal-to-

luminal differentiation and causes invasive prostate cancer in mice. *Am. J. Pathol.* **182**, 975–991.

Luche, H., Weber, O., Nageswara Rao, T., Blum, C. and Fehling, H. J. (2007). Faithful activation of an extra-bright red fluorescent protein in "knock-in" Cre-reporter mice ideally suited for lineage tracing studies. *Eur. J. Immunol.* **37**, 43–53.

Luchman, H. A., Friedman, H. C., Villemaire, M. L., Peterson, A. C. and Jirik, F. R. (2008). Temporally controlled prostate epithelium-specific gene alterations. *Genesis* **46**, 229–234.

Luecken, M. D. and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology* **15**, e8746.

Luo, J., Zha, S., Gage, W. R., Dunn, T. A., Hicks, J. L., Bennett, C. J., Ewing, C. M., Platz, E. A., Ferdinandusse, S., Wanders, R. J., et al. (2002). Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. *Cancer Res.* **62**, 2220–2226.

Lutterbach, B., Westendorf, J. J., Linggi, B., Isaac, S., Seto, E. and Hiebert, S. W. (2000). A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia. *J. Biol. Chem.* **275**, 651–656.

Lux, C. T., Yoshimoto, M., McGrath, K., Conway, S. J., Palis, J. and Yoder, M. C. (2008). All primitive and definitive hematopoietic progenitor cells emerging before E10 in the mouse embryo are products of the yolk sac. *Blood* **111**, 3435–3438.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214.

Magnen, C. L., Virk, R. K., Dutta, A., Kim, J. Y., Panja, S., Lopez-Bujanda, Z. A., Califano, A., Drake, C. G., Mitrofanova, A. and Abate-Shen, C. (2018). Cooperation of loss of NKX3.1 and inflammation in prostate cancer initiation. *Disease Models & Mechanisms* **11**,.

Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. and Berns, A. (2000). Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev* **14**, 994–1004.

Marker, P. C., Donjacour, A. A., Dahiya, R. and Cunha, G. R. (2003). Hormonal, cellular, and molecular control of prostatic development. *Developmental Biology* **253**, 165–174.

Marks, L. S. (2004). 5alpha-reductase: history and clinical importance. *Rev Urol* **6 Suppl 9**, S11-21.

Martin, R. M., Donovan, J. L., Turner, E. L., Metcalfe, C., Young, G. J., Walsh, E. I., Lane, J. A., Noble, S., Oliver, S. E., Evans, S., et al. (2018). Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. *JAMA* **319**, 883.

Martinez, M., Hinojosa, M., Trombly, D., Morin, V., Stein, J., Stein, G., Javed, A. and Gutierrez, S. E. (2016). Transcriptional Auto-Regulation of RUNX1 P1 Promoter. *PLoS One* **11**,.

Masse, I., Barbollat-Boutrand, L., Molina, M., Berthier-Vergnes, O., Joly-Tonetti, N., Martin, M. T., Caron de Fromentel, C., Kanitakis, J. and Lamartine, J. (2012). Functional interplay between p63 and p53 controls RUNX1 function in the transition from proliferation to differentiation in human keratinocytes. *Cell Death Dis* **3**, e318.

Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M. S. H., Ko, S. B. H., Gouda, N., Hayashi, T. and Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**, 2314–2321.

McAuley, E., Moline, D., VanOpstall, C., Lamperis, S., Brown, R. and Vander Griend, D. J. (2019). Sox2 Expression Marks Castration-Resistant Progenitor Cells in the Adult Murine Prostate. *Stem Cells* **37**, 690–700.

McCabe, A., Dolled-Filhart, M., Camp, R. L. and Rimm, D. L. (2005). Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis. *J. Natl. Cancer Inst.* **97**, 1808–1815.

McCarroll, C. S., He, W., Foote, K., Bradley, A., Mcglynn, K., Vidler, F., Nixon, C., Nather, K., Fattah, C., Riddell, A., et al. (2018). Runx1 Deficiency Protects Against Adverse Cardiac Remodeling After Myocardial Infarction. *Circulation* **137**, 57–70.

McDonald, L., Ferrari, N., Terry, A., Bell, M., Mohammed, Z. M., Orange, C., Jenkins, A., Muller, W. J., Gusterson, B. A., Neil, J. C., et al. (2014). RUNX2 correlates with subtype-specific breast cancer in a human tissue microarray, and ectopic expression of Runx2 perturbs differentiation in the mouse mammary gland. *Dis Model Mech* **7**, 525–534.

McGinnis, C. S., Murrow, L. M. and Gartner, Z. J. (2019a). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e4.

McGinnis, C. S., Patterson, D. M., Winkler, J., Conrad, D. N., Hein, M. Y., Srivastava, V., Hu, J. L., Murrow, L. M., Weissman, J. S., Werb, Z., et al.

(2019b). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nature Methods* **16**, 619–626.

McGrath, K. E., Frame, J. M., Fegan, K. H., Bowen, J. R., Conway, S. J., Catherman, S. C., Kingsley, P. D., Koniski, A. D. and Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Rep* **11**, 1892–1904.

McInnes, L., Healy, J. and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*.

McNeal, J. E. (1968). Regional morphology and pathology of the prostate. *Am J Clin Pathol* **49**, 347–357.

McNeal, J. E. (1981). The zonal anatomy of the prostate. *Prostate* **2**, 35–49.

McNeal, J. E. and Bostwick, D. G. (1986). Intraductal dysplasia: a premalignant lesion of the prostate. *Hum Pathol* **17**, 64–71.

McNeal, J. E., Redwine, E. A., Freiha, F. S. and Stamey, T. A. (1988). Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. *Am J Surg Pathol* **12**, 897–906.

Medvinsky, A. and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897–906.

Mehmood, A., Laiho, A., Venäläinen, M. S., McGlinchey, A. J., Wang, N. and Elo, L. L. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. *Briefings in Bioinformatics* **21**, 2052–2065.

Method of the Year 2013 (2014). *Nature Methods* **11**, 1–1.

Method of the Year 2019: Single-cell multimodal omics (2020). *Nature Methods* **17**, 1–1.

Mevel, R., Draper, J. E., Lie-a-Ling, M., Kouskoff, V. and Lacaud, G. (2019). RUNX transcription factors: orchestrators of development. *Development* **146**, dev148296.

Mevel, R., Steiner, I., Mason, S., Galbraith, L. C., Patel, R., Fadlullah, M. Z., Ahmad, I., Leung, H. Y., Oliveira, P., Blyth, K., et al. (2020). RUNX1 marks a luminal castration-resistant lineage established at the onset of prostate development. *eLife* **9**, e60225.

Michaud, J., Wu, F., Osato, M., Cottles, G. M., Yanagida, M., Asou, N., Shigesada, K., Ito, Y., Benson, K. F., Raskind, W. H., et al. (2002). In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet

disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis. *Blood* **99**, 1364–1372.

Mikkola, H. K. A. and Orkin, S. H. (2006). The journey of developing hematopoietic stem cells. *Development* **133**, 3733–3744.

Mitchell, T. and Neal, D. E. (2015). The genomic evolution of human prostate cancer. *Br. J. Cancer* **113**, 193–198.

Mitsuda, Y., Morita, K., Kashiwazaki, G., Taniguchi, J., Bando, T., Obara, M., Hirata, M., Kataoka, T. R., Muto, M., Kaneda, Y., et al. (2018). RUNX1 positively regulates the ErbB2/HER2 signaling pathway through modulating SOS1 expression in gastric cancer cells. *Sci Rep* **8**, 1–13.

Miyagawa, K., Sakakura, C., Nakashima, S., Yoshikawa, T., Kin, S., Nakase, Y., Ito, K., Yamagishi, H., Ida, H., Yazumi, S., et al. (2006). Down-regulation of RUNX1, RUNX3 and CBFbeta in hepatocellular carcinomas in an early stage of hepatocarcinogenesis. *Anticancer Res.* **26**, 3633–3643.

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y. and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10431–10434.

Miyoshi, H., Ohira, M., Shimizu, K., Mitani, K., Hirai, H., Imai, T., Yokoyama, K., Soeda, E. and Ohki, M. (1995). Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. *Nucleic Acids Res.* **23**, 2762–2769.

Moad, M., Hannezo, E., Buczacki, S. J., Wilson, L., El-Sherif, A., Sims, D., Pickard, R., Wright, N. A., Williamson, S. C., Turnbull, D. M., et al. (2017). Multipotent Basal Stem Cells, Maintained in Localized Proximal Niches, Support Directed Long-Ranging Epithelial Flows in Human Prostates. *Cell Rep* **20**, 1609–1622.

Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., Bonkhofer, F., Kirmizitas, A. and Patient, R. (2016). Transforming Growth Factor β Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. *Dev Cell* **38**, 358–370.

Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., Yim, K., Elzen, A. van den, Hirn, M. J., Coifman, R. R., et al. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology* **37**, 1482–1492.

Morita, K., Suzuki, K., Maeda, S., Matsuo, A., Mitsuda, Y., Tokushige, C., Kashiwazaki, G., Taniguchi, J., Maeda, R., Noura, M., et al. (2017a). Genetic

regulation of the RUNX transcription factor family has antitumor effects. *J. Clin. Invest.* **127**, 2815–2828.

Morita, K., Maeda, S., Suzuki, K., Kiyose, H., Taniguchi, J., Liu, P. P., Sugiyama, H., Adachi, S. and Kamikubo, Y. (2017b). Paradoxical enhancement of leukemogenesis in acute myeloid leukemia with moderately attenuated RUNX1 expressions. *Blood Adv* **1**, 1440–1451.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628.

Mottet, N., Bellmunt, J., Bolla, M., Briers, E., Cumberbatch, M. G., De Santis, M., Fossati, N., Gross, T., Henry, A. M., Joniau, S., et al. (2017). EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* **71**, 618–629.

Mukai, K., BenBarak, M. J., Tachibana, M., Nishida, K., Karasuyama, H., Taniuchi, I. and Galli, S. J. (2012). Critical role of P1-Runx1 in mouse basophil development. *Blood* **120**, 76–85.

Mulholland, D. J., Tran, L. M., Li, Y., Cai, H., Morim, A., Wang, S., Plaisier, S., Garraway, I. P., Huang, J., Graeber, T. G., et al. (2011). Cell Autonomous Role of PTEN in Regulating Castration-Resistant Prostate Cancer Growth. *Cancer Cell* **19**, 792–804.

Murphy, W. M., Soloway, M. S. and Barrows, G. H. (1991). Pathologic changes associated with androgen deprivation therapy for prostate cancer. *Cancer* **68**, 821–828.

Muus, C., Luecken, M. D., Eraslan, G., Waghray, A., Heimberg, G., Sikkema, L., Kobayashi, Y., Vaishnav, E. D., Subramanian, A., Smilie, C., et al. (2020). Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. *bioRxiv* 2020.04.19.049254.

Na, R., Zheng, S. L., Han, M., Yu, H., Jiang, D., Shah, S., Ewing, C. M., Zhang, L., Novakovic, K., Petkewicz, J., et al. (2017). Germline Mutations in ATM and BRCA1/2 Distinguish Risk for Lethal and Indolent Prostate Cancer and are Associated with Early Age at Death. *Eur Urol* **71**, 740–747.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* **320**, 1344–1349.

Nagata, T., Gupta, V., Sorce, D., Kim, W. Y., Sali, A., Chait, B. T., Shigesada, K., Ito, Y. and Werner, M. H. (1999). Immunoglobulin motif DNA recognition and

heterodimerization of the PEBP2/CBF Runt domain. *Nat. Struct. Biol.* **6**, 615–619.

Naji, L., Randhawa, H., Sohani, Z., Dennis, B., Lautenbach, D., Kavanagh, O., Bawor, M., Banfield, L. and Profetto, J. (2018). Digital Rectal Examination for Prostate Cancer Screening in Primary Care: A Systematic Review and Meta-Analysis. *Ann Fam Med* **16**, 149–154.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L. B., Martin, S., Wedge, D. C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54.

North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M. and Speck, N. A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development* **126**, 2563–2575.

North, T. E., de Bruijn, M. F. T. R., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E. and Speck, N. A. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661–672.

North, T. E., Stacy, T., Matheny, C. J., Speck, N. A. and de Bruijn, M. F. T. R. (2004). Runx1 is expressed in adult mouse hematopoietic stem cells and differentiating myeloid and lymphoid cells, but not in maturing erythroid cells. *Stem Cells* **22**, 158–168.

Novick, A. and Weiner, M. (1957). Enzyme Induction as an All-or-None Phenomenon. *PNAS* **43**, 553–566.

Nowotschin, S., Setty, M., Kuo, Y.-Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795–801.

Nyberg, T., Frost, D., Barrowdale, D., Evans, D. G., Bancroft, E., Adlard, J., Ahmed, M., Barwell, J., Brady, A. F., Brewer, C., et al. (2020). Prostate Cancer Risks for Male BRCA1 and BRCA2 Mutation Carriers: A Prospective Cohort Study. *Eur Urol* **77**, 24–35.

Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigesada, K. and Ito, Y. (1993). PEBP2/PEA2 represents a family of transcription factors homologous to the products of the Drosophila runt gene and the human AML1 gene. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6859–6863.

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. and Downing, J. R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321–330.

Osato, M. (2004). Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. *Oncogene* **23**, 4284–4296.

Osorio, K. M., Lee, S. E., McDermitt, D. J., Waghmare, S. K., Zhang, Y. V., Woo, H. N. and Tumbar, T. (2008). Runx1 modulates developmental, but not injury-driven, hair follicle stem cell activation. *Development* **135**, 1059–1068.

Osorio, K. M., Lilja, K. C. and Tumbar, T. (2011). Runx1 modulates adult hair follicle stem cell emergence and maintenance from distinct embryonic skin compartments. *J. Cell Biol.* **193**, 235–250.

Otto, F., Thornell, A. P., Crompton, T., Denzel, A., Gilmour, K. C., Rosewell, I. R., Stamp, G. W. H., Beddington, R. S. P., Mundlos, S., Olsen, B. R., et al. (1997). Cbfa1, a Candidate Gene for Cleidocranial Dysplasia Syndrome, Is Essential for Osteoblast Differentiation and Bone Development. *Cell* **89**, 765–771.

Ousset, M., Van Keymeulen, A., Bouvencourt, G., Sharma, N., Achouri, Y., Simons, B. D. and Blanpain, C. (2012). Multipotent and unipotent progenitors contribute to prostate postnatal development. *Nat Cell Biol* **14**, 1131–1138.

Owens, T. W., Rogers, R. L., Best, S. A., Ledger, A., Mooney, A.-M., Ferguson, A., Shore, P., Swarbrick, A., Ormandy, C. J., Simpson, P. T., et al. (2014). Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer. *Cancer Res.* **74**, 5277–5286.

Palanisamy, N., Yang, J., Shepherd, P. D. A., Li-Ning-Tapia, E. M., Labanca, E., Manyam, G. C., Ravoori, M. K., Kundra, V., Araujo, J. C., Efstathiou, E., et al. (2020). The MD Anderson Prostate Cancer Patient-derived Xenograft Series (MDA PCa PDX) Captures the Molecular Landscape of Prostate Cancer and Facilitates Marker-driven Therapy Development. *Clin Cancer Res* **26**, 4933–4946.

Palis, J., Robertson, S., Kennedy, M., Wall, C. and Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* **126**, 5073–5084.

Park, J. W., Lee, J. K., Phillips, J. W., Huang, P., Cheng, D., Huang, J. and Witte, O. N. (2016). Prostate epithelial cell of origin determines cancer differentiation state in an organoid transformation assay. *PNAS*.

Peehl, D. M., Sellers, R. G. and McNeal, J. E. (1996). Keratin 19 in the adult human prostate: tissue and cell culture studies. *Cell Tissue Res* **285**, 171–176.

Pegg, H. J., Harrison, H., Rogerson, C. and Shore, P. (2019). The RUNX Transcriptional Coregulator, CBFβ, Suppresses Migration of ER+ Breast Cancer Cells by Repressing ERα-Mediated Expression of the Migratory Factor TFF1. *Mol. Cancer Res.* **17**, 1015–1023.

Peixoto, A., Monteiro, M., Rocha, B. and Veiga-Fernandes, H. (2004). Quantification of Multiple Gene Expression in Individual Cells. *Genome Res.* **14**, 1938–1947.

Pelletier, N., Champagne, N., Stifani, S. and Yang, X.-J. (2002). MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2. *Oncogene* **21**, 2729–2740.

Pencovich, N., Jaschek, R., Tanay, A. and Groner, Y. (2011). Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models. *Blood* **117**, e1-14.

Pencovich, N., Jaschek, R., Dicken, J., Amit, A., Lotem, J., Tanay, A. and Groner, Y. (2013). Cell-Autonomous Function of Runx1 Transcriptionally Regulates Mouse Megakaryocytic Maturation. *PLOS ONE* **8**, e64248.

Pereira, B., Chin, S.-F., Rueda, O. M., Vollan, H.-K. M., Provenzano, E., Bardwell, H. A., Pugh, M., Jones, L., Russell, R., Sammut, S.-J., et al. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **7**, 11479.

Pernar, C. H., Ebot, E. M., Wilson, K. M. and Mucci, L. A. (2018). The Epidemiology of Prostate Cancer. *Cold Spring Harb Perspect Med* **8**,.

Peterson, L. F. and Zhang, D.-E. (2004). The 8;21 translocation in leukemogenesis. *Oncogene* **23**, 4255–4262.

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G. and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10**, 1096–1098.

Pietrzak, K., Kuzyakiv, R., Simon, R., Bolis, M., Bär, D., Aprigliano, R., Theurillat, J.-P., Sauter, G. and Santoro, R. (2020). TIP5 primes prostate luminal cells for the oncogenic transformation mediated by PTEN-loss. *Proc Natl Acad Sci U S A* **117**, 3637–3647.

Pignon, J.-C., Grisanzio, C., Geng, Y., Song, J., Shivdasani, R. A. and Signoretti, S. (2013). p63-expressing cells are the stem cells of developing prostate, bladder, and colorectal epithelia. *Proc Natl Acad Sci U S A* **110**, 8105–8110.

Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L., et al.

(2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495.

Pimanda, J. E., Donaldson, I. J., de Bruijn, M. F. T. R., Kinston, S., Knezevic, K., Huckle, L., Piltz, S., Landry, J.-R., Green, A. R., Tannahill, D., et al. (2007). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proceedings of the National Academy of Sciences* **104**, 840–845.

Planagumà, J., Díaz-Fuertes, M., Gil-Moreno, A., Abal, M., Monge, M., García, A., Baró, T., Thomson, T. M., Xercavins, J., Alameda, F., et al. (2004). A differential gene expression profile reveals overexpression of RUNX1/AML1 in invasive endometrioid carcinoma. *Cancer Res.* **64**, 8846–8853.

Planagumà, J., Gonzalez, M., Doll, A., Monge, M., Gil-Moreno, A., Baró, T., García, A., Xercavins, J., Alameda, F., Abal, M., et al. (2006). The up-regulation profiles of p21WAF1/CIP1 and RUNX1/AML1 correlate with myometrial infiltration in endometrioid endometrial carcinoma. *Hum. Pathol.* **37**, 1050–1057.

Plass, M., Solana, J., Wolf, F. A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F. J., Kocks, C. and Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**,.

Ploussard, G., Nicolaiew, N., Marchand, C., Terry, S., Vacherot, F., Vordos, D., Allory, Y., Abbou, C.-C., Salomon, L. and de la Taille, A. (2014). Prospective Evaluation of an Extended 21-Core Biopsy Scheme as Initial Prostate Cancer Diagnostic Strategy. *European Urology* **65**, 154–161.

Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* **32**, 1053–1058.

Potter, G. A., Barrie, S. E., Jarman, M. and Rowlands, M. G. (1995). Novel steroidal inhibitors of human cytochrome P45017 alpha (17 alpha-hydroxylase-C17,20-lyase): potential agents for the treatment of prostatic cancer. *J Med Chem* **38**, 2463–2471.

Pozner, A., Goldenberg, D., Negreanu, V., Le, S. Y., Elroy-Stein, O., Levanon, D. and Groner, Y. (2000). Transcription-coupled translation control of AML1/RUNX1 is mediated by cap- and internal ribosome entry site-dependent mechanisms. *Mol. Cell. Biol.* **20**, 2297–2307.

Pozner, A., Lotem, J., Xiao, C., Goldenberg, D., Brenner, O., Negreanu, V., Levanon, D. and Groner, Y. (2007). Developmentally regulated promoter-switch

transcriptionally controls Runx1 function during embryonic hematopoiesis. *BMC Developmental Biology* **7**, 84.

Preisser, F., Chun, F. K. H., Pompe, R. S., Heinze, A., Salomon, G., Graefen, M., Huland, H. and Tilki, D. (2019). Persistent Prostate-Specific Antigen After Radical Prostatectomy and Its Impact on Oncologic Outcomes. *Eur Urol* **76**, 106–114.

Prins, G. S. and Lindgren, M. (2015). Chapter 18 - Accessory Sex Glands in the Male. In *Knobil and Neill's Physiology of Reproduction (Fourth Edition)* (ed. Plant, T. M.) and Zeleznik, A. J.), pp. 773–804. San Diego: Academic Press.

Pritchard, C. C., Mateo, J., Walsh, M. F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *https://doi.org/10.1056/NEJMoa1603144*.

Ptasinska, A., Assi, S. A., Mannari, D., James, S. R., Williamson, D., Dunne, J., Hoogenkamp, M., Wu, M., Care, M., McNeill, H., et al. (2012). Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* **26**, 1829–1841.

Puig-Kröger, A., Aguilera-Montilla, N., Martínez-Nuñez, R., Domínguez-Soto, A., Sánchez-Cabo, F., Martín-Gayo, E., Zaballos, A., Toribio, M. L., Groner, Y., Ito, Y., et al. (2010). The novel RUNX3/p33 isoform is induced upon monocyte-derived dendritic cell maturation and downregulates IL-8 expression. *Immunobiology* **215**, 812–820.

Purysko, A. S., Magi-Galluzzi, C., Mian, O. Y., Sittenfeld, S., Davicioni, E., du Plessis, M., Buerki, C., Bullen, J., Li, L., Madabhushi, A., et al. (2019). Correlation between MRI phenotypes and a genomic classifier of prostate cancer: preliminary findings. *Eur Radiol* **29**, 4861–4870.

Putz, G., Rosner, A., Nuesslein, I., Schmitz, N. and Buchholz, F. (2006). AML1 deletion in adult mice causes splenomegaly and lymphomas. *Oncogene* **25**, 929–939.

Putzi, M. J. and De Marzo, A. M. (2000). Morphologic transitions between proliferative inflammatory atrophy and high-grade prostatic intraepithelial neoplasia. *Urology* **56**, 828–832.

Qin, J., Liu, X., Laffin, B., Chen, X., Choy, G., Jeter, C., Calhoun-Davis, T., Li, H., Palapattu, G. S., Pang, S., et al. (2012). The PSA−/lo prostate cancer cell population harbors self-renewing long-term tumor-propagating cells that resist castration. *Cell Stem Cell* **10**, 556–569.

Qu, J., Tanis, S. E. J., Smits, J. P. H., Kouwenhoven, E. N., Oti, M., van den Bogaard, E. H., Logie, C., Stunnenberg, H. G., van Bokhoven, H., Mulder, K. W., et al.

(2018). Mutant p63 Affects Epidermal Cell Identity through Rewiring the Enhancer Landscape. *Cell Rep* **25**, 3490-3503.e4.

Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226.

Ramaswamy, S., Ross, K. N., Lander, E. S. and Golub, T. R. (2003). A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54.

Ran, D., Shia, W.-J., Lo, M.-C., Fan, J.-B., Knorr, D. A., Ferrell, P. I., Ye, Z., Yan, M., Cheng, L., Kaufman, D. S., et al. (2013). RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells. *Blood* **121**, 2882–2890.

Raveh, E., Cohen, S., Levanon, D., Negreanu, V., Groner, Y. and Gat, U. (2006). Dynamic expression of Runx1 in skin affects hair structure. *Mech. Dev.* **123**, 842–850.

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. *eLife* **6**, e27041.

Rennert, J., Coffman, J. A., Mushegian, A. R. and Robertson, A. J. (2003). The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms. *BMC Evolutionary Biology* 11.

Reuter, V. E. (1997). Pathological changes in benign and malignant prostatic tissue following androgen deprivation therapy. *Urology* **49**, 16–22.

Richardson, G. D., Robson, C. N., Lang, S. H., Neal, D. E., Maitland, N. J. and Collins, A. T. (2004). CD133, a novel marker for human prostatic epithelial stem cells. *Journal of Cell Science* **117**, 3539–3545.

Riggio, A. I. and Blyth, K. (2017). The enigmatic role of RUNX1 in female-related cancers - current knowledge & future perspectives. *FEBS J.* **284**, 2345–2362.

Rini, D. and Calabi, F. (2001). Identification and comparative analysis of a second runx3 promoter. *Gene* **273**, 13–22.

Robinson, D., Van Allen, E. M., Wu, Y.-M., Schultz, N., Lonigro, R. J., Mosquera, J.-M., Montgomery, B., Taplin, M.-E., Pritchard, C. C., Attard, G., et al. (2015). Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* **162**, 454.

Rody, A., Karn, T., Liedtke, C., Pusztai, L., Ruckhaeberle, E., Hanker, L., Gaetje, R., Solbach, C., Ahr, A., Metzler, D., et al. (2011). A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Research* **13**, R97.

Rooney, N., Mason, S. M., McDonald, L., Däbritz, J. H. M., Campbell, K. J., Hedley, A., Howard, S., Athineos, D., Nixon, C., Clark, W., et al. (2020). RUNX1 Is a Driver of Renal Cell Carcinoma Correlating with Clinical Outcome. *Cancer Res* **80**, 2325–2339.

Roy, N. C., Altermann, E., Park, Z. A. and McNabb, W. C. (2011). A comparison of analog and Next-Generation transcriptomic tools for mammalian studies. *Briefings in Functional Genomics* **10**, 135–150.

Rubin, M. A., Dunn, R., Strawderman, M. and Pienta, K. J. (2002). Tissue microarray sampling strategy for prostate cancer biomarker analysis. *Am J Surg Pathol* **26**, 312–319.

Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology* **37**, 547–554.

Saito, M., Iwawaki, T., Taya, C., Yonekawa, H., Noda, M., Inui, Y., Mekada, E., Kimata, Y., Tsuru, A. and Kohno, K. (2001). Diphtheria toxin receptor–mediated conditional and targeted cell ablation in transgenic mice. *Nature Biotechnology* **19**, 746–750.

Sakakura, C., Hagiwara, A., Miyagawa, K., Nakashima, S., Yoshikawa, T., Kin, S., Nakase, Y., Ito, K., Yamagishi, H., Yazumi, S., et al. (2005). Frequent downregulation of the runt domain transcription factors RUNX1, RUNX3 and their cofactor CBFB in gastric cancer. *Int. J. Cancer* **113**, 221–228.

Sakr, W. A., Grignon, D. J., Haas, G. P., Heilbrun, L. K., Pontes, J. E. and Crissman, J. D. (1996). Age and racial distribution of prostatic intraepithelial neoplasia. *Eur Urol* **30**, 138–144.

Salinas, C. A., Tsodikov, A., Ishak-Howard, M. and Cooney, K. A. (2014). Prostate cancer in young men: an important clinical entity. *Nat Rev Urol* **11**, 317–323.

Samokhvalov, I. M., Samokhvalova, N. I. and Nishikawa, S. (2007). Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. *Nature* **446**, 1056–1061.

Scheitz, C. J. F. and Tumbar, T. (2013). New insights into the role of Runx1 in epithelial stem cell biology and pathology. *J. Cell. Biochem.* **114**, 985–993.

Scheitz, C. J. F., Lee, T. S., McDermitt, D. J. and Tumbar, T. (2012). Defining a tissue stem cell-driven Runx1/Stat3 signalling axis in epithelial cancer. *EMBO J.* **31**, 4124–4139.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* **270**, 467–470.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Liu, S., Lin, S., Berube, P., Lee, L., et al. (2017). Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming. *bioRxiv* 191056.

Schröder, F. H., Hugosson, J., Roobol, M. J., Tammela, T. L. J., Zappa, M., Nelen, V., Kwiatkowski, M., Lujan, M., Määttänen, L., Lilja, H., et al. (2014). Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* **384**, 2027–2035.

SEER Cancer Statistics Review, 1975-2016 *SEER*.

Selman, S. H. (2011). The McNeal prostate: a review. *Urology* **78**, 1224–1228.

Seo, W., Tanaka, H., Miyamoto, C., Levanon, D., Groner, Y. and Taniuchi, I. (2012). Roles of VWRPY motif-mediated gene repression by Runx proteins during T-cell development. *Immunol. Cell Biol.* **90**, 827–830.

Seppälä, E. H., Ikonen, T., Mononen, N., Autio, V., Rökman, A., Matikainen, M. P., Tammela, T. L. J. and Schleutker, J. (2003). CHEK2 variants associate with hereditary prostate cancer. *Br J Cancer* **89**, 1966–1970.

Shannon, B. A., McNeal, J. E. and Cohen, R. J. (2003). Transition zone carcinoma of the prostate gland: a common indolent tumour type that occasionally manifests aggressive behaviour. *Pathology* **35**, 467–471.

Shappell, S. B., Thomas, G. V., Roberts, R. L., Herbert, R., Ittmann, M. M., Rubin, M. A., Humphrey, P. A., Sundberg, J. P., Rozengurt, N., Barrios, R., et al. (2004). Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. *Cancer Res.* **64**, 2270–2305.

Sharp, A., Coleman, I., Yuan, W., Sprenger, C., Dolling, D., Rodrigues, D. N., Russo, J. W., Figueiredo, I., Bertan, C., Seed, G., et al. (2019). Androgen receptor splice variant-7 expression emerges with castration resistance in prostate cancer. *J Clin Invest* **129**, 192–208.

Shen, M. M. and Abate-Shen, C. (2010). Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev.* **24**, 1967–2000.

Sheng, H. Z., Lin, P. X. and Nelson, P. G. (1994). Analysis of Multiple Heterogeneous mRNAs in Single Cells. *Analytical Biochemistry* **222**, 123–130.

Siegel, R. L., Miller, K. D. and Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70**, 7–30.

Siiteri, P. K. and Wilson, J. D. (1974). Testosterone formation and metabolism during male sexual differentiation in the human embryo. *J Clin Endocrinol Metab* **38**, 113–125.

Sinnott, J. A., Rider, J. R., Carlsson, J., Gerke, T., Tyekucheva, S., Penney, K. L., Sesso, H. D., Loda, M., Fall, K., Stampfer, M. J., et al. (2015). Molecular differences in transition zone and peripheral zone prostate tumors. *Carcinogenesis* **36**, 632–638.

Skene, P. J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856.

Skene, P. J., Henikoff, J. G. and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006–1019.

Sokol, E. S., Sanduja, S., Jin, D. X., Miller, D. H., Mathis, R. A. and Gupta, P. B. (2015). Perturbation-Expression Analysis Identifies RUNX1 as a Regulator of Human Mammary Stem Cell Differentiation. *PLOS Computational Biology* **11**, e1004161.

Song, W. J., Sullivan, M. G., Legare, R. D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I. C., Haworth, C., Hock, R., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175.

Sood, R., Kamikubo, Y. and Liu, P. (2017). Role of RUNX1 in hematological malignancies. *Blood* **129**, 2070–2082.

Spender, L. C., Whiteman, H. J., Karstegl, C. E. and Farrell, P. J. (2005). Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells. *Oncogene* **24**, 1873–1881.

Squire, J. A. (2009). TMPRSS2-ERG and PTEN loss in prostate cancer. *Nat. Genet.* **41**, 509–510.

Sramkoski, R. M., Pretlow, T. G., Giaconia, J. M., Pretlow, T. P., Schwartz, S., Sy, M. S., Marengo, S. R., Rhim, J. S., Zhang, D. and Jacobberger, J. W. (1999). A new human prostate carcinoma cell line, 22Rv1. *In Vitro Cell. Dev. Biol. Anim.* **35**, 403–409.

Sroczynska, P., Lancrin, C., Kouskoff, V. and Lacaud, G. (2009). The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood* **114**, 5279–5289.

Stabile, A., Giganti, F., Rosenkrantz, A. B., Taneja, S. S., Villeirs, G., Gill, I. S., Allen, C., Emberton, M., Moore, C. M. and Kasivisvanathan, V. (2020).

Multiparametric MRI for prostate cancer diagnosis: current status and future directions. *Nature Reviews Urology* **17**, 41–61.

Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82.

Stanfel, M. N., Moses, K. A., Carson, J. A., Zimmer, D. B., DeMayo, F., Schwartz, R. J. and Zimmer, W. E. (2006). Expression of an Nkx3.1-CRE gene using ROSA26 reporter mice. *Genesis* **44**, 550–555.

Stefanska, M., Batta, K., Patel, R., Florkowska, M., Kouskoff, V. and Lacaud, G. (2017). Primitive erythrocytes are generated from hemogenic endothelial cells. *Sci Rep* **7**, 6401.

Stifter, S. A. and Greter, M. (2020). STOP floxing around: Specificity and leakiness of inducible Cre/loxP systems. *European Journal of Immunology* **50**, 338–341.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R. and Smibert, P. (2017). Large-scale simultaneous measurement of epitopes and transcriptomes in single cells. *Nat Methods* **14**, 865–868.

Stone, K. R., Mickey, D. D., Wunderli, H., Mickey, G. H. and Paulson, D. F. (1978). Isolation of a human prostate carcinoma cell line (DU 145). *Int. J. Cancer* **21**, 274–281.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21.

Sugimura, Y., Cunha, G. R. and Donjacour, A. A. (1986a). Morphogenesis of ductal networks in the mouse prostate. *Biol. Reprod.* **34**, 961–971.

Sugimura, Y., Cunha, G. R. and Donjacour, A. A. (1986b). Morphological and histological study of castration-induced degeneration and androgen-induced regeneration in the mouse prostate. *Biol. Reprod.* **34**, 973–983.

Sun, C., Dobi, A., Mohamed, A., Li, H., Thangapazham, R. L., Furusato, B., Shaheduzzaman, S., Tan, S.-H., Vaidyanathan, G., Whitman, E., et al. (2008). TMPRSS2-ERG fusion, a common genomic alteration in prostate

cancer activates C-MYC and abrogates prostate epithelial differentiation. *Oncogene* **27**, 5348–5353.

Sun, M., Choueiri, T. K., Hamnvik, O.-P. R., Preston, M. A., De Velasco, G., Jiang, W., Loeb, S., Nguyen, P. L. and Trinh, Q.-D. (2016). Comparison of Gonadotropin-Releasing Hormone Agonists and Orchiectomy: Effects of Androgen-Deprivation Therapy. *JAMA Oncol* **2**, 500–507.

Sun, C.-C., Li, S.-J., Chen, Z.-L., Li, G., Zhang, Q. and Li, D.-J. (2019). Expression and Prognosis Analyses of Runt-Related Transcription Factor Family in Human Leukemia. *Mol Ther Oncolytics* **12**, 103–111.

Svensson, V., Vento-Tormo, R. and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols* **13**, 599–604.

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., et al. (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun* **4**, 2924.

Szczyrba, J., Niesen, A., Wagner, M., Wandernoth, P. M., Aumüller, G. and Wennemuth, G. (2017). Neuroendocrine Cells of the Prostate Derive from the Neural Crest. *J Biol Chem* **292**, 2021–2031.

Takayama, K., Suzuki, T., Tsutsumi, S., Fujimura, T., Urano, T., Takahashi, S., Homma, Y., Aburatani, H. and Inoue, S. (2015). RUNX1, an androgen- and EZH2-regulated gene, has differential roles in AR-dependent and -independent prostate cancer. *Oncotarget* **6**, 2263–2276.

Tan, M. H. E., Li, J., Xu, H. E., Melcher, K. and Yong, E. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol. Sin.* **36**, 3–23.

Tang, Y. Y., Crute, B. E., Kelley, J. J., Huang, X., Yan, J., Shi, J., Hartman, K. L., Laue, T. M., Speck, N. A. and Bushweller, J. H. (2000). Biophysical characterization of interactions between the core binding factor alpha and beta subunits and DNA. *FEBS Lett.* **470**, 167–172.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382.

Tay, L. S., Krishnan, V., Sankar, H., Chong, Y. L., Chuang, L. S. H., Tan, T. Z., Kolinjivadi, A. M., Kappei, D. and Ito, Y. (2018). RUNX Poly(ADP-Ribosyl)ation and BLM Interaction Facilitate the Fanconi Anemia Pathway of DNA Repair. *Cell Reports* **24**, 1747–1755.

Taylor, R. A., Toivanen, R. and Risbridger, G. P. (2010). Stem cells in prostate cancer: treating the root of the problem. *Endocr Relat Cancer* **17**, R273–R285.

Taylor, R. A., Fraser, M., Rebello, R. J., Boutros, P. C., Murphy, D. G., Bristow, R. G. and Risbridger, G. P. (2019). The influence of BRCA2 mutation on localized prostate cancer. *Nature Reviews Urology* **16**, 281–290.

Telfer, J. C. and Rothenberg, E. V. (2001). Expression and function of a stem cell promoter for the murine CBFalpha2 gene: distinct roles and regulation in natural killer and T cell development. *Dev. Biol.* **229**, 363–382.

Telfer, J. C., Hedblom, E. E., Anderson, M. K., Laurent, M. N. and Rothenberg, E. V. (2004). Localization of the domains in Runx transcription factors required for the repression of CD4 in thymocytes. *J. Immunol.* **172**, 4359–4370.

Teo, M. Y., Rathkopf, D. E. and Kantoff, P. (2019). Treatment of Advanced Prostate Cancer. *Annu Rev Med* **70**, 479–499.

Terry, A., Kilbey, A., Vaillant, F., Stewart, M., Jenkins, A., Cameron, E. and Neil, J. C. (2004). Conservation and expression of an alternative 3' exon of Runx2 encoding a novel proline-rich C-terminal domain. *Gene* **336**, 115–125.

Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Möröy, T., et al. (2016). GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat. Cell Biol.* **18**, 21–32.

Thomson, A. A. and Marker, P. C. (2006). Branching morphogenesis in the prostate gland and seminal vesicles. *Differentiation* **74**, 382–392.

Tietjen, I., Rihel, J. M., Cao, Y., Koentges, G., Zakhary, L. and Dulac, C. (2003). Single-Cell Transcriptional Analysis of Neuronal Progenitors. *Neuron* **38**, 161–175.

Tijssen, M. R., Cvejic, A., Joshi, A., Hannah, R. L., Ferreira, R., Forrai, A., Bellissimo, D. C., Oram, S. H., Smethurst, P. A., Wilson, N. K., et al. (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell* **20**, 597–609.

Tika, E., Ousset, M., Dannau, A. and Blanpain, C. (2019). Spatiotemporal regulation of multipotency during prostate development. *Development* dev.180224.

Timms, B. G. (2008). Prostate development: a historical perspective. *Differentiation* **76**, 565–577.

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the

multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196.

Tober, J., Yzaguirre, A. D., Piwarzyk, E. and Speck, N. A. (2013). Distinct temporal requirements for Runx1 in hematopoietic progenitors and stem cells. *Development* **140**, 3765–3776.

Toivanen, R. and Shen, M. M. (2017). Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. *Development* **144**, 1382–1398.

Tolis, G., Ackman, D., Stellos, A., Mehta, A., Labrie, F., Fazekas, A. T., Comaru-Schally, A. M. and Schally, A. V. (1982). Tumor growth inhibition in patients with prostatic carcinoma treated with luteinizing hormone-releasing hormone agonists. *Proc Natl Acad Sci U S A* **79**, 1658–1662.

Tolkach, Y. and Kristiansen, G. (2018). The Heterogeneity of Prostate Cancer: A Practical Approach. *PAT* **85**, 108–116.

Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648.

Tomlins, S. A., Alshalalfa, M., Davicioni, E., Erho, N., Yousefi, K., Zhao, S., Haddad, Z., Den, R. B., Dicker, A. P., Trock, B. J., et al. (2015). Characterization of 1577 Primary Prostate Cancers Reveals Novel Biological and Clinicopathologic Insights into Molecular Subtypes. *European Urology* **68**, 555–567.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**, 381–386.

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. A. and Quake, S. R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375.

Tsang, J. C. H., Yu, Y., Burke, S., Buettner, F., Wang, C., Kolodziejczyk, A. A., Teichmann, S. A., Lu, L. and Liu, P. (2015). Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. *Genome Biology* **16**, 178.

Tsujimura, A., Koikawa, Y., Salm, S., Takao, T., Coetzee, S., Moscatelli, D., Shapiro, E., Lepor, H., Sun, T.-T. and Wilson, E. L. (2002). Proximal location of mouse prostate epithelial stem cells. *J Cell Biol* **157**, 1257–1265.

Tsuzuki, S. and Seto, M. (2012). Expansion of functionally defined mouse hematopoietic stem and progenitor cells by a short isoform of RUNX1/AML1. *Blood* **119**, 727–735.

Tsuzuki, S., Hong, D., Gupta, R., Matsuo, K., Seto, M. and Enver, T. (2007). Isoform-specific potentiation of stem and progenitor cell engraftment by AML1/RUNX1. *PLoS Med.* **4**, e172.

Umansky, K. B., Gruenbaum-Cohen, Y., Tsoory, M., Feldmesser, E., Goldenberg, D., Brenner, O. and Groner, Y. (2015). Runx1 Transcription Factor Is Required for Myoblasts Proliferation during Muscle Regeneration. *PLOS Genetics* **11**, e1005457.

Umkehrer, C., Holstein, F., Formenti, L., Jude, J., Froussios, K., Neumann, T., Cronin, S. M., Haas, L., Lipp, J. J., Burkard, T. R., et al. (2020). Isolating live cell clones from barcoded populations using CRISPRa-inducible reporters. *Nature Biotechnology* 1–5.

Uzgare, A. R., Xu, Y. and Isaacs, J. T. (2004). In vitro culturing and characteristics of transit amplifying epithelial cells from human prostate tissue. *J Cell Biochem* **91**, 196–205.

Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S. and Marioni, J. C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nature Methods* **14**, 565–571.

van Bragt, M. P. A., Hu, X., Xie, Y. and Li, Z. (2014). RUNX1, a transcription factor mutated in breast cancer, controls the fate of ER-positive mammary luminal cells. *Elife* **3**, e03881.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., Burdziak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e27.

VanHorn, S. and Morris, S. A. (2020). Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. *Developmental Cell*.

Varambally, S., Dhanasekaran, S. M., Zhou, M., Barrette, T. R., Kumar-Sinha, C., Sanda, M. G., Ghosh, D., Pienta, K. J., Sewalt, R. G. A. B., Otte, A. P., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629.

Ventura, A., Kirsch, D. G., McLaughlin, M. E., Tuveson, D. A., Grimm, J., Lintault, L., Newman, J., Reczek, E. E., Weissleder, R. and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**, 661–665.

Verze, P., Cai, T. and Lorenzetti, S. (2016). The role of the prostate in male fertility, health and disease. *Nat Rev Urol* **13**, 379–386.

Voon, D. C.-C., Hor, Y. T. and Ito, Y. (2015). The RUNX complex: reaching beyond haematopoiesis into immunity. *Immunology* **146**, 523–536.

Voronov, D., Gromova, A., Liu, D., Zoukhri, D., Medvinsky, A., Meech, R. and Makarenkova, H. P. (2013). Transcription Factors Runx1 to 3 Are Expressed in the Lacrimal Gland Epithelium and Are Involved in Regulation of Gland Morphogenesis and Regeneration. *Invest Ophthalmol Vis Sci* **54**, 3115–3125.

Vu, L. P., Perna, F., Wang, L., Voza, F., Figueroa, M. E., Tempst, P., Erdjument-Bromage, H., Gao, R., Chen, S., Paietta, E., et al. (2013). PRMT4 blocks myeloid differentiation by assembling a methyl-RUNX1-dependent repressor complex. *Cell Rep* **5**, 1625–1638.

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G. and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987.

Wang, S. W. and Speck, N. A. (1992). Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. *Mol. Cell. Biol.* **12**, 89–102.

Wang, S., Wang, Q., Crute, B. E., Melnikova, I. N., Keller, S. R. and Speck, N. A. (1993). Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. *Mol Cell Biol* **13**, 3324–3339.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H. and Speck, N. A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3444–3449.

Wang, Y., Hayward, S., Cao, M., Thayer, K. and Cunha, G. (2001). Cell differentiation lineage in the prostate. *Differentiation* **68**, 270–279.

Wang, X.-P., Åberg, T., James, M. J., Levanon, D., Groner, Y. and Thesleff, I. (2005a). Runx2 (Cbfa1) Inhibits *Shh* Signaling in the Lower but not Upper Molars of Mouse Embryos and Prevents the Budding of Putative Successional Teeth. *Journal of Dental Research* **84**, 138–143.

Wang, X., Blagden, C., Fan, J., Nowak, S. J., Taniuchi, I., Littman, D. R. and Burden, S. J. (2005b). Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. *Genes Dev* **19**, 1715–1722.

Wang, X.-D., Wang, B.-E., Soriano, R., Zha, J., Zhang, Z., Modrusan, Z., Cunha, G. R. and Gao, W.-Q. (2007). Expression profiling of the mouse prostate after

castration and hormone replacement: implication of H-cadherin in prostate tumorigenesis. *Differentiation* **75**, 219–234.

Wang, Z., Gerstein, M. and Snyder, M. (2009a). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63.

Wang, X., Kruithof-de Julio, M., Economides, K. D., Walker, D., Yu, H., Halili, M. V., Hu, Y.-P., Price, S. M., Abate-Shen, C. and Shen, M. M. (2009b). A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**, 495–500.

Wang, L., Gural, A., Sun, X.-J., Zhao, X., Perna, F., Huang, G., Hatlen, M. A., Vu, L., Liu, F., Xu, H., et al. (2011a). The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science* **333**, 765–769.

Wang, L., Brugge, J. S. and Janes, K. A. (2011b). Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. *Proceedings of the National Academy of Sciences* **108**, E803–E812.

Wang, Z. A., Mitrofanova, A., Bergren, S. K., Abate-Shen, C., Cardiff, R. D., Califano, A. and Shen, M. M. (2013). Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nature Cell Biology* **15**, 274–283.

Wang, C. Q., Krishnan, V., Tay, L. S., Chin, D. W. L., Koh, C. P., Chooi, J. Y., Nah, G. S. S., Du, L., Jacob, B., Yamashita, N., et al. (2014). Disruption of Runx1 and Runx3 Leads to Bone Marrow Failure and Leukemia Predisposition due to Transcriptional and DNA Repair Defects. *Cell Reports* **8**, 767–782.

Wang, B., Wang, X., Long, J. E., Eastham-Anderson, J., Firestein, R. and Junttila, M. R. (2015). Castration-Resistant Lgr5+ Cells Are Long-Lived Stem Cells Required for Prostatic Regeneration. *Stem Cell Reports* **4**, 768–779.

Wang, X., Xu, H., Cheng, C., Ji, Z., Zhao, H., Sheng, Y., Li, X., Wang, J., Shu, Y., He, Y., et al. (2020a). Identification of a Zeb1 expressing basal stem cell subpopulation in the prostate. *Nature Communications* **11**, 706.

Wang, Z., Song, Y., Ye, M., Dai, X., Zhu, X. and Wei, W. (2020b). The diverse roles of SPOP in prostate cancer and kidney cancer. *Nature Reviews Urology* **17**, 339–350.

Warren, L., Bryder, D., Weissman, I. L. and Quake, S. R. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *PNAS* **103**, 17807–17812.

Watson, P. A., Arora, V. K. and Sawyers, C. L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer* **15**, 701–711.

Wegner, K. A., Cadena, M. T., Trevena, R., Turco, A. E., Gottschalk, A., Halberg, R. B., Guo, J., McMahon, J. A., McMahon, A. P. and Vezina, C. M. (2017). An immunohistochemical identification key for cell types in adult mouse prostatic and urethral tissue sections. *PLoS One* **12**,.

Wei, X., Zhang, L., Zhou, Z., Kwon, O.-J., Zhang, Y., Nguyen, H., Dumpit, R., True, L., Nelson, P., Dong, B., et al. (2019). Spatially Restricted Stromal Wnt Signaling Restrains Prostate Epithelial Progenitor Growth through Direct and Indirect Mechanisms. *Cell Stem Cell* **24**, 753-768.e6.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. and Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**,.

Welch, H. G. and Black, W. C. (2010). Overdiagnosis in cancer. *J Natl Cancer Inst* **102**, 605–613.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544.

Wilt, T. J., MacDonald, R., Rutks, I., Shamliyan, T. A., Taylor, B. C. and Kane, R. L. (2008). Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Ann Intern Med* **148**, 435–448.

Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L. and Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology* **20**, 59.

Wu, X., Wu, J., Huang, J., Powell, W. C., Zhang, J., Matusik, R. J., Sangiorgi, F. O., Maxson, R. E., Sucov, H. M. and Roy-Burman, P. (2001). Generation of a prostate epithelial cell-specific Cre transgenic mouse model for tissue-specific gene ablation. *Mech. Dev.* **101**, 61–69.

Wu, X., Gong, S., Roy-Burman, P., Lee, P. and Culig, Z. (2013). Current mouse and cell models in prostate cancer research. *Endocr Relat Cancer* **20**,.

Xiao, Z. S., Thomas, R., Hinson, T. K. and Quarles, L. D. (1998). Genomic structure and isoform expression of the mouse, rat and human Cbfa1/Osf2 transcription factor. *Gene* **214**, 187–197.

Yan, J., Liu, Y., Lukasik, S. M., Speck, N. A. and Bushweller, J. H. (2004). CBFβ allosterically regulates the Runx1 Runt domain via a dynamic conformational equilibrium. *Nature Structural & Molecular Biology* **11**, 901.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural & Molecular Biology* **20**, 1131–1139.

Yang, X., Kui, L., Tang, M., Li, D., Wei, K., Chen, W., Miao, J. and Dong, Y. (2020). High-Throughput Transcriptome Profiling in Drug and Biomarker Discovery. *Front. Genet.* **11**,.

Yarmus, M., Woolf, E., Bernstein, Y., Fainaru, O., Negreanu, V., Levanon, D. and Groner, Y. (2006). Groucho/transducin-like Enhancer-of-split (TLE)-dependent and -independent transcriptional regulation by Runx3. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7384–7389.

Ye, M., Iwasaki, H., Laiosa, C. V., Stadtfeld, M., Xie, H., Heck, S., Clausen, B., Akashi, K. and Graf, T. (2003). Hematopoietic stem cells expressing the myeloid lysozyme gene retain long-term, multilineage repopulation potential. *Immunity* **19**, 689–699.

Yeh, H.-Y., Cheng, S.-W., Lin, Y.-C., Yeh, C.-Y., Lin, S.-F. and Soo, V.-W. (2009). Identifying significant genetic regulatory networks in the prostate cancer from microarray data based on transcription factor analysis and conditional independency. *BMC Med Genomics* **2**, 70.

Yokomizo, T., Yamada-Inagawa, T., Yzaguirre, A. D., Chen, M. J., Speck, N. A. and Dzierzak, E. (2012). Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nat Protoc* **7**, 421–431.

Yoo, Y. A., Roh, M., Naseem, A. F., Lysy, B., Desouki, M. M., Unno, K. and Abdulkadir, S. A. (2016). Bmi1 marks distinct castration-resistant luminal progenitor cells competent for prostate regeneration and tumour initiation. *Nature Communications* **7**, 12943.

Yoshida, C. A. (2004). Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes & Development* **18**, 952–963.

Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M. J., Porayette, P., Shelley, W. C., Conway, S. J., Dorshkind, K. and Yoder, M. C. (2011). Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1468–1473.

Yoshimoto, M., Porayette, P., Glosson, N. L., Conway, S. J., Carlesso, N., Cardoso, A. A., Kaplan, M. H. and Yoder, M. C. (2012). Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood* **119**, 5706–5714.

Zaidi, S. K., Javed, A., Choi, J. Y., van Wijnen, A. J., Stein, J. L., Lian, J. B. and Stein, G. S. (2001). A specific targeting signal directs Runx2/Cbfa1 to subnuclear domains and contributes to transactivation of the osteocalcin gene. *J. Cell. Sci.* **114**, 3093–3102.

Zambidis, E. T., Peault, B., Park, T. S., Bunz, F. and Civin, C. I. (2005). Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development. *Blood* **106**, 860–870.

Zaret, K. S. and Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes & Development* **25**, 2227–2241.

Zeigler, B. M., Sugiyama, D., Chen, M., Guo, Y., Downs, K. M. and Speck, N. A. (2006). The allantois and chorion, when isolated before circulation or chorio-allantoic fusion, have hematopoietic potential. *Development* **133**, 4183–4192.

Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142.

Zeng, C., McNeil, S., Pockwinse, S., Nickerson, J., Shopland, L., Lawrence, J. B., Penman, S., Hiebert, S., Lian, J. B., Wijnen, A. J. van, et al. (1998). Intranuclear targeting of AML/CBFα regulatory factors to nuclear matrix-associated transcriptional domains. *PNAS* **95**, 1585–1589.

Zhang, L., Fried, F. B., Guo, H. and Friedman, A. D. (2008). Cyclin-dependent kinase phosphorylation of RUNX1/AML1 on 3 sites increases transactivation potency and stimulates cell proliferation. *Blood* **111**, 1193–1200.

Zhang, Y., Xie, R.-L., Croce, C. M., Stein, J. L., Lian, J. B., van Wijnen, A. J. and Stein, G. S. (2011). A program of microRNAs controls osteogenic lineage progression by targeting transcription factor Runx2. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9863–9868.

Zhang, D., Zhao, S., Li, X., Kirk, J. S. and Tang, D. G. (2018). Prostate Luminal Progenitor Cells in Development and Cancer. *Trends in Cancer* **4**, 769–783.

Zhang, Y., Zhou, C. K., Rencsok, E. M., Fall, K., Lotan, T. L., Loda, M., Giunchi, F., Platz, E. A., Marzo, A. M. D., Mucci, L. A., et al. (2019). A prospective study

of intraprostatic inflammation, focal atrophy, and progression to lethal prostate cancer. *Cancer Epidemiol Biomarkers Prev*.

Zhao, X., Jankovic, V., Gural, A., Huang, G., Pardanani, A., Menendez, S., Zhang, J., Dunne, R., Xiao, A., Erdjument-Bromage, H., et al. (2008). Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. *Genes Dev.* **22**, 640–653.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049.

Zhou, T., Luo, M., Cai, W., Zhou, S., Feng, D., Xu, C. and Wang, H. (2018). Runt-Related Transcription Factor 1 (RUNX1) Promotes TGF-β-Induced Renal Tubular Epithelial-to-Mesenchymal Transition (EMT) and Renal Fibrosis through the PI3K Subunit p110δ. *EBioMedicine* **31**, 217–225.

Zhu, X., Yeadon, J. E. and Burden, S. J. (1994). AML1 is expressed in skeletal muscle and is regulated by innervation. *Mol. Cell. Biol.* **14**, 8051–8057.

Zong, Y. and Goldstein, A. S. (2013). Adaptation or selection—mechanisms of castration-resistant prostate cancer. *Nature Reviews Urology* **10**, 90–98.

Zong, Y., Xin, L., Goldstein, A. S., Lawson, D. A., Teitell, M. A. and Witte, O. N. (2009). ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells. *Proc Natl Acad Sci U S A* **106**, 12465–12470.

Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., Zanetta, L., Dejana, E., Gasson, J. C., et al. (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell* **3**, 625–636.