Investigating the tumour suppressor function of RUNX1 in breast cancer

Kerri Louise Sweeney

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Cancer Sciences
College of Medical, Veterinary & Life Sciences
University of Glasgow

Cancer Research UK Beatson Institute
Garscube Estate
Switchback Road
Glasgow
G61 1BD
United Kingdom

October 2022

Abstract

Despite the remarkable advances in the understanding of breast cancer since the 1940s, and therefore in the treatment of the disease, there are still patients whose cancer either cannot be addressed using current targeted therapies or whose cancer does not respond to therapy in the desired manner. These issues highlight the need to identify novel prognostic/therapeutic targets in breast cancer in order that these unmet clinical needs can be addressed.

Various sequencing studies have facilitated the identification of driver mutations in breast cancer and mapping of the breast cancer genomic landscape. Such studies resulted in the identification of *RUNX1* as a gene of interest in breast cancer as tumours frequently harboured loss-of-function mutations in *RUNX1*, particularly those classed as oestrogen receptor(ER)-positive. Previously, RUNX1 has been shown to be an important player in leukaemias where it elicits both a tumour suppressor and oncogenic role depending on context. Likewise, studies focusing on its role in breast cancer have revealed that while ER-positive breast cancers often have RUNX1 mutations, indicating a tumour suppressive role, high levels of RUNX1 expression in the triple-negative (ER-/PR-/HER2-) subtype has been correlated with poor outcome in patients.

Using genetically engineered mouse models (GEMM) for mammary tumorigenesis (with or without the additional mammary-specific deletion of Runx1), our laboratory provided the first definitive *in vivo* confirmation of RUNX1 acting to restrict tumour development in preclinical models of breast cancer. Accelerated emergence and increased numbers of multifocal and multicentric lesions were observed upon depletion of Runx1, indicating a possible role for RUNX1 in regulating the stem-like potential of the mammary epithelium. Evidence from one of these models suggests the early expansion of a tumour-initiating sub-population of cells with an elevated stem/progenitor-like potential, which may at least partly explain this phenotype.

In this thesis, the mechanisms behind the tumour suppressive functions of RUNX1 in the mammary epithelium were explored, with a particular focus on the transcriptional alterations that are initiated upon deleting this transcription factor. RNA-Sequencing highlighted a myriad of biological functions and pathways that this protein is essential for and offers insights into the various potential mechanisms by which it exerts its tumour-suppressing functions. Notably, the data indicates altered expression of stem-related genes in RUNX1-deleted cells, and thereby offers potential mechanisms that may underpin RUNX1-dependent control over the stem cell compartment (including the regulation of stem-associated genes such as *Aldh1a1*, *Aldh1a7*, and *Bcl11b*), thereby offering some explanation as to why tumours, in which RUNX1 is deleted, exhibit accelerated onset.

Using an immortalised mammary cell line, these findings were validated using mammosphere and 3D colony formation assays for the analysis of stem/progenitor-like potential. While CRISPR/Cas9-mediated deletion of RUNX1 in cells elevates the stem-like potential of mammary epithelial cells, ectopic expression of RUNX1 conversely decreased mammosphere and colony formation capabilities. Additionally, loss of RUNX1 function appears to enhance the stemness-promoting effects of WNT3A-treatment, while its overexpression dampens the stemness-promoting abilities of WNT3A. Furthermore, stem-like genes such as *Aldh1a1* and *Aldh1a7* were enriched in *Runx1*-deleted 3D colonies. Together, this work indicates that RUNX1 is an essential factor for the regulation of the stem-like

compartment of mammary epithelial cells, and its loss may result in the expansion of a cellular subpopulation that is more vulnerable to transformation following oncogenic insult.

Collectively, this research offers some valuable insights into the potential functional and molecular mechanisms behind the tumour suppressive function of RUNX1 in the mammary gland, and provides promising avenues to explore in order to extend our understanding of this transcription factor in breast cancer.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to offer my sincerest thanks to my supervisor, Karen Blyth for offering me the opportunity to work as part of her amazing team. Her complete devotion and unfiltered enthusiasm for her work has been inspiring, and I feel very lucky to have been given her valuable guidance and support over the past 5 years. I would also like to extend my gratitude to my second supervisor, Ewan Cameron, for contributing his immense knowledge during our illuminating, and often very long, discussions. I am thankful for the time he committed to our meetings. Thank you to my advisors, Jim Norman and Seth Coffelt, for their helpful input given to me throughout my PhD studies.

To all Y35 members, past and present, I would like to offer my sincere thanks for the support and advice you have all given me, and for helping to make my time at the Beatson more enjoyable. I am extremely grateful to Alessandra Riggio for teaching me so much during my first year at the Beatson, and for entrusting me to continue on hers and Nicola Ferrari's fascinating work. Thanks to Susan Mason and Sandeep Dhayade for being extremely patient while teaching me how to perform the various animal experiments that were so key to my PhD project, and to Nicholas Rooney for all of the help he gave me in the lab and for answering my many questions about CRISPR. A special thanks to Dimitris Athineos for all of his hard work on the RNA-Seq project — it would not have been possible without your help — and for his advice on qPCR. I would also like to extend my gratitude to Kirsteen Campbell for the extremely valuable advice that she has offered me throughout my studies, and for always going out of her way to make herself available to chat to if I ever need help or advice. And, of course, thank you to my fellow Y35 PhD students, Narisa Phinichkusolchit, Adiba Khan, and Matthew Winder for all the help and support you've given me.

To Breast Cancer Now and CRUK, who both funded my PhD studies, and to the University of Glasgow, who facilitated this degree, I extend my sincere gratitude.

The work carried out by Core Facilities is essential for so much of the research we do at the Beatson. Firstly, I would like to thank Colin Nixon and his team in Histology for helping with my many questions and requests, Tom Gilbey and Yi-Hsia Liu for their help and advice with flow cytometry analysis and sorting, the BSU/BRU staff for supporting my animal work, and Margaret O'Prey for her expert advice on all things microscopy. Special thanks go to Billy Clark and Robin Shaw for all their amazing help with RNA-Sequencing and bioinformatics, and for being patient and helpful with my many questions.

I would also like to take this opportunity to sincerely thank Farah Hughes, who helped me with DNA sequencing of my CRISPR clones and answered at least a hundred questions. She is always a joy to grab a coffee/chat with, and never fails to cheer me up.

Thank you also to Jackie Beesley, Catherine Winchester and Katharina Schraut for all their hard work and for the support they have extended to me throughout my studies. I would like to thank Heather Jørgensen for her continued advice and support she has provided to me, particularly over the past year. Huge thanks to my viva examiners, Helen Wheadon and Rob Clarke, and viva convenor, Heather Jørgensen, for making my viva such an enjoyable experience.

To my former undergraduate supervisor, Frances Fuller-Pace, I would like to offer my genuine thanks for encouraging me to pursue a PhD. Without your support and teachings, I would not have considered Cancer Sciences as a career path.

To my mentor and good friend Natalie Harries, thank you for all of our lovely chats and for always giving me a boost when I need it. To my Dundee pals Amy, Charlie, and Nicola — I could not have gotten through undergrad without you all and I'm so glad I can always count on you for moral support. To my awesome friend Colin for always checking in on me and cheering me up when I'm in a bad mood. To my fellow Trainee Clinical Scientists at the Queen Elizabeth for tolerating my many rants. And to my Glasgow friend, Amy, for being an all-round amazing person.

Massive thank yous go to my amazing family: my mum, big sis Rebecca, wee brother Liam, and brother-in-law Craig. Thank you for always believing in me and being there when I need you. An extra special thank you to my fab nieces Ariane, Robin, and Amelia who never fail to make me happy when I see them. I also want to take this opportunity to acknowledge the family who sadly were not able to see me finish my PhD, but had a huge impact on my life and were always supportive of me: my amazing "Huntly Grannie" Lena who gave the best hugs and made the best tablet; my brilliantly weird and awesome Auntie Aileen who was taken from us far too soon; and my caring and encouraging grandparents from Bridgend, Granny Barbara and Grandad Cliff. I miss you all and know you would be so proud of how far I've come.

Finally, I have reserved my biggest thank you to my partner, and best friend, Chris. You have supported me through my bad moods, my even worse moods, AND we managed to survive through a pandemic and 2 PhDs. I could not have done this without you — I love you.

THANK YOU!

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Kerri Louise Sweeney

October 2022

Definitions/Abbreviations

A	area
ABC	ATP binding cassette
Actb	actin beta
AKT	protein kinase B
ALDH	aldehyde dehydrogenase
ALDH1	aldehyde dehydrogenase 1 family
ALDH1A1	aldehyde dehydrogenase 1 family, member A1
Aldh1a7	aldehyde dehydrogenase family 1, subfamily A7
Amp	ampicillin
ANOVA	analysis of variance
AP1	activator protein 1
Apc	adenomatous polyposis coli
AXIN	axis inhibition protein
AXIN1	axis inhibition protein 1
AXIN2	axis inhibition protein 2
Axl	AXL receptor tyrosine kinase
BCA	bicinchoninic acid
Bcl11b	B cell leukemia 11b
BCL2	B-cell lymphoma 2
BLAST	Basic Local Alignment Search Tool
Blg	beta-lactoglobulin
BRCA	breast cancer gene
BRCA1	breast cancer gene 1
BRCA2	breast cancer gene 2
BSA	bovine serum albumin
Cas9	CRISPR associated protein 9
Catnb	beta catenin
CBFβ	core-binding factor subunit beta
CCD	cleidocranial dysplasia
Ccnd1	cyclin D1

CD24	cluster of differentiation 24, aka heat stable antigen CD24
CD29	integrin beta-1
CD31	cluster of differentiation 31, aka platelet endothelial cell adhesion molecule (PECAM-1)
CD45	leucocyte common antigen, aka protein tyrosine phosphatase, receptor type, C (PTPRC)
CD49f	integrin α6
CDK	cyclin-dependent kinase
cDNA	complementary deoxyribonucleic acid
ChIP-Seq	chromatin immunoprecipitation sequencing
CK14	cytokeratin 14
*CK8/*18	cytokeratin 8/18
Col4a2	collagen type IV alpha 2 chain
CRC	colorectal cancer
Cre	carbapenem-resistant Enterobacterales
CRISPR	clustered regularly interspaced short palindromic repeats
D	days
Da	dalton
DAPI	4’,6-diamidino-2-phenylindole
Dhh	desert hedgehog
Dkk1	dickkopf WNT signaling pathway inhibitor 1
Dkk2	dickkopf WNT signaling pathway inhibitor 2
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DON	date of tumour notice
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
eGFP	enhanced green fluorescent protein
EGFR	epidermal growth factor receptor
EMT	epithelial-to-mesenchymal transition
EP	clinical end-point
ER	oestrogen receptor

ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ERK	extracellular signal-regulated kinases
ERT2	estrogen ligand-binding domain
Erα	oestrogen receptor alpha
ESR1	oestrogen receptor 1
Esr2	oestrogen receptor 2
FACS	fluorescence-activated cell sorting
Fas	Fas cell surface death receptor
FBS	foetal bovine serum
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded
FGF-2	fibroblast growth factor 2
fl	floxed
FMO	fluorescence minus one
FOXO	forkhead box O transcription factors
FOXO1	forkhead box protein O1
FSC	forward scatter
Fzd7	frizzled class receptor 7
g	gram
G	gauge
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
gDNA	genomic DNA
GEMM	genetically engineered mouse model
Gli	GLI family zinc finger
gRNA	guide RNA
GRP	green fluorescent protein
GSEA	gene set enrichment analysis
Gt(ROSA)26Sortm1Hjf	gene trap ROSA 26, Philippe Soriano; targeted mutation 1, Hans Jorg Fehling
HCC	hepatocellular carcinoma
HER2	human epidermal growth factor receptor 2
HFSC	hair follicle stem cell
HIER	heat-induced epitope retrieval

HPV	human papillomavirus
HPV	human papillomavirus
HSC	haematopoietic stem cell
IF	immunofluorescence
IHC	Immunohistochemistry
Ihh	indian hedgehog
kDa	kilodalton
Ki-67	marker Of proliferation Ki-67
KO	knock-out
LB	lysogeny broth
LC	liver cirrhosis
LGR	leucine-rich repeat-containing G-protein coupled receptor
Lgr4	leucine-rich repeat-containing G-protein coupled receptor 4
LGR5	leucine-rich repeat-containing G-protein coupled receptor 5
Lgr6	leucine-rich repeat-containing G-protein coupled receptor 5
Lin	lineage
MDPs	multi-drug-resistance proteins
MEC	mammary epithelial cell
METABRIC	Molecular Taxonomy Of Breast Cancer International Consortium
Min	multiple intestinal neoplasia
ml	millilitre
mm	millimeter
MMEC	mouse mammary epithelial cell
Mmp12	matrix metallopeptidase 12
Mmp9	matrix metallopeptidase 9
MMTV	mouse mammary tumour virus
MP	multiparous
MSigDB	The Molecular Signatures Database
mTOR	mammalian target of rapamycin
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
NBF	neutral buffered formalin

Neo	neomycin
NES	normalised enrichment score
NF-kB	nuclear factor-kB
ng	nanogram
NGS	normal goat serum
NK cells	natural killer cells
nm	nanometer
NOD/SCID	nonobese diabetic/severe combined immunodeficiency
NOTCH	neurogenic locus notch homolog protein
Notum	notum, palmitoleoyl-protein carboxylesterase
NP	nulliparous
PARP	poly(ADP-ribose) polymerase
PARPi	poly(ADP-ribose) polymerase inhibitors
PBS	phosphate buffered saline
PBS-T	PBS-Tween
PC	principal component
PCA	principle component analysis
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PD-L1	programmed death-ligand 1
PEACS	perturbation-expression analysis of cell states
P-gp	P-glycoprotein
PI3K	Phosphatidylinositol-3-kinase
PN	pre-neoplastic
PPARγ	peroxisome proliferator-activated receptor gamma
PR	progesterone receptor
PSG	penicillin/streptomycin/glutamine
Ptch	protein patched homolog
PTM	post-translational modification
Puro	puromycin
pX-459-Puro	pSpCas9(BB)-2A-Puro
PyMT	polyoma virus middle T antigen

qPCR	quantitative PCR
Rag1	recombination activating gene 1
RB	retinoblastoma gene
RFP	red fluorescent protein
RHD	runt-homology domain
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-seq	RNA sequencing
Rnf43	ring finger protein 43
RTK	receptor tyrosine kinase
RT-qPCR	real-time quantitative reverse transcription PCR
RUNX1	runt-related transcription factor 1
RUNX2	runt-related transcription factor 2
RUNX3	runt-related transcription factor 3
SCC	squamous cell carcinoma
Sfrp2	secreted frizzled-related protein 2
Shh	sonic hedgehog
shRNA	short/small hairpin RNA
Smo	smoothened, frizzled class receptor
Snai2	snail family transcriptional repressor 2
Sox9	SRY-box transcription factor 9
SpCas9	Cas9 variant isolated from the bacterium Streptococcus pyogenes
SSC	side scatter
STAT1	signal transducer and activator of transcription 1
SV40	simian virus 40
TBS	tris buffered saline
TBS-T	TBS-Tween
TCGA	The Cancer Genome Atlas
TCs	thymocytes
tdRFP	tandem-dimer red fluorescent protein
TGFβ	transforming growth factor beta
TMA	tumour microarray

TNBC	triple-negative breast cancer
TNM	tumour, node, metastasis
TrkC	tropomyosin receptor kinase C
TRP53	transformation-related protein 53
WAP	whey acidic protein
Wif1	WNT inhibitory factor 1
wk	week
wt	wild-type
YAP	yes-associated protein 1
YFP	yellow fluorescent protein
Znrf3	zinc and ring finger 3
µg	microgram
µl	microlitre
µM	micrometer
∩	intersection

1 Introduction

Cancer is described as the uncontrolled growth and proliferation of abnormal cells, which often invade surrounding tissues and even spread to distant organs in a process known as metastasis. Human cells progress from normalcy to neoplastic states through the acquisition of a variety of functional capabilities that increase their ability to form malignant tumours. This set of underlying principles, which are essential for the neoplastic transformation of cells, are described in The Hallmarks of Cancer (Hanahan and Weinberg 2000, Hanahan and Weinberg 2011, Hanahan 2022). Despite their shared name, “cancer” is the name given to a large collection of diseases often with distinct genetic and pathological characteristics.

1.1 The ever-increasing prevalence of cancer

Worldwide cancer incidence has been growing significantly over the years and is predicted to increase a further ~57% between 2020 and 2040 (Ferlay J , Bray F 2006, Ferlay, Colombet et al. 2021). Estimations of the lifetime risk for developing cancer have also changed gradually over time, with 1 in 2 people born in the UK after 1960 now expected to develop a form of cancer at some point in their life (Ahmad, Ormiston-Smith et al. 2015). Although different cancer types have their own associated risk factors for the development of that specific disease, it is possible to explain much of the overall increased cancer risk and incidence with three main factors: increased population growth, ageing populations, and increased exposure to risk factors (You and Henneberg 2018, Sung, Ferlay et al. 2021).

1.2 Risk factors for cancer development

Ageing is considered one of the biggest risk factors for the development of most cancers due to accumulations in genetic anomalies during one’s lifetime and, due to increased population sizes with longer overall life expectancy and larger numbers of people living into old age, this increases the cancer development risk for the overall population (Smetana, Lacina et al. 2016).

Other risk factors are often more variable between countries and cancer types, however lifestyle changes among many populations have increased exposure to a variety of risk factors for several cancer types. These risk factors include, but are not limited to, the following: tobacco use; being overweight or obesity; unhealthy diet (diets high in red and processed meats, low fruit and vegetable intake); lack of physical activity; alcohol use; carcinogenic infections (e.g. Helicobacter pylori, Hepatitis B and C virus, Human papillomavirus); ultraviolet radiation exposure; urban air pollution (Institute of Medicine Committee on Cancer Control in 2007, 2022). Prevention strategies therefore involve avoidance of risk factors, reduced exposure to radiation, and vaccinating against carcinogenic viruses (HPV and hepatitis B).

Another contributor to the significant increase in new cancer cases is the improved and earlier detection and diagnosis of cancers due to increased awareness, upgraded and more standardised methods of testing, and screening of particular demographics within populations (some of whom may otherwise appear healthy, but may have some asymptomatic abnormality or pre-cancer) (Public Health England 2020, Nelson October 2019).

1.3 Changes in cancer-related deaths

As well as being a particularly prevalent disease, cancer is also still one of the leading causes of death globally, second only to cardiovascular diseases (2016, Ritchie 2018). Cancer-related deaths have continued to increase in spite of improved survival rates, and can again be explained by the growing and ageing global population (Ritchie 2018). With a larger population size often comes increased total death numbers and, due to massive progress being made with tackling causes of death that would have previously caused mortality at an earlier age, more people reach older ages where they are afflicted with illness or disease, such as cancer, that tend to cause death in older individuals.

1.4 Breast cancer

1.4.1 Breast cancer survival

Breast cancer is one such cancer type where there have been vast improvements made in relation to survival rates, with five-year relative survival rates in the UK around 80-85% in recent years (Cancer Research UK). However, because breast cancer is the second most common cancer type both in the UK and worldwide (Cancer Research UK , Ferlay J , Ferlay, Colombet et al. 2021), even survival rates of 85% translate into significant numbers of deaths. In the UK, breast cancer is still the fourth most common cause of cancer-related death overall and the second most common in women, resulting in an average of 32 deaths per day in the UK alone between 2017 and 2019 (Cancer Research UK). It is for this reason that there are still efforts and resources being put into continuing breast cancer research. In order for breast cancer survival statistics to be raised further, it is important to consider how such significant improvements have been achieved since the 1940s.

1.4.2 Breast cancer survival improvements

1.4.2.1 Improved breast cancer detection

Earlier detection of breast cancers in patients has been a key factor for reducing breast cancer-related deaths. This early detection has been facilitated through increased public knowledge of signs and symptoms, due to several charity-lead awareness campaigns, and routine screening programmes that are offered to post-menopausal women and other demographics who are particularly at risk, that have allowed for the improved detection of even asymptomatic breast cancers (Burton and Bell 2013, Mandelblatt, van Ravesteyn et al. 2013, Tabár, Dean et al. 2019). Detecting breast cancer at these early stages is important, as there are significantly more treatment options available, meaning there is an increased chance for survival. Treatment options are also often less aggressive at these earlier stages, leading to improved quality of life for patients. Once detected, decisions are managed and implemented by a multi-disciplinary team of healthcare professionals, who are well-trained and experienced in their respective fields . Compared to previously implemented approaches to breast cancer care, this highly stratified approach has significantly reduced the amount and variability of breast cancer-related mortality in patients (Kalager, Haldorsen et al. 2009, Kesson, Allardice et al. 2012).

1.4.2.2 Targeted therapies for breast cancers subtypes

The emergence of routine immunohistochemical (IHC) and molecular subtyping of breast cancers, and the development of therapies targeted to each of these subtypes, has provided more options and helped to inform healthcare professionals' decisions on the most appropriate therapeutic approach to take for each individual (Yersal and Barutca 2014, Zaha 2014). Breast tumours are often tested for the presence of oestrogen and progesterone hormone receptors (ER and PR) and human epidermal growth factor receptor 2 (HER2). ER and PR are expressed in the vast majority of breast cancers (with ~75% of cases being ER/PR-positive (Onitilo, Engel et al. 2009, Kohler, Sherman et al. 2015). Tumour cells often experience increased sensitivity to oestrogen, thereby leading to aberrant ER signalling and overexpression of a number of downstream target genes that fuel the growth and progression of the tumour (Gross and Yee 2002). Endocrine therapies, such as tamoxifen that binds to ERs to block oestrogen binding and signalling; fulvestrant that blocks and damages ER; and aromatase inhibitors (e.g. anastrozole and letrozole) that block oestrogen production, have been imperative to improving survival in patients with hormone-positive breast cancers (Meisel, Venur et al. 2018).

Breast cancers with *ERBB2* gene amplification or overexpression of its associated HER2 protein (described as HER2-positive tumours) account for ~15% of all cases and are associated with increased tumour aggression (Onitilo, Engel et al. 2009, Rakha, Pinder et al. 2015). Targeted therapies have also been developed to block HER2 function in breast cancer, including: monoclonal antibodies (e.g. trastuzumab and pertuzumab) to block HER2 protein function; antibody-drug conjugates (e.g. ado-trastuzumab emtansine and fam-trastuzumab deruxtecan) to target chemotherapy specifically towards the HER2 protein on cancer cells; and kinase inhibitors (e.g. lapatinib and neratinib) to block the kinase activity of HER2 (Callahan and Hurvitz 2011, Iqbal and Iqbal 2014, Meisel, Venur et al. 2018).

One of the most aggressive breast cancer subtypes, triple negative breast cancer (TNBC), accounts for ~15% of cases and is characterised simply by its lack of expression of both hormone receptors and *ERBB2*/HER2, rather than by the

presence of specific markers (Newman, Reis-Filho et al. 2015). Although TNBC cannot be specifically targeted through ER or HER2, recent advances have produced targeted therapies for this subtype in the form of poly(ADP-ribose) polymerase inhibitors (PARPi) for treating cancers harbouring BRCA mutations (Lyons 2019, McCann, Hurvitz et al. 2019), that disproportionately appear in TNBCs (Newman, Reis-Filho et al. 2015, Chen, Wu et al. 2018), and checkpoint inhibitors to enhance the anti-tumour immunity in this immunogenic subtype (Lyons 2019, McCann, Hurvitz et al. 2019). The results of the clinical trials that lead to the FDA approval of these targeted therapies for the treatment of breast cancer are shown in Table 1.1 (PARPi) and Table 1.2 (immune checkpoint inhibitors). There have also been recent investigations into the efficacy of combined PARPi and immune checkpoint inhibitor therapy for the treatment of patients with TNBC and germline or somatic BRCA mutations, some of which are still ongoing. These details of these investigations are summarised in Table 1.3.

Table 1.1 Clinical trials for PARP inhibitors for the treatment of breast cancer.

Trial Name (ClinicalTrials.gov identifier)	OlympiAD (NCT02000622)	EMBRACA (NCT01945775)
Drug Name	Olaparib (AZD-2281, trade name Lynparza)	Talazoparib (BMN 673)
Drug Specificity	PARP1 and PARP2	PARP1 and PARP2
Patients Recruited	Patients with a germline BRCA mutation and HER2-negative metastatic breast cancer who had received ≤2 prior chemotherapy regimens.	Patients with advanced breast cancer and a germline *BRCA1* and/or *BRCA2* mutation who had received ≤3 prior chemotherapy regimens.
Treatment Groups	Olaparib administered orally twice daily (BID) at 300 mg (2 x 150 mg tablets) N=205; Standard single-agent chemotherapy of physician's choice of capecitabine, vinorelbine, or eribulin (TPC) N=97.	Talazoparib 1 mg oral capsules once daily for 21 continuous days N=287; Standard single-agent chemotherapy of physician's choice of capecitabine, eribulin, gemcitabine, or vinorelbine N=144.
Results	Median overall survival: Olaparib = 19.3 months; TPC = 17.1 months (P = 0.513). 6 months overall survival: Olaparib = 93.1%; TPC = 85.8%. 12 months overall survival: Olaparib = 72.7%; TPC = 69.2%. 18 months overall survival: Olaparib = 54.1%; TPC = 48.0%.	Median progression-free survival: Talazoparib = 8.6 months; standard therapy = 5.6 months (P < 0.001). Objective response rate: Talazoparib = 62.6%; standard therapy = 27.2% (P < 0.001).
FDA Recommendation	Approved for the adjuvant treatment of adult patients with deleterious or suspected deleterious germline *BRCA*-mutated (confirmed by Myriad's BRACAnalysis CDx test), HER2-negative, high-risk early breast cancer who have been treated with neoadjuvant or adjuvant chemotherapy.	Approved for patients with deleterious or suspected deleterious germline *BRCA*-mutated (confirmed by Myriad's BRACAnalysis CDx test), HER2-negative locally advanced or metastatic breast cancer.

Table 1.2 Clinical trials for immune checkpoint inhibitors for the treatment of breast cancer.

Trial Name (ClinicalTrials.gov identifier)	IMpassion130 (NCT02425891)	KEYNOTE-355 (NCT02819518)
Drug Name	Atezolizumab (Trade name Tecentriq)	Pembrolizumab (Trade name Keytruda)
Drug Specificity	PD-L1	PD-L1
Patients Recruited	Patients with locally advanced or metastatic triple-negative breast cancer (TNBC) who have not received prior systemic therapy for metastatic breast cancer.	Patients with untreated locally recurrent inoperable or metastatic triple-negative breast cancer.
Treatment Groups	Atezolizumab (840 mg via IV infusion on Days 1 and 15 of each 28-day cycle until disease progression or unacceptable toxicity) in combination with nanoparticle albumin-bound(nab)-Paclitaxel (100 mg per square meter via IV infusion on Days 1, 8, and 15 of each 28-day cycle. Administered for a target of at least 6 cycles, with no maximum in the absence of disease progression or unacceptable toxicity) N=451. Placebo (administered via IV infusion on Days 1 and 15 of each 28-day cycle until disease progression or unacceptable toxicity) with nab-Paclitaxel (as previous) N=451.	Pembrolizumab (200 mg via IV infusion every 3 weeks) plus chemotherapy of physician's choice of nab-paclitaxel, paclitaxel, or gemcitabine plus carboplatin N=566. Saline placebo plus chemotherapy N=281.
Results	Median progression-free survival: Atezolizumab plus nab-Paclitaxel = 7.2 months; placebo plus nab-Paclitaxel = 5.5 months (P = 0.002). Median overall survival: Atezolizumab plus nab-Paclitaxel = 21.3 months; placebo plus nab-Paclitaxel = 17.6 months (P = 0.08).	Median progression-free survival among patients with CPS of 10 or more: Pembrolizumab-chemotherapy = 9·7 months; placebo-chemotherapy = 5·6 months (P = 0.0012).
FDA Recommendation	Accelerated approval in 2019 for the treatment of adults with unresectable locally advanced or metastatic triple-negative breast cancer whose tumors express PD-L1, as determined by the FDA-approved VENTANA PD-L1 (SP142) Assay. Approval withdrawn in 2021 based on FDA assessment of the metastatic TNBC treatment landscape at the time, and in accordance with the requirements of the accelerated approval program.	Approved in 2020 in combination with chemotherapy for the treatment of patients with locally recurrent unresectable or metastatic triple-negative breast cancer (TNBC) whose tumors express PD-L1 (CPS ≥10), as determined by an FDA approved Dako PD-L1 IHC 22C3 pharmDx diagnostic assay.

Table 1.3 Clinical trials for the combined PARPi and immune checkpoint inhibitor treatment of breast cancer.

Trial Name (ClinicalTrials.gov identifier)	(NCT03101280)	TOPACIO/KEYNOTE-162 (NCT02657889)	MEDIOLA (NCT02734004)	(NCT02849496)
Trial Phase	Phase I	Phase I/II	Phase I/II	Phase II
Study Start	27/04/2017	15/04/2016	17/03/2016	15/11/2016
Trial Status	Completed	Completed	Active, not recruiting	Active, not recruiting
Study Completion Date (actual/estimated)	11/08/2020 (actual)	17/09/2021 (actual)	30/12/2022 (estimated)	31/08/2023 (estimated)
Drug Combination Tested	Rucaparib and Atezolizumab	Niraparib and Pembrolizumab	MEDI4736 (Durvalumab) and Olaparib	Olaparib and Atezolizumab
Patients Recruited	Patients with previously treated advanced ovarian or endometrial cancer (Part 1) and platinum-sensitive ovarian cancer or triple-negative breast cancer (TNBC) (Part 2). Patients with previously treated TNBC with a BRCA mutation or BRCA-like molecular signature and have not been exposed to cancer immunotherapies.	Patients with advanced or metastatic TNBC and patients with recurrent ovarian cancer. Patients with advanced or metastatic breast cancer must have TNBC and may have received up to 2 lines of cytotoxic therapy.	PARPi- and immunotherapy-naïve patients with relapsed small cell lung cancer (SCLC), germline BRCA mutated (gBRCAm) metastatic human epidermal growth factor receptor 2 (HER2)-negative breast cancer, gBRCAm platinum-sensitive relapsed ovarian cancer, and gastric cancer.	Patients with a locally advanced and unresectable, or with a metastatic, HER2-negative breast cancer and a known germline or somatic BRCA1/2 mutation.
Treatment Groups	All participants received the combination of Rucaparib (600 mg oral administration twice per day) and Atezolizumab (1200 mg administered by IV infusion once every 3 weeks) in 21-day cycles.	All participants received Niraparib (300mg/day orally) in combination with Pembrolizumab (200mg IV on Day 1 of each 21-day cycle).	All participants given Olaparib twice daily starting on week 1 day 1, and MEDI4736 every 4 weeks starting on week 5 day 1.	Arm 1: Olaparib administered orally twice daily on days 1-21 of each cycle. Cycles repeat every 21 days in the absence of disease progression or unacceptable toxicity. Arm 2: Olaparib administered as in Arm I and Atezolizumab administered by IV on day 1 of each cycle. Cycles repeat every 21 days in the absence of disease progression or unacceptable toxicity.

1.4.2.3 Improved therapeutic strategies

Targeted therapies, such as those described above, are usually administered in combination with chemotherapy or radiotherapy, either before or after surgery, meaning that there are a wide range of options and potential combinations that can be administered. An example of one such therapeutic strategy made possible is adjuvant therapy, whereby primary surgery is followed by an additional therapy (such as chemotherapy, radiotherapy, and targeted therapy). Adjuvant therapy is administered to ensure minimal risk of local and distant recurrence, caused by escaped breast cancer cells or micro-metastases, after surgical removal of the primary tumour (Chew 2001). Implementation of adjuvant therapy in patients who were identified to be likely to benefit from this strategy has resulted in significant survival benefits (including improved recurrence-free survival and overall survival) to patients with various disease subtypes (Rossi, Stevens et al. 2015, Park, Han et al. 2019, Tolaney, Guo et al. 2019). Neoadjuvant therapy, on the other hand, is the pre-operative administration of chemotherapy, radiotherapy, and targeted therapy. This therapeutic strategy is implemented with the aim of shrinking the tumour prior to surgery, potentially allowing for a lumpectomy (or breast-conserving surgery) to be performed rather than a mastectomy, or for easier removal of any affected lymph nodes (Trimble, Ungerleider et al. 1993, Ikeda, Jinno et al. 2002, Briest and Stearns 2011). In addition, a neoadjuvant approach can be used to refine or inform clinicians on the best adjuvant regimen to employ and thereby potentially maximise the effectiveness of the therapy given.

1.4.3 The need for novel therapeutic targets for breast cancer

In spite of these vast improvements, there is an ongoing requirement for the identification of novel cancer drivers and the associated development of targeted therapies in breast cancer research. One reason for this is that many patients' cancers cannot currently be addressed using therapies specifically targeted to that particular subtype. For example, only ~15% of TNBC cases have the required BRCA mutations required for PARPi therapeutic efficacy (Sharma, Klemp et al. 2014, Engel, Rhiem et al. 2018), and a large portion of tumours lack the PD-L1 expression required for successful treatment with checkpoint inhibitors (Mittendorf, Philips et al. 2014, Doğukan, Uçak et al. 2019), leaving very limited therapeutic options for TNBC patients.

In a study looking at long-term recurrence risk in hormone positive breast cancer, patients who were considered cancer-free at the end of their treatment, in the form of 5 years of endocrine therapy, experienced a steady rate of recurrence throughout the study period, from 5 years until as late as 20 years following the initial diagnosis (Pan, Gray et al. 2017). The risk of recurrence was directly correlated with the diameter, nodal status (including the number of cancerous nodes), and characteristics (such as tumour grade and Ki-67 status) of the primary tumour prior to therapy (Pan, Gray et al. 2017). It is therefore essential that follow-up studies be conducted so that the identification of individuals who are most at risk of even long-term recurrence is made possible. Based on this information, healthcare providers may consider lengthening the duration of hormone therapy past the usual 5-10 years for patients with higher risks of relapse. It may also be necessary for patients' tumours to be further investigated for other therapeutic targets and novel therapies that may benefit them more than current/traditional therapies, particularly if they offer more benefit than long-term endocrine therapies that would extend exposure to their associated side effects and toxicities.

Further to these challenges, it should also be noted that there is no absolute guarantee that chemotherapy/radiotherapy or even targeted therapies will be 100% effective. This is, at least in part, due to intrinsic resistance (innate resistance to therapy) and acquired resistance (the obtained ability to resist therapy following exposure) to treatments (Wang, Zhang et al. 2019). There have been several studies conducted that sought to identify mechanisms behind resistance to currently available therapies. One such mechanism involves increased efflux of drugs from cancer cells caused by the overexpression of ATP binding cassette (ABC) transporters, including P-glycoprotein (P-gp) and the multi-drug-resistance proteins (MRPs) subfamily of ABC transporters that, as the name suggests, confer resistance to multiple drugs (Housman, Byler et al. 2014, Ji, Lu et al. 2019, Wang, Zhang et al. 2019). In breast cancer, a significant proportion of tumours express P-gp, whose expression has been shown to increase following chemotherapy or hormone therapy (Rudas, Filipits et al. 2003). Pre-chemotherapy MRP1 expression was found to be significantly associated with shorter progression-free and overall survival in patients (Rudas, Filipits et al. 2003). Modifications of

drug targets has also been linked to intrinsic and acquired resistance to several breast cancer therapies, among other factors (Housman, Byler et al. 2014, Ji, Lu et al. 2019, Wang, Zhang et al. 2019).

Investigations specifically into mechanisms of endocrine resistance have revealed several factors involved, including: loss of ERα expression (Johnston, Saccani-Jotti et al. 1995, Musgrove and Sutherland 2009); mutations in ERα (Musgrove and Sutherland 2009, Szostakowska, Trębińska-Stryjewska et al. 2019); expression of truncated variants of ERα and ERß (Musgrove and Sutherland 2009, Haque and Desai 2019); increased expression of mediators of ERα activity, including activator protein 1 (AP1) and nuclear factor-κB (NF-κB) (Musgrove and Sutherland 2009, Haque and Desai 2019); post-translational modifications of ERα (Ali and Coombes 2002, Jordan and O'Malley 2007, Musgrove and Sutherland 2009); constitutive ERα expression via deregulations in the activity of ERα co-activators (Ali and Coombes 2002, Ring and Dowsett 2004, Musgrove and Sutherland 2009); overexpression of several members of the epidermal growth factor receptor (*Egfr*) family, resulting in increased receptor tyrosine kinase signalling and Erk and PI3K pathway activation (Ali and Coombes 2002, Musgrove and Sutherland 2009, Haque and Desai 2019); and deregulations in the cell cycle and apoptotic machinery (particularly increased expression of anti-apoptotic molecules, such as BCL2, and decreased expression of pro-apoptotic molecules, including caspase 9) (Musgrove and Sutherland 2009).

Resistance mechanisms against HER2-targeted therapies have also been identified. Impaired binding of drugs to HER2 can negatively impact their function and can be due to HER2 gene mutations (Rexer and Arteaga 2012), low or heterogeneous expression of HER2 (Hou, Nitta et al. 2017, Vernieri, Milano et al. 2019), expression of splicing variants of HER2 that compromise drug binding (for example, a lack of an extracellular domain impairs the activity of drugs whose binding site is located in this domain) (Pohlmann, Mayer et al. 2009, Rexer and Arteaga 2012, Vernieri, Milano et al. 2019), or masking of the binding domain (Pohlmann, Mayer et al. 2009, Rexer and Arteaga 2012, Vernieri, Milano et al. 2019). Resistance can also occur due to altered or constitutive activation of

parallel or downstream pathways, including constitutive activation of the PI3K/AKT/mTOR axis, caused by various alterations in pathway genes (Pohlmann, Mayer et al. 2009, Rexer and Arteaga 2012, Vernieri, Milano et al. 2019). Additionally, although therapies such as trastuzumab prevent HER2-driven signalling by blocking homodimerisation, HER2 molecules that remain unbound by the drug may still heterodimerise with other receptor tyrosine kinases (RTKs) and initiate downstream pathways that are similar to those activated upon homodimerisation, thus preserving some of its tumorigenic functions (Nahta, Yuan et al. 2005, Pohlmann, Mayer et al. 2009, Rexer and Arteaga 2012, Vernieri, Milano et al. 2019).

This limited availability of therapies for some subtypes and significantly reduced options for some patients following resistance/recurrence highlights the need for the identification of additional/novel therapeutic targets for breast cancer. This would offer more options and would mean that the most appropriate therapy could be administered based on the genetic makeup of the particular disease.

1.4.4 Identifying novel biomarkers and therapeutic targets in breast cancer

In order to develop new therapies to effectively target and treat various breast cancers, there first needs to be novel and promising biological targets identified in patients. Promising targets are those whose activity can be modified in a clinical setting (in other words, they are a druggable target) and with limited or manageable off-target effects and toxicities (Gashaw, Ellinghaus et al. 2011). The discovery of new drug targets involves looking deeper into the disease by investigating large breast tumour sample numbers, collected from many different patients (with various disease stages, characteristics, and subtypes), and comparing to untransformed breast cells. This would allow for the identification of genetic changes that most significantly correlate with, or accurately predict, factors such as disease initiation, progression, and response to current therapy. These genetic alterations are known as driver mutations, which are causally implicated in oncogenesis and confer a selective growth advantage on the cells carrying them (Haber and Settleman 2007, Merid, Goranskaya et al. 2014, Pon and Marra 2015). Many of the mutations found in cancers, however, are described as

passenger mutations that are not selected for, do not confer selective advantage to the cell, but happened to be present in an ancestor of the cancer cell when it acquired one of its drivers (Haber and Settleman 2007, Merid, Goranskaya et al. 2014, Pon and Marra 2015).

The emergence of several next-generation and high-throughput sequencing studies have been key for identifying cancer drivers and mapping out the genomic landscape of various cancers, including breast cancer, (Aherne, McDonald et al. 2002, Sjöblom, Jones et al. 2006, Greenman, Stephens et al. 2007, Hansen and Bedard 2013, Jeon, Nim et al. 2014) from its "mountains", that are mutated in large proportions of tumours, to its numerous "hills", that individually do not appear as frequently as the "mountains" but collectively dominate the cancer landscape (Wood, Parsons et al. 2007, Vogelstein, Papadopoulos et al. 2013). These next-generation and high-throughput sequencing studies have not only confirmed the cancer driving potential of genes previously implicated in breast cancer, but have also identified numerous novel significantly mutated genes that were not previously identified in clinical breast cancer samples. *RUNX1* was one such novel gene that was found to harbour genomic alterations, often resulting in loss-of-function of the protein, in relatively high frequencies (Banerji, Cibulskis et al. 2012, Ellis, Ding et al. 2012, Koboldt, Fulton et al. 2012, Cornen, Guille et al. 2014, Nik-Zainal, Davies et al. 2016, Pereira, Chin et al. 2016). Interestingly, these *RUNX1* mutations appeared to associate predominantly with the ER+ tumours (Ellis, Ding et al. 2012, Cornen, Guille et al. 2014, Nik-Zainal, Davies et al. 2016, Pereira, Chin et al. 2016). The *RUNX1* gene, and the protein it encodes, are of particular interest to the Blyth lab, particularly for its role in solid tumours, and forms the basis of much of the research conducted, including the work described in this thesis.

1.5 The RUNX family of transcription factors

1.5.1 The essential roles of RUNX transcription factors in development

The RUNX family of transcription factors consists of three individual members – RUNX1, RUNX2, and RUNX3 – each with distinct functions that are essential for several cellular and developmental processes including, but not limited to,

proliferation, differentiation, and cell fate (Ito 2008, Mevel, Draper et al. 2019). The function of the RUNX proteins is dependent upon their interactions with their obligate co-factor, core-binding factor subunit beta (CBFβ), which can form a heterodimeric complex with each member, enhance their DNA binding affinity, and stabilise the RUNX proteins from proteasome degradation such that the CBF complex is able to bind to the DNA and either activate or repress their downstream targets (Gu, Goetz et al. 2000, Tang, Crute et al. 2000, Tang, Shi et al. 2000). These interactions between RUNX1 family members and CBFβ, and between the CBF complex and DNA, is facilitated by an evolutionarily conserved runt-homology domain (RHD) in the N-terminus that is characteristic of the RUNX transcription factors (Bäckström, Wolf-Watz et al. 2002). Shown in Figure 1.1 are the schematic representations of each *RUNX1* gene, with the RHD and transactivation domain shown. Figure 1.1 also shows that each of the *RUNX* genes is under the control of two promoters, a distal (P1) promoter and a proximal (P2) promotor, which leads to the generation of distinct mRNA isoforms.

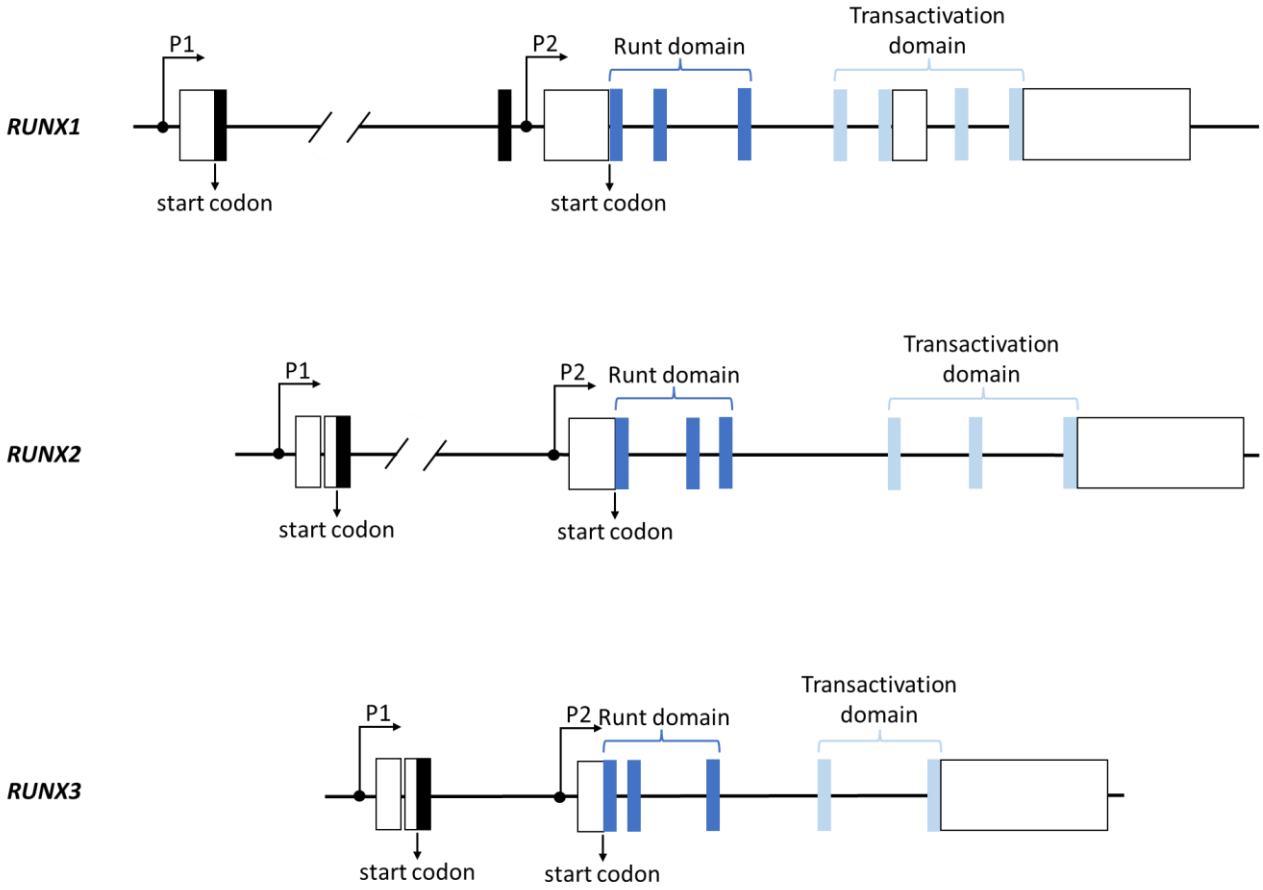


Figure 1.1 Schematic representations of *RUNX1*, *RUNX2*, and *RUNX3* gene structures. Structures of each of the human RUNX genes are shown in the above schematic, which highlights their similar genomic organization including their highly conserved Runt domains and two promoters (P1 and P2).

In the RUNX1 gene, the proximal promoters generate the protein isoforms RUNX1a and RUNX1b, while the distal promoter regulates RUNX1c. The structures of these isoforms are shown in the schematic in Figure 1.2. All protein isoforms contain the Runt domain, however the absence of a transactivation domain in the RUNX1c isoform, which is present in the RUNX1b and RUNX1c isoforms, suggests an antagonistic role for this specific isoform.

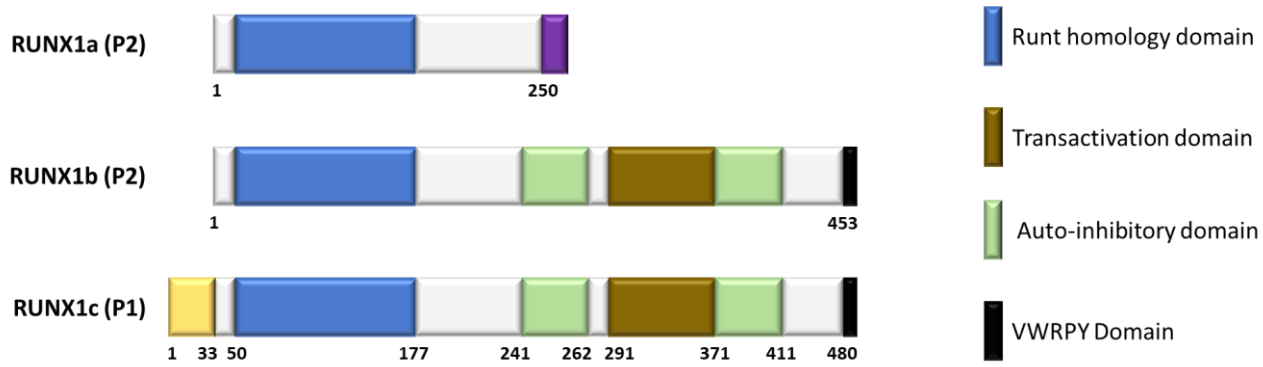


Figure 1.2 Schematic representations of the three major isoforms of RUNX1.

Structures of the three major isoforms of RUNX1 (RUNX1a, RUNX1b, and RUNX1c) are summarised above, with the major functional domains and isoform-specific regions indicated in the key.

Gene knockout studies in mice have been essential for uncovering the discrete, and sometimes overlapping, functions of each *Runx* family member in various systems. *RUNX1* was identified as a key factor for definitive haematopoiesis and the generation of haematopoietic stem cells, and homozygous loss of *Runx1* function during murine development results in embryonic lethality due to a lack of definitive haematopoiesis (Okuda, van Deursen et al. 1996, Okada, Watanabe et al. 1998, North, Gu et al. 1999, Yokomizo, Ogawa et al. 2001, Bäckström, Wolf-Watz et al. 2002). Homozygous loss of *Runx2* function in mice results in neonatal lethality caused by an inability to breathe due to a complete lack of ossification, indicating an essential role for *RUNX2* in osteogenesis (Komori, Yagi et al. 1997, Otto, Thornell et al. 1997). In addition to this, mutations in *RUNX2* have been linked to a genetic condition in humans known as cleidocranial dysplasia (CCD) that affects the development of bones (Mundlos, Otto et al. 1997, Otto, Thornell et al. 1997, Peng, Chen et al. 2017, Xu, Chen et al. 2017). Observations from loss of *Runx3* in genetic models indicate that this gene is involved in a wide variety of tissues. Studies have indicated that RUNX3 is important in the normal development of the gastric epithelium (although this is highly disputed) (Levanon, Brenner et al. 2001, Li, Ito et al. 2002, Levanon, Brenner et al. 2003, Brenner, Levanon et al. 2004, Carvalho, Milne et al. 2005, Ito, Inoue et al. 2009, Levanon, Bernstein et al. 2011), TrkC dorsal root ganglia neurons (Levanon, Bettoun et al. 2002, Inoue, Ozaki et al. 2003), chondrocytes (Yoshida, Yamamoto et al. 2004), hair follicle formation (Raveh, Cohen et al. 2005), and T cell and NK cell

development and regulation (Taniuchi, Osato et al. 2002, Woolf, Xiao et al. 2003, Djuretic, Levanon et al. 2007, Levanon, Negreanu et al. 2014).

1.5.2 The involvement of RUNX1 in cancer

1.5.2.1 RUNX1 in haematological malignancies

Given how essential the RUNX transcription factors are in a variety of fundamental processes in the body, it does not come as a surprise that perturbations in their expression are often associated with a variety of cancers. The RUNX proteins' roles in cancer are often paradoxical as both tumour suppressive and oncogenic properties have been identified in each member of the RUNX family, with each role dependent upon the specific context. This is perfectly exemplified by the role of *RUNX1* in haematological cancers, as *RUNX1* mutations are among the most common perturbations observed in a variety of haematological malignancies (Niini, Kanerva et al. 2000, Osato 2004, De Braekeleer, Férec et al. 2009, De Braekeleer, Douet-Guilbert et al. 2011) while, rather contradictorily, in some leukaemia subtypes *RUNX1* has been observed as a dominant oncogene and the expression of wildtype *RUNX1* is relied upon for the survival of acute myeloid leukaemia cells (Ben-Ami, Friedman et al. 2013, Goyama, Schibler et al. 2013, Choi, Illendula et al. 2017). Because *RUNX1* aberrations have been so frequently observed in haematological malignancies, and due to the (previously described) dramatic phenotype witnessed in *Runx1* KO mice linked to its essential role in haematopoiesis, it is in these contexts that the function of *RUNX1* has been most extensively studied. However, there is increasing evidence, from sequencing studies and conditional knockout mouse models, to suggest that *RUNX1* plays a role in various cancers of non-haematological origin, with its expression being altered in numerous solid tumours.

1.5.2.2 The role of RUNX1 in solid tumours

Several studies have indicated that *RUNX1* plays an important tumour suppressor role in multiple components of the gastrointestinal tract. Recurrent deletions of *RUNX1* were found in oesophageal cancers, which was shown to be functionally significant as reintroduction into an oesophageal carcinoma cell line resulted in a 69% reduction in anchorage-independent growth (Dulak, Schumacher et al. 2012).

RUNX1 was also discovered to have been frequently, and significantly, downregulated in both gastric cancer cell lines and primary samples from gastric cancer patients (Sakakura, Hagiwara et al. 2005). The generation of an Apc^{Min} mouse model, for intestinal tumorigenesis, with conditional knockout of *Runx1* in the epithelial cells of the GI tract resulted in enhanced tumorigenesis (Fijneman, Anderson et al. 2012). Even more interestingly, *Runx1* deletion on an Apc^{wt} background was adequate to initiate tumorigenesis (Fijneman, Anderson et al. 2012). In another investigation, tagging for *RUNX1* single nucleotide polymorphisms (SNPs) showed significant associations with colon and rectal cancer (Slattery, Lundgreen et al. 2011). Other SNP studies have also implicated *RUNX1* SNPs in prostate cancer progression, revealing significant associations with increased risk of progression and lymph node metastasis, resulting in poorer prostate-specific antigen (PSA)-free survival (Huang, Lan et al. 2011). There is also evidence from work in the Blyth laboratory that *RUNX1* is a tumour suppressor in prostate cancer (McKillop, Edwards et al. 2017). *RUNX1* was also revealed to be significantly downregulated in samples of liver cirrhosis (LC) and hepatocellular carcinoma (HCC), with the prevalence of *RUNX1* downregulations increasing with progression from LC to HCC, and with the progression of HCC (Miyagawa, Sakakura et al. 2006).

As with haematological malignancies, the role of *RUNX1* in solid tumours is often complex as it can also function as an oncogene in some contexts. In a 2019 publication, analyses of gene expression microarray datasets (TCGA and GSE106582) was supported with *in vitro* and *in vivo* experiments involving colorectal cancer (CRC) cell lines generated with either RUNX1 overexpression or silencing (Li, Lai et al. 2019). The analyses demonstrated that *RUNX1* (and its associated protein) expression was upregulated in CRC tissues, compared to normal or non-neoplastic tissues, and CRC patients with high RUNX1 expression levels exhibited poorer disease-free and overall survival statistics compared to those with lower levels of expression (Li, Lai et al. 2019). *In vitro* (wound healing and Transwell) assays using these cell lines revealed that increased RUNX1 expression was associated with enhanced migration and invasion capabilities, while *in vivo* orthotopic injection of these cell lines into nude mice demonstrated that RUNX1 overexpression increased the metastatic ability of the cells (Li, Lai et

al. 2019). Further analysis of the The Cancer Genome Atlas (TCGA) and GSE17538 datasets and analysis of the RUNX1 overexpressing and silenced cell lines, demonstrated that RUNX1 expression was associated with an epithelial-to-mesenchymal-transition(EMT)-associated phenotype and was positively correlated with multiple components of the EMT process (Li, Lai et al. 2019). Following the observation that RUNX1 was an enhancer of the Wnt/ß-catenin pathway activation, through direct interactions with ß-catenin and the promoter and enhancer regions of KIT, it was theorised that this could be a potential mechanism for enhanced metastasis and EMT in RUNX1-overexpressing cells, leading to poorer patient prognosis (Li, Lai et al. 2019). Overexpression of *RUNX1* was demonstrated to cause neoplastic transformation when overexpressed in fibroblasts (Kurokawa, Tanaka et al. 1996). There is significant evidence in the skin of *RUNX1* acting as an oncogene, with key publications revealing overexpression of *RUNX1* in human skin squamous cell carcinoma (SCC) and human head and neck SCC samples (Scheitz, Lee et al. 2012). *Runx1* was also highly expressed in mouse models of skin papilloma and SCC, and its deficiency resulted in reduced SCC formation in the mouse skin (Hoi, Lee et al. 2010). Further to this, it was demonstrated that *Runx1* is required for tumour initiation and lineage-tracing studies revealed that *Runx1*-expressing hair follicle stem cells (HFSCs) are the cells of origin for most papilloma (Scheitz, Lee et al. 2012). Interestingly, the removal of *Runx1* in already established tumours in mice resulted in remarkable levels of tumour regression, suggesting *Runx1* may also be required for tumour maintenance in epithelial skin tumours (Scheitz, Lee et al. 2012). Studied conducted by the Blyth laboratory identified a novel protumorigenic role for the RUNX genes in clear cell renal cell carcinoma (ccRCC), whereby tissue microarray (TMA) analysis revealed significantly poorer patient survival with increased RUNX1 (and RUNX2) expression (Rooney, Mason et al. 2020). Investigations using ccRCC cell lines revealed that deletion of RUNX1 reduced tumour cell growth and viability in both *in vitro* and *in vivo* based investigations, while in a genetically engineered mouse model (GEMM) of kidney cancer with RUNX1 deletion exhibited increased overall survival and reduced tumour cell proliferation.

1.5.2.3 Analysis of cBioportal data

The duality and far-reaching consequences of *RUNX1* in cancer is perfectly illustrated when exploring the TCGA PanCancer Atlas Studies (made available through cBioportal) (cBioportal for Cancer Genomics , Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013). *RUNX1* was altered in 24 out of 35 (68.57%) cancer types that were explored in these studies. Looking more closely at the types of alterations (Figure 1.3), this data appears to be in agreement with the previously described paradoxical roles of *RUNX1* in cancer types, as the alterations in oesophageal carcinomas and colorectal cancers, for example, are primarily deep deletions or mutations, while the alterations present in leukaemias are comparatively more frequent and varied. Interestingly, the 2nd most common cancer type possessing alterations in RUNX1, invasive breast carcinoma, appears to also possess a high frequency of *RUNX1* alterations, the nature of which are relatively diverse. Using data from the slightly more in-depth METABRIC exploration of primary breast tumours (available through cBioportal) (cBioportal for Cancer Genomics , Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013), by separating tumours with *RUNX1* alteration into ER-positive and ER-negative populations (a feature commonly used to stratify breast cancer patients), it can be seen that mutations are the most common *RUNX1* alteration type affecting the ER-positive breast cancers, while amplifications are most common in the ER-negative group (Figure 1.4). This indicates that, similar to in leukaemias, the role of *RUNX1* in breast cancer may be highly context-dependent and differ between disease subtypes. There is considerable evidence in the wider literature to suggest that this may indeed be the case, and is discussed in greater detail below.

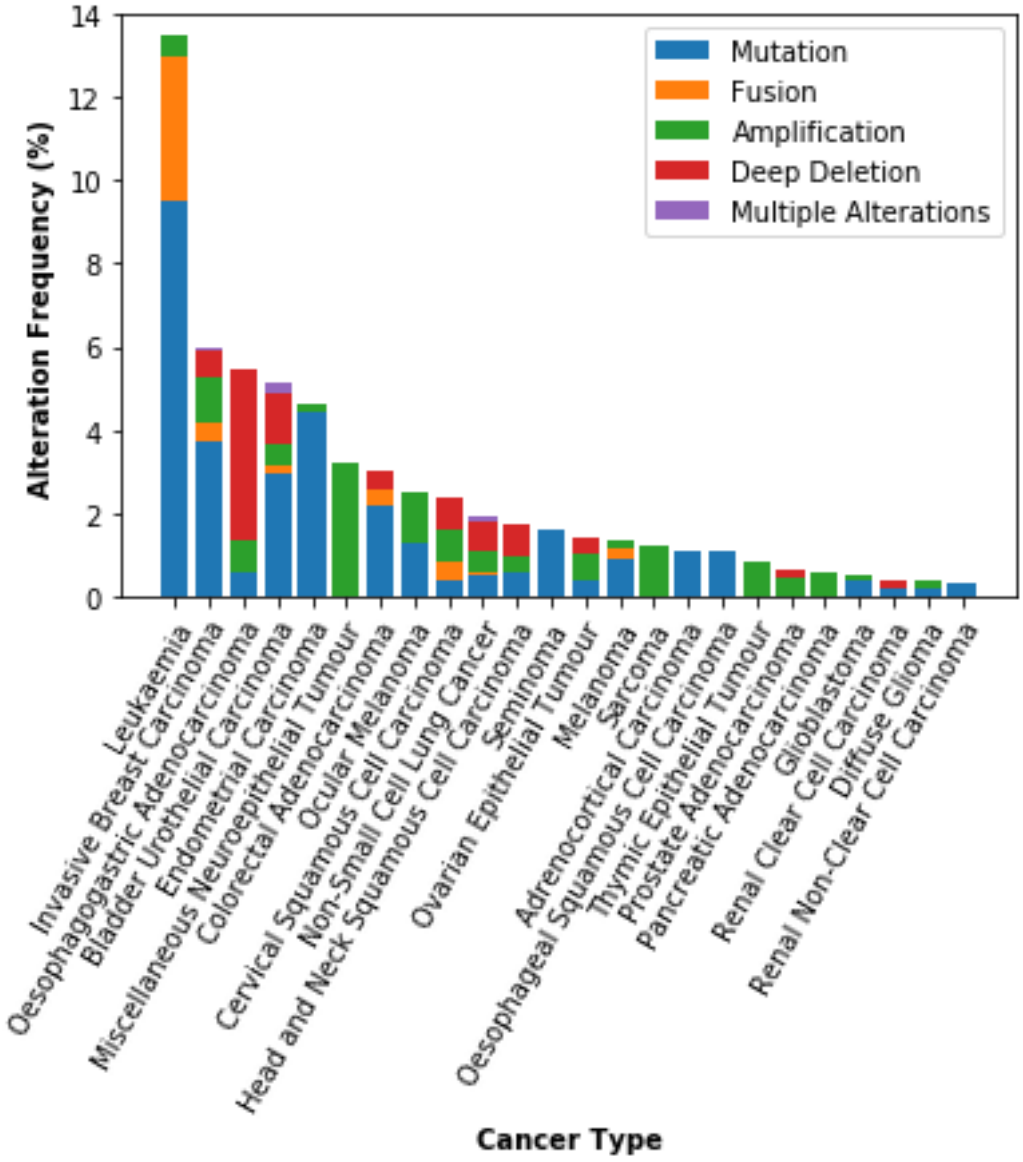


Figure 1.3 Frequencies of various *RUNX1* alteration types in the TCGA PanCancer atlas.

Represented in this figure are the 24 cancer types in which *RUNX1* was altered in the TCGA PanCancer Atlas Studies. The frequency of *RUNX1* alteration is shown for each cancer type and this is broken down further to illustrate the types of *RUNX1* alterations that are detected in each of the types of cancer studied. All data for this figure was obtained through cBioPortal for Cancer Genomics using the TCGA PanCancer Atlas Studies (cBioportal for Cancer Genomics , Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013).

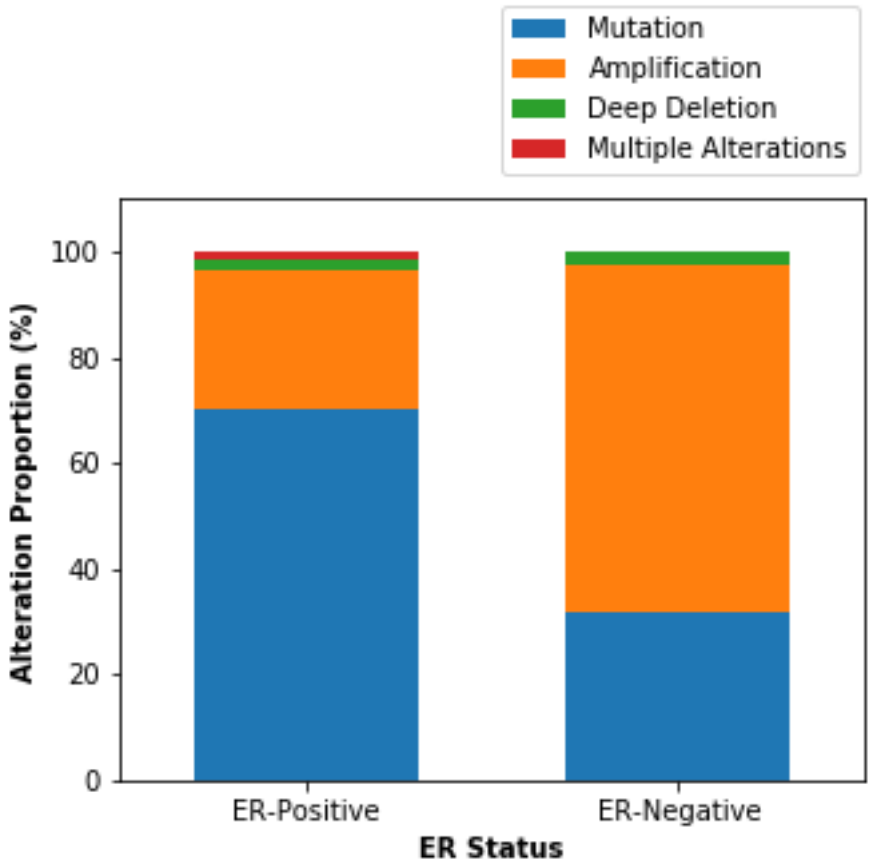
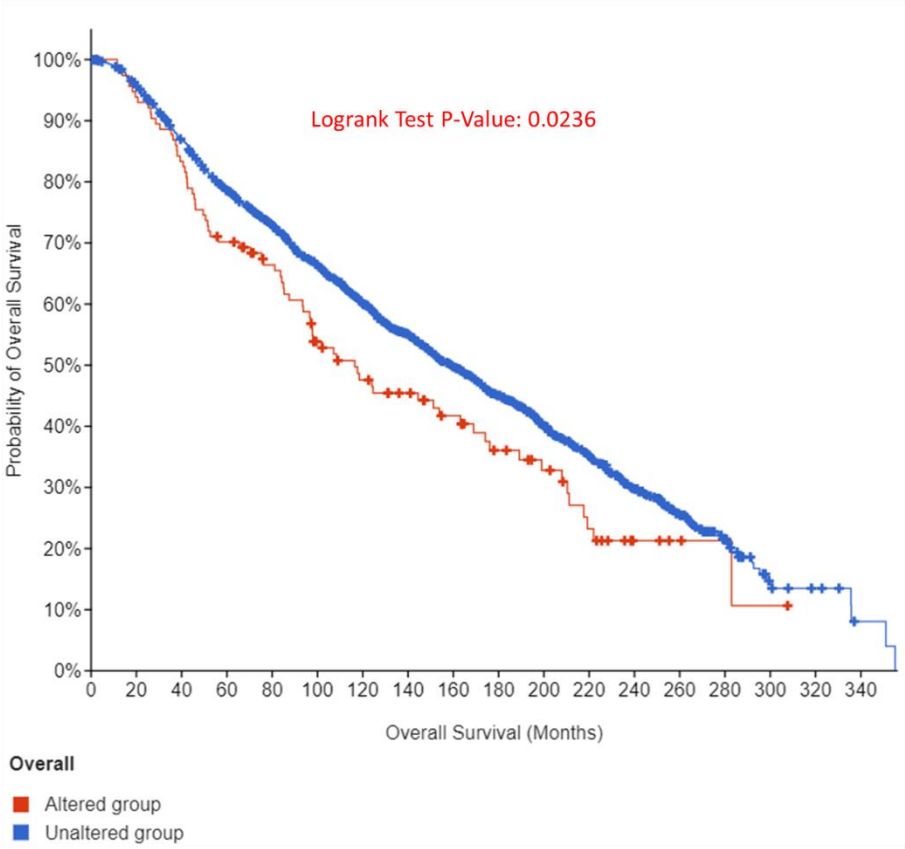


Figure 1.4 *RUNX1*-Altered Invasive Breast Carcinomas, Sub-Divided into ER-Positive and ER-Negative Groups, Reveal Subtype-Dependent Alteration Signatures.

Depicted in this figure is data from the METABRIC study of invasive breast carcinomas, of which the *RUNX1*-altered tumours were selected and subsequently separated into ER-positive (n=81) and ER-negative (n=41) groups. The proportion of the various types of *RUNX1* alterations present in the two primary breast tumour subgroups are represented as a percentage of the total number of breast cancer cases in which *RUNX1* alterations appeared within their respective group. This allows a direct comparison to be made between the *RUNX1* alteration signatures of each group. All data for this figure was obtained through cBioPortal for Cancer Genomics using the METABRIC study (cBioportal for Cancer Genomics , Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013).

1.5.3 The paradoxical role of *RUNX1* in breast cancer

It is clear, from the previously described data obtained from next-generation and high-throughput genome sequencing studies, that disturbed expression of *RUNX1* occurs relatively frequently in invasive breast carcinomas. Data from the METABRIC sequencing study demonstrate that these alterations are also clinically relevant, as altered expression is associated with a significant reduction in overall survival, as shown in Figure 1.5 below. However, detailed analyses have revealed that, much like in haematological malignancies, the function of *RUNX1* in breast cancer is not quite as simple as this and, as suggested by the previous observation of subtype-dependent *RUNX1* alteration signatures in invasive breast cancers, context is key, as both pro-tumour and anti-tumour roles have been observed. Investigated together, many of these studies implicate the breast cancer subtype as a determining factor for the specific role of *RUNX1* in the progression and outcome of that tumour.



	Number of Cases, Total	Number of Events	Median Months Overall (95% CI)
Altered group	**115**	**75**	**116.43 (96.63 - 174.13)**
Unaltered group	**1866**	**1069**	**159.00 (148.87 - 170.63)**

Figure 1.5 Alterations in *RUNX1* are associated with shorter overall survival in breast cancer patients, according to METABRIC sequencing study data.

Shown in the above figure is a survival curve, taken from cBioportal's Comparison/Survival function (cBioportal for Cancer Genomics , Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013) that allows comparisons to be made between the survival outcomes of two or more groups of patients. The above curve, generated by the website, compares overall survival between patients whose primary tumours had altered expression of *RUNX1* (n=115) and those whose primary tumour showed unaltered *RUNX1* expression (n=1866). A logrank test was used to directly compare the survival distributions of the two groups, and the P-value of 0.0236 ($P<0.05$) indicates a significant difference. Overall survival is demonstrably shorter in individuals possessing tumours with *RUNX1* alterations with median survival being 116.43 months, compared to 159 months with unaltered *RUNX1* expression.

1.5.3.1 RUNX1 as a tumour suppressor in hormone-dependent tumours

There is, indeed, a body of evidence, some of which predates these sequencing studies, indicating that *RUNX1* does in fact function as a tumour suppressor in breast cancer. In one study, researchers compared the gene expression profiles of adenocarcinoma metastases of multiple origin (including lung, breast, and prostate) with unmatched primary adenocarcinomas (Ramaswamy, Ross et al. 2003). This analysis facilitated the discovery and refinement of a metastases-

associated gene signature that could be used to predict the metastatic potential of primary tumours, although none of these genes were individual predictors of metastasis. This metastases-associated gene signature was comprised of 17 genes, 8 of which were upregulated and 9 of which were downregulated (Table 1.4). Among the downregulated genes in this 17-gene signature was *RUNX1*, indicating that downregulation of *RUNX1*, while alone is insufficient to predict the metastatic potential of individual tumours, is an important component of the 17-gene signature that could be used to predict metastasis and outcome in solid tumours.

Table 1.4 The 17-gene signature associated with tumour metastasis.

Gene	Upregulated/Downregulated	Function
ACTG2	Downregulated	Encodes actin gamma 2; a smooth muscle actin found in enteric tissues.
CNN1	Downregulated	Encodes a shin filament-associated protein implicated in the regulation and modulation of smooth muscle contraction. Enables actin binding activity.
COL1A1	Upregulated	Encodes a fibril-forming collagen found in most connective tissues and abundant in bone, cornea, dermis and tendon.
COL1A2	Upregulated	Encodes a fibril-forming collagen found in most connective tissues and abundant in bone, cornea, dermis and tendon.
DHP5	Upregulated	Encodes a component of the diphthamide (a unique post-translationally modified histidine residue) synthesis pathway.
EIF4EL3	Upregulated	Encodes a protein that is part of the mRNA cap binding activity complex and facilitates the processing and translation of mRNA.
HLA-DPB1	Downregulated	Encodes protein expressed in antigen presenting cells, which are essentual to the immune system function.
HNRPAB	Upregulated	Belongs to the subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs) associated with pre-mRNAs in the nucleus, and influence pre-mRNA processing and other aspects of mRNA metabolism and transport.
LMNB1	Upregulated	Encodes Lamin B1, a scaffolding component of the nuclear envelope.
MT3	Downregulated	A member of the metallothionein family of genes involved in zinc and copper homeostasis, and is induced under hypoxic conditions.
MYH11	Downregulated	Encodes smooth muscle myosin heavy chain 11, a major contractile protein involved in cell movement and the transport of materials within and between cells.
MYLK	Downregulated	Encodes myosin light chain kinase, a calcium/calmodulin dependent enzyme that facilitates myosin interaction with actin filaments to produce contractile activity.
NR4A1	Downregulated	Encodes protein involved in cell cycle mediation, inflammation and apoptosis.

This data is in agreement with a later publication, which found that RUNX1 could serve as a positive regulator of E-cadherin (an epithelial cell marker whose expression is decreased during EMT in cancer) and thereby limit the migratory potential of breast cancer cells with a metastatic capacity (Liu, Lee et al. 2005). Further to this, an MCF10A series of cell lines – ranging from an immortalised, but otherwise untransformed, human breast epithelial model to transformed premalignant human breast cells with a potential for neoplastic progression to fully malignant breast cancer lines capable of metastasis – was used to explore various genomic alterations involved in breast cancer progression (Kadota, Yang et al. 2010). In the poorly differentiated malignant MCF10CA1a cell line, intragenic *RUNX1* deletions were found and confirmed by DNA and RNA analyses. RUNX1 protein expression was also reduced in high-grade primary breast tumours when compared to low/mid-grade tumours, an observation that was backed-up by several Oncomine datasets showing that RUNX1 was increasingly lost from low- to mid- to high-grade breast tumours. Notably, these analyses allowed for the identification of a series of genetic events that characterise the progression of breast cancer from early to late stages. *RUNX1* deletion was one such genomic alteration that was found to delineate poorly differentiated malignant tumours.

Another group, who also used the MCF10A cell line as their model, discovered that lentiviral shRNA knockdown of *RUNX1* resulted in hyperproliferation (which could be suppressed using roscovitine, thereby implicating sustained CDK activity potentiated by increased/unobstructed transcriptional activity of FOXO) and abnormal morphogenesis in the cells (Wang, Brugge et al. 2011). This phenotype could be reversed by providing the cells with RNAi-resistant murine *Runx1*, which re-established the reduced proliferative phenotype as witnessed in normal MCF10A cells, indicating the phenotype was specifically induced by loss of *RUNX1* and not a compensatory mechanism. By investigating the loss of *RUNX1* in this model further, it was also revealed that FOXO expression was essential for mediating the shRUNX1-induced hyperproliferative phenotype as expression of a dominant-negative truncated variant of FOXO also restored the cells to a reduced proliferative phenotype. Interestingly, a retrospective analysis of microarray data revealed a significant negative correlation between *RUNX1* and *FOXO1* expression specifically in TNBCs (a subset that enriches for basal-like cancers, a molecular

subtype similar to MCF10A cells in their molecular expression profiles) that was not observed in other subtypes. This strong association between *RUNX1* downregulation and *FOXO1* upregulation again indicates a critical role for FOXO in the maintenance of breast cancers initiated by low or impaired expression of *RUNX1*.

Later publications illuminated associations between *RUNX1* and ER in breast cancers. In one such publication, van Bragt and associates first explored the expression of *Runx1* and its associated protein in the murine mammary gland (van Bragt, Hu et al. 2014). Results from analyses using RT-qPCR, IHC, and published microarray datasets revealed *Runx1* is expressed in all mammary epithelial cell (MEC) subsets, except the secretory alveolar luminal cells that emerge during pregnancy (van Bragt, Hu et al. 2014). Examinations of the missense mutations of *RUNX1*, identified in the previously described sequencing studies of human breast cancers, showed that such mutations have significant consequences for *RUNX1* function and lead to loss-of-function due to perturbed DNA binding. Following on from this discovery, van Bragt and associates sought to explore the consequences of this loss-of-function on the development of normal MECs using a *Runx1* conditional knockout model, in which the mouse mammary tumour virus (*MMTV*) promoter directed Cre expression in the mammary epithelium, thereby driving MEC-specific deletion of *Runx1*. The introduction of a YFP reporter allowed MEC population changes to be traced in this model. Flow cytometry-based analyses of the YFP-positive mammary epithelial cells revealed a significant reduction in the luminal MECs, particularly in the ER-positive mature luminal subpopulation, in *Runx1*-null mammary glands. Mammary tumours, however, were not detected in any of these females, which were followed for at least 18 months. Although this data may appear contradictory to the results found in previous sequencing studies, it was hypothesised that the ER+ luminal MECs were the cell of origin for *RUNX1* mutant cancers and that, while this cell of origin for this cancer is lost with perturbed *Runx1* expression alone, there may a requirement for additional collaborating oncogenic events to initiate luminal breast cancers from this cell of origin. Further to this, and consistent with its role in haematopoietic cell fate determination, various *in vitro* and *in vivo* expression analyses revealed that

RUNX1 positively regulates genes involved in the ER program whilst repressing genes that regulate the alveolar cells.

In another key study, a combination of *in vitro* and *in vivo* techniques were implemented to explore possible mechanisms behind the tumour suppression function of *RUNX1* in an ER+ context (Chimge, Little et al. 2016). One particularly interesting revelation from this study was the ability of RUNX1 to indirectly regulate the canonical Wnt/ß-catenin signalling pathway, a key player in various tumorigenic events in breast cancer (including cancer stem cell generation and maintenance) (Yin, Wang et al. 2018). RUNX1 was found to potentially mediate its tumour suppressive role in the context of oestrogen receptor-positive (ER+) breast cancer by antagonising oestrogen-mediated suppression of *AXIN1*, a negative regulator of the Wnt signalling pathway due to its role in the ß-catenin destruction complex. The second intron of the *AXIN1* gene was discovered to be co-occupied by RUNX1 and ERα, resulting in the combinatorial regulation of *AXIN1* by these proteins, whereby oestrogen signalling negatively regulates its function and RUNX1 abrogates this oestrogen-mediated suppression. This ER-dependent regulation of AXIN by RUNX1 was confirmed in the clinical setting as analysis of breast cancer patient data from the TCGA cohort indicated a correlation between the RUNX1 inhibitory index and *AXIN1* mRNA in ER-positive tumours and not in ER-negative tumours. Immunostaining of a TMA showed that associations between RUNX1 and AXIN1 at the protein level were dependent on the expression levels of ER, with ER^{high} tumours displaying significant correlations between the proteins and no significant association exhibited in ER^{low} tumours. Functionally, the loss of RUNX1 in an ER+ context resulted in decreased *AXIN1* expression, upregulated active ß-catenin, stimulated cell proliferation and stem cell marker expression, which was at least partially restored upon stabilising *AXIN1*. This indicates that ER+ breast cancers with deregulated ß-catenin activity due to loss of RUNX1 function could be targeted by either restoring RUNX1 function or stabilising *AXIN1* against ER-mediated suppression.

Other, more recent, publications have more closely examined the effect of *RUNX1* on EMT at a mechanistic level. The Stein group initially compared *RUNX1* mRNA

and associated protein expressions between a control breast cell line and various breast cancer cell lines, finding marked reductions in expression in the luminal-like cancer cell lines compared to the control, and less so in the triple-negative cancer line (Hong, Messier et al. 2017). In support of data from Wang and associates, mRNA and protein analysis showed a reduction of *RUNX1*/RUNX1 expression with cancer progression and, additionally, showed a correlation between the loss of *RUNX1* and the emergence of an EMT phenotype. Both TGFß-dependent and TGFß-independent induction of EMT resulted in decreased *Runx1* expression in MCF10A cells, whilst direct depletion of *RUNX1* in both non-tumorigenic MCF10A cells and MCF7 breast cancer cells caused them to lose their epithelial morphology and initiate EMT. Rescuing RUNX1 expression in MCF10A cells with TGFß-dependent induction of EMT and in the tumorigenic MCF10AT1 cell line drove both cell types back to an epithelial-like state. The functional consequences of *RUNX1* loss were demonstrated in TMAs of primary breast cancers and their matched lymph metastatic sites, where quantification of RUNX1 levels showed significantly lower levels in lymph samples than primary tumour samples. Low levels of RUNX1 also appeared to correlate with higher-grade tumours. The subtype dependency of *RUNX1* expression was additionally demonstrated when mining the TCGA database, as subtypes with poorer prognoses (luminal B, HER2-enriched, and basal-like) all displayed reduced *RUNX1* mRNA levels, while levels in less aggressive luminal A tumours were comparable to that of normal-like breast tissues. Further data from the Gene Expression Omnibus database showed that the overall survival was significantly poorer for patients with tumours expressing low levels of *RUNX1* mRNA compared to individuals whose tumour exhibited high levels of expression. Taken together, this data suggests that *RUNX1* acts to inhibit EMT in breast cancers, thereby limiting their progression and metastatic capacity, and *RUNX1* levels could serve as a prognostic indicator to predict patient outcome.

Kulkarni and collaborators offered another novel insight into the relationship between RUNX1 and EMT in breast cancer, this time identifying RUNX1 and RUNX3 as regulators of YAP-mediated EMT (Kulkarni, Tan et al. 2018). RUNX1 or RUNX3 co-expression with YAP (a core component of the Hippo pathway) attenuated many of the YAP-induced oncogenic effects on mammary epithelial cells – which included increased proliferation, enhanced trans-well migration, elevated

expression of migration and mesenchymal markers, enhanced mammosphere formation, and aberrant acini development – in a manner that was highly dependent on direct interactions between YAP and the RUNX factors. These interaction-dependent effects of RUNX1 and RUNX3 on YAP were shown, using microarray gene expression analysis with RT-PCR validation, to be facilitated by their ability to alter YAP-Signature gene expression in an interaction-dependent manner. Enrichment analyses allowed for the discovery that the YAP-Signature genes altered by RUNX1 and RUNX3 were most significantly representative of EMT and mammary stem cell signatures, with both RUNX1 members inhibiting these YAP-mediated functions. Analysing whole genome expression profiles of breast cancers allowed assessments to be made for the clinical significance of the various regulatory functions of RUNX1 and RUNX3 on YAP-mediated oncogenic functions. Unsupervised hierarchical clustering of the breast cancer patient expression data segregated the data into 4 groups based on their similarity with the YAP-Signature. The subgroups with higher enrichment for the YAP-Signature were denoted as YAP_{sig}^{high} while those with lower enrichment were YAP_{sig}^{low}, and each of these subgroups was further distinguished based on their expression of RUNX1/RUNX3 by subdividing into RUNX1/RUNX3^{high} and RUNX1/RUNX3^{low}. High expression of RUNX1/RUNX3 significantly improved overall and disease-free survival of patients with higher enrichment for the YAP-Signature. While both YAP_{sig}^{high} groups scored higher enrichment for EMT and stemness signatures compared to the YAP_{sig}^{low} groups, there was a lower enrichment for these signatures when RUNX1/RUNX3 expression was high, again indicating a protective role for RUNX1 and RUNX3 against the oncogenic functions of YAP through modulating EMT and stemness gene expression.

1.5.3.2 The oncogenic function of RUNX1 in ER-negative breast cancers

Paradoxically, there is also increasing evidence indicating a pro-oncogenic role for *RUNX1* in breast cancer, particularly in the ER-negative and TNBC subtypes. Notably, the triple negative breast cancer subgroups exhibited upregulated *RUNX1* mRNA expression in independent transcriptome studies (Karn, Pusztai et al. 2011, Rody, Karn et al. 2011). In an additional study, focusing on super-enhancers, H3K27ac ChIP-Seq data was used to identify super-enhancers in various human cancers along with their associated genes, which consisted of a large and varied

range of known oncogenes. *Oct4*, *Sox2*, and *Nanog* are examples of transcription factors that are enriched at super-enhancers and contribute to super-enhancer formation by binding directly to their known DNA sequence motifs (Hnisz, Abraham et al. 2013). These murine embryonic stem cell transcription factors were previously shown to occupy super-enhancers (add reference – Whyte). *RUNX1* was another such oncogenic driver that was associated with super-enhancers, particularly in an ER-negative breast cancer cell line (Hnisz, Abraham et al. 2013).

In a comprehensive TMA analysis of RUNX1 protein expression and associated clinical outcome in primary breast tumours, immunohistochemical staining of invasive ductal breast cancer biopsies showed a statistically significant association between RUNX1 expression and poorer cancer-specific survival in patients with ER-negative breast cancer and patients with TNBC, but not patients with ER-positive breast cancer (Ferrari, Mohammed et al. 2014). This was the first study to identify RUNX1 as an independent prognostic marker in the TNBC subtype, which correlates with poor patient outcome.

Another study focused on the role of *Runx1* within the *MMTV-PyMT* transgenic mouse model (Browne, Taipaleenmäki et al. 2015). Expression of *Runx1* RNA and RUNX1 protein was shown to increase gradually with disease progression in the mammary tissue (the primary tumour site), and was further elevated in tumours and distal metastases in the lung in later stages of progression. Further *in vitro* analyses were conducted on an epithelial tumour cell line, derived from the tumour of a 13 week old *MMTV-PyMT* mouse. This cell line showed elevated *Runx1*/RUNX1 levels compared to a normal mouse mammary gland epithelial cell line (NMuMG) and demonstrated increased migration and invasion capabilities, which were significantly abrogated upon inhibition of Runx1 by RNA interference. Overall, this data demonstrated an oncogenic role for *Runx1* in the *MMTV-PyMT* mouse model, and indicates it may contribute to the progression and metastasis of the model. It is important to consider these findings within the context of the human disease as, while the majority of tumours that develop in *MMTV-PyMT* mice are subtyped as a luminal/ER+ subtype (with gene expression profiling revealing that *PyMT* tumours cluster with the luminal B subtype of human breast cancers),

they also exhibit loss of ER expression with disease progression (Attalla, Taifour et al. 2021). Within this study, the most significant increases in *Runx1* expression were observed in the later stages of disease progression, when ER expression was likely very low or non-existent. Notably, the *MMTV* tumour cells (derived from a mouse in a late disease stage) used for *in vitro* investigations were shown to be negative for the *Esr1* gene, encoding for the ERα protein (Browne, Taipaleenmäki et al. 2015). It is therefore likely that this potential oncogenic role for *Runx1* can be more closely associated with progression and aggressive behaviours in mammary tumours that lack ER expression.

1.6 Mammary stem cells and breast cancer stem cells

1.6.1 Mammary stem cells

The mammary gland is a highly specialised organ whose development is distinctive from other organs due to the fact its morphogenesis predominantly occurs during postnatal development, in puberty (Sternlicht 2006). The mammary gland is also highly dynamic and, even post-puberty, continues to undergo extensive morphogenesis with oestrus cycling and with each occurrence of pregnancy. The processes that underpin this dynamic nature of the mammary gland involve a complex interplay of systemic (hormonal) and local (growth factor) cues, which promote the various intricate molecular and cellular mechanisms necessary for these cellular and structural changes. In particular, this plasticity of the mammary epithelium is accompanied by various cellular changes including proliferation, differentiation, and apoptosis, which ultimately lead to tissue remodelling. It was therefore often theorised that this was indicative of the existence of a renewable stem or progenitor cell population that is capable of differentiating into the various mammary cell types, thus ensuring the growth and development of the gland during its development/remodelling cycle.

It was not until the development of the cleared mammary fat pad technique (Deome, Faulkin et al. 1959) that it became possible to explore a cell population's self—renewal capacity and its ability to reconstitute a fat pad, cleared of its epithelium, following transplantation. One particular study, using an adapted version of this cleared fat pad assay, was instrumental in revealing that mammary

stem cells are distributed throughout the ductal network of the mouse mammary gland at any developmental stage throughout their lifetime (Daniel, Aidells et al. 1975). This was indicated by the fact that any portion of the ductal tree, taken at any age and at any developmental stage, transplanted into a cleared fat pad, had the capacity to regenerate the entire structure and could differentiate into various specialised functional cell types that were responsive and adaptive to the changing hormonal environment. This is an interesting contrast to other structures in the human body, such as the intestine, in which the stem cells are contained within specific regions in the base of the crypts of the adult intestine (Barker, van de Wetering et al. 2008, Potten, Gandara et al. 2009).

Further advances in mammary stem cell research were facilitated through the establishment of techniques for the dissociation of mammary tissues and experimental methodologies for the isolation of particular cell populations. In some publications, this isolation of specific cell populations was facilitated using flow-assisted cell sorting (FACS) of cell populations labelled with specific antibodies against cell surface antigens (Smalley, Titley et al. 1998, Stingl, Eaves et al. 1998, Stingl, Eaves et al. 2001), an approach that was typically implemented in the analysis of haematopoietic stem cells. For example, basal cells are identifiable by their $CD49f^{High};EPCAM^{+};K14^{+}$ expression pattern, luminal progenitors are described as $CD49f^{+};EPCAM^{+};ER^{-};K8/18^{+}$ populations, and mature luminal cells show a $CD49f^{-};EPCAM^{+};ER^{+};K8/18^{+}$ expression pattern. Various isolated stem/progenitor cell populations were phenotypically and functionally characterised using *in vitro* colony formation assays and *in vivo* limiting dilution transplantation assays into the cleared fat pads of mice (Stingl, Eaves et al. 2001, Welm, Tepera et al. 2002). Though these studies differed in their precise methodologies, including *in vivo* transplantation sites (Stingl, Eirew et al. 2006, Eirew, Stingl et al. 2008), they collectively demonstrate the similarities that exist between the stem cell hierarchies of the mouse and human mammary epithelium (Shehata, Teschendorff et al. 2012). It was not until the publication of a crucial study in 2006 that the full potential of the mammary stem cell population was realised, whereby a single mammary stem cell was demonstrated to have the capacity to reconstitute a complete mammary gland structure following transplantation into a cleared fat pad (Shackleton, Vaillant et al. 2006). Serial

transplantation of the resulting clonal outgrowths also demonstrated the self-renewal capabilities of these cells. These results had been previously predicted in a much earlier publication from 1998, which presented early evidence predicting that each of the individual multipotent stem cells (positioned throughout the fully developed post-pubertal mammary gland) may individually possess the ability to differentiate into sufficient progeny in order to recapitulate an entire functional gland over several transplant generations (Kordon and Smith 1998).

1.6.2 Lineage tracing studies

In addition to these transplantation studies, lineage tracing experiments have been a rich source of evidence for the existence of a hierarchy of stem and progenitor cells within the mammary epithelium. Lineage tracing studies differ from transplantation studies in that they can be used to track individual stem cells and their progeny *in vivo* through expression of a reporter gene, whereas transplantation studies allow for the quantification of the repopulating capacity of distinct mammary subpopulations. A tissue- or cell-specific promoter is often used to drive expression of a Cre recombinase to cause recombination at *loxP* sites in targeted cells, thereby causing excision of the stop cassette preceding the reporter gene (*GFP*, *RFP* etc.). The reporter gene therefore becomes expressed within the cells of interest and becomes a permanent, heritable identifier for these cells and their progeny. This approach was previously used for the identification of intestinal stem cells whereby a tamoxifen inducible *CreERT2*, along with *EGFP*, were knocked into the *Lgr5* locus in order to confirm predictions that this gene is marker of intestinal stem cells (Barker, van Es et al. 2007). An early mammary-specific inducible cell fate mapping study resulted in the identification of long-lived unipotent stem cells in both the luminal and myoepithelial lineages that independently drive expansion of the ductal tree in puberty and pregnancy, and maintain ductal homeostasis in adulthood (Van Keymeulen, Rocha et al. 2011). Another study showed that, in addition to the distinct long-lived progenitor cells, there also exists bipotent mammary stem cells within the mammary gland that contribute to both its morphogenesis and maintenance (Rios, Fu et al. 2014). These bipotent mammary epithelial cells were traced *in situ* using a novel three-dimensional imaging strategy in combination with clonal cell-fate mapping studies using a doxycycline-inducible multicolour

(*Confetti*) reporter system, generated by the Clevers lab (Snippert, van der Flier et al. 2010), in two separate mouse models. Interestingly, an additional mouse model was generated in order to trace *Lgr5*-expressing cells, which demonstrated *Lgr5* to be a marker of bipotent stem cells that are distributed throughout the epithelial tree of the adult gland (Rios, Fu et al. 2014).

1.6.3 *In vitro* stem cell assays

Additional key progress in the mammary stem cell field has been achieved through the development of the mammosphere assay (Dontu, Abdallah et al. 2003), an *in vitro* single cell propagation assay (based on the neurosphere assay protocol) used to test for various stem cell properties (including self-renewal and differentiation) in cell populations, as well as enabling the isolation and characterisation of these stem cells. Since its initial introduction into the field of mammary research, this method has been further refined (Shaw, Harrison et al. 2012). The mammosphere assay also allows for the relative quantification of stem cell/progenitor activity in given mammary epithelial cell populations by comparing the numbers of 3D mammary epithelial cell-derived spheres formed after seeding single cell suspensions in low density, free-floating, and growth-limiting conditions. It is thought that only cells with stem-like features are capable of forming mammospheres, and additional stem-associated behaviours such as self-renewal capacity can be further probed for by harvesting and dissociating primary mammospheres and re-seeding to determine their ability to form secondary and tertiary mammospheres (Shaw, Harrison et al. 2012).

1.6.4 Breast cancer stem cells

Combining together much of the knowledge obtained from the various investigations into mammary stem and progenitor cells, there is clear evidence that the mammary stem cell compartment is highly heterogeneous. Studies suggest the existence of multiple molecularly distinct populations of stem cells, each with their own distinctive (and often context-dependent) capabilities and functions (Colacino, Azizi et al. 2018, French and Tornillo 2019). Of note, plasticity and self-renewal are essential stem cell capabilities that are required for the morphogenesis, remodelling/regeneration, and maintenance of the

mammary epithelium throughout the various developmental stages and cycles. Interestingly, these features are also thought to be acquired by subpopulations of transformed mammary cells, known as cancer stem cells in order to drive tumour initiation, heterogeneity, and resistance or recurrence following therapy. These breast cancer stem cells are thought to have three individual possible causes of their emergence. One such source is the transformation of normal mammary stem cells through the acquisition of multiple genetic mutations. It is also possible that breast cancer stem cells arise due to an acquisition of self-renewal capabilities in normal mammary epithelial (non-stem) cells, through the accumulation of genetic mutations. A third possible source is (due to the high degree of cellular plasticity within mammary tumours) de-differentiation of cancer cells back towards a stem- or progenitor-like state.

Notably, various studies have shown that adult mammary stem cells and breast cancer stem cells share some common molecular regulators and markers. For example, the numbers of cells expressing the cell surface markers ($Lin^{-}CD24^{+}CD29^{high}$), which can be used to enrich for mammary stem cells, were found to be expanded in pre-neoplastic tissues from Wnt-driven mouse models of mammary tumorigenesis (Shackleton, Vaillant et al. 2006, Wan, Lu et al. 2014). Subpopulations of cells expressing these surface markers were also identified in *Brca1*-mutant cancer cell lines (derived from mammary tumours from *Brca1*$^{Ko/Co}$*;p53*$^{+/-}$*;WAP-Cre* mice) and *BRCA1*-mutant primary mammary tumours (Vassilopoulos, Wang et al. 2008). These subpopulations were also shown to increase the tumorigenic ability of these cell lines, implying that breast neoplasia and tumour-initiation are at least partly driven through expansion of the normal stem cell phenotype following oncogenic transformation of the mammary gland. Another more recent example is *Lgr5*, wherein lineage tracing studies indicated that cells expressing *Lgr5* contribute to the formation of both luminal and myoepithelial cell subsets of the duct, in addition to the alveoli, which established their bipotent potential (Rios, Fu et al. 2014). FACS sorting mouse mammary epithelial cells (MMECs) for those expressing *Lgr5*, followed by *in vitro* mammary colony formation assays and *in vivo* mammary repopulating assays, demonstrated that *Lgr5* marks epithelial cell populations with increased regenerative potential compared to those not expressing *Lgr5* (Plaks, Brenot et al. 2013, Trejo, Luna et

al. 2017). Interestingly, *LGR5* (and its protein) was overexpressed in primary breast cancer tissues compared to paired adjacent non-tumour tissues, in both RT-qPCR and TMA studies (Yang, Tang et al. 2015). There were several clinical variables, including tumour size, TNM (tumour, node, metastasis) stage, recurrence and metastasis, and a bias towards more aggressive subtypes, that LGR5 expression was correlated with (Yang, Tang et al. 2015, Hou, Chen et al. 2018, Lee, Myung et al. 2021), in addition to also having associations with poorer patient outcomes, including shorter overall and disease free-survival (Yang, Tang et al. 2015, Hou, Chen et al. 2018). Additionally, *LGR5* was shown to promote a wide range of stem cell-related aggressive behaviours in breast cancer cells (migration, invasion, EMT, metastasis), using a variety of *in vitro* and *in vivo* investigations, through its activation of the canonical Wnt/β-catenin signalling pathway (Yang, Tang et al. 2015). The Wnt signalling pathway is important for the regulation of mammary stem cell differentiation in the normal gland and, in the context of breast cancer, its dysregulation is associated with putative cancer stem cells and the expansion of a progenitor/stem-like cell population during tumour progression (Valkenburg, Graveel et al. 2011). This indicates that this dysregulation of the Wnt/β-catenin signalling pathway is a critical event for both the initiation and progression of mammary tumorigenesis. The *ALDH* genes, particularly *ALDH1A1*, have also been identified as relevant markers of stem cells in both the normal and malignant mammary gland (Ginestier, Hur et al. 2007). Isolating ALDHEFLUOR-positive and -negative populations from the normal mammary epithelium and analysing using a variety of *in vitro* and *in vivo* assays demonstrated that only those with *ALDH1* expression exhibited characteristics associated with stemness, such as the ability to form mammospheres or 3D colonies, and the ability to reconstitute a cleared mammary fat pad. Human breast cancers that were FACS sorted based on ALDH expression showed that only ALDEFLUOR-positive cells were capable of reproducibly generating tumours when limiting dilutions were transplanted orthotopically in the humanized cleared fat-pad of NOD/SCID mice, indicating its importance in tumorigenesis. Independent studies, analysing ALDH1 expression by IHC in TMAs, revealed that ALDH1 expression was associated with more aggressive breast cancer subtypes, with the highest expression levels seen in the hormone receptor-negative subtypes (Ginestier, Hur et al. 2007, Khoury, Ademuyiwa et al. 2012), in addition to increased tumour size, more advanced tumour stages, and metastasis to lymph

nodes (Khoury, Ademuyiwa et al. 2012). It was also correlated with poorer clinical outcome for patients, including shorter overall and disease-free survival in patients treated with neoadjuvant therapy (Ginestier, Hur et al. 2007, Khoury, Ademuyiwa et al. 2012). Furthermore, ALDH1A1 protein was expressed in a significant percentage of tumours from patients whose disease showed a recurrence following neoadjuvant therapy (Khoury, Ademuyiwa et al. 2012), a process that stem cells are essential in.

In addition to various stem cell markers being common to both the normal and tumorigenic mammary epithelium, there is much evidence that mutations of components or pathways critical for the normal regulation of mammary stem cells can cause the formation of breast cancer stem cells, and thereby initiate breast tumorigenesis. This includes various components of the Wnt/β-catenin (β-catenin, *Wnt1*, *Wnt3a*, *Wnt10b*), Notch (*Notch1-Notch4*), and Hedgehog (*Shh*, *Ihh*, *Dhh*, *Ptch1*, *Ptch2*, *Smo*, *Gli1-3*) signalling pathways (Farnie and Clarke 2007, Lamb, Ablett et al. 2013, Yang, Wang et al. 2017). Many of these previous studies have focused on the role of proto-oncogenes in the formation of cancer stem cells, however there have also been potential links made with loss of function mutations in tumour suppressor genes and the initiation of breast cancer stem cells. For example, loss of function mutations in the retinoblastoma gene (*RB*), a prototypic tumour suppressor that is altered in breast cancer, promotes the proliferation of cancer cells, is associated with poorer patient outcomes, and has links with hormone therapy resistance (Hwang-Verslues, Chang et al. 2008). This indicates that loss of *RB* function in the breast, due to its mutation, may at least partly contribute to the process by which normal mammary stem cells transform into breast cancer stem cells. Additionally, during investigations into the mechanisms of the tumour suppressor functions of *PPARγ*, it was found that loss of functional *PPARγ* in a *Wnt1*-driven mouse model of basal breast cancer resulted in earlier tumour emergence and faster tumour growth (Kramer, Wu et al. 2016). This was accompanied by the expansion of a pool of mammary stem cells expressing $CD24^{+}/CD49f^{hi}$ cell surface markers, and increased expression of the Proliferating Cell Nuclear Antigen (PCNA) proliferative marker by IHC, within these tumours, indicating a role for *PPARγ* in limiting the cancer stem cell potential in the context of *Wnt1*-driven mammary tumorigenesis.

1.6.5 A possible role for RUNX1 in mammary stem cells and breast cancer stem cells

RUNX1 is a putative tumour suppressor gene in the mammary gland whose expression has been linked with regulation of stem cells in the normal and tumorigenic breast, and its loss in already established breast tumour cells has been associated with the expansion of cell populations possessing progenitor-/stem-like properties. One such publication identified *RUNX1* as a potential regulator of mammary stem cell differentiation and mammary morphogenesis in 3D culture models, whereby shRNA-mediating *RUNX1* silencing in the MCF10A cell line resulted in a significantly reduced ability to produce mature lobules, or form ducts in 3D cultures. Rescuing *RUNX1* expression reversed this phenotype (Sokol, Sanduja et al. 2015). The Blyth lab finds indications of *Runx1* being involved in restricting the stem/progenitor cell functions in the tumorigenic mammary gland within two independent mouse models (Riggio 2017), which are discussed in detail in this thesis. This supports publications from the Stein group, who used the MCF10A series of cell lines to interrogate the functions of *RUNX1* (and *RUNX2*) in the context of breast cancer stem cells (Hong, Fritz et al. 2018, Fritz, Hong et al. 2020). MCF10CA1a cells, injected into the mammary fat pads of SCID mice, exhibited reduced expression of *RUNX1* in the resulting tumours. Overexpressing *RUNX1* in these cells reduced tumorigenesis and resulted in smaller tumours being formed. *In vitro* assays were also performed using *RUNX1*-overexpressing MCF10AT1 and MCF10CA1a cell lines, which demonstrated reduced migration and invasion capacities (Hong, Fritz et al. 2018). This data, coupled with observations that loss of *RUNX1* leads to epithelial-to-mesenchymal transition (EMT) that can be reversed by rescuing its expression (Hong, Messier et al. 2017), suggests that *RUNX1* could be repressing the breast cancer stem cell phenotype within established tumours, which forms at least part of its role as a tumour suppressor in the mammary gland.

1.7 The importance of genetically engineered mouse models (GEMMs) as tools to study breast cancer

While transplantation models have been key to understanding the mouse mammary epithelium, particularly its stem cell populations, this experimental approach has also received some criticism in the past. Based on observations in

lineage tracing studies, it has been argued that cells, which were isolated from mammary tissues using dissociation techniques and subsequently transplanted into cleared fat pads, may acquire properties or behaviours they would not typically exhibit in the context of the intact tissue (Visvader and Stingl 2014). There are also indications that limiting dilution assays into cleared fat pads and cancer cell transplantation studies do not fully recapitulate the complexity of the mammary gland landscape and the various intricate cellular interactions occurring therein. Notably, tumours of mice that were genetically engineered to develop primary tumours in the gastrocnemius muscle reacted differently to therapy than transplantation-derived tumours, which was indicated to be at least partly influenced by the distinct immune landscapes within each of their tumour environments (Wisdom, Mowery et al. 2020). Within this same vein, transplantation experiments involving human-derived cells often necessitates the use of immunodeficient mice, which could potentially impact on tumour seeding and growth capabilities in addition to altering the response to therapies.

Many of the loss-of-function studies used to explore the biological functions of proteins have relied upon the use of immortalised cell lines derived from already-established tumours, meaning it has mainly been studied within the context of late tumour progression as opposed to tumour initiation and early tumorigenesis. These breast cancer cell lines, and any tumours they form in transplantation studies, are mostly comprised of cells that have accumulated various genetic alterations in order to facilitate their tumourigencity, proliferative capacity, and more aggressive-like behaviours (such as an EMT phenotype, propensity for metastasis, etc.) and therefore do not represent the early transformation events that occur in the breast to facilitate tumour emergence.

Genetically engineered mouse models offer a potentially more relevant or suitable context in which to decipher the tumour-suppressive or pro-oncogenic functions of proteins within the normal mammary gland and during the various stages of tumorigenesis. Mice are the primary animal model used for the study of the human breast and breast cancer, due to the mouse gland being developmentally/structurally and functionally similar to that of the human breast

(McNally and Stein 2017). It is also particularly relevant for the study of *RUNX1*, and its functional loss, in the mammary gland due to it being expressed rather highly in mouse mammary epithelial cells, as demonstrated in RT-qPCR analysis (Ferrari 2013, McDonald, Ferrari et al. 2014) and by co-immunofluorescence (IF), shown in Figure 1.6, conducted by myself.

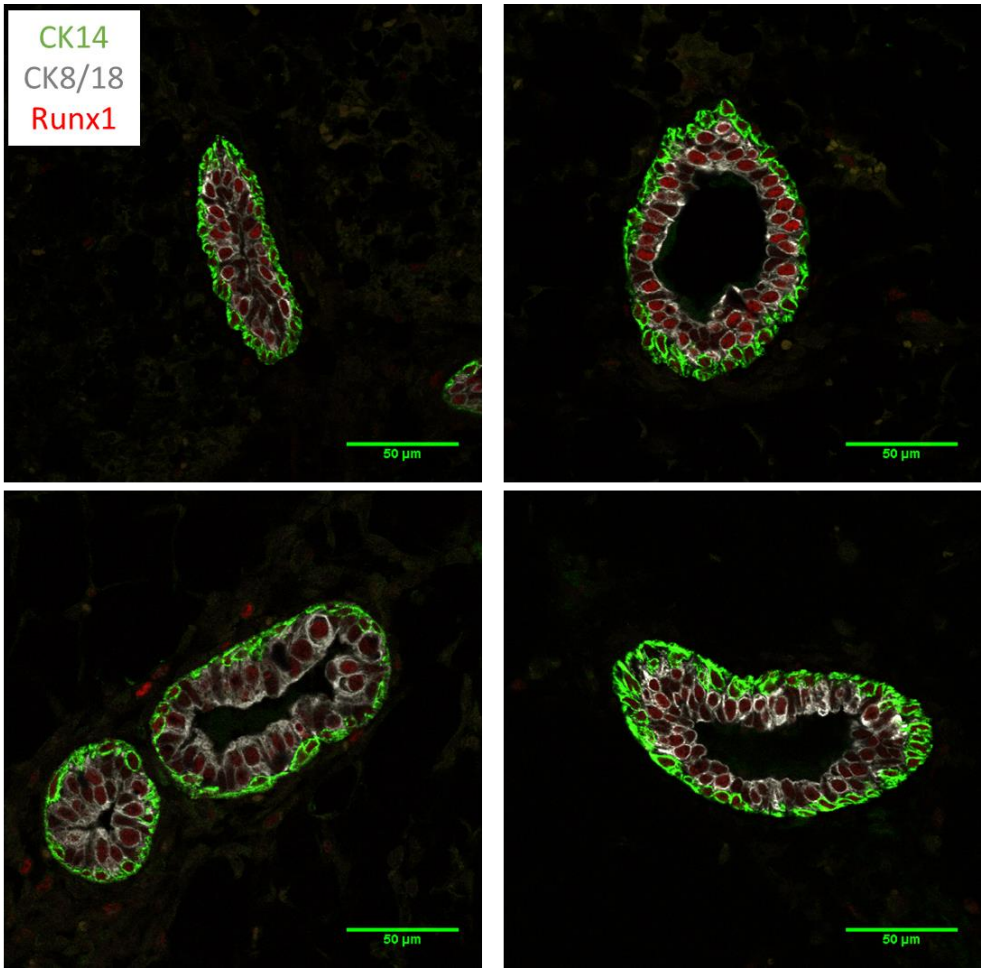


Figure 1.6 RUNX1 protein expression in ducts of post-pubertal mouse mammary glands.

Co-immunofluorescence of Cytokeratin 14 (CK14 – green, marking basal cells), Cytokeratin 8/18 (CK8/18 – grey, marking luminal cells), and RUNX1 proteins (red) in formalin-fixed paraffin-embedded (FFPE) tissues extracted from 6 week-old (post-pubertal) mice. Scale bar=50µM. Images were taken using a Zeiss LSM 710 microscope and processed (scale bar added) using ImageJ.

Whole-body loss-of-function studies for *Runx1* lead to embryonic lethality in mice, due to its absolute requirement for definitive haematopoiesis, making it impossible to study *RUNX1* in breast cancer in these models. However, this

limitation can be circumvented by using conditional knockout mice whereby gene knockout can be both spatially and temporally controlled using site-specific or inducible recombinases (Hall, Limaye et al. 2009). This thesis used two such independent GEMM models to study mammary-specific gene deletion. Specifically, the *Blg-Cre* (Selbert, Bentley et al. 1998) and *MMTV-Cre* (Wagner, Wall et al. 1997) promoters were used to drive mammary-specific gene knockout in models of mammary tumorigenesis described below.

1.7.1 The *MMTV-Cre;MMTV-PyMT* mouse model

The *MMTV-Cre* mouse line was developed by the Hennighausen laboratory, whose work on this model was first published in 1997 (Wagner, Wall et al. 1997). It has since been estimated, from a subsequent paper in 2011, that this mouse line has been used by at least 40 different labs to explore the functions of various genes (using loss-of-expression or loss-of-function experiments) in the mammary epithelium (Robinson and Hennighausen 2011), a number which has likely increased in the years subsequent to this. Three distinct experiments were used to characterise *MMTV*-driven *Cre* expression: the tissue distribution of *Cre* mRNA was evaluated by RT-PCR analysis of various adult mouse tissues; delivery of an adenovirus carrying a reporter gene was used in order to functionally analyse *Cre* recombinase at the single-cell level; the *MMTV-Cre* line was crossed onto a reporter strain carrying a lox-stop-lox transgene and a subsequent recombination assay performed in order to examine *Cre* activity. By these means it was discovered that expression of the *MMTV-Cre* transgene was confined to striated ductal cells of the salivary gland and mammary epithelial cells of adult (virgin and lactating) mice (Wagner, Wall et al. 1997).

The *MMTV-PyMT* transgenic mouse model, originally developed in 1992 by the Muller laboratory (Guy, Cardiff et al. 1992), is well established and widely used in the field of breast cancer research due to its potential to study various aspects of tumorigenesis in the breast (initiation, progression, and metastasis). Although *PyMT* is not a human oncogene, it interacts with various signal transducers and thereby mimics an activated receptor tyrosine kinase (Attalla, Taifour et al. 2021), a class of receptors that play an essential role in the progression of several cancer

types, including breast cancer (Butti, Das et al. 2018). Given that the *MMTV-PyMT* transgene can effectively mimic such a broad and essential spectrum of components relating to human cancer progression, it is no surprise that this model can recapitulate the progression of the human disease relatively accurately. Notably, and as previously mentioned, the *PyMT*-driven tumours were shown to cluster with the luminal B subtype of human breast cancers in gene expression profiling experiments (Herschkowitz, Simin et al. 2007, Lim, Wu et al. 2010). Similar to observations within the human disease, *PyMT*-driven tumours often also exhibit loss of ER expression with disease progression (Attalla, Taifour et al. 2021).

Since the initial generation of the *MMTV-Cre* mouse line, it has been discovered that this transgene can adversely affect lactation (Robinson and Hennighausen 2011, Yuan, Wang et al. 2011). The inability of these mice to nurse was correlated with the appearance of the tissue within the mammary gland structures, including incomplete filling of the fat pads. The *MMTV-Cre* line also exhibits some expression mosaicism, which may result in the selection of cells containing undeleted genes, though this is highly dependent on the nature of the targeted gene and whether cells containing deletions are at a selective disadvantage as compared with their undeleted counterparts. Additionally, where the MMTV-PyMT tumour model is used in combination with other conditional genetic alterations regulated by the *MMTV-Cre*, it is important to note that both *Cre* and *PyMT* are driven by independent *MMTV* promotors sitting in different insertion site. This may result in *MMTV-PyMT*-driven oncogenesis and *MMTV-Cre*-mediated gene alteration occurring asynchronously and independently. Finally, the *MMTV-Cre* promoter was shown to not be entirely specific to the mammary gland as it was found to be expressed in other epithelia and in haematopoietic lineages, which has implications for *Runx1* deletion due to its essential role in definitive haematopoiesis.

1.7.2 The *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ mouse model

The *Blg-Cre* promoter offers a potential alternative to the *MMTV-Cre* promoter which, in addition to being highly tissue-specific (with very little background expression in non-mammary tissues at each stage of development), has a certain

amount of temporal regulation that can be achieved (Selbert, Bentley et al. 1998). Relatively low levels (7%) of *Cre*-mediated deletion were exhibited in the cells of the virgin gland, compared to the lactating mammary gland in which the vast majority of cells (70-80%) exhibited *Cre*-mediated recombination. Characterisation of the *Blg-Cre* driver, conducted by myself, has been carried out using a reporter allele (Luche, Weber et al. 2007) and glands assessed by flow cytometry. These analyses have nicely demonstrated an increased expression in the luminal lineage and an almost complete expression in the pregnant gland (Figure 1.7). Notably there was little expression in the non-mammary lineage (lineage/Lin positive population). This analysis also supported previous data and indicated that parity was a necessary procedure for potentiating sufficient expression of *Blg-Cre*, and therefore the desired genetic alterations, in the mammary epithelium as even a single pregnancy in these mice resulted in highly elevated levels of RFP as compared with their virgin counterparts.

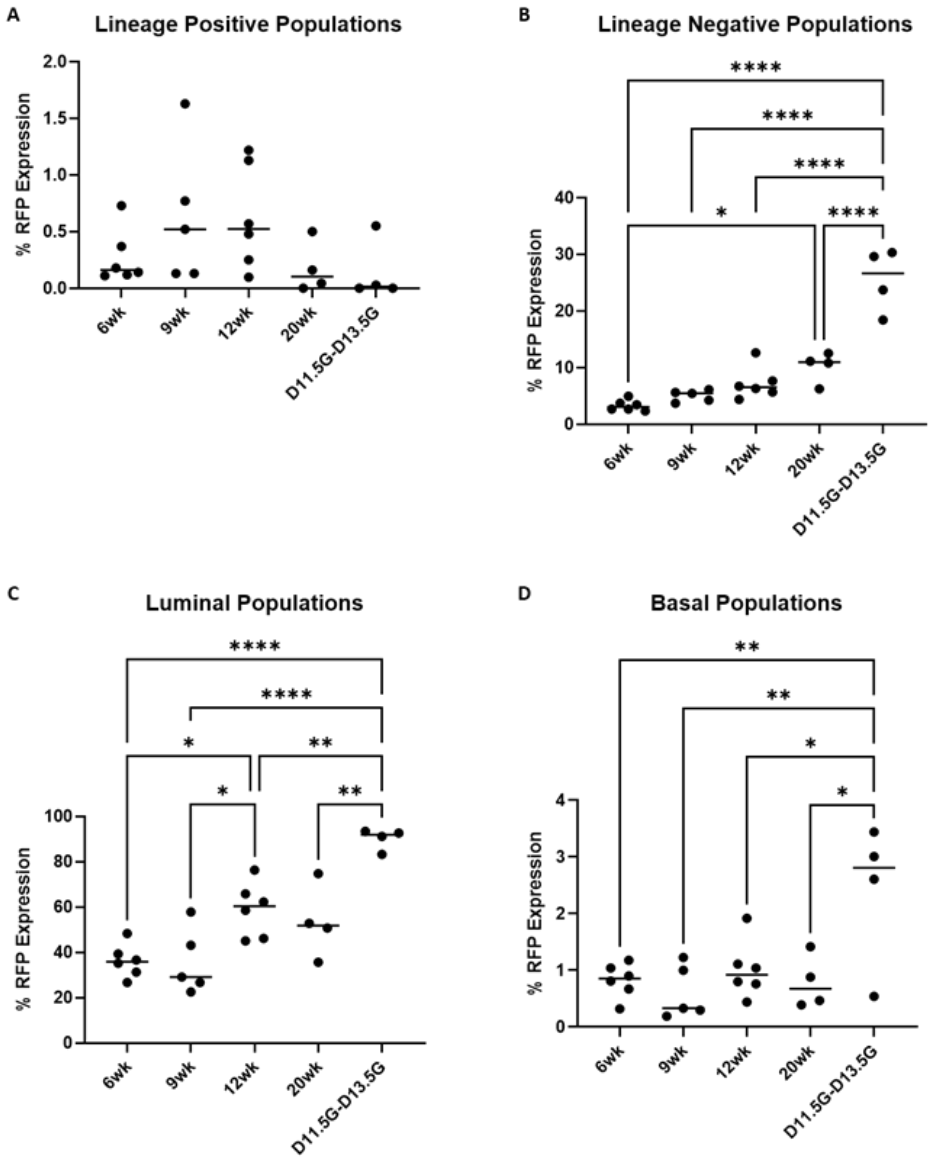


Figure 1.7 Characterisation of *Blg-Cre* expression in the mouse mammary gland using surrogate RFP reporter expression.

Scatter plots from flow-cytometry based investigations showing the proportion of RFP+ (*Blg-Cre*-expressing) cells in various populations of mammary cells, isolated from mammary glands of *Blg-Cre;tdRFP* virgin mice at 6 (n=6), 9 (n=5), 12 (n=6) and 20 (n=4) weeks of age and pregnant mice (n=4) at day 11.5-13.5 gestation (D11.5G-D13.5G). Graphs were plotted using GraphPad Prism 9, with each data point representing the percentage of cells expressing RFP within each of the specified Lin positive [A], luminal [C], and basal [D] sample populations at each of the timepoints. For statistical comparisons, one-way ANOVA was performed, with Tukey's Post-Hoc Test to correct for multiple comparisons [* $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$].

The *Catnb*$^{wt/lox(ex3)}$ mouse line is a line that was generated to carry a stabilised form of ß-catenin (Harada, Tamai et al. 1999). This stabilisation of ß-catenin is achieved through the insertion of two *loxP* sites into the introns flanking exon 3 of the gene, which encodes the phospho-acceptor residues that promote ß-catenin

degradation. By crossing this mouse line with *Blg-Cre* mice, the Blyth laboratory was able to create a novel conditional breast cancer model mimicking mammary-specific constitutive activation of the Wnt/ß-catenin pathway, which has been shown to be a common feature of various solid tumours, including breast cancers (Zhan, Rindtorff et al. 2017), and is potentially involved in the pathogenesis of some acute myeloid leukaemias (Simon, Grandage et al. 2005). This model has a long latency of mammary tumourigenesis with incomplete penetrance in virgin animals and increased incidence in parous mice (Riggio **Unpublished**). The usage of *loxP* sites to generate a mutant ß-catenin allele in the *Catnb*$^{wt/lox(ex3)}$ mouse line results in this model having an advantage over the *MMTV-PyMT* line, in that the *Blg-Cre* promoter can be used to simultaneously and synchronously drive mammary-specific oncogenic activation of the Wnt/ß-catenin alongside any other desirable *Cre-loxP*-driven genetic alterations. This GEMM model therefore provided an opportunity to explore the implications of mammary specific loss of *Runx1* function in the context of Wnt-driven mammary tumorigenesis.

1.8 Aims of the thesis

While previous studies relating to the role of *RUNX1* provide strong evidence for a context-dependent role in breast cancer, and offers evidence of its EMT- and potential stem-like behaviour-restricting properties in the normal and tumorigenic mammary gland, there is still much work to be done to fully comprehend the tumour suppressive role of *RUNX1* in breast cancer. For example, there is very limited data pertaining to the molecular mechanisms behind the tumour suppressive role of *RUNX1*. In particular, there is a dearth of information relating to the potential molecular mechanisms behind the regulatory role for *RUNX1* in the potential stemness of the normal and tumorigenic breast. Additionally, many of these studies relied on evidence from *in vitro* and *ex vivo* protocols and much of the *in vivo* evidence was collected using transplantation models as opposed to genetically engineered mouse models (GEMMs).

Using a novel GEMM for mammary cancer, the Blyth laboratory demonstrated for the first time that *Runx1* is capable of restricting the oncogenic action of activated ß-catenin in the mammary gland *in vivo*. This supports previous observations of

RUNX1 loss in ER-positive breast cancers. Additionally, the stem-associated phenotype observed within this model indicates that RUNX1 may protect the mammary gland from ER-positive breast tumours by controlling its development and by inhibiting the expansion of subpopulations of cells that are particularly vulnerable to transformation events. It is also possible that RUNX1 could be acting to limit the tumorigenic potential of the oncogenes driving breast cancer by antagonising the various downstream targets that facilitate their tumour-promoting functions.

1.8.1 Aim 1: Uncovering the mechanisms of RUNX1 tumour suppression in the mammary gland

One specific aim of this work was to explore the potential molecular mechanisms behind the tumour suppressor role of *RUNX1* in breast cancer, by exploring the means by which loss of functional *Runx1* facilitates earlier mammary oncogenesis in the mouse mammary gland. A relatively unbiased molecular approach was taken to achieve this, using RNA-sequencing analysis of mammary epithelial cells extracted from *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;tdRFP* mice, in order to uncover the various molecular mechanisms underlying earlier tumour emergence in mammary glands with loss of functional *Runx1*.

1.8.2 Aim 2: Exploring the role of RUNX1 in the "stemness" of the normal and tumorigenic mammary gland

A second aim was to investigate the effects of RUNX1 on the mammary cell populations and determine whether its loss results in the expansion of a subpopulation of stem or progenitor cells that are more vulnerable to oncogenic insult. The role of *RUNX1* in stemness was investigated within the context of normal mammary epithelial cells and at various stages of tumorigenesis, in order to explore the possibility that the tumour-suppressive role of *RUNX1* may be at least partially mediated by its control over the stem-associated behaviours. Phenotypic and molecular approaches were taken, within *in vitro* and *in vivo* contexts. *In vivo* approaches additionally explored the effects of RUNX1 in stemness within the context of two distinct oncogenes, and at several stages of cancer progression in one model, in order to explore whether this tumour-suppressive mechanism might be context or stage-specific.

2 Materials and Methods

2.1 Animal procedures

All animal work was carried out with ethical approval from the University of Glasgow Animal Welfare and Ethical Review Board and according to the UK Home Office regulations. Work was carried out under Project Licenses 70/8645 and PP6345023. Mice were housed in a pathogen-free animal facility within individually ventilated cages. Staff from the Biological Service Unit (BSU) within the CRUK Beatson Institute carried out general husbandry (daily checks) and ensured the mice were provided with sufficient food (irradiated diet), water (in sterilised pouches), and clean/enriching cage conditions. At the time of weaning, BSU staff ear notched the mice for identification purposes and sent the excess tissue to Transnetyx (Cordova, TN, USA) for genotyping. Mouse health checks were carried out at least twice weekly under a laminar flow hood. Immediately following the onset of clinical symptoms (observation of palpable tumours), males and non-experimental females were culled (according to the conditions set out in the Project Licence) and experimental females were monitored for tumour growth until clinical end-point. Animals were sacrificed using a combination of at least two different ASPA (Animals (Scientific Procedures) Act 1986) Schedule 1 methods. Exposure to rising concentrations of carbon dioxide was predominantly employed as a primary method (overdose of sodium pentobarbital in pregnant females), followed by a confirmatory method of either cervical dislocation or onset of rigor mortis.

2.1.1 Generation of Genetically Engineered Mouse Model (GEMM) cohorts

2.1.1.1 Blg-Cre;Catnb$^{wt/lox(ex3)}$ cohorts

In order to study the effects of mammary-specific conditional deletion of *Runx1* and/or *Runx2* in a ß-catenin-dependent model of mammary tumorigenesis, the following mouse lines were crossed together to produce various combinations of *Runx1* and *Runx2* alterations: the betalactoglobulin (*BLG*)-*Cre* line, kindly provided by AR Clarke (Selbert, Bentley et al. 1998); the stabilised ß-catenin (*Catnb*$^{wt/lox(ex3)}$) line, obtained from OJ Sansom (Harada, Tamai et al. 1999); the *Runx1*$^{fl/fl}$ conditional knockout mice, originally generated by the Nancy Speck lab (Growney, Shigematsu et al. 2005) and given to the Blyth lab by Marella De Bruijn

(University of Oxford); and the *Runx2*$^{fl/fl}$ line, from Professor Michael Owen (Imperial Cancer Research Fund London UK) (Ferrari, Riggio et al. 2015). The red fluorescent protein *Gt(ROSA)26Sor*^{tm1Hjf} knock-in *Cre*-reporter (Luche, Weber et al. 2007) (henceforth known as *tdRFP*), obtained from the EMMA archive, was also crossed onto the cohort to trace *Cre* expression. The final cohorts were maintained on a mixed background, derived from a mix of various strains of mice, including: C57BL/6J; FVB/NJ; CBA/Ca; and 129/Sv. The cohorts generated from these crosses were each given abbreviations, as shown in Table 2.1.

Table 2.1 *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ cohort abbreviations.

Cohort	Abbreviation
Blg-Cre;Catnb $^{wt/lox(ex3)}$	BCAT
Blg-Cre;Catnb $^{wt/lox(ex3)}$ *;Runx1* $^{fl/fl}$	BCAT_1
Blg-Cre;Catnb $^{wt/lox(ex3)}$ *;Runx2* $^{fl/fl}$	BCAT_2
Blg-Cre;Catnb $^{wt/lox(ex3)}$ *;Runx1* $^{fl/fl}$ *;Runx2* $^{fl/fl}$	BCAT_12

2.1.1.2 MMTV-Cre;MMTV-PyMT cohorts

The *MMTV-Cre;MMTV-PyMT* model was initially used to study the effects of mammary-specific conditional deletion of *Runx1* alone within the context of mammary tumorigenesis. For these experiments, the following mouse lines were crossed together: the *MMTV-Cre* line, obtained from the WJ Muller lab (Andrechek, Hardy et al. 2000); the *MMTV-PyMT* line, also from the WJ Muller lab (Guy, Cardiff et al. 1992); and the *Runx1*$^{fl/fl}$ conditional knockout mice, generated by the Nancy Speck lab (Growney, Shigematsu et al. 2005) and provided by Marella De Bruijn. The *tdRFP* transgenic reporter line (Luche, Weber et al. 2007) was also crossed onto the cohort to trace *Cre* expression. These mice were finally backcrossed for up to 10 generations onto a pure FVB background. Subsequent experiments, investigating the additional loss of *Runx2* within the context of *MMTV-PyMT*-driven tumorigenesis, were conducted on these mice after crossing with *Cre*-negative *Runx2*$^{fl/fl}$ mice from the *Blg-Cre* cohort and backcrossing onto a pure FVB background. The cohorts generated from these crosses were each given abbreviations, as shown in Table 2.2.

Table 2.2 *MMTV-Cre;MMTV-PyMT* cohort abbreviations.

Cohort	Abbreviation
MMTV-Cre;tdRFP	MC
MMTV-Cre;tdRFP;Runx1$^{fl/fl}$	MC_1
MMTV-Cre;MMTV-PyMT;tdRFP	MCMP
MMTV-Cre;MMTV-PyMT;tdRFP;Runx1$^{fl/fl}$	MCMP_1

2.1.2 Survival analyses and samples

Enhanced monitoring of experimental mice was performed by palpating at least twice a week to assess for the formation of mammary gland lesions. Upon tumour notice, the initiation date and tumour size(s) were noted and tumours were subsequently measured twice a week using callipers until clinical end-point, at which point mice were sacrificed. The end-point, stipulated in the Project Licence, was the point at which the largest tumour diameter of any tumour reached 15mm, or at tumour ulceration. After sacrifice, various measurements were taken including body weight, the cumulative mammary gland weight, and lung weight. These measurements were later used to calculate tumour and lung burdens, which are the tumour or lung tissue weights represented as percentages of the respective total mouse weight. End-point mammary tumours were sampled for analyses, with the majority being fixed in 10% neutral buffered formalin (NBF), several small pieces fresh frozen, and several small pieces stored in RNAlater (ThermoFisher, #AM7021). Lung tissue was also fixed in 10% NBF and several small pieces were fresh frozen. For survival analyses, 3 distinct parameters were used to compare mouse cohorts: tumour onset (time from birth to tumour notice); tumour progression (time from tumour notice to end point); and overall survival (time from birth to end point). This data was plotted as Kaplan-Meier curves for comparison. Mice culled due to non-mammary tumour-related pathologies (cystic lesions) were censored on the survival curves.

2.1.3 Samples from *Blg-Cre;tdRFP* mice

To analyse RFP expression levels (as a surrogate of *Blg-Cre* expression) in the mammary epithelium of *Blg-Cre;tdRFP* mice during their development, virgin female mice were sacrificed at 6, 9, 12, and 20 weeks of age. Pregnant females were also taken between 11.5 and 13.5 days gestation (D11.5G-D13.5G). For each

mouse, one of the #4 mammary glands was excised for formalin fixation (2.2.1) and paraffin embedding for IF analysis (2.2.3). The rest of the glands were collected (with lymph nodes removed from the remaining #4 mammary gland) and processed as a single sample into a single-cell suspension of isolated mouse mammary epithelial cells (2.3.1), which were subsequently analysed for RFP expression within the different mammary cell lineages by flow cytometry (2.4.2).

2.1.4 Samples from *Blg-Cre;Catnb$^{wt/lox(ex3)}$;tdRFP* mice

To analyse mouse mammary epithelial cells (MMECs), isolated from 9 week-old (pre-neoplastic) *Blg-Cre;Catnb$^{wt/lox(ex3)}$;tdRFP* mammary glands, by RNA sequencing (RNA-seq), virgin mice were sacrificed at 9 weeks of age, prior to the formation of palpable lesions within the mammary glands. Following removal of the lymph nodes contained in the #4 mammary glands, all mammary gland tissue was excised from each sacrificed mouse. These were then processed as a single sample into a single-cell suspension of mouse mammary epithelial cells that were isolated by a series of enzymatic digestions (Collagenase/Hyaluronidase, Trypsin) and red blood cell lysis steps (Ammonium chloride) as described in a later section (2.3.1). RFP-expressing cells were isolated from each single cell suspension using a flow-assisted cell sorting method (2.4.3) and stored at -80°C in 100µl Lysis Solution from the RNAqueous™-Micro Total RNA Isolation Kit (Invitrogen, #AM1931) prior to RNA extraction and subsequent analysis by RNA-seq (2.5).

2.1.5 Samples from *MMTV-Cre;MMTV-PyMT;tdRFP* mice

2.1.5.1 Initial RFP analysis and tumoursphere/qPCR analyses

For the initial analysis of RFP expression in mammary glands from *MMTV-Cre;MMTV-PyMT;tdRFP* mice, RFP expression was analysed in both *MMTV-PyMT*-expressing and *MMTV-PyMT*-negative females, either with or without *MMTV-Cre*-driven loss of *Runx1* expression. *MMTV-PyMT*-expressing mice were sacrificed either upon initial date of tumour notice (DON) or at clinical end-point (EP). *MMTV-PyMT*-negative females were sacrificed at a DON equivalent timepoint of ~60 days and an EP equivalent timepoint of ~90 days. These timepoints were chosen due to the average DON and EP timepoints, calculated for *Runx1$^{wt/wt}$* and *Runx1$^{fl/fl}$* cohorts combined, being 56.4 days and 90 days, respectively. After

removing the lymph nodes within both #4 mammary glands, all mammary gland tissue (including any mammary lesions) was excised from each sacrificed mouse and processed as a single sample into a single-cell suspension of isolated mouse mammary epithelial cells (2.3.1). These cell suspensions were analysed for RFP expression by flow cytometry (2.4.2). For samples that were being set up in mammosphere/tumoursphere experiments, glands were sampled and processed in the same manner to isolate MMECs, and subsequently set up in mammosphere/tumoursphere experiments (described in 2.13.3). Some of the *MMTV-PyMT*-expressing samples that were set up in these 3D assays were also analysed by RT-qPCR (2.11) by taking a sub-sample of MMECs and freezing down as a cell pellet.

2.1.5.2 Analysis of RFP in individual glands

For subsequent analysis of RFP expression within individual mammary glands from *MMTV-Cre;MMTV-PyMT;tdRFP* mice, RFP expression was again analysed in both *MMTV-PyMT*-expressing and *MMTV-PyMT*-negative females, either with or without *MMTV-Cre*-driven loss of *Runx1* expression. As with the initial analysis, *MMTV-PyMT*-expressing mice were sacrificed upon initial date of tumour notice (DON) or clinical end-point (EP). *MMTV-PyMT*-negative females were sacrificed at a DON equivalent timepoint of ~60 days and an EP equivalent timepoint of ~90 days. An additional pre-neoplastic (PN) timepoint of ~μ days was introduced for both the *MMTV-PyMT*-expressing and –negative cohorts to investigate RFP expression at an earlier point in development. This was chosen as it was the earliest possible post-pubertal timepoint occurring prior to the formation of any palpable lesions. For each mouse that was sacrificed for this analysis, mammary gland #4 on the right body axis was taken for formalin fixation (2.2.1). After removing the lymph node from mammary gland #4 on the left body axis, the remaining mammary glands were excised and each collected as individual samples for RFP analysis, with the exception of the thoracic glands (#2 and #3) that were combined together on each side due to their overlap and indistinguishability. For the *MMTV-PyMT*-negative samples, the glands at positions #1 and #5 contained very few MMECs relative to the other glands, and compared to the #1 and #5 mammary glands in the *MMTV-PyMT*-positive glands (due to the presence of lesions in the latter). For this reason, the left and right glands were combined into a single sample for each of these

mammary gland positions. This meant that 7 individual samples were obtained from each *MMTV-PyMT*-expressing mouse, and 5 were collected from each *MMTV-PyMT*-negative mouse, as shown in Figure 2.1. These samples were processed into single-cell suspensions of isolated mouse mammary epithelial cells (2.3.1) and subsequently analysed for RFP expression by flow cytometry (2.4.2).

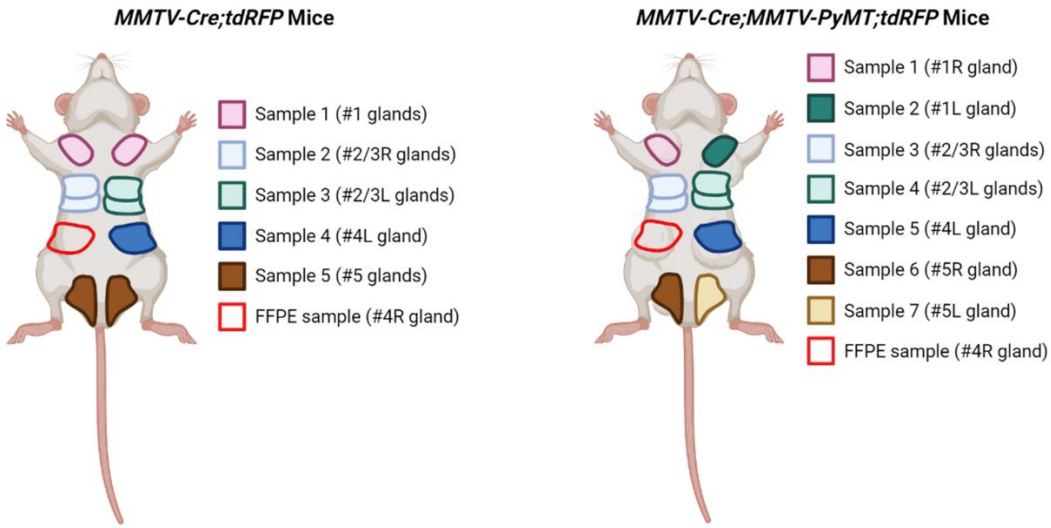


Figure 2.1 Samples taken from *MMTV-Cre;tdRFP* and *MMTV-Cre;MMTV-PyMT;tdRFP* mice for RFP expression analysis.

Diagram depicting the glands that were used to form each of the final samples used for the analysis of RFP expression in *MMTV-Cre;tdRFP* and *MMTV-Cre;MMTV-PyMT;tdRFP* mouse mammary glands.

2.2 Histology

2.2.1 Tissue fixation

Tissue samples were fixed in 10% NBF for at least 48 hours before transferring into 70% ethanol. Samples were then embedded by the CRUK Beatson Institute Histology Department in paraffin wax, within embedding cassettes.

2.2.2 Antibodies

Table 2.3 Lists of antibodies used for immunofluorescence and immunohistochemical analysis of FFPE samples.

Antibodies for Immunofluorescence Analysis					
Primary Antibody Target	**Primary Antibody (Supplier)**	**Primary Antibody Concentration**	**Primary Antibody Host Species**	**Secondary Antibody (Supplier)**	**Secondary Antibody Concentration**
CK14 (Basal cells)	Anti-Cytokeratin 14 antibody [LL002] (Abcam, #ab7800)	1:2000	Mouse	Goat anti-Mouse IgG (H+L), Alexa Fluor 488 (Invitrogen, #A-11001)	1:500
CK8/18 (Luminal Cells)	Cytokeratin 8 +18 antibody (Fitzgerald Industries, #20R-CP004)	1:1000	Guinea Pig	Goat anti-Guinea Pig IgG (H+L), Alexa Fluor 647 (Invitrogen, #A-21450)	1:500
RUNX1	AML1 (D4A6) Rabbit mAb (Mouse Preferred) (Cell Signaling Technology, #8529)	1:50	Rabbit	Goat anti-Rabbit IgG (H+L), Alexa Fluor 594 (Invitrogen, # A-11012)	1:500
Antibodies for Immunohistochemistry Analysis					
Primary Antibody Target	**Primary Antibody Name/Supplie**	**Primary Antibody Concentration**	**Primary Antibody Host Species**	**Secondary Antibody Name/Supplier**	**Secondary Antibody Concentration**
Ki67 Proliferating Cells	Ki-67 (D3B5) Rabbit mAb (Mouse Preferred; IHC Formulated) (Cell Signaling, #12202)	1:1000	Rabbit	EnVision+ Rabbit secondary antibody (Agilent, #K400311-2)	N/A (ready-to-use)
Cleaved Caspase 3 (Apoptotic Cells)	Cleaved Caspase 3 (Asp175) Antibody (Cell Signaling, #9661)	1:500	Rabbit	EnVision+ Rabbit secondary antibody (Agilent, #K400311-2)	N/A (ready-to-use)

2.2.3 Immunofluorescence (IF)

2.2.3.1 Immunofluorescence Staining

The CRUK Beatson Institute Histology Department cut blank 4µM sections from formalin-fixed paraffin-embedded (FFPE) tissue and mounted onto microscope slides. The tissue was de-waxed by submerging in Xylene (Genta Medical, #XYL050) for 3x5 minute washes, followed by a series of washes with decreasing concentrations of ethanol (2x5 minute washes in 100% ethanol, 1x5 minute wash in 70% ethanol, and a final wash in de-ionised water) to gradually rehydrate. Heat-induced epitope retrieval (HIER) was performed using a PT Module Epitope Retrieval Solution (ThermoFisher, #TA-250-PM1X) in a Thermo Scientific PT Module, which was run using for 30 minutes at 98˚C. Slides were subsequently washed in phosphate buffered saline (PBS) solution with 0.05% Tween (PBS-T). A Dako Pen was used to draw a water-repelling circle around tissue sections in order to obtain more uniform IF staining results and reduce the volume of reagents required. Sections were blocked by incubating with 5% Normal Goat Serum (NGS, Dako, #X090710-8) in PBS for 30 minutes at room temperature, to reduce non-specific binding of antibodies. Slides were then washed in PBS-T for 2x5 minute washes before incubating with primary antibody in 5% NGS/PBS overnight at 4˚C in a humidified container. Slides were washed in PBS-T for 3x5 minute washes, and then incubated in the dark with an appropriate secondary antibody in 5% NGS/PBS for 1 hour at room temperature, before 3x5 minute final washes in PBS-T. Coverslips were then mounted onto each slide using the ProLong™ Gold Antifade Mountant with DAPI (Invitrogen, #P36931), which was allowed to dry overnight at room temperature.

2.2.3.2 Immunofluorescence Analysis

Images were taken of IF-stained mammary glands manually, using a Zeiss LSM 710 microscope. These images were uploaded to the ImageJ software, which was used to analyse the images and add auto-calibrated scale bars.

2.2.4 Immunohistochemistry (IHC)

2.2.4.1 Immunohistochemical Staining

IHC staining was conducted by the CRUK Beatson Institute Histology Department. Blank 4µM sections were cut from formalin-fixed paraffin-embedded (FFPE) tissues or cell plugs and mounted onto microscope slides by the Histology Department, who then stained slides on a Leica Bond Rx Autostainer, using a Leica Bond Intense R Detection System (Leica, #DS9263) with protocol IHC Env Rb. The slides were first dry heated at 60˚C for 2 hours before being de-waxed in Bond Dewax Solution (Leica, #AR9222) at 72˚C for 30 minutes. Slides were then washed in Bond Wash Solution (Leica, #AR9590) for 3x5 minute washes and HIER was performed at 100˚C in Bond Epitope Retrieval Solution 2 (Leica, #AR9640). A further 3x5 minute washes in Bond Wash Solution were conducted both before and after a 5 minute incubation in a Peroxidase Block (Leica, #RE7101-CE) containing 3% Hydrogen Peroxide with stabilisers, which used to neutralise endogenous peroxidases. Slides were incubated with primary antibody at a previously optimised dilution (described in Table 2.2) in a Bond Primary Antibody Diluent (Leica, #AR9352) for 35 minutes. Slides were then washed in Bond Wash Solution for 8x5 minute washes to remove any excess/unbound primary antibody. An EnVision+ Rabbit secondary antibody (Table 2.2) was applied for 30 minutes, before slides were then washed using Bond Wash Solution for 5x5 minute washes. DAB chromogen (Leica, provided exclusively in Bond Intense R Detection System #DS9263) was then applied to the slides for 5 minutes, which were washed in Bond Wash Solution for 3x5 minute washes. The slides were counterstained using Haematoxylin (Leica, provided exclusively in Bond Intense R Detection System #DS9263) and final 5 minute washes in Bond Wash Solution and Deionised Water were performed. After removing the slides from the Autostainer, they were finally dehydrated, cleared, and a coverslip was mounted using DPX Mountant. Digital images were captured on an Aperio AT2 slide scanner at x20 magnification.

2.2.4.2 Immunohistochemical Analysis

The digital images, captured using the Aperio AT2 slide scanner at x20 magnification, were uploaded to the HALO image analysis software (Indica Labs) for quantification of IHC staining. The software was first trained to identify cells based on the presence of the Haematoxylin counterstain in the cell nuclei, and

then trained to identify which of these additionally stained positive for the target antigen, as indicated by the presence of a brown precipitate in the cell. This trained software was then used to analyse images of mammospheres for the percentage of cells that stained positive within the structures.

2.3 Murine Cell Extraction

2.3.1 Mouse mammary epithelial cell (MMEC) extraction

Mammary gland samples were dissected from experimental mice as described in sections 2.1.3 to 2.1.5 and stored in 10ml Ham's F-12 Nutrient Mix (ThermoFisher, #11765-054) supplemented with 1% Penicillin/Streptomycin (ThermoFisher, #15140-122) and 2mM L-glutamine (ThermoFisher, #25030-024) (PSG) in a 50ml Falcon tube (Merck, #CLS430291), stored on ice. For each sample, the tissue was chopped into a fine paste or slurry using a McIlwain Tissue Chopper, transferred into a 50ml Falcon tube containing Ham's F12 supplemented with PSG and 10% Collagenase/Hyaluronidase (Sigma, #C9891/Sigma, #H3506), and incubated at 37°C for 1.5 hours at 100rpm in a shaking incubator. The purpose of this was to dissociate the mammary gland tissue and digest any fat. The dissociated tissue was centrifuged at 350g for 5 minutes. The supernatant was collected in a fresh 50ml Falcon tube and spun down at 350g for 5 minutes before discarding the supernatant. The purpose of this was to remove the digested fat, while also collecting as many MMECs as possible. For each sample, both MMEC pellets were combined into a single Corning 15ml Falcon tube (Merck, #CLS430053) by resuspending together in 5ml of Ham's F12 (PSG), which were subsequently centrifuged at 350g for 5 mins. Red blood cells in the sample pellet were lysed by resuspending in 0.75-1.5ml (specific volume depending upon the size of the pellet) NH_4Cl (0.8%, Sigma-Aldrich, #A9434-500G) and incubating at room temperature for 10 minutes. The reaction was stopped using 5ml of Ham's F12 (PSG) and the lysed red blood cells removed by centrifuging the sample at 350g for 5 minutes and aspirating away the supernatant. To selectively remove stromal cells from the samples, each was plated in a Corning 10cm cell culture dish (Merck, #CLS430167) in 10ml of Ham's F12 medium, supplemented with PSG and 10% foetal bovine serum (FBS, Gibco, #10270-106), and incubated at 37°C for 1 hour. During this time, stromal cells (including fibroblasts) attach to the culture surface, whilst very few of the epithelial cells attach within this timeframe. The unattached

mammary epithelial cells were collected by removing the medium and gently washing them out of the dish using PBS. A cell pellet was collected by centrifugation at 350g for 5 minutes, which was washed with PBS and spun down again to remove any traces of FBS, thus allowing for efficient trypsinisation. Cell pellets were resuspended in 0.5-1ml (volume again dependent on pellet size) of TEG cell dissociation solution composed of: 1X PBS; 0.25% Trypsin solution 10X (Gibco, #15090-046]; 1mM EGTA; and 0.1mg/ml DNase I recombinant, RNase-free (Roche, #04716728001). The cells were incubated in the TEG solution for 10 minutes in a 37°C waterbath to dissociate cells and produce a single cell suspension. Addition of Ham's F12 medium (10% FBS, PSG) was used to stop the enzymatic digestion and each sample was passed through a 40µM cell strainer (Greiner, #542040) to obtain a purified population of single cells. A viable cell count was carried out by Trypan Blue exclusion on a haemocytometer, with a 1:10 dilution of Trypan Blue Stain (ThermoFisher, #T10282). The isolated cells were then used in various analyses.

2.3.2 Mouse thymocyte extraction

Thymus tissue was extracted and stored in 10ml Roswell Park Memorial Institute (RPMI) 1640 Medium (Gibco, #31870-025), supplemented with PSG, in a 50ml Falcon tube and stored on ice. The tissue was chopped up in a 10cm cell culture dish using sterile scalpel blades (Swann-Morton, #0570) and the resulting paste was resuspended in 5ml RPMI medium, supplemented with PSG, FBS, and 2-Mercaptoethanol (Sigma-Aldrich, #63689-100ML-F) at a final concentration 0.05mM. This suspension was filtered through a 70µM cell strainer (Greiner, #542070) into a 15ml Falcon tube. A further 3ml of medium was used to wash away any remaining tissue from the culture dish and filtered through a 70µM cell strainer into the same 15ml Falcon tube. 5ml of Ficoll-Paque Plus (Sigma-Aldrich, #GE17-1440-03) was pipetted into a new 15ml centrifuge tube, and the cell suspension was very gently pipetted on top. This was then centrifuged at 1000g for 10 minutes. The interphase layer, containing live thymocytes, was extracted and washed in 10ml RPMI medium (+PSG, +FBS, +2-Mercaptoethanol) and centrifuged at 250g for 5 minutes. The supernatant was discarded and the thymocytes resuspended in 10ml of culture medium. A viable cell count was carried out by Trypan Blue exclusion on a haemocytometer, with a 1:10 dilution of Trypan Blue

Stain. The cells were then frozen down as pellets of $1x10^6$ cells for use in Western Blot Analysis (2.9).

2.4 Flow cytometry

2.4.1 Antibodies

Table 2.4 List of antibodies used for flow cytometry-based analysis of mammary epithelial cells.

Target	Fluorochrome	Antibody Concentration	Laser	Supplier
Fc (CD16/32)	N/A (Fc block only)	1:100	N/A (Fc block only)	Biolegend, #101319
CD31	Allophycocyanin (APC)	1:200	RED 1 (RL1)	BD Pharmingen, #551262
CD45	APC	1:200	RED 1 (RL1)	BD Pharmingen, 559864
Ter119	APC	1:200	RED 1 (RL1)	BD Pharmingen, #557909
CD24	Fluorescein isothiocyanate (FITC)	1:100	BLUE 1 (BL1)	BD Pharmingen, #553261
CD49f (Integrin alpha 6)	PerCP-eFluor™ 710	1:300	BLUE 3 (BL3)	Thermo Fisher Scientific, #1943473
Dead cells	DAPI	1:1000	VIOLET 1 (VL1)	BD Pharmingen, #564907
RFP	N/A	N/A	YELLOW 1 (YL1)	N/A

2.4.2 Flow cytometry analysis

2.4.2.1 Analysis of Blg-Cre;tdRFP samples

Mammary glands were extracted from virgin *Blg-Cre;tdRFP* mice (of various ages) and pregnant *Blg-Cre;tdRFP* mice (2.1.3) and each sample processed into a single cell suspension of isolated MMECs (2.3.1) that were transferred to Falcon 5 mL Round Bottom polypropylene test tubes (Fisher Scientific, #14-959-11A) for antibody binding and analysis. These samples were analysed for RFP expression within different mammary cell types using flow cytometry analysis. Flow cytometry analysis relies on the specific binding of antibodies (each with different fluorescent labels) to unique regions in specific cellular subsets. However, the Fragment crystallisable (Fc) region of the antibody interacts with a variety of Fc receptors that are present on the cell surface of a wide range of immune cells, and can therefore result in some non-specific binding. This unwanted, or non-specific binding, to Fc receptors was prevented by saturating the receptors with an anti-Fc receptor antibody prior to staining with labelled antibodies. The Fc receptor antibody used was a TruStain FcX™ (anti-mouse CD16/32) Antibody (Biolegend, #101319), which was diluted 1:100 in FACS buffer (composed of PBS, 2mM EDTA pH 8, and 1% FBS) and incubated with cells on ice for 10 minutes. Cells were then washed with 1ml of FACS buffer, centrifuged at 350g for 5 minutes, and the supernatant was removed. MMECs were then labelled with a mastermix containing various anti-mouse antibodies, diluted in FACS buffer to their optimum concentrations, listed in Table 2.4 (2.4.1). Cells were covered and incubated on ice for 30 minutes. Cells were again washed with 1ml of FACS buffer and centrifuged at 350g for 5 minutes to remove the supernatant. Cells were suspended in an appropriate volume of FACS buffer and the DAPI antibody, also listed in Table 2.4 (2.4.1), was added to the samples 5 minutes prior to analysing using the Attune NxT Flow Cytometer (Invitrogen), using the indicated laser settings in Table 2.4 (2.4.1).

For each experiment, in addition to the test samples there were also various controls, including single antibody-labelled controls using UltraComp eBeads™ Compensation Beads (Invitrogen, #01-2222-42) which were washed and resuspended in an eBioscience™ Flow Cytometry Staining Buffer (Invitrogen, #00-4222-26) after staining and prior to analysis. Unstained and fluorescence minus

one (FMO) control samples were used for each experiment in order to appropriately set the gates for analysis. Samples were then run through the Flow Cytometer and FlowJo software was used to analyse the data obtained. MMECs were analysed using a series of gating steps. The first of these involved excluding any cell debris using the SSC-A vs FSC-A axis settings (Figure 2.2A). Cell doublets were then excluded using FSC-H vs FSC-A (Figure 2.2B) and SSC-A vs SSC-H (Figure 2.2C) axis settings. Any dead/DAPI-positive cells were then removed (Figure 2.2D), as were haematopoietic (CD45- and TER119-positive) and endothelial (CD31-positive) cells, collectively referred to as lineage positive (Lin^+) cells (Figure 2.2E). The remaining Lin^- cells were analysed for CD24 and CD49f expression (Figure 2.2F) to separate them into luminal ($Lin^-CD24^+CD49f^{low}$) and basal/myoepepithelial ($Lin^-CD24^+CD49f^{high}$) subpopulations. Each of these populations was analysed for their respective RFP expression levels (Figure 2.2G and Figure 2.2H), expressed as a percentage of the total numbers of luminal or basal cells being analysed.

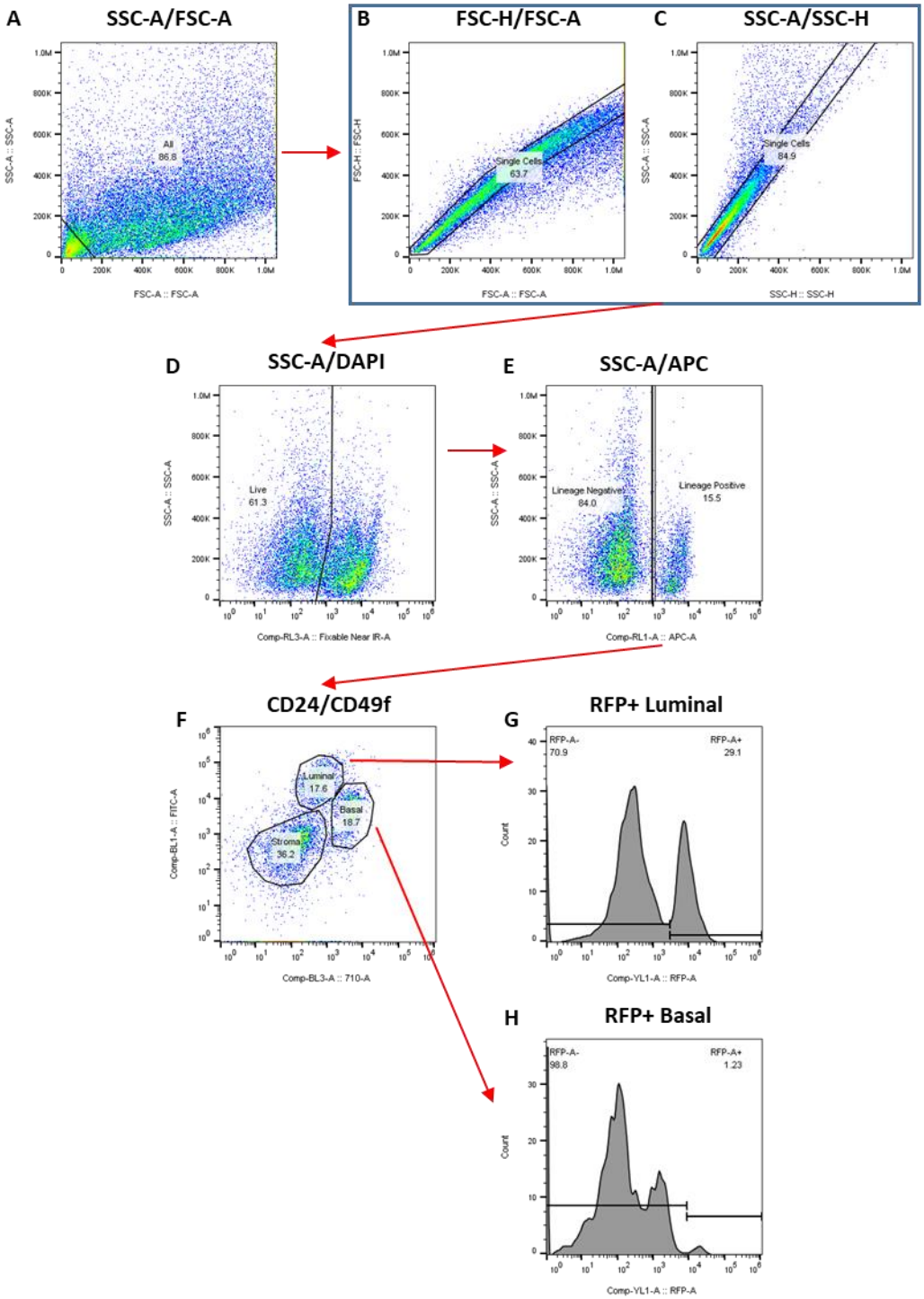


Figure 2.2 Gating strategy for flow cytometry analysis of *Blg-Cre;tdRFP* mammary epithelial cells.

Mammary epithelial cells from *Blg-Cre;tdRFP* mice were analysed by flow cytometry using the gating strategy shown. Forward and side scatter settings were used to exclude cell debris (SSC-A/FSC-A) [A] and cell doublets (FSC-H/FSC-A and SSC-A/SSC-H) [B and C]. Any dead/DAPI-positive cells were then removed from analysis (SSC-A/DAPI) [D], as were Lin^+ ($CD31^+/CD45^+/Ter119^+$) cells (SSC-A/APC) [E]. The remaining Lin^- cells were then separated into luminal ($Lin^-CD24^+CD49f^{low}$), basal/myoepepithelial ($Lin^-CD24^+CD49f^{high}$), and stromal ($Lin^-CD24^-CD49f^{low}$) subpopulations [F]. Each of these populations was analysed for their respective RFP expression levels [G and H].

2.4.2.2 RFP Analysis of MMTV-Cre;MMTV-PyMT;tdRFP samples

Mammary samples were extracted from MMTV-Cre;tdRFP and MMTV-Cre;MMTV-PyMT;tdRFP mice at various timepoints (2.1.5) and each was processed into a single cell suspension of isolated MMECs (2.3.1). Cells were suspended in an appropriate volume of FACS buffer, transferred into a Falcon 5 mL Round Bottom polypropylene test tube and the DAPI antibody, the details of which are in Table 2.4 (2.4.1), was added to the samples 5 minutes prior to analysing using the Attune NxT Flow Cytometer (Invitrogen), using the indicated laser settings in Table 2.4 (2.4.1).

For each experiment, in addition to the test samples, MMECs were also extracted from a mouse not expressing *tdRFP*, to use for an unstained control and a DAPI only control. The purpose of these samples was to appropriately set the gates for the analysis of data from each experiment. Samples were then run through the Flow Cytometer and FlowJo software was used to analyse the data obtained. MMECs were analysed using a series of gating steps. The first of these involved excluding any cell debris using the SSC-A vs FSC-A axis settings (Figure 2.3A). Cell doublets were then excluded using FSC-H vs FSC-A (Figure 2.3B) and SSC-A vs SSC-H (Figure 2.3C) axis settings. Any dead/DAPI-positive cells were then removed (Figure 2.3D), and the remaining live cells were analysed for their RFP expression (Figure 2.3E) expressed as a percentage of the total number of live cells).

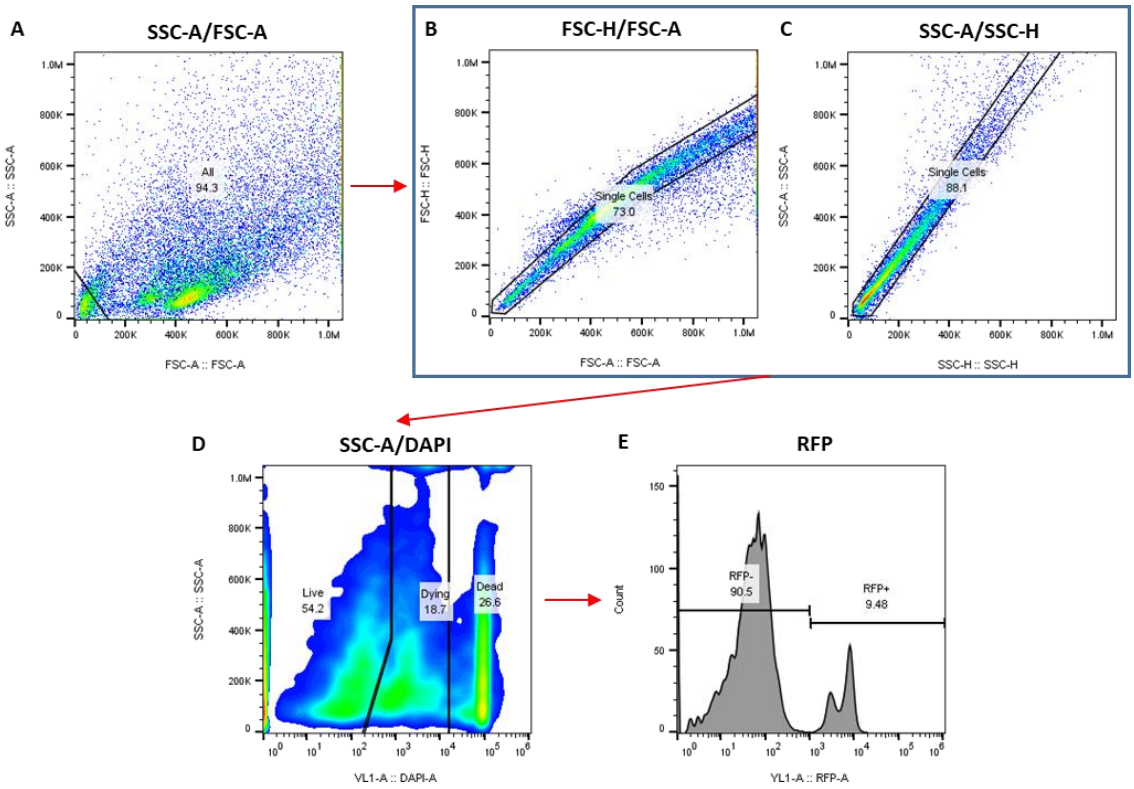


Figure 2.3 Gating strategy for flow cytometry analysis of *MMTV-Cre;tdRFP* and *MMTV-Cre;MMTV-PyMT;tdRFP* mammary epithelial cells.

Mammary epithelial cells from *MMTV-Cre;tdRFP* and *MMTV-Cre;MMTV-PyMT;tdRFP* mice were analysed by flow cytometry using the gating strategy shown. Forward and side scatter settings were used to exclude cell debris (SSC-A/FSC-A) [A] and cell doublets (FSC-H/FSC-A and SSC-A/SSC-H) [B and C]. Any dead/DAPI-positive cells were then removed from analysis (SSC-A/DAPI) [D] and the live cells were analysed for RFP expression levels [E].

2.4.3 Flow-assisted cell sorting (FACS)

Mammary glands were extracted from 9 week-old (pre-neoplastic) virgin *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;tdRFP* mice (2.1.4) and each sample processed into a single cell suspension of isolated MMECs (2.3.1). Cells were suspended in an appropriate volume of FACS buffer and the DAPI antibody, listed in Table 2.4 (2.4.1), was added to the samples 5 minutes prior to sorting. Samples were sorted using a BD FACSAria™ III Cell Sorter, with the indicated laser settings detailed in Table 2.4 (2.4.1).

For each experiment, in addition to the test samples, MMECs were also extracted from a mouse not expressing *tdRFP*, to use for an unstained control and a DAPI only control. The purpose of these samples was to appropriately set the gates to

set up accurate parameters and sort gates for each experiment. MMECs were sorted using the same gating strategy described in Figure 2.3. Sorting ensured the exclusion of cell debris using the SSC-A vs FSC-A axis settings (Figure 2.3A). Cell doublets were also excluded using FSC-H vs FSC-A (Figure 2.3B) and SSC-A vs SSC-H (Figure 2.3C) axis settings. Any dead/DAPI-positive cells were then removed (Figure 2.3D), and the remaining live cells were sorted based on RFP expression (Figure 2.3E), where only RFP-positive cells from each sample were collected in a 5ml Falcon® Round-Bottom Polypropylene Tube containing 2ml of MMEC culture media. Cells were washed with PBS before resuspending in 100µl Lysis Buffer from the RNAqueous™-Micro Total RNA Isolation Kit (Invitrogen, #AM1931) and storing at -80°C. RNA was subsequently extracted and samples were analysed by RNA sequencing, as described in 2.5.

2.5 RNA sequencing (RNA-Seq)

2.5.1 RNA extraction

RNA was extracted from RFP-sorted MMECs, isolated from 9wks-old *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;tdRFP* mice, stored at -80°C in Lysis Buffer (described in 2.1.4, 2.3.1, and 2.4.3). RNA extractions were performed by Dimitris Athineos, a colleague in Y35, using the RNAqueous™-Micro Total RNA Isolation Kit (Invitrogen, #AM1931) according to the manufacturer's protocol, provided with the kit. Residual genomic DNA was removed using the DNA-free™ system, also included in the kit. Several lysed MMEC samples were combined together into a single 20µl RNA sample by processing them through the same spin column. The MMECs isolated from individual mice that were combined together to create the final pilot and additional samples are detailed in Tables 3.1 and 3.3 (Chapter 3), respectively. To combine MMECs from several mice into a single RNA sample, 50µl of ethanol was added to each sample (according to the first step of the extraction protocol) and then passing each the lysate/ethanol mixture through a single Micro Filter Cartridge Assembly, one sample at a time, until RNA from all samples were bound to the same Micro Filter Cartridge for further processing. A NanoDrop 2000 Spectrophotometer was used to quantify and assess the purity of the extracted RNA prior to RNA-seq analysis.

2.5.2 Library preparation, RNA-seq and analysis

The RNA samples, extracted as described in 2.5.1, were given to Billy Clark in the Molecular Technology Services for quality assessment and sequencing. RNA quality was first assessed using the RNA integrity number (RIN) obtained from analysis using the Agilent 2200 TapeStation System with an RNA ScreenTape Analysis. A higher RIN value is indicative of a higher degree of RNA integrity, and levels of 7.5 or above are generally accepted as sufficient for library construction. Samples with RNA values exceeding 7.5 were chosen for RNA-Seq wherever possible, however it was necessary to analyse some samples with RIN values ~3 (particularly for the extended analysis – see 3.2.1 and 3.2.2) in order to obtain at least N=3 per test group, and N=2 for the BCAT_2 cohort. RNA-seq libraries were prepared using the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio, #634411). The quality of these libraries assessed on an Agilent 2200 TapeStation System with a High Sensitivity D1000 ScreenTape Analysis (Agilent, #5067-5584 and #5067-5585), and the quantity was measured using a Qubit Fluorometer (Invitrogen). Libraries were sequenced using the Illumina NextSeq 500 System with paired-end (2x36) sequencing using the MiniSeq High Output Reagent Kit (75-cycles, Illumina, #FC-420-1001). The raw RNA-Seq data was processed by Robin Shaw, a trained bioinformatician in the Computational Biology Department of the CRUK Beatson Institute. The raw files were checked for quality using FastQC and FastQ Screen. The paired-end reads were then aligned to mouse genome build GRCm38.98 using TopHat2 with Bowtie2. Expression levels of each gene were determined and statistically between each of the test groups and their control group. This was achieved using a combination of packages and analysis suites in Python (HTSeq package) and R Studio (Bioconductor and DESeq2) that were specifically developed for the analysis of high-throughput sequencing data and for differential gene expression analysis based on the negative binomial distribution. Subsequent Gene Set Enrichment Analysis (GSEA) was also performed by Robin Shaw. The resulting processed data was subsequently analysed by myself, as described in Chapter 3, and the MetaCore analysis suite was additionally used to interpret some of this processed data.

2.6 Cell Culture

2.6.1 T6i lymphoma cells

T6i cells were derived from a mouse T-cell lymphoma, which expressed high levels of endogenous *Runx1* due to proviral insertional mutagenesis (Wotton, Stewart et al. 2002). They were used for this thesis as a known positive control for murine RUNX1 protein expression, and were a kind gift from the Cameron lab. T6i cells are suspension cells, and were maintained in an incubator at 37°C and 5% CO_2. Cells were maintained in RPMI 1640 medium supplemented with 10% FBS, PSG, and 2-Mercaptoethanol at a final concentration of 0.05mM. Cells were passaged at a 1:10-1:15 dilution when they reached ~80-90% confluence.

2.6.2 HC11 mouse mammary epithelial cells

The HC11 cell line (ATCC, #CRL-3062) is an immortalised cell line originating from the normal mammary gland of a pregnant Balb/c mouse (Danielson, Oborn et al. 1984, Ball, Friis et al. 1988). It is an adherent cell line that is predominantly composed of luminal stem- or progenitor-like epithelial cells. HC11 cells were grown in T75 flasks at 37°C and 5% CO_2. They were maintained in RPMI 1640 medium supplemented with 10% FBS, PSG, Insulin solution (Sigma-Aldrich, #I9278-5ML) to a final concentration of 5µg/ml, and Mouse Recombinant Epidermal Growth Factor (EGF, STEMCELL Technologies, #78016.1) to a final concentration of 10ng/ml. Cells were provided with fresh medium every 2-3 days and passaged upon reaching ~80-90% confluence. To split cells, they were trypsinised in 2ml of Trypsin/EDTA at 37°C for ~3-4 minutes and passaged at a 1:15-1:20 dilution.

2.6.3 HEK293T embryonic human kidney cells

HEK293T cells (ATCC, #CRL-3216) are immortalized human embryonic kidney cells, derived from the HEK293 parent cell line. These cells are commonly used for retroviral production due to their ability to produce recombinant proteins within plasmid vectors containing the SV40 promoter, which is due to the incorporation of the SV40 large T antigen in their genome. HEK293T cells are semi-adherent in nature and require growth conditions of 37°C and 5% CO_2. They were maintained in Dulbecco's Modified Eagle Medium (DMEM, Gibco, #21969-035) supplemented with 10% FBS and PSG. Cells were provided with fresh medium every 2-3 days and

passaged upon reaching ~80-90% confluence. They were used for retrovirus and lentivirus production as described in 2.7 and 2.8. To split cells, they were trypsinised in 2ml of Trypsin/EDTA at 37°C for ~2 minutes and passaged at a 1:10-1:15 dilution.

2.6.4 Freezing Cell lines

To freeze cells, pellets of at least $1x10^6$ cells were each resuspended in 1ml of freezing medium, composed of 40% culture medium, 10% DMSO, and 50% FBS. These 1ml suspensions were each transferred into 1.8ml cryogenic tubes (Thermo Scientific, #377267) and frozen in a -80°C freezer in a CoolCell Freezing Container (Corning, #432001) for 24 hours before transferring the vials to liquid nitrogen for long-term storage.

2.7 *Runx1* Overexpression

2.7.1 pBABE-Puro vector construct for *Runx1* overexpression

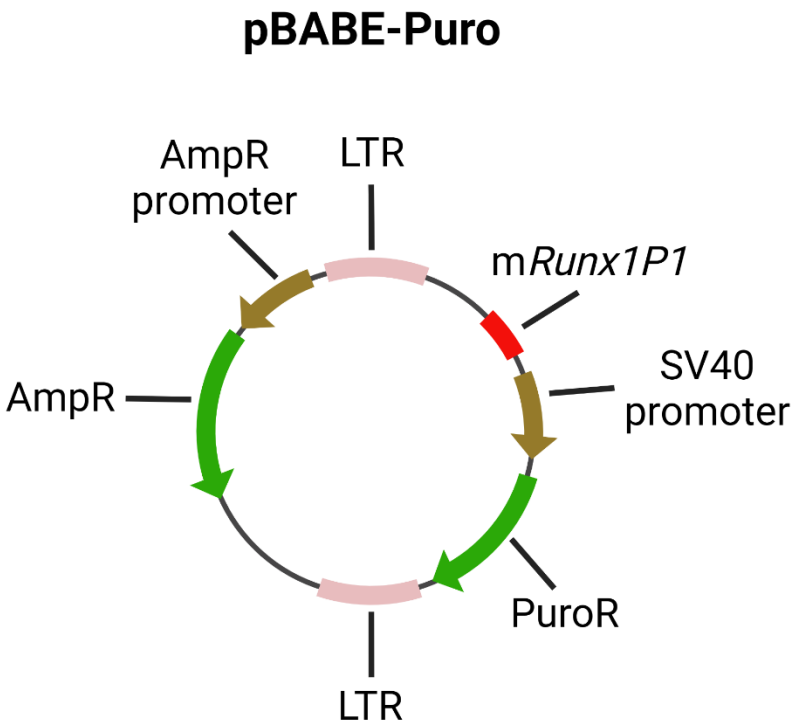


Figure 2.4 pBABE-Puro construct used to induce Runx1P1 overexpression.

Schematic representation of pBABE-Puro construct, containing a mouse Runx1P1 sequence. Plasmid construct was used to made retrovirus that was used for viral transduction in HC11 cell line.

2.7.2 Transformation of Stable Competent Escherichia coli (E.coli) with pBABE-Puro vector

In order to replicate the pBABE-Puro constructs (Empty and *mRunx1P1*) for use in overexpression assays, NEB® Stable Competent E. coli (NEB, #C3040H) was transformed with the constructs by following the High Efficiency Transformation Protocol provided with the kit. Transformed colonies were selected for by growing on Lysogeny Broth (LB) agar plates supplemented with 100µg/ml Ampicillin (Amp), made in-house by the CRUK Beatson Institute Central Services. Starter cultures were made by transferring a colony into a universal tube (Fisher Scientific, #07-000-475) containing 5ml of selective LB broth (+100µg/ml Amp (Sigma-Aldrich, #A5354-10ML)) and incubating in a shaking incubator at 37°C, at 225rpm, for several hours until turbid. 1ml from each transformation was transferred into its own respective conical flask with ~50ml selective LB broth (+100µg/ml Amp) and incubated overnight in a shaking incubator at 37°C, at 225rpm, until turbid. The bacterial cells were collected by centrifugation for 30 minutes at 2000g at 4°C and by pouring off the supernatant. The cell pellet was then given to Molecular Technology Services for Maxipreps, to isolate the plasmid construct DNA from the transformed bacteria. This was performed using a PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen, #K2100-17).

2.7.3 Overexpression of Runx1 using retroviral infection of pBABE-Puro constructs

2.7.3.1 Retrovirus production

HEK293T cells were transfected with the experimental plasmid constructs (encoding the altered viral genome), along with packaging and envelope plasmids (encoding the components of the viral capsid and envelope). The purpose of this was to produce retrovirus for the pBABE-Puro-mRunx1P1 construct (and pBABE-Puro Empty Vector as a control) to transduce HC11 cells with and induce Runx1 overexpression within the cell line. HEK293T cells were first split into new flasks in preparation for their transfection as they should ideally be ~50% confluent when transfected. Cells were transfected the following day using a calcium phosphate method. For each viral construct, the DNA was prepared for the transfection by combining 10µg of pBABE-Puro plasmid (mRunx1P1-expressing construct or empty vector), 7.5µg of psPAX2 (packaging vector, Addgene, #12260), and 4µg of pVSV-G (envelope protein, Addgene, #138479). This DNA mix was then made up to a

total volume of 440µl using sterile H_2O. 500µl of 2X HEPES Buffered Saline (HBS, Sigma-Aldrich, #51558-50ML) was added and mixed in thoroughly before adding 60µl of 2M $CaCl_2$ (Sigma-Aldrich, #C7902-500G) and mixing again using a pipette. This transfection mixture was incubated at 37°C for 30 minutes, mixed by pipetting up and down several times with a P1000, and gently added drop by drop to the flask containing HEK293T cells, which was gently rotated/swirled to disperse the DNA mix thoroughly. The cells were incubated for 3 days at 37°C and 5% CO_2. Medium was replaced after 24 hours with fresh medium, and the virus-containing medium was removed from the HEK293T cells on days 2 and 3, filtered through a 0.45µM filter (Fisher Scientific, #15216869) using a 10ml syringe (Becton Dickinson, #302995), and the filtrate collected in a Falcon tube. Fresh medium was added to the HEK293T cells on day 2 to allow for an additional virus harvest on day 3. Viral titres for days 2 and 3 were combined together at the end. The virus produced was then used to infect HC11 cells.

2.7.3.2 Retroviral infection (transduction)

The HC11 cells that were being transduced were split into a fresh flask 24 hours prior to infection so that they were be ~30-40% confluent on the day of infection. On the day of transfection, the HC11 flasks were brought into CATII, along with the retroviruses and Hexadimethrine bromide (aka Polybrene, Sigma-Aldrich, #TR-1003-G). Polybrene was added to each flask of cells to a final concentration of 10µg/ml. Flasks were swirled to mix and incubated for ~15 minutes in a 37°C incubator. While the cells were incubating with Polybrene, the infection medium was prepared by mixing ~4-6ml of collected virus to equal parts cell culture medium, with Polybrene added to a final concentration of 10µg/ml. The HC11 cells were removed from the incubator and the Polybrene-containing medium was removed from the flasks and replaced with viral medium. Cells were incubated with viral medium overnight at 37°C and 5% CO_2 before replacing with culture medium supplemented with Polybrene to a final concentration of 10µg/ml. After culturing for a further 24 hours, transduced HC11 cells were selected by culturing with fresh medium supplemented with an appropriate concentration of the relevant antibiotic. HC11 cells required 1.5µg/ml of Puromycin (InvivoGen, #ant-pr-1) for efficient antibiotic selection, as was determined in pilot experiments. Puromycin treatment of a non-transduced HC11 cell line was used as a control to

determine at which point the untransduced cells would likely all have been killed by this antibiotic selection. Cells were continuously grown in antibiotic for up to 2 weeks before performing an EnzChek assay to remove cells from CATII (2.7.3.3). Cells were then immediately analysed for protein expression (2.9) and various growth assays performed (2.12 to 2.14).

2.7.3.3 EnzChek Assay to confirm virus clearance

To assess whether cells generated in Containment Level 2 (CL2) could be moved to Containment Level 1 (CL1), an EnzChek Reverse Transcriptase Assay Kit (Invitrogen, #E22064) was used to confirm whether there was complete clearance of virus from the transduced cells. Media was collected from the transduced cell lines, and from the parental/non-transduced cell line to be used as a negative control. Both were spun down at 350g for 5 minutes to remove any dead cells/debris and several mls of supernatant was collected. Samples were then set up in an EnzChek assay according to the manufacturer's protocol, using dilution series of 6 known amounts of reverse transcriptase (15U, 1.5U, 0.5U, 0.15U. 0.1U and 0U) to generate a standard curve for comparison. Sample fluorescence was measured using a Tecan Safire 2 multi-detection microplate reader, using the EnzChek assay settings to analyse at the standard fluorescein wavelengths (excitation ~480 nm, emission ~520 nm) and calculate the normalised levels of reverse transcriptase activity relative to the standard curve generated from the reverse transcriptase dilution series. The resulting data was sent to the CRUK Beatson Institute's Health & Safety Officer to verify that there was indeed clearance of the virus from the cells.

2.8 CRISPR-Cas9-mediated *Runx1* and/or *Runx2* deletion

2.8.1 pX-459-Puro/lentiCRISPRv2-Neo vector constructs for *Runx1* or *Runx2* deletion

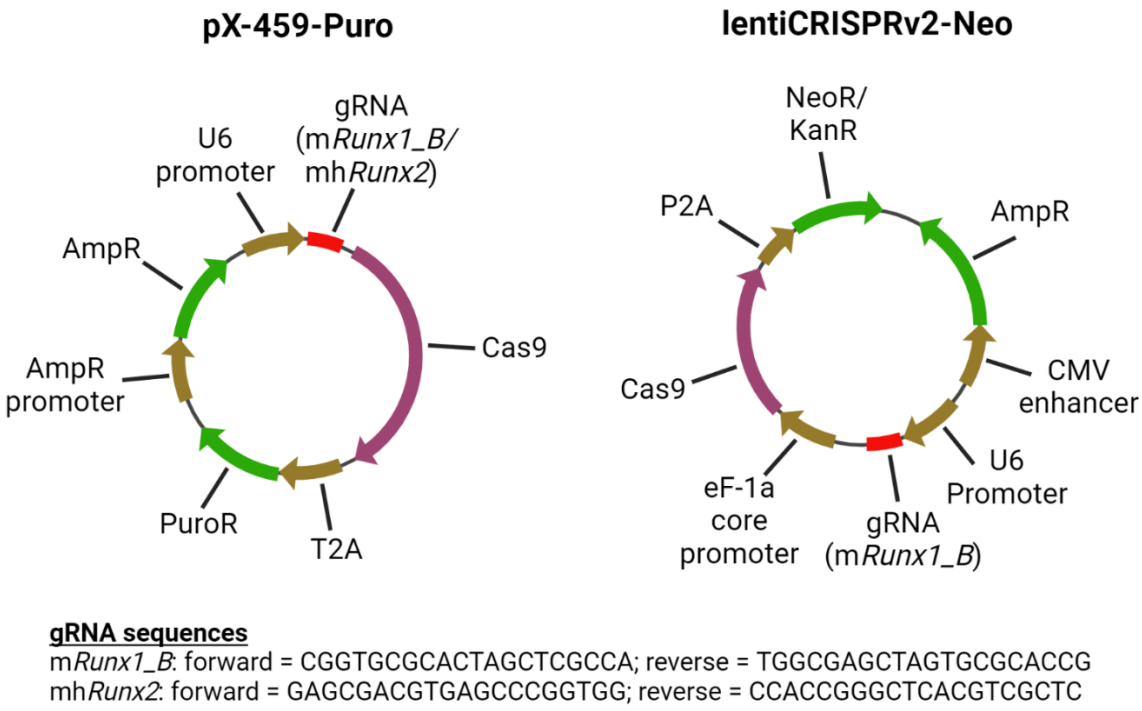


Figure 2.5 Vector constructs used for the CRISPR-Cas9-mediated deletion of Runx1 and/or Runx2 in HC11 cells.

Schematic representations of the pX-459-Puro and lentiCRISPRv2-Neo vector constructs used for the depletion of Runx1 and/or Runx2 expression in HC11 mammary cells. pX-459-Puro construct transfected, while lentivirus made using lentiCRISPRv2-Neo that was used for the transduction of cells.

2.8.2 Transformation of Stable Competent E.coli with pX-459-Puro/lentiCRISPRv2-Neo vectors

The various pX-459-Puro (empty vector, or with either *Runx1*- or *Runx2*-targeting gRNA) and lentiCRISPRv2-Neo (empty vector, or with *Runx1*-targeting gRNA) constructs were replicated using NEB® Stable Competent E. coli cells, and plasmid DNA extracted using the protocol detailed in 2.7.2.

2.8.3 Transfection of pX-459-Puro vector constructs into HC11 cells to induce *Runx1* or *Runx2* deletion

24 hours prior to transfecting HC11 cells with *Runx1*- or *Runx2*-targeting plasmid constructs (each generated from a modified pX-459-Puro plasmid), HC11 cells were seeded at an appropriate density in 6-well culture plates (Corning, #353046) so that they would be ~30-40% confluent at transfection. A Lipofectamine™ 2000 Transfection Reagent (Invitrogen, #11668-019) was used to transfect cells with these plasmid constructs. For each CRISPR construct being transfected, 36µl of Transfection Reagent was diluted in 450µl of Opti-MEM medium (Gibco, #31985-062), or any other serum-free medium available. 14µg of each plasmid construct was also diluted in 700µl of medium, in a separate Eppendorf tube. For each CRISPR construct, 450µl of DNA mix was combined with 450µl of Lipofectamine 2000 mix. These were then incubated at room temperature for ~15 minutes. 250µl of DNA/Lipofectamine mix was then added (drop-by-drop) to each of the wells containing the HC11 cells to be transfected. Cells were incubated at 37°C for 24 hours before the medium was replaced with fresh culture medium supplemented with 1.5µg/ml of Puromycin. Successfully transfected cells were selected for using the antibiotic selection method described in 2.7.3.2 before single-cell cloning was performed for CRISPR lines (2.8.5) to ensure complete/homogenous gene knockout.

2.8.4 Deletion of *Runx1* using lentiviral infection of lentiCRISPRv2-Neo construct

Lentivirus was made for the lentiCRISPR-Neo-mRunx1_B construct (and its Empty Vector counterpart as a control) for CRISPR-induced knockout of Runx1. For lentivirus production and transduction of HC11 cells by infecting with the virus produced, the methods used for pBABE-Puro retrovirus production and infection (detailed in 2.7.3) were applied. The same protocol was followed, with the exception of the antibiotic used for the selection of successfully transduced cells. HC11 cells required treatment with 800ug/ml of G418 Sulphate/Geneticin™ Selective Antibiotic (an analogue of Neomycin sulphate, ThermoFisher, #10131035) for 2-3 weeks to ensure efficient antibiotic selection, as was determined in pilot experiments. An EnzChek assay was used to confirm clearance of the virus from the HC11 cells, so they could be moved back to CL1 (2.7.3.3).

Single-cell cloning was performed (2.8.5) in order to ensure that relatively homogeneous HC11 cell lines, with complete *Runx1* knockout, were produced.

2.8.5 Single-cell cloning of *Runx1* and/or *Runx2* CRISPR cell lines

Transduction and transfection of the HC11 cells was not necessarily 100% efficient. If the loss of *Runx1* and/or *Runx2* did not provide a selective advantage to the cells, the untransduced/untransfected cells within the cell populations may have eventually outgrown any genetically altered cells. It was therefore necessary for a single-cell cloning method to be used, whereby the transfected cells were diluted down to such an extent that single cells were isolated in a well. If one of these single cells plated was a CRISPR-altered cell, any cell lines that arose from this original cell would also exhibit CRISPR-mediated gene deletion, and therefore the transduced/transfected cell could be grown up into a homogenous *Runx1*- and/or *Runx2*-depleted cell line. In order to dilute the cells enough to obtain single-cell clones, the selected transduced/transfected HC11 cells needed to be serial-diluted in a 96-well culture plate (Corning, #353072). Cells from each CRISPR cell line (and their empty vector control) were trypsinised, collected, and diluted to $2x10^4$ cells/ml. This suspension of cells was serially diluted at a 1:10 dilution each time, initially down the first column and then along each of the rows. The plates were incubated (37°C 5% CO_2) and wells containing single cells were identified the following day before returning to the incubator. Single cells were monitored for growth and given fresh medium every 2-3 days. Once the HC11 cells became confluent, they were trypsinised and all cells were transferred into the well of a 24-well culture plate (Corning, #353047). This was continued, with cells being trypsinised and transferred into gradually larger plates/flasks (to 6-well plates (Corning, #353046), then T25 flasks (Corning, #353109), then T75 flasks (Corning, #353136)) once they become confluent. After cells had grown to sufficient numbers for analysis and assay set-up, cell pellets were taken for Western Blot analysis (2.9) and sequencing (2.10) to confirm knockout before setting up in various experiments (2.12 to 2.14). *Runx1* loss was determined by protein expression (Western Blot) analysis (5.2.1) while loss of *Runx2* expression was confirmed using both Western Blot (5.2.7) and sequencing (Appendix) methods.

2.9 Western blot analysis

2.9.1 Protein extraction

Protein was extracted from cell pellets containing ~1x10^6 cells each. First, a lysis and extraction buffer mastermix was made by diluting a 100X Halt™ Protease Inhibitor Cocktail (Thermo Scientific, #78429) in a RIPA Lysis and Extraction Buffer (Thermo Scientific, #89900). Each cell pellet was resuspended in 100µl of the lysis and extraction buffer mastermix. The samples were stored on ice for ~30 mins and vortexed every few minutes to ensure efficient cell lysis. Lysates were centrifuged at 19,000g on a bench-top centrifuge for 30 mins at 4°C and the supernatant was transferred into a fresh 1.5ml Eppendorf tube. The protein was quantified prior to setting up a Western Blot or freezing.

2.9.2 Protein quantification

Protein samples were quantified using a Pierce™ BCA Protein Assay Kit (Thermo Scientific, #23225), with Pierce™ Bovine Serum Albumin Standard Ampules (Thermo Scientific, #23209) used to generate a standard curve for relative comparisons of protein concentrations. The 2000µg/ml stock was diluted in sterile water to get BSA standards of 80,100, 200, 400, 1000 µg/ml. The 2000µg/ml stock was also used neat as a protein standard. 10µl of each standard solution, and sterile water as a blank, were pipetted in triplicate into a 96-well plate. Each sample was diluted 1:10 using sterile water before being pipetted into the same 96-well plate in triplicate. A BCA protein quantification solution was made up by diluting Reagent B 1/50 in Reagent A. 200µl of the BCA solution was added into each well of the plate, which was then incubated at 37°C for 30 minutes.

A SpectraMax ABS Plus Microplate Reader was used to measure the absorbance at 595nm in each sample. The SoftMax Pro Software plotted the standard curve, using the absorbance measurements from the BSA standards, and then calculated the concentrations of cell lysates from these standard curves. These protein lysate concentrations were used to calculate the volume of sample to be loaded for Western Blot Analysis.

2.9.3 Western blot protocol

2.9.3.1 Antibodies

Table 2.5 List of antibodies used for Western Blot analyses.

Primary Antibody Target	Primary Antibody Name/Supplier	Primary Antibody Concentration	Secondary Antibody Name/Supplier	Secondary Antibody Dilution
GAPDH (loading control)	GAPDH (14C10) Rabbit mAb (HRP Conjugate) Cell Signaling Technology #3683	1:1000	Not required (Primary antibody is HPR conjugated)	N/A
RUNX1	AML1 (D4A6) Rabbit mAb (Mouse Preferred) Cell Signaling Technology #8529	1:1000	Anti-rabbit IgG, HRP-linked Antibody (Cell Signaling Technology #7074)	1:10,000
RUNX2	RUNX2 (D1H7) Rabbit mAb Cell Signaling Technology #8486	1:1000	Anti-rabbit IgG, HRP-linked Antibody (Cell Signaling Technology #7074)	1:10,000

2.9.3.2 Sample Preparation and gel electrophoresis

The volume of lysate to be loaded into each well was calculated using the relative concentrations of the lysates obtained from the BCA assay. 20µg of protein from each sample was loaded into the gel for analysis by Western Blot. To prepare each sample for analysis, a mastermix of lysate, sample reducing agent, sample buffer, and sterile water was made up to a total volume of 20µl, as this was the maximum loading volume for the precast gels used. A volume of lysate corresponding to 20µg of protein was transferred into a clean Eppendorf tube. 2µl of 10X NuPAGE™ Sample Reducing Agent (Invitrogen, #NP0004) was added to obtain a final 1X concentration. 5µl of 4X NuPAGE™ LDS Sample Buffer (Invitrogen, #NP0007) was then added to obtain a final 1X concentration. The sample was then made up to the final 20µl volume using sterile water. Samples were heated on a heat block at ~90˚C for 5-10 minutes before they were vortexed to thoroughly mix them and centrifuged to collect the samples at the bottom of their tubes. Samples were then loaded into precast NuPAGE™ 4-12% Bis-Tris 1.0mm Mini Protein Gels (Invitrogen, #NP0321BOX or #NP0322BOX) along with a Full range Amersham™ ECL™ Rainbow™ Marker (Cytiva, #GERPN800E) as a molecular weight marker. Gel

electrophoresis was carried out at 180V at maximum current, in a 20X NuPAGE™ MOPS SDS Running Buffer (Invitrogen, #NP0001) diluted to 1X in sterile water, until the MW Marker reached the bottom of the gel.

2.9.3.3 Protein Transfer

Once the lysates were each run on the gel and separated into their constituent proteins based on their molecular weights, the proteins were transferred onto a Nitrocellulose membrane. A wet transfer was performed within 20X NuPAGE™ Transfer Buffer (Invitrogen, #NP0006-1), diluted 1X in water and methanol (to a final methanol concentration of 20%). The gel and Nitrocellulose membrane were sandwiched between stacks of filter papers and sponges (as shown in Figure 2.6), all of which were equilibrated prior to sandwich assembly by soaking in 1X Transfer Buffer for 5 minutes. The transfer cassette that this sandwich was assembled in was inserted into the transfer tank, filled with 1X Transfer Buffer, and electrophoresed for 1 hour 40 minutes at full current and 30V.

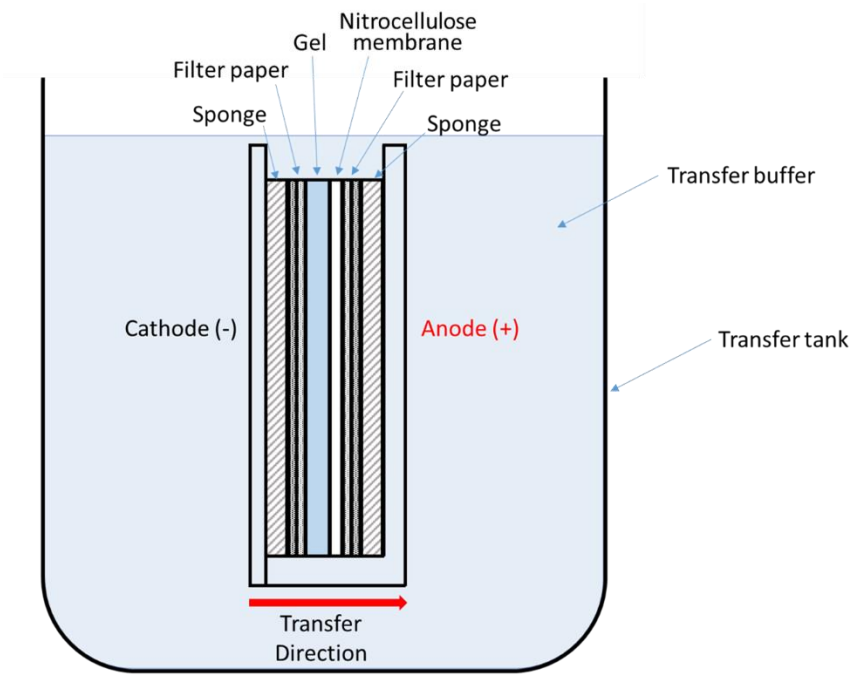


Figure 2.6 Western Blot Transfer setup.

Annotated diagram depicting the setup used for the transfer of proteins from a gel to a Nitrocellulose membrane.

2.9.3.4 Immunoblotting and imaging

A wash buffer was prepared by diluting a 10X Tris buffered saline (TBS) solution containing Tween (TBS-T), made up by the CRUK Beatson Institute's Central Services, to a 1X working concentration. Following protein transfer, the Nitrocellulose membrane was blocked in 5% non-fat milk (Marvel)/TBS-T blocking buffer at room temperature for 1 hour. The primary antibody was diluted to the desired concentration in 5ml of 5% milk/TBS-T blocking buffer. The membrane was transferred into a 50ml Falcon tube along with the primary antibody solution and incubated overnight in a 4˚C cold room on a tube roller. Following this, the membrane was transferred into an incubation box and washed using 3x5 minute washes of TBS-T to remove any excess or unbound primary antibody. The appropriate secondary antibody was then made at a 1:10,000 concentration and the membrane was incubated with this at room temperature for 1 hour on an orbital shaker. 3 more 5 minute washes were performed in TBS-T. A SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Scientific, #34095) was used for signal development. This substrate kit contains two separate components, which were mixed together 1:1. The Nitrocellulose membranes were soaked with the chemiluminescent substrate for 1 minute before being transferred into a clear plastic sleeve and chemiluminescent and colorimetric images were taken of each blot using the ChemiDoc Imaging System.

2.9.3.5 Stripping membrane for re-blotting

In order to probe a single blot for multiple similarly sized proteins, the Nitrocellulose membrane was stripped between each immunoblot. After blotting with an antibody corresponding to one of the proteins of interest and capturing images, the membranes were stripped of any bound antibody using a 10X ReBlot Plus Mild Antibody Stripping Solution (Millipore, #2502). This Stripping Solution as diluted to a 1X solution in sterile water and the membrane was incubated with the solution in an incubation box for 15 minutes at room temperature on an orbital shaker. The membrane was then washed using 3x5 minute TBS-T washes and re-blocked for 1 hour at room temperature in 5% milk/TBS-T blocking solution. The membrane could then be probed for different proteins using the same immunoblotting method as described in 2.9.3.4.

2.9.4 Western blot analysis

Images taken on the ChemiDoc Imaging System were uploaded to the Bio-Rad Image Lab software for analysis. For each blot being analysed, the chemiluminescent and colorimetric images were merged in order that the molecular weight of the protein bands (in the chemiluminescent image) could be determined using the protein ladder (in the colorimetric image). Relative densitometry analysis was also performed using the Image Lab software in order to compare protein abundance between samples.

2.10 Confirmation of *Runx2* knockout

2.10.1 T7 Endonuclease I mismatch assay

Results generated from Western Blot Analysis of *Runx2* CRISPR clones did not provide a definitive answer as to whether there was successful or complete knockout of *Runx2* and its corresponding protein. Farah Hughes, a colleague in the Strathdee lab at the CRUK Beatson Institute, kindly carried out mismatch assays and sequencing analysis of genomic DNA isolated from these samples in order to answer this question. First, gDNA was isolated from both *Runx2* CRISPR clones, and an empty vector-treated clone as a control sequence, using a QIAamp DNA Mini Kit (Qiagen, #51304). Forward (GACTCTGTCCGGTCTCCAGT) and reverse (ACAGGAAGTTGGGACTGTCG) primer sequences were designed for the target region of the gRNA. These primers selectively amplified the region of interest in PCR reactions for gDNA isolated from test and control HC11 cell lines, resulting in products of 500-600bp. The PCR products were purified, eluted in nuclease-free water, and the concentration and purity measured using a NanoDrop 2000. Purified PCR products from each *Runx2* CRISPR gDNA sample were annealed with purified PCR product from the Empty Vector control gDNA sample. A T7 Endonuclease I digestion was then performed on these annealed sequences, before the reaction was stopped and the products purified. The DNA fragments were then eluted in nuclease-free water and the fragmented PCR products were analysed by running on an agarose gel. The T7 endonuclease I recognised and cleaved structural deformities (DNA mismatches) in the generated DNA heteroduplexes of the *Runx2* CRISPR (test) and Empty Vector-treated (control) PCR amplicons. By running the cleavage products on an agarose gel, both the full length and cleavage products were resolved, indicating gene editing was successful. The intensity of the

respective bands can also allow for the calculation of the gene editing percentage that occurred, however this was not necessary for this experimental setup as sequencing was also performed on the PCR products.

2.10.2 Sequencing of *Runx2* CRISPR genomic DNA

The purified, un-annealed PCR products generated in 2.10.1 were also sequenced in order to determine the size(s), location(s), and therefore the implications of any CRISPR-mediated gene alterations. A linearised version of the pUC19 cloning vector (Addgene, #50005) was generated by PCR (using oligos: forward TCTAGAGGATCCCCGGGTAC; reverse CTGCAGGCATGCAAGCTTGG) and then purified by running on an agarose gel. An In-Fusion® HD Cloning Plus kit (Takara Bio, #638911) was used to carry out an In-Fusion cloning reaction (using oligos: forward CGGGGATCCTCTAGAGACTCTGTCCGGTCTCCAGT; reverse CTTGCATGCCTGCAGACAGGAAGTTGGGACTGTCG) to clone the purified PCR product, from the *Runx2* CRISPR gDNA, into the linearised pUC19 vector in the correct orientation, whilst also producing minimal empty vector product. This *Runx2* CRISPR/pUC19 vector construct was then transformed into Stellar Competent Cells (Takara Bio, #636763), provided in the Cloning Plus kit, which were plated onto LB (+Amp) plates for selection. Colonies were picked and grown in 5ml of selective LB broth (+Amp) before cells were spun down and given to the Molecular Technology Services for Minipreps and sequencing using M13 forward (GTAAAACGACGGCCAGT) and reverse (GGAAACAGCTATGACCATG) primers. The sequences that were generated from the *Runx2* CRISPR cell lines were compared with a reference *Runx2* sequence using a Basic Local Alignment Search Tool (BLAST) analysis. This identified any sequence variations in each of the sequenced *Runx2* CRISPR clones, and provided information about their sizes and locations.

2.11 Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

2.11.1 Primers

2.11.1.1 Bio-Rad Primers

Table 2.6 List of Bio-Rad primer sets used for qPCR analysis of cDNA samples.

Gene	Assay ID	Amplicon Context Sequence
Actb	qMmuCED0027505	TGCTCGAAGTCTAGAGCAACATAGCACAGCTTCTCTTTGATGTCACGCACGATTTCCCTCTCAGCTGTGGTGGTGAAGCTGTAGCCACGCTCGGTCAGGATCTTCATGAGGTAGTCTGTCAGGTCCCGGCCAGCCAGGT
Aldh1a1	qMmuCID0005196	GGTTAACTGCTATATGATGTTGTCAGCCCAGTGCCCCTTCGGTGGATTCAAGATGTCTGGAAATGGAAGAGAACTGGGTGAACATGGTCTTTATGAATACACTGAGCTCAAGACAGTCGCAATGAAGATATCTCAGAAGAACTCCTA
Aldh1a7	qMmuCID0005202	ATGCTTTTATGCTGATCTCTACATCCAACAGGTATGCATGAGCAAAGACTTTCCCAGCATTCATCGATTCCATTGTAGCCAGCAGCAGACGATCTCTCTCCATTAAGTCAGCCAGCTT
Ccnd1	qMmuCID0023518	CTGGCGCAGGCTTGACTCCAGAAGGGCTTCAATCTGTTCCTGGCAGGCACGGAGGCAGTCCGGGTCACACTTGATGACTCTGGAAAGAAAGTGCGTTGTGCGGTAG
Dkk1	qMmuCID0020261	CTTTGGTGTGATATATTTTTGAAGTCAGTGTGGTTCTTCTGGGATATCCATCCCCCGCGGCGGCGTTGTGGTCATTACCAAGGTTTTCAATGATGCTTTCCTCAATTTCCCCTCGAGGAAAATGGCTGTGGTCAGAGGGCATGCATATTCCATTTTTGCAGTAGTTCCCGGGGCAG
Esr1	qMmuCID0018069	AGCATCGCCGCCTAGCTCAGCTCCTTCTCATTCTTTCCCATATCCGGCACATGAGTAACAAAGGCATGGAGCATCTCTACAACATGAAATGCAAGAACGTTGTG
Esr2	qMmuCED0044295	TGTACCCTCGAAGCGTGTGAGCATTCAGCATCTCCAGCAGCAGGTCGTACACCGGGACCACATTTTTGCACTTCATGCTGAGCAGATGTTCCATGCCCTTGTTACTG
Fzd7	qMmuCED0048431	GCCACGAGGCCATCGAGGCCAACTCGCAGTACTTTCATCTGGCCGCGTGGGCTGTGCCAGCGGTCAAGACAATCACCATTTTGGCCATGGGCCAGGTGGATGGTGACCTACTCAGTGGAGTGTGCTACGTGGGCCTGTCTAGTG
Ly6a	qMmuCED0003761	Not provided. Product discontinued.

*continued on next page

Table 2.6 (continued) List of Biorad primer sets used for qPCR analysis of cDNA samples.

Gene	Assay ID	Amplicon Context Sequence
Myc	qMmuCID0006528	CCTGAGCCCCTAGTGCTGCATGAGGAGACACCGCC CACCACCAGCAGCGACTCTGAAGAAGAGCAAGAAGA TGAGGAAGAAATTGATGTGGTGTCTGTGGAGAAGA GGCAAACCCCTGC
Nanog	qMmuCID0005399	TCCAGCAGATGCAAGAACTCTCCTCCATTCTGAACC TGAGCTATAAGCAGGTTAAGACCTGGTTTCAAAACC AAAGGATGAAGTGCAAGCGGTGGCAGAAAAACCAG TGGTTGAAGACTAGCAATGGTCTGATTCAGAAGGG CT
Rnf43	qMmuCID0020292	AGAGGGCAGGAGATGTCCTTACTCACATTGCCAGG TGCTGCCAGCCCAACCTGGCTCGGAAGAGGAGCTG GAGGAGCTGTGCGAGCAGGCTGTGTGACGTGAGCT GGTCAGATGAAGCTTCGATGGCGAGTGTGCTCCAG ATG
Runx2	qMmuCID0005205	GCGATCAGAGAACAAACTAGGTTTAGAGTCATCAAG CTTCTGTCTGTGCCTTCTTGGTTCCCGGGGACCGTC CACTGTCACTTTAATAGCTCTGTGGTAAGTGGCCAC TTGGGGAGGATTTGTGAAGACTGTTATGGTCAAGG TGAAA
Sox9	qMmuCED004468 5	TGAGCAGCGAGCCAGGCCAGTCCCAGCGAACGCAC ATCAAGACGGAGCAGCTGAGCCCCAGCCACTACAGC GAGCAGCAGCAGCACTCCCCGCAACAGATCTCCTAC AGCCCCTTCAA
Trp63	qMmuCID0006015	CGTCAGAATACACACGGAATCCAGATGACTTCCATC AAGAAACGGAGATCCCCAGATGATGAGCTGCTGTAC CTACCAGTGAGAGGTCGTGAGACGTACGAGATGTT GCTGAAGATCAAAGAGTCACTGGAGCTCATGCAGTA CCTC
Wif1	qMmuCID0010813	AATGCCAGCCATTCCTGTCAATATCCACTCCATGAA TTTTACCTGGCAAGCTGCGGGGCAGGCAGAATACT TCTACGAGTTCCTGTCTCTGCGCTCCCTGGATAAAG GCATCAT
Wnt3a	qMmuCID0005162	GCACAGAGAATGGGCTGAGTGCTCAGAGAGGAGTA CTGGGGTCCCACAGCCAAGGACCACCAGATCGGGT AGCTGCCCAGAGCCTGCTTCAGGCTGCAGAGCACT AAGAGGTATCCGAGAGGAGCCATCGCCGGCT
Znrf3	qMmuCID0006616	TAGCTCCCCGCTGAACTGCTCTCTTGGCCTTGCCCA GGACAGTGAGACATGGTTTGGGGTCCAATTCTGGC TGTTCAAGCTTCACCACTCCTACCCAGCCATATTCAT ACAAGTCCTCTTCGTCATTGTTATTACATAGGCCTA GTGGGT

2.11.1.2 Other Primer Sequences

Table 2.7 List of other primer sequences used for qPCR analysis of cDNA samples.

Gene	Forward Primer Sequence	Reverse Primer Sequence
Axin2	5′-GCGACGCACTGACCGACGAT-3′	5′-GCAGGCGGTGGGTTCTCGGA-3′
Gapdh	5′-GAAGGCCGGGGCCCACTTGA-3′	5′-CTGGGTGGCAGTGATGGCATGG-3′
Lgr5	5′-GACAATGCTCTCACAGAC-3	5′-GGAGTGGATTCTATTATTATGG-3′
Notum	5′-CTGCGTGGTACACTCAAGGA-3′	5′-CCGTCCAATAGCTCCGTATG-3′

2.11.2 RNA extraction

The RNAqueous™-Micro Total RNA Isolation Kit (Invitrogen, #AM1931) was used to isolate and purify RNA from samples for analysis by RT-qPCR. A NanoDrop 2000 Spectrophotometer was used to quantify and assess the purity of the extracted RNA prior to this analysis. 40ng/µl was the minimum required concentration of RNA required for a minimum total input of 500ng for each cDNA synthesis reaction. The ratio of absorbance at 260nm to 280nm (the 260/280 ratio), calculated by the NanoDrop, was used to assess the purity of RNA, while the 260/230 ratio was used as a secondary measure of nucleic acid purity. According to the NanoDrop T042 Technical Bulletin, a 260/280 ratio of ~2.0 is generally accepted as “pure” RNA, though anything above 1.7 is considered acceptable for analysis by PCR. The expected 260/230 ratios are expected to fall within a slightly higher range of ~2.0-2.2, although 1.7 is also considered sufficient for PCR analysis. If either ratios are significantly lower than these expected values, it may indicate the presence of contaminants. For some samples, the extracted RNA had initial 260/230 ratios that were appreciably lower than the expected values. Significant absorbance at 230nm (and therefore low 260/230 ratios) can indicate the presence of guanidine isothiocyanate (used for RNA isolations), EDTA, carbohydrates, or phenol, which can interfere with downstream cDNA synthesis. Ethanol precipitation of the RNA was used in order to purify it further and remove these contaminants (2.11.3).

2.11.3 RNA Purification

Ethanol precipitation of the RNA was used in order to remove the impurities that were causing low 260/230 ratios in any of the samples. For each RNA sample with

a 260/230 ratio lower than 1.7, the RNA was first made up to a 100µl total volume using nuclease-free water (Qiagen, #129115). 10µl of 3M sodium acetate solution (Sigma-Aldrich, #71196-100ML) was then added. The salt in this solution neutralises the charge on the RNA, which causes it to become less hydrophilic and facilitates its efficient precipitation. To further maximise RNA recovery from this reaction, 1µl of a 20mg/ml glycogen solution (Thermo Scientific, #R0551) was also added. 2.5-3 volumes of ice cold 100% ethanol was then added to each sample, which were vortexed to mix thoroughly. The precipitation reaction was then allowed to carry out overnight at -20°C. The RNA was pelleted by centrifuging for 30 mins at 19,000g in a benchtop centrifuge, chilled to 4°C, and the supernatant removed using a micropipette. The pellet was washed twice with 0.5ml ice cold 75% ethanol, spinning at 4°C for 10 mins each time and then removing the supernatant each time to re-pellet. Samples were spun briefly at top speed to remove as much of the trace amount of Ethanol as possible before air drying the pellet for ~3 minutes. The resulting RNA pellet was finally re-suspended in an appropriate volume of nuclease-free water (tailored to the expected yield), and quantified using a NanoDrop 2000 Spectrophotometer. This RNA was used to synthesise cDNA (2.11.4) before analysing by qPCR (2.11.5).

2.11.4 Reverse transcription

cDNA was synthesised from the extracted RNA using a Qiagen QuantiTect Reverse Transcription Kit (Qiagen, #205311) according to the protocol, specified by the manufacturer, that was supplied with the kit. Samples were each diluted 1:10 in nuclease-free water before proceeding to qPCR. At least 500µg of template RNA was used to synthesise cDNA for each sample, and wherever possible (i.e. where there was sufficient RNA concentration and total quantity), cDNA was synthesised from a total of 1µg of RNA by setting up 2 individual reverse transcription reactions for 500ng of RNA and combining the resulting cDNA at the end. This resulted in all samples having the same final concentrations of cDNA regardless of the total RNA input.

2.11.5 qPCR protocol

2.11.5.1 qPCR Plate Setup

For each qPCR experimental setup, each sample being tested for a particular gene was set up in triplicates, where the final gene expression result for that sample was representative of the average of the three technical replicates. A nuclease-free water sample was set up in triplicate as a negative control. In addition, the inclusion of at least 1 or 2 housekeeping genes was essential in order to ensure accurate comparisons between samples by normalising quantification data. For the qPCR experiments presented throughout this thesis, housekeeping genes *Actb* and *Gapdh* were used. For each gene, a primer mastermix was prepared for each sample as follows: 10µl of SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma-Aldrich, #S4438-500RXN), 1µl of primer mix (primers listed in 2.11.1), and 4µl of nuclease-free water. This was scaled up depending on the number of samples to be tested using that primer, for that particular experimental setup. 15µl of primer mastermix was pipetted into the wells of a 96-well PCR plate (Bio-Rad, #HSL9645) and 5µl of each sample or water control was pipetted (in triplicate) into the correct corresponding wells with the primer master mix. The plates were sealed with MicroAmp™ Optical Adhesive Film (Applied Biosystems, #4360954) and the plate was briefly centrifuged using an Axygen Axyspin Mini Plate Spinner Centrifuge.

2.11.5.2 qPCR Reaction and Data Analysis

Plates were run on the Bio-Rad CFX96 Touch Real-Time PCR Detection System. Samples were first denatured at 95°C for 2 minutes before performing 40 cycles of the following denaturation, annealing, and extension steps: 95°C for 15 seconds; 60°C for 1 minute; and a plate read. Following this, a melt curve was generated by heating from 65°C and 95°C with 0.5°C increments, 5 seconds dwell time, and a plate read at each temperature. For each sample being tested, the Bio-Rad CFX Maestro software calculated the normalised relative expression values of the given test gene using a $\Delta\Delta C_q$ method. The normalised expression value obtained from this calculation represents the relative quantity of the target gene normalised to the relative quantity of the housekeeping gene(s). In order to facilitate statistical analyses of the normalised expression data between test

groups, the data was log-normally distributed prior to performing statistical analyses, as described in 2.16.

2.12 2D growth assays

2.12.1 Plate setup for 2D growth assays

HC11 cells to be set up in 2D growth assays were trypsinised to detach cells from the culture flask. This enzymatic reaction was then stopped and cells collected using RPMI medium (+FBS, +PSG) and the suspension was passed through a 40µM cell filter into a 50ml falcon tube to disaggregate the cells. Cells were then spun down at 350g for 5 minutes, resuspended in ~10-15ml of culture medium, and passed 2-3 times through a 23g Blue 1.25 inch BD Microlance Needle (Becton Dickinson, #300700) attached to a syringe to obtain a single cell suspension. A viable cell count was taken using either a CASY cell counter, or a DeNovix CellDrop™ Automated Cell Counter with Trypan Blue exclusion. The cell suspensions were then diluted to a final concentration of $5x10^3$ cells/ml, and 1ml of this suspension was added to each well of standard 24-well culture plates. Plates were cultured in a 37°C incubator with 5% CO_2 for 7 days, and supplemented with fresh medium every 2-3 days. If cells were additionally being treated with recombinant mouse WNT3A protein (Abcam, # ab81484), this was also added to the initial culture medium at the point of seeding cells and supplemented every 3-4 days, as described in 2.15.

2.12.2 2D cell counts (*Runx1* overexpression/CRISPR cells)

For each *Runx1*-altered cell line (and their respective Empty Vector counterparts) set up in 2D growth assays, 3 independent cell counts were taken every 24 hours by Trypsinising cells from 3 different wells, collecting and processing each into a single cell suspension, and calculating the total cell count using an average of 2 live cell counts obtained from the CASY Cell Counter. Results were plotted as line graphs, used to represent the mean average of the 3 total cell counts taken for each cell line at each timepoint, with error bars used to display population standard deviations. Statistical analyses were used to compare the average cell counts for each cell line at each timepoint. Statistical analyses are described in 2.16.

2.12.3 2D cell counts (*Runx1*/*Runx2* double CRISPR cells)

The HC11 cell lines generated to explore the simultaneous knockout of Runx1 and Runx2 were analysed for 2D growth using a slightly altered method, as the CASY Counter was no longer available. For each of these *Runx1* and/or *Runx2* CRISPR cell lines (and Empty Vectors control line), 4 independent cell counts were taken every 24 hours by Trypsinising cells from 4 different wells, collecting and processing each into a single cell suspension, and calculating the total cell count using an average of 2 live cell counts obtained from the DeNovix CellDrop™ Automated Cell Counter, with Trypan Blue exclusion. Results were plotted as line graphs, used to represent the mean average of the 4 total cell counts taken for each cell line at each timepoint, with error bars used to display population standard deviations. The average cell counts for each cell line were statistically compared at each timepoint as described in 2.16.

2.13 Stemness/3D growth assays

2.13.1 HC11 mammosphere assays

2.13.1.1 Primary mammosphere setup and counting

Medium for mammosphere culture was composed of RPMI 1640 Medium supplemented with PSG, mouse EGF to a final concentration of 20ng/ml, recombinant mouse basic fibroblast growth factor (FGF-2, Sigma-Aldrich, #SRP4038-50UG) to a final concentration of 20ng/ml, heparin (Sigma-Aldrich, #H3149-50KU) to a final concentration of 4µg/ml, and 50X B27 Supplement (Gibco, #17504-044) to a final 1X concentration. HC11 cells to be set up in primary mammosphere assays were trypsinised to detach cells from the culture flask. This enzymatic reaction was then stopped and cells collected using RPMI medium (+FBS, +PSG) and the suspension was passed through a 40µM cell filter into a 50ml falcon tube to disaggregate the cells. Cells were spun down at 350g for 5 minutes, resuspended in ~10-15ml of mammosphere medium, and passed 2-3 times through a 23G Blue 1.25 inch BD Microlance Needle attached to a syringe to obtain a single cell suspension. A viable cell count was taken using either a CASY cell counter, or a DeNovix CellDrop™ Automated Cell Counter with Trypan Blue exclusion. The cell suspensions were diluted to a final concentration of $5x10^3$ cells/ml, and 1ml of this suspension was added to each well of 24-well plates with Ultra-Low Attachment surfaces (VWR, #734-1584). Plates were cultured in a 37°C, 5% CO_2

incubator for 7 days, and removed every 2-3 days to supplement with mouse EGF and mouse FGF-2 (made up as a growth factor mastermix by diluting in PBS), which were added to the mammosphere cultures at final concentrations of 20ng/ml in order to maintain the mammosphere cultures. If mammospheres were additionally being treated with recombinant mouse WNT3A protein, this was also added to the initial mammosphere medium and then supplemented every 3-4 days, as described in 2.15. Following 7 days of incubation, the plates were removed from the incubator and analysed using an Olympus CKX41 Inverted Phase Contrast Microscope at x10 magnification. The squared graticule within the eyepiece of the microscope was used to determine which of the 3D spherical structures within the cultures were to be counted as mammospheres. The structures that were at least the size of one of the small squares within the graticule were counted as a mammosphere and anything smaller than this was not counted. Mammospheres were manually counted for each well. Primary mammospheres were imaged (2.13.1.3). Some were set up in subsequent secondary mammosphere assays (2.13.1.2), or embedded in agarose plugs (2.13.1.4) for subsequent IHC analysis (2.2.4), or collected as cell pellets for subsequent RT-qPCR analysis (2.11).

2.13.1.2 Secondary mammosphere setup and counting

For mammospheres being set up in secondary mammosphere assays, the primary mammospheres were all collected together into a 50ml Falcon tube and spun down to remove the mammosphere medium. Cells were washed in PBS, transferred to a 15ml Falcon tube, and spun down again to obtain a cell pellet. The mammospheres contained within the pellet were trypsinised, resuspended in RPMI medium (+FBS, +PSG), and passed through a 40µM cell filter into a 50ml falcon tube to disaggregate the cells. Cells were spun down, resuspended in ~5-10ml of mammosphere medium, and passed 2-3 times through a 23G Blue 1.25 inch BD Microlance Needle attached to a syringe to obtain a single cell suspension. A viable cell count was taken using either a CASY cell counter, or a DeNovix CellDrop™ Automated Cell Counter with Trypan Blue exclusion. The cell suspensions were then diluted to a final concentration of $5x10^3$ cells/ml, and 1ml of this suspension was added to each well of 24-well plates with Ultra-Low Attachment surfaces. Plates were cultured in a 37°C incubator for 7 days, and removed every 2-3 days to supplement with further mouse EGF and mouse FGF-2 (diluted as a growth

factor mastermix in PBS), which were added to the mammosphere cultures at final concentrations of 20ng/ml in order to maintain the mammosphere cultures. Mammospheres were visualised and manually counted using an Olympus CKX41 Inverted Phase Contrast Microscope at x10 magnification, as described in 2.13.1.1 for primary mammospheres. Secondary mammospheres were imaged (2.13.1.3) and collected as cell pellets for subsequent RT-qPCR analysis (2.11).

2.13.1.3 Mammosphere imaging and size analysis

Images of mammospheres were taken using an Olympus CKX41 Inverted Phase Contrast Microscope at x10 magnification. An image was also taken of using a 1mm stage micrometer, using the same settings as were used to capture the mammosphere images. All of the images were analysed for mammosphere sizes using ImageJ. Firstly, the image of the stage micrometer was used to calibrate the scale bar within ImageJ, in order to facilitate a more accurate size analysis. This was done by using the Straight Line tool within ImageJ to draw a line of known distance on the stage micrometer image, and then transferring the pixel distance and the known distance (μM) into the Set Scale function. These settings were used to calibrate the scale bar in ImageJ for the analysis of all images taken using those settings. The images taken for mammosphere experiments were then uploaded and individual mammospheres were sized by manually drawing around each using the Freehand drawing tool, and then using the Measure function to obtain their respective areas (μM^2). For each experimental setup, the mammosphere sizes were statistically compared between the experimental groups as described in 2.16.

2.13.1.4 Agarose plugs of mammospheres

To analyse mammospheres using IHC methods, they were first embedded in agarose before paraffin embedding and IHC staining by the Histology Department. For each sample, all mammospheres were collected together into a 50ml Falcon tube and centrifuged at 200g for 5 minutes. The culture medium was aspirated away and the mammospheres were re-suspended in PBS and transferred into a 1.5ml Eppendorf tube. The tubes were spun down at 200g for 5 minutes and the PBS supernatant removed. The tubes were again spun down briefly at 200g for ~10

seconds and residual PBS removed. The mammospheres were re-suspended in ~100-200µl of 2% agarose depending on the pellet size. The mammosphere/agarose mixture was pipetted into an Eppendorf lid and allowed to solidify at room temperature. Agarose plugs were stored in PBS before giving to the Histology Department for embedding and IHC staining (described in 2.2.4). The resulting slides were finally imaged and analysed (also described in 2.2.4).

2.13.2 HC11 colony formation assays

Medium for colony formation was composed of RPMI 1640 Medium supplemented with PSG, mouse EGF to a final concentration of 20ng/ml, and mouse FGF-2 to a final concentration of 20ng/ml. HC11 cell lines were trypsinised, dissociated, and processed into a single-cell suspension that was counted, as described in 2.13.1.1. The cell counts for each cell line were used to calculate the volume of each cell suspension that was required to obtain 200 cells/well for the number of wells being set up. These volumes were transferred into the wells of a 96-well cone-bottom plate (Greiner, #780271) and centrifuged at 300g for 5 minutes to collect cells at the bottom. The supernatant was removed by quickly inverting the plate, and cells were resuspended in ice-cold Growth-factor-reduced Matrigel (Corning, #356231) to obtain a cell concentration of $1x10^4$ cells/ml. 20µl of each suspension was carefully pipetted into the centre of each well of a 24-well plate (200 cells/well), which was transferred into an incubator for 5 minutes to allow the Matrigel to solidify. 1ml of the colony formation medium was then pipetted into each well and cultures were incubated (37°C, 5% CO_2) for 7 days. Every 3-4 days, the colony formation medium was removed and replaced with fresh colony formation medium in order to keep the cells properly supplemented with growth factors. After 7 days, Matrigel colonies were manually counted and imaged using an Olympus CKX41 Inverted Phase Contrast Microscope at x10 magnification, as described in 2.13.1.1 for primary mammospheres.

2.13.3 PyMT mammosphere/tumoursphere assays

Mammosphere/tumoursphere medium, for primary cells extracted from *MMTV-Cre* and *MMTV-Cre;MMTV-PyMT* mice, was composed of Ham's F12 Medium supplemented with PSG, Cholera Toxin (Sigma-Aldrich, #C8052-.5MG) to a final

concentration of 5ng/ml, mouse EGF to a final concentration of 20ng/ml, recombinant mouse FGF-2 to a final concentration of 20ng/ml, heparin to a final concentration of 4µg/ml, and 50X B27 Supplement to a final 1X concentration. PyMT mice were sacrificed and mammary gland samples taken (2.1.5.1), which were processed into single cell suspensions of isolated MMECs (2.3.1). After obtaining a viable cell count, as described in 2.3.1, cells were spun down at 350g for 5 minutes and diluted to a final concentration of $1x10^4$ cells/ml in mammosphere medium. 1ml of this cell suspension was added to each well of 24-well plates with Ultra-Low Attachment surfaces. Plates were incubated (37°C, 5% CO_2) for 10 days, and removed every 3 days to supplement with mouse EGF and mouse FGF-2 (diluted as a growth factor mastermix in PBS), which were added to the mammosphere cultures to final concentrations of 20ng/ml in order to maintain the mammosphere cultures. Following 10 days of incubation, the plates were removed from the incubator and analysed using an Olympus CKX41 Inverted Phase Contrast Microscope at x10 magnification. The squared graticule within the eyepiece of the microscope was used to determine which of the 3D spherical structures within the cultures were to be counted as mammospheres. The structures that were at least the size of one of the small squares within the graticule were counted as a mammosphere and anything smaller than this was not counted. Mammospheres/tumourspheres were manually counted for each well.

2.14 MTS cell viability and metabolic activity assay

2.14.1 Plate Setup for MTS assay

The MTS assay is a sensitive and accurate colorimetric assay, which is used for the relative quantification of metabolically active cells within given populations, with a particular focus on cells' responses to growth factors and drug treatments. The MTS assay protocol is based on the conversion of the MTS tetrazolium compound into a coloured soluble formazan dye, whose presence can be quantified by measuring absorbance at ~490nm. The NAD(P)H-dependent cellular oxidoreductase enzyme, which is present in metabolically active cells, is responsible for this reduction of the MTS tetrazolium compound. The ability of cells to initiate this conversion is therefore used to relatively assess their metabolic activity, which is often used as a surrogate for cell viability or proliferative capacity within cell populations. To prepare HC11 cell lines for the

MTS assay, they were each processed into single cell suspensions and counted using a CellDrop, as described in 2.13.1. Each of the cell lines was diluted into 4 different single cell suspensions, which had 2 different final concentrations (2 suspensions of $1x10^4$ cells/ml and $2x10^4$ cells/ml), and for each of these concentrations of cells one suspension was supplemented with sterile water (vehicle control) while the other was supplemented with 100ng/ml of recombinant WNT3A protein. 100µl of each of the cell suspensions (and medium as a blank/control) was plated in triplicate into 6 separate Greiner CELLSTAR® 96-well plates with flat bottom black polystyrene wells and micro-clear bottoms (Sigma-Aldrich, #M0562-32EA).

2.14.2 MTS assay and data analysis

Plates were then incubated at 37°C for 3 different timepoints: 24 hours, 48 hours, and 72 hours. At each of these timepoints, 2 of the 96-well plates were removed from the incubator and 20µl of MTS Reagent from the CellTiter 96® AQ_{ueous} One Solution Cell Proliferation Assay was added to each well containing cells or blank medium. Absorbance at 490nm was measured using a Tecan Sunrise Microplate Reader and results exported in the form of Excel sheets for analysis. The media blank/control results were averaged and removed from each sample result, in order to remove any background absorbance. For each cell line and timepoint, the average absorbance value of the vehicle control-treated cells was used to normalise each of the absorbance values obtained for that cell line. Each of the normalised values for the vehicle- and Wnt3a-treated replicates was expressed as a percentage of the average absorbance of the vehicle treated cells. Results were plotted using Graphpad Prism 9, which was also used to perform the appropriate statistical analyses described in 2.16 and the appropriate figure legends.

2.15 Wnt3a treatment

Recombinant Mouse WNT3A Protein (Abcam, #ab81484) was added to culture medium for 2D growth assays (2.12), mammosphere assays (2.13.1), and in MTS cell viability and metabolic activity assays (2.14). WNT3A was added to the initial culture medium to a final concentration of 100ng/ml, and cells were supplemented every 3-4 days with WNT3A to a final concentration of 100ng/ml.

Cells set up in MTS assays did not require supplementation with further Wnt3a as these experiments lasted 3 days.

2.16 Statistical analyses

All of the statistical analyses for this thesis were performed using Graphpad Prism 9. The details of the specific statistical analyses used for each individual experiment are described in the appropriate/corresponding figure legends.

3 Investigating the molecular mechanisms of early tumour emergence in *Runx1*-deficient mice.

3.1 Introduction

The Wnt/ß-catenin signalling pathway is one of several important cancer-related pathways implicated in the development and progression of breast cancer (Song, Cui et al. 2014). Wnt signalling is dysregulated in a large proportion of breast tumours, and aberrant expression of ß-catenin (a hallmark of canonical Wnt activation) has been observed in over half of all invasive breast cancers analysed by IHC in independent studies (Karayiannakis, Nakopoulou et al. 2001, Li, Li et al. 2014). In addition, this aberrant activation of ß-catenin was significantly correlated with poor prognosis for breast cancer patients, irrespective of the molecular subtype of their disease (Lin, Xia et al. 2000, Li, Li et al. 2014).

The Blyth lab use genetically engineered mouse models (GEMMs) of breast cancer as *in vivo* tools to recapitulate the human disease and interrogate specific biological processes within a physiologically relevant model. Due to the clear significance of the Wnt/ß-catenin pathway in the context of breast cancer, a GEMM with mammary-specific oncogenic activation of ß-catenin is one such model used for these investigations. This model was used as the basis for the generation of several conditional knock-out cohorts (described in 2.1.1.1) in order to study the effects of *Runx1* and/or *Runx2* deletion on Wnt/ß-catenin-driven mammary tumorigenesis. The *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ cohort shall henceforth be referred to as BCAT, *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;Runx1*$^{fl/fl}$ as BCAT_1, *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;Runx2*$^{fl/fl}$ as BCAT_2, and the *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ cohort as BCAT_12 (as described in Table 2.1). The analysis of these cohorts, discussed below, was led by Alessandra Riggio as part of her PhD studies (Riggio 2017, Riggio Unpublished), and showed the striking impacts of *Runx1* loss in this mammary model.

3.1.1 *Runx1* loss accelerates Wnt-driven mammary tumorigenesis

BCAT, BCAT_1, and BCAT_2 cohort mice were put through 2 rounds of parity in order to maximise expression of the activated ß-catenin oncogene in the mammary gland (as described in 1.6.2). By monitoring all BCAT, BCAT_1, BCAT_2,

and BCAT_12 mice for the formation of palpable mammary tumours, and routinely checking and measuring these until clinical end-point, cohorts could be compared using 3 distinct parameters (Figure 3.1A): tumour onset (time from birth to tumour notice); tumour progression (time from tumour notice to end point); and overall survival (time from birth to end point). The Kaplan-Meier curves shown in Figure 3.1 demonstrate that tumours emerged in BCAT_1 mice at a significantly earlier timepoint than their wild-type (BCAT) controls and resulted in a reduction in their overall survival. Onset in BCAT_12 mice occurred at a time dramatically earlier than other cohorts and as a result, it was not possible for the female mice to go through parity to maximise *Blg-Cre* expression. The overall survival of these mice was also significantly shorter than other cohorts. Surprisingly, tumour initiation was not significantly changed in the BCAT_2 cohort and they showed survival patterns similar to that of their wild-type counterparts.

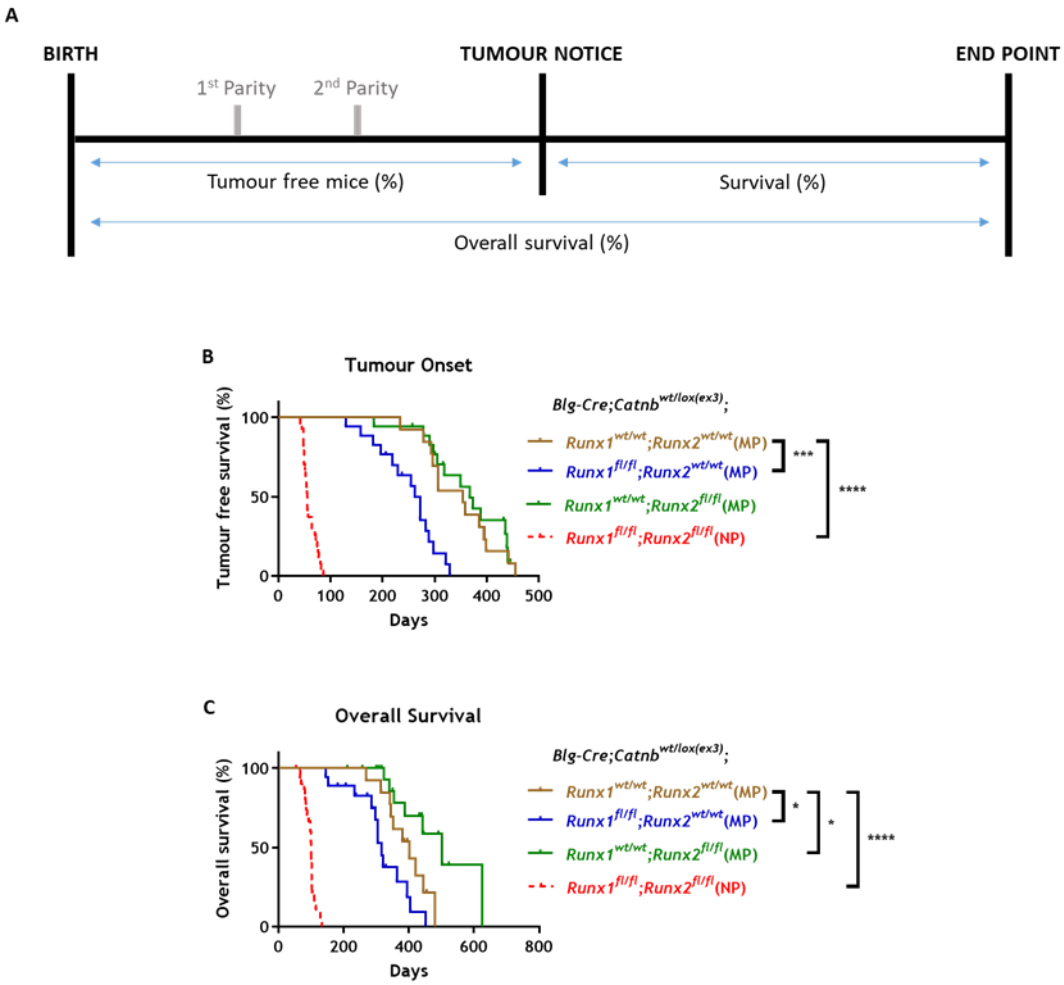


Figure 3.1 *Runx1* deletion accelerated onset of mammary tumours in Wnt-driven mouse model, and even more dramatically with simultaneous loss of *Runx2*.

* Data provided by Alessandra Riggio and shown here with permission*

[A] A schematic representation of the parameters used to assess the impact of *Runx1* and/or *Runx2* loss on Wnt-driven mammary tumorigenesis. BCAT, BCAT_1, and BCAT_2 females underwent multiple parities (MP) and were monitored for tumour initiation by recording the first instances of palpable tumours. BCAT_12 mice remained nulliparous (NP) due to the rapid rate of tumour development. Following tumour notice, size measurements were taken 2-3 times per week using callipers. Mice were sacrificed once a mammary tumour reached the clinical end-point.

[B] Kaplan-Meier curve comparing time from birth to tumour notice observed in BCAT (n=13), BCAT_1 (n=17), BCAT_2 (n=18), and BCAT_12 (n=27) mice.

[C] Kaplan-Meier curve comparing time from birth to clinical end point recorded for BCAT (n=13), BCAT_1 (n=18), BCAT_2 (n=19), and BCAT_12 (n=33) mice.

Statistical analyses performed for each, using the Log-rank (Mantel-Cox) test [* $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$].

3.1.2 *Runx* deletion dramatically increases tumour burden in a Wnt-driven mouse model of breast cancer

In addition to these changes in tumour initiation and progression observed in BCAT_1 and BCAT_12 cohorts, the tumours were also significantly different at endpoint, both qualitatively and quantitatively. Representative images of each cohort at end-point are shown in Figure 3.2A. The gross pathology of BCAT_2 tumours appear very similar to that of BCAT (wild-type) tumours, with some but not all mammary glands bearing single lesions. A greater proportion of mammary glands within the BCAT_1 mice appear to bear tumours, and some bearing more than a single lesion. The most remarkable phenotypic change was observed in the BCAT_12 cohort, in which many multifocal and multicentric tumours were present within all (or almost all) glands at clinical end-stage. Quantification and statistical comparisons of the number of affected glands in each cohort are summarised in Figure 3.2B, and demonstrate that the visual differences between the cohorts were only statistically significant when comparing the BCAT_12 mice with the BCAT wild-types. Although not significantly altered in BCAT_1 mice, there did appear to be a trend towards increased numbers of affected glands. Tumour burdens (cumulative mammary gland weight measured as a percentage of the total body weight) were plotted for each cohort mouse, although only showed statistically significant changes within BCAT_12 glands, albeit there was a potential trend towards increased tumour burdens in BCAT_1 mice too.

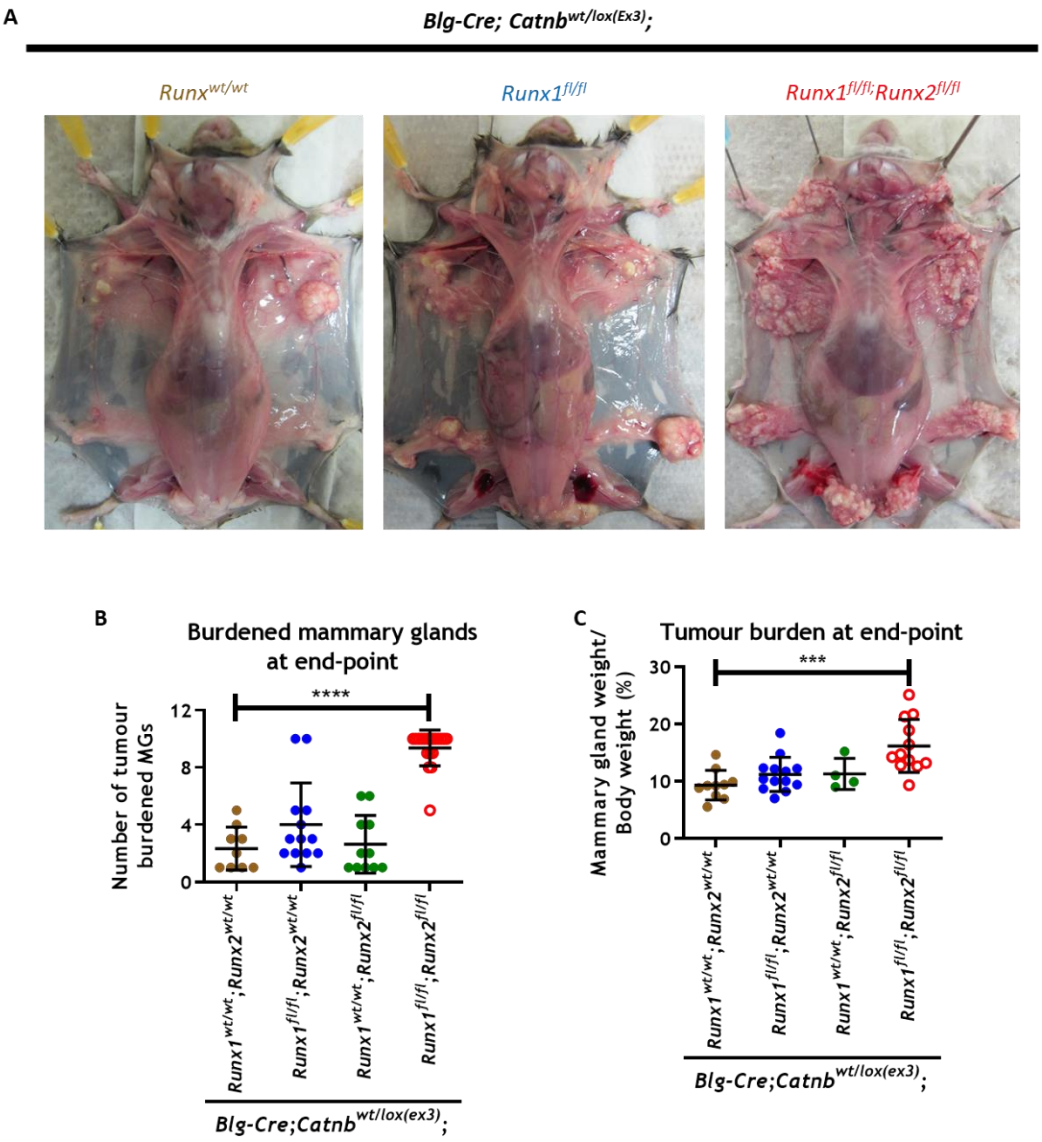


Figure 3.2 Effect of mammary-specific *Runx1* and/or *Runx2* loss on tumour burden of mice with oncogenic activation of ß-catenin.

Data provided by Alessandra Riggio and shown here with permission

[A] Representative images of *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ mice, being dissected at clinical endpoint, showing the phenotypic changes within Wnt-driven mouse models upon homozygous loss of *Runx1* alone and simultaneous loss of *Runx1* and *Runx2*.

Numbers of affected mammary glands [B] and tumour burdens (expressed as a percentage of the total body weight) [C] were noted at end-point for each mouse and plotted for each cohort as above. Each data point represents the number of burdened glands and tumour burden obtained from a single mouse. One-way ANOVA, with Dunnett's multiple comparisons test [*** $p < 0.001$; **** $p < 0.0001$].

[B] *Runx*$^{wt/wt}$ (n=9), *Runx1*$^{fl/fl}$ (n=13), *Runx2*$^{fl/fl}$ (n=11), and *Runx1*$^{fl/fl}$;*Runx2*$^{fl/fl}$ (n=19).

[C] *Runx*$^{wt/wt}$ (n=10), *Runx1*$^{fl/fl}$ (n=13), *Runx2*$^{fl/fl}$ (n=4), and *Runx1*$^{fl/fl}$;*Runx2*$^{fl/fl}$ (n=12).

3.1.3 Immune signalling altered in *Runx*-deleted mammary glands

This study provided the first *in vivo* evidence of *RUNX1* functioning as a tumour suppressor within the context of the mammary gland. The earlier emergence of higher numbers of independently arising tumours also hinted towards a possible role for *Runx1* in the stem-like behaviours of the mammary gland cells and in cancer stem cells. The next objective for the Blyth lab was therefore to explore the molecular mechanisms behind the dramatic phenotypes observed in the mammary gland tissues as a result of *Runx1* (and simultaneous *Runx2*) loss. RNA-sequencing (RNA-seq) uses high-throughput sequencing methods to provide relatively unbiased insights into the transcriptomic profiles of analysed cells. It offers much higher coverage and resolution than more traditional Sanger and microarray technologies. RNA-seq was therefore used to analyse mammary gland cell populations that were isolated from cohort mice prior to the formation of palpable tumours (9wks), in order to determine the molecular events within the glands at these early stages that facilitate the earlier emergence of independently arising tumours. Due to the overwhelmingly high numbers of genes that were found to be differentially expressed between the 4 cohort sample groups, the MetaCore bioinformatics suite was used to analyse the data and identify the most relevant pathways, networks, and cellular processes (Figure 3.3). Interestingly, these analyses identified enhancements of multiple pro-inflammatory/pro-tumorigenic immune signatures and ECM remodelling signatures in BCAT_12 samples.

Due to the RNA-seq samples originating from preparations of bulk mammary glands, there were various cell types present within them, in addition to the mammary epithelial cells. This cellular heterogeneity meant it was not possible to deduce the cell of origin for the immune and ECM-remodelling signatures. Since it could not be assumed if or which these changes were originating from the *Blg-Cre*-expressing mouse mammary epithelial cells (MMECs) or from their microenvironment, it was not possible to determine which changes were occurring as a direct result of *Runx1* (and simultaneous *Runx2*) loss or which were due to indirect changes, such as immune cell infiltration. Another factor introduced by this cellular heterogeneity was that, due to the nature of traditional bulk RNA-seq experiments, gene expression was being averaged across all cell types. This, combined with the fact that the immune and ECM-remodelling signatures were

largely dominating the alteration profiles of these samples, was potentially diluting or drowning out some cell type-specific molecular signatures originating from within the genetically altered MMECs.

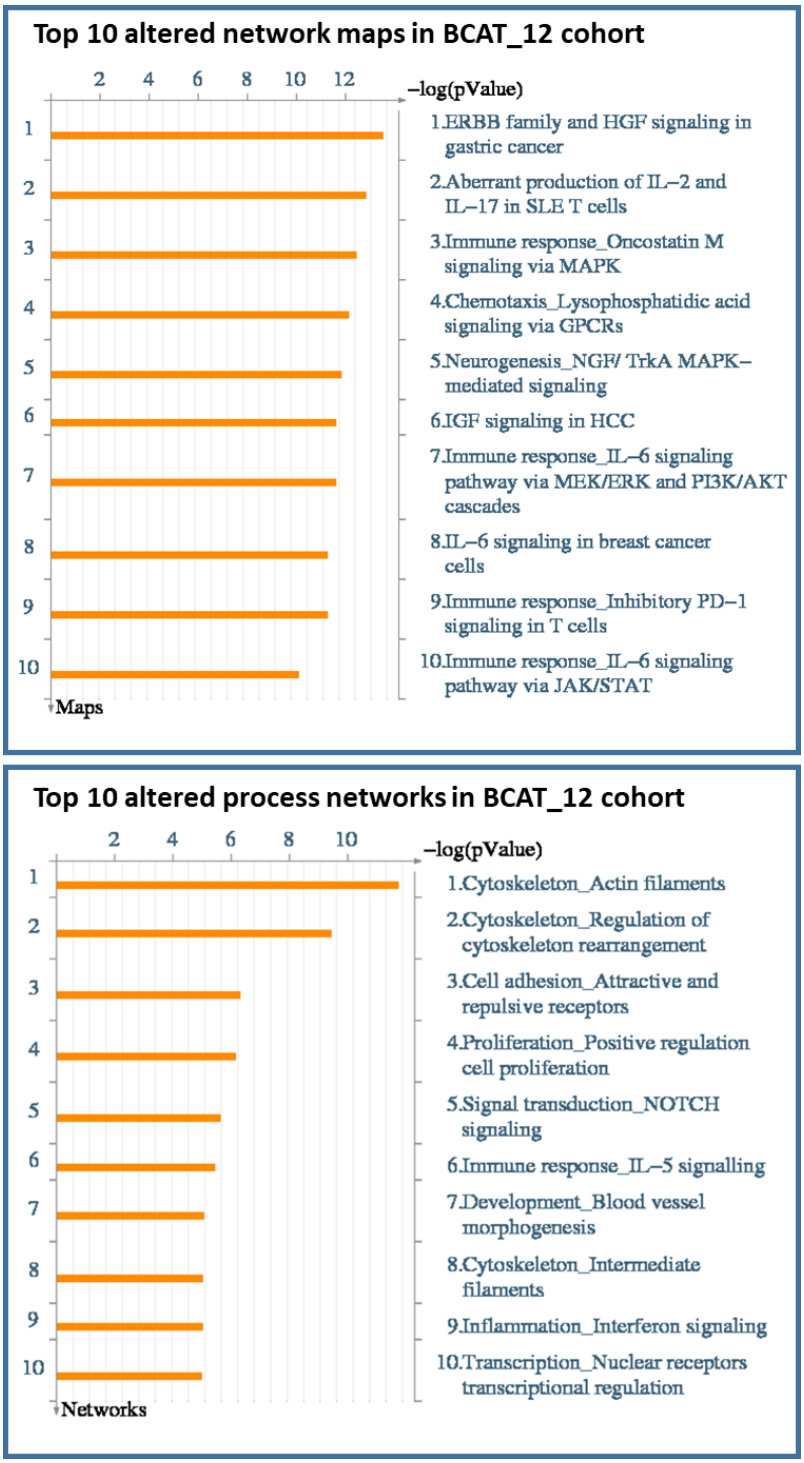


Figure 3.3 Enrichment of immune responses and cytoskeleton rearrangement in BCAT_12 glands in bulk mammary cell RNA-seq.

Data provided by Alessandra Riggio and shown here with permission

RNA-seq data was processed by Robin Shaw and analysed by Alessandra Riggio. Figure shows select results from these analyses, namely the top 10 MetaCore network maps and process networks that were significantly altered in BCAT_12 (n=4) mammary glands compared to wild-type (n=5) counterparts, using significance cutoffs of Fold Change ≤−1.5 and ≥1.5, and Adjusted p value of <0.05.

3.1.4 Experimental Aims

One of the aims for the work presented in this chapter was to carry out a more focused analysis, specifically identifying the molecular changes within the *Runx*-deleted mammary cells. In order to isolate MMECs, with simultaneous constitutive activation of ß-catenin alongside conditional knock-out of *Runx1* and/or *Runx2*, from the mammary glands of cohort mice, an RFP reporter was introduced into this model as a surrogate for oncogenic ß-catenin expression and *Runx*-deletion. FACS of single cell suspensions (described in 2.4.2) obtained from mammary glands of 9wk old cohort mice (processed according to the method in 2.3.1) facilitated this isolation of genetically altered MMECs. The RNA-seq analysis of these RFP-sorted cells forms the basis of this chapter.

3.2 Results

3.2.1 Analysis of pilot RNA-seq data reveals significantly altered Wnt signalling and elevated stem cell-related signature.

In order to understand the transcriptional changes occurring in the *Runx*-deleted mammary cells to facilitate the earlier emergence of palpable tumours, an RFP reporter allele was introduced into the ß-catenin mammary model. RFP expression served as a surrogate for oncogenic ß-catenin expression and *Runx*-deletion, and flow-assisted cell sorting (FACS) could be used to isolate these RFP-positive (ß-catenin expressing, *Runx*-deleted) cells. In order to collect sufficient amounts of RNA from RFP-sorted MMECs (isolated from 9 week-old BCAT, BCAT_1, BCAT_2, and BCAT_12 mice) for analysis by RNA-seq, MMECs isolated from multiple mice were combined together to form each individual sample. Therefore, in an effort to apply the Three Rs of Humane Animal Experimentation to this experimental setup (particularly the reduction and refinement principles) a pilot RNA-seq study was conducted. 3 *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;tdRFP* experimental groups were analysed for the purposes of this pilot investigation: MMECs expressing wild-type Runx (BCAT); MMECs with mammary-specific loss of *Runx1* (BCAT_1); MMECs with simultaneous mammary-specific depletion of *Runx1* and *Runx2* (BCAT_12). A BCAT_2 group was not used in this initial investigation due to a lack of availability of mice/samples, and due to there being a relatively unremarkable phenotype observed in this group compared to the wild-type controls. 2 samples were analysed within each of the experimental groups, and RFP-sorted MMECs isolated

from the mammary gland tissue of 3 separate mice were combined together to generate each sample. The details of each sample, including RNA integrity numbers (RIN), are listed in Table 3.1. The samples were processed and analysed according to the methods described in 2.5 for RNA-seq.

Table 3.1 Samples, generated from pooled RFP-sorted (BCAT, BCAT_1, and BCAT_12) MMECs, used for pilot RNA-seq analysis.

Pilot Sample of Pooled RFP+ MMECs Individual Mice	**Total Number of Pooled RFP+ MMECs** Number of MMECs Collected Per Mouse	**RIN Value**
BCAT sample 1 — 3 mice	**~145,500**	**7.7**
YDG138.4h	50,664	
YDG138.4i	79,869	
YDG153.8h	~15,000	
BCAT sample 2 — 3 mice	**~120,000**	**8.7**
YDG153.4h	101,451	
YDG153.6j	Not provided	
YDG154.5h	16,407	
BCAT_1 sample 1 — 3 mice	**~109,000**	**8.4**
YDD99.4f	2,335	
YDD99.4g	26,146	
YDD99.5e	80,393	
BCAT_1 sample 2 — 3 mice	**~114,500**	**6.4**
YDD106.3g	30,349	
YDD110.2i	61,851	
YDD110.2j	22,296	
BCAT_12 sample 1 — 3 mice	**~201,000**	**6.3**
YDE93.3a	109,778	
YDE93.3b	89,530	
YDE95.2c	1,642	
BCAT_12 sample 2 — 3 mice	**~228,500**	**10**
YDE95.2b	41,987	
YDE96.1a	76,350	
YDE96.1e	110,305	

Initial analyses of the resulting data involved a principle component analysis (PCA) of the samples. PCA takes a large multidimensional data set (such as that from an RNA-seq experiment) and flattens it into 2 or 3 dimensions in order to make the analysis less complex and more manageable. It tries to find the most meaningful way to flatten the data by focusing on the factors (or Principal Components) that are different between the given samples. PC1 is the name given to the principal

component (and axis in the PCA plot) that spans the most variation, while PC2 spans the second most variation. The PCA plot in Figure 3.4 demonstrates that the samples clustered relatively well into their respective groups, with the BCAT samples showing the least amount of variability, and samples within the BCAT_1 and BCAT_12 groups clustering further away from one another. In addition to the samples clustering into their respective groups, each of the groups seemed to cluster distinctly from one another. Interestingly, the BCAT_12 samples appeared to cluster furthest from the wild-type controls, and the BCAT_1 samples were clustered at an intermediate position between the two groups. These results appear to reflect observations in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mouse models, where the phenotypic changes exhibited in mice with simultaneous depletion of *Runx1* and *Runx2* appeared even more drastic than those observed with loss of *Runx1* alone.

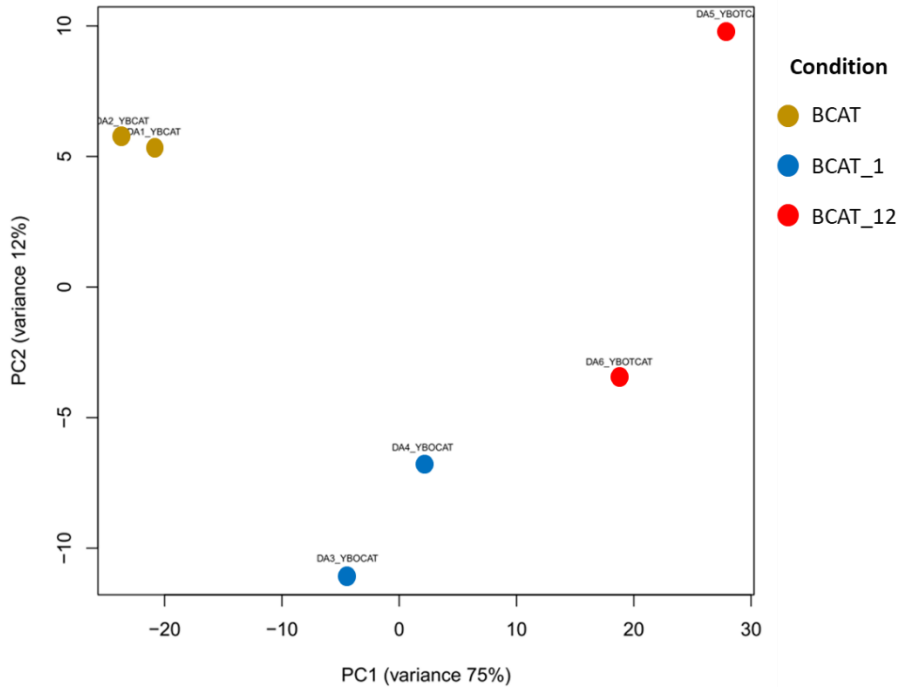


Figure 3.4 RFP-sorted BCAT_1 and BCAT_12 samples cluster distinctly from wild-type controls, and each other, in pilot experiment.

A PCA was plotted for the RNA-seq data obtained from RFP-sorted MMECs (isolated from BCAT, BCAT_1, and BCAT_12 cohort mice) set up in a pilot experiment. There were n=2 samples for each experimental group. Each data point represents a pool of RFP-positive MMECs isolated from 3 individual mice, detailed in Table 3.1.

Following on from PCA analysis, the gene expression data of the MMECs was also examined to determine whether *Runx1* alteration initiated significant changes in their gene expression profiles. Initially, the data was interrogated to find and compare the numbers of altered genes within each of the *Runx1*-depleted test groups compared with the wild-type control group. Venn diagrams, shown in Figure 3.5, summarise the results from this analysis and demonstrate that both the BCAT_1 and BCAT_12 groups display significant changes in the expression of substantial numbers of genes compared to the BCAT MMEC samples. Figure 3.5A shows the total number of significantly altered genes, irrespective of the direction of change, while Figures 3.5B and 3.5C illustrate the genes shown to be upregulated and downregulated, respectively. Each Venn diagram displays the number of genes altered uniquely within the BCAT_1 and BCAT_12 groups (labelled A and B, respectively) and the significantly altered genes shared between the two test groups (A∩B). Both test groups show significant changes in the expression of many genes, in both directions. While both groups display unique gene changes, many of the total changes seen in the BCAT_1 group are shared with those differentially expressed in the BCAT_12 samples, and the latter possesses considerably more unique alterations. Looking at the numbers of altered genes in Figure 3.5A, while 1148 genes were differentially expressed in the BCAT_1 group compared to the BCAT control cohort, a relatively small proportion of these (305 genes) were unique to the BCAT_1 group and the rest (843 genes) were also altered in the BCAT_12 samples. In contrast, the majority of the 2898 genes that were altered in the BCAT_12 group, as compared with their wild-type counterparts, were unique changes exhibited within these samples (2055 genes).

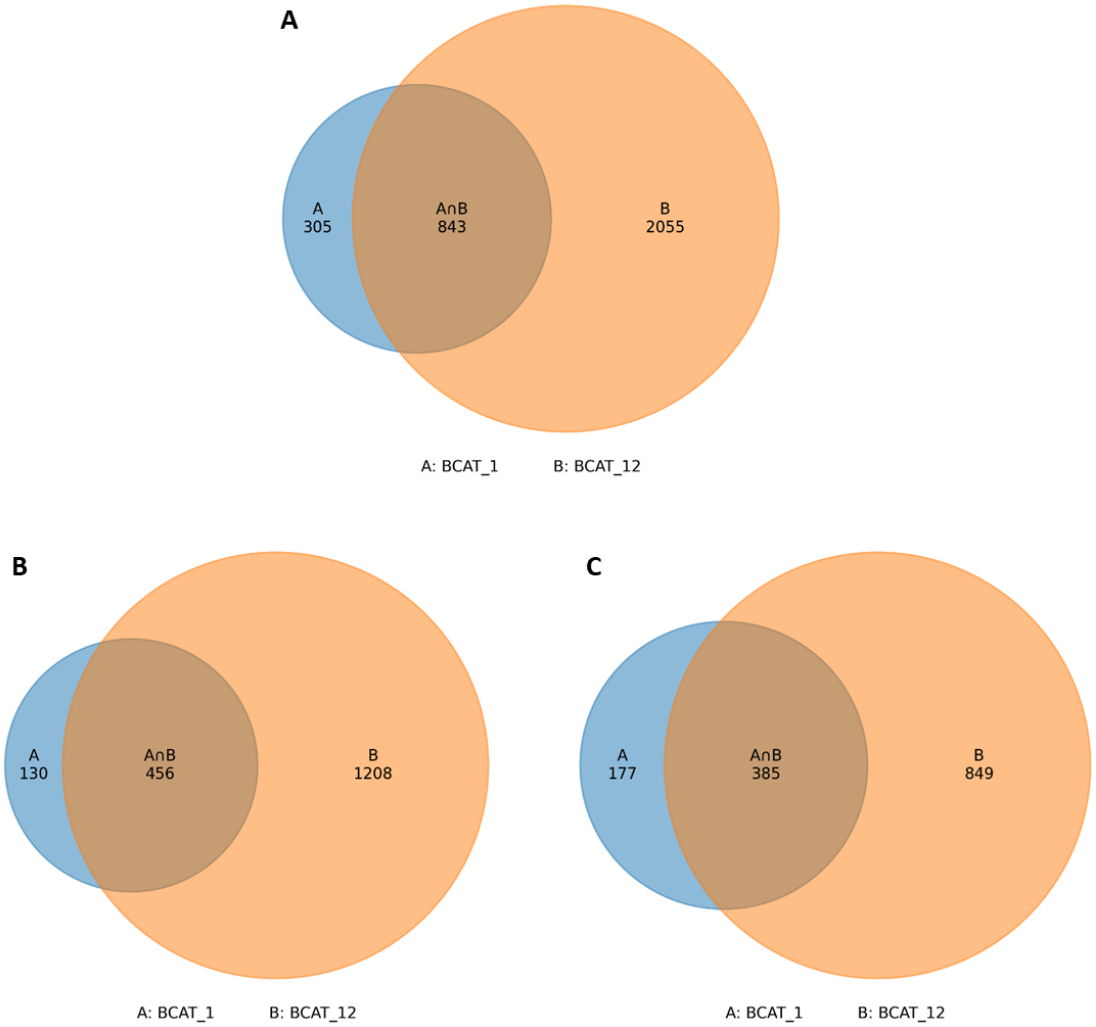


Figure 3.5 Venn diagrams of altered genes from RNA-seq analysis of *Runx1* and *Runx1/Runx2* deleted MMECs.

Pilot RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=2), BCAT_1 (n=2), and BCAT_12 (n=2) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from 3 individual mice (detailed in Table 3.1). Data obtained from the pilot RNA-seq experiment was analysed for significantly altered genes (genes with Fold Change ≤−1.5 or ≥1.5, and Adjusted p value of <0.05) in each of the experimental groups (BCAT_1 and BCAT_12) compared with the wild-type (BCAT) control group, and plotted in Venn diagrams. A = genes uniquely altered in BCAT_1 samples. B = genes uniquely altered in BCAT_12 samples. A∩B = genes altered in both sample groups.

[A] Analysis of all significantly altered genes (up or down) in the BCAT_1 and BCAT_12 groups, summarised in a Venn diagram.

[B] Venn diagram summarising the analysis of significantly upregulated genes in the BCAT_1 and BCAT_12 experimental groups.

[C] Venn diagram of significantly downregulated genes in the BCAT_1 and BCAT_12 test groups.

While analysing the gene expression data, it was noted that there were changes in a variety of genes relating to pathways and functions relevant to the *Blg-Cre; Catnb*$^{wt/lox(Ex3)}$*;tdRFP* mouse cohorts and their phenotypes. *Runx1* loss in this model appeared to be associated with differential expression of a number of genes associated with the Wnt pathway. These pathway changes were further explored to determine if they could be related to the more aggressive phenotypes in BCAT_1 and BCAT_12 mice. Previous work from the Sansom lab produced a list of definitive Wnt target genes that were well-characterised within colorectal cancer models (Sansom, Meniel et al. 2007). This list of Wnt target genes forms a gene set within the GSEA database. As there was no analogous list available for breast cancer, an investigation was conducted into which Wnt targets might be relevant within this context by validating a selection of targets from the Sansom Wnt target list in mammary-specific mouse models of Wnt pathway activation. To achieve this, MMECs, isolated from pregnant *Blg-Cre;Catnb*$^{wt/wt}$ and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice, were analysed by qPCR for their relative expression of several Wnt pathway-related genes, which were selected from the SANSOM_WNT_PATHWAY_REQUIRE_MYC gene set from the GSEA database. The results of this investigation into relevant Wnt targets in the mammary gland are summarised in Table 3.2 below, and box plots of gene expression shown in Appendix 1.

Table 3.2 List of Wnt target genes whose expression was compared in MMECs isolated from pregnant Blg-Cre;*Catnb*$^{wt/wt}$ and Blg-Cre;*Catnb*$^{wt/lox(Ex3)}$ mice, by qPCR.

Gene	Significantly altered (Y/N)	Direction of alteration (Up/Dn)	P-value	Significance
Axin2	Y	Up	0.0051	**
Ccnd1	N	N/A	0.3299	N/A
Lgr5	Y	Up	0.0076	**
Notum	Y	Up	0.0007	***
Sox9	N	N/A	0.2365	N/A
Znrf3	N	N/A	0.75	N/A
Dkk1	Y	Up	0.0393	*
Rnf43	N	N/A	0.7582	N/A
Wif1	N	N/A	0.1044	N/A
Wnt3a	N	N/A	0.6777	N/A

The aforementioned list of validated Wnt target genes was investigated in the RNA-seq pilot data. Normalised counts were plotted for each gene, for each of the

3 experimental groups, in box plots shown in Figures 3.6A to 3.6J, below. Expression data of additional Wnt pathway-associated genes selected from GSEA Wnt pathway-related gene sets (including the Sansom Wnt target gene set), which were not included in the initial validation experiments, are shown in Figures 3.6K to 3.6O.

Several of the investigated genes are demonstrably upregulated in either one of both of the BCAT_1 and BCAT_12 groups compared to their *Runx* wild-type controls (BCAT). Generally, the expression of these genes was more significantly upregulated in the BCAT_12 samples compared to BCAT_1 samples. Notably, *Lgr5* (whose expression was upregulated in MMECs upon oncogenic activation of ß-catenin, shown in Figure 3.2 and Appendix 1) was found to be significantly upregulated upon simultaneous loss of *Runx1* and *Runx2* function (Figure 3.6C). *Lgr5* is known to potentiate the Wnt/ß-catenin signalling pathway in various cell types and thereby promote several tumorigenic properties (tumour formation; cancer stem cell proliferation and self-renewal; cancer cell mobility; and epithelial-mesenchymal transition (EMT)) (Xu, Lin et al. 2019). Analyses also revealed elevated expression levels of a Wnt pathway target gene, *Mmp9* (Figure 3.6N), whose expression is induced by Wnt activation (Wu, Crampton et al. 2007, Ingraham, Park et al. 2011) and is a negative prognostic marker in colorectal cancer (Lee, Park et al. 2014) and breast cancer (Joseph, Alsaleem et al. 2020, Jiang and Li 2021). *Mmp9* expression was significantly enhanced in BCAT_1 and BCAT_12 test groups. Gene expression of *Mmp12*, which promotes ß-catenin expression (Li, Zhou et al. 2021) and correlates with poor prognosis in multiple tumour types (Ng, Qi et al. 2011, Klupp, Neumann et al. 2016, Ella, Harel et al. 2018), was also upregulated in the BCAT_12 group (Figure 3.6O).

Notum and *Sox9*, genes shown to be upregulated or unchanged in MMECs with oncogenic Wnt pathway activation in Table 3.2 and Appendix 1, exhibited reduced expression in BCAT_1 MMECs (Figures 3.6D and 3.6E). Both of these genes have been shown to repress the Wnt signalling pathway in a negative feedback loop (Giráldez, Copley et al. 2002, Topol, Chen et al. 2009, Sinha, Fan et al. 2021),

indicating a dysfunction of the regulatory mechanisms of the canonical Wnt pathway upon loss of functional *Runx1* in the mammary gland.

This data suggests that the oncogenic dysregulation of the canonical Wnt/ß-catenin pathway is exacerbated upon loss of *Runx1*, and even more so with the additional dysfunction of *Runx2*, due to various alterations in the genetic landscape of the signalling pathway that both enhance its activity and perturb its regulatory mechanisms. These alterations may be contributing to the more aggressive phenotype in both of these genetically altered mouse models of mammary tumorigenesis.

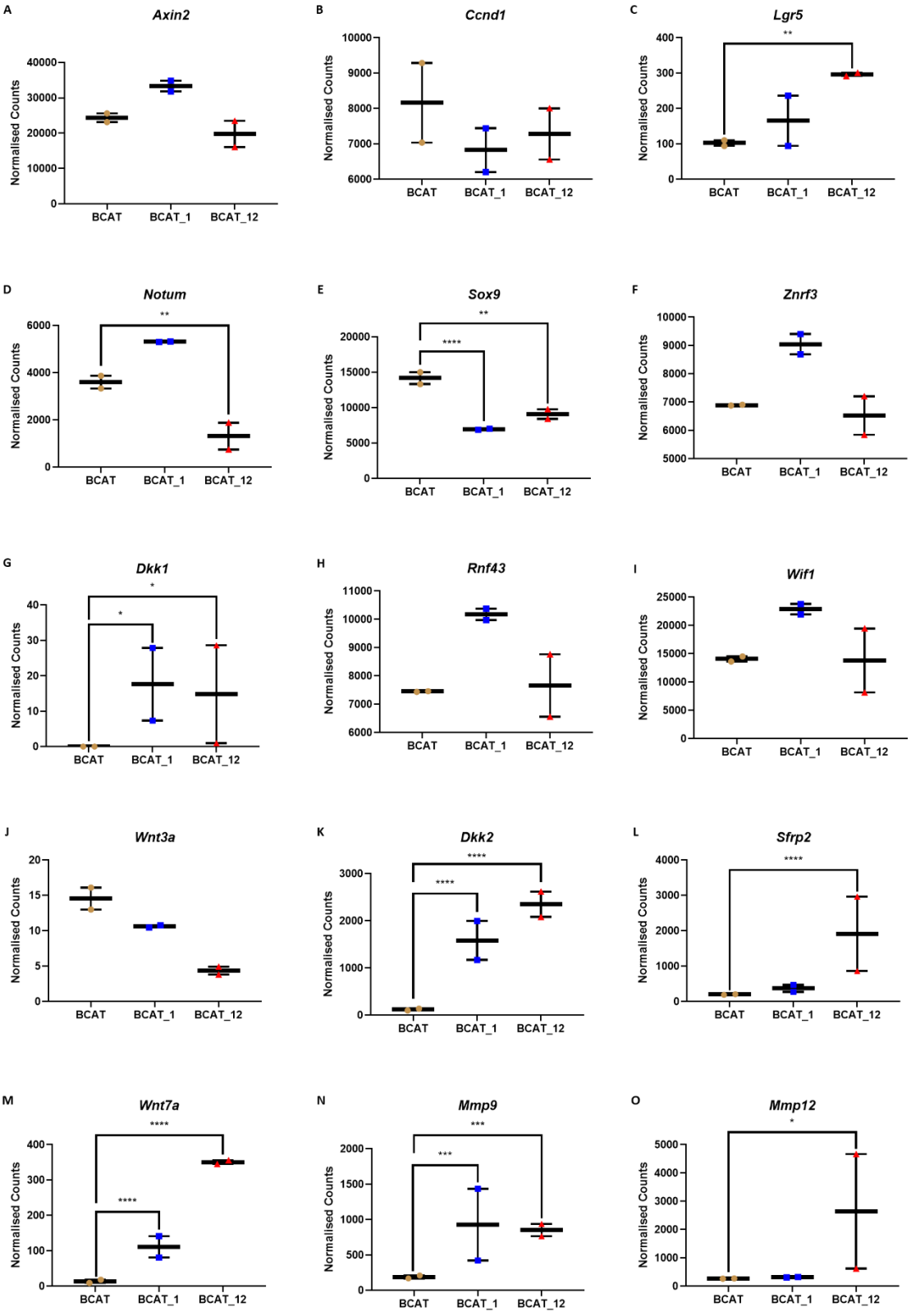

A
Axin2
Normalised Counts
40000
30000
20000
10000
0
BCAT
BCAT_1
BCAT_12
B
Ccnd1
Normalised Counts
10000
9000
8000
7000
6000
BCAT
BCAT_1
BCAT_12
C
Lgr5
**
Normalised Counts
400
300
200
100
0
BCAT
BCAT_1
BCAT_12
D
Notum
**
Normalised Counts
6000
4000
2000
0
BCAT
BCAT_1
BCAT_12
E
Sox9
**
****
Normalised Counts
20000
15000
10000
5000
0
BCAT
BCAT_1
BCAT_12
F
Znrf3
Normalised Counts
10000
9000
8000
7000
6000
5000
BCAT
BCAT_1
BCAT_12
G
Dkk1
*
*
Normalised Counts
40
30
20
10
0
BCAT
BCAT_1
BCAT_12
H
Rnf43
Normalised Counts
11000
10000
9000
8000
7000
6000
BCAT
BCAT_1
BCAT_12
I
Wif1
Normalised Counts
25000
20000
15000
10000
5000
0
BCAT
BCAT_1
BCAT_12
J
Wnt3a
Normalised Counts
20
15
10
5
0
BCAT
BCAT_1
BCAT_12
K
Dkk2
****
****
Normalised Counts
3000
2000
1000
0
BCAT
BCAT_1
BCAT_12
L
Sfrp2
****
Normalised Counts
4000
3000
2000
1000
0
BCAT
BCAT_1
BCAT_12
M
Wnt7a
****
****
Normalised Counts
400
300
200
100
0
BCAT
BCAT_1
BCAT_12
N
Mmp9
***
***
Normalised Counts
2000
1500
1000
500
0
BCAT
BCAT_1
BCAT_12
O
Mmp12
*
Normalised Counts
5000
4000
3000
2000
1000
0
BCAT
BCAT_1
BCAT_12

Figure 3.6 Significant alterations in Wnt signalling genes of ß-catenin-activated MMECs with loss of *Runx* genes.

RNA-seq analysis of RFP-positive MMECs isolated from BCAT, BCAT_1, and BCAT_12 mice at 9 weeks of age. Normalised gene counts and adjusted p values were calculated for each of the genes analysed for the pilot RNA-seq experiment. Various Wnt pathway-related genes were analysed and plotted in the above box and whisker plots, with each data point representative of the normalised gene counts from a single sample. n=2 samples per cohort and each sample is a pool of RFP-positive MMECs isolated from 3 mice. Significance of comparisons with the BCAT group are shown in each plot [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$]. Wnt targets that were previously analysed in MMECs isolated from pregnant Blg-Cre;*Catnb*$^{wt/lox(Ex3)}$ mice (Table 3.2 and Appendix 1) are shown in [A-J], and additional Wnt pathway-associated genes are plotted in [K-O].

RNA-seq data analysis also revealed significant changes in genes known to be involved in the regulation of stem cells in both the normal and tumorigenic breast (Figure 3.7). Similar to gene changes observed in Wnt pathway targets, most of the investigated stem-associated genes appeared to be more significantly altered in BCAT_12 MMECs compared to BCAT_1 samples. Expression of *Snai2* was found to be highly elevated in MMECs with the simultaneous depletion of *Runx1* and *Runx2*, as in the BCAT_12 mice (Figure 3.7D). This gene was previously demonstrated to be involved in normal stem cell function and various cancer stem cell-like behaviours in the breast (Bhat-Nakshatri, Appaiah et al. 2010, Guo, Keckesova et al. 2012). *Axl* was also expressed at significantly higher levels in BCAT_1 and BCAT_12 MMECs compared to their BCAT wild-type counterparts (Figure 3.7B). *Axl* is a promoter of stemness in the normal mammary gland and its expression is required for the reconstitution of cleared mammary fat pads. This gene was also indicated to drive stemness and metastasis of breast cancer cells, and thus reduce overall patient survival (Gjerdrum, Tiron et al. 2010, Engelsen, Wnuk-Lipinska et al. 2020). Additionally, the *Aldh1a1* gene and its rodent-specific paralogue, *Aldh1a7*, were highly upregulated in both *Runx1*-depleted, ß-catenin-activated experimental groups compared to the BCAT wild-type controls (Figure 3.7E and 3.7F). Previous publications revealed that *Aldh1a1* expression was correlated with tumour initiation within populations of breast cancer stem cells and promoted invasion and metastasis of breast cancer cells both *in vitro* and *in vivo* (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016). High mRNA levels of *Aldh1a1* in breast tumours were also correlated with poorer overall survival in breast cancer patients (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016). A further gene alteration of particular note was the significant increase in *Bcl11b* expression. This gene is involved in maintenance and self-renewal of mammary stem cell populations (Cai, Kalisky et al. 2017, Miller, Jin et al. 2018). *Bcl11b* is a close paralogue of *Bcl11a*, a gene that has been implicated in the functions of stem and progenitor cells within the triple negative subtype of breast cancer and has been shown to promote various aggressive behaviours in cancer cells (Khaled, Choon Lee et al. 2015, Seachrist, Hannigan et al. 2020, Wang, Xu et al. 2020).

These findings reinforce the theory that the earlier emergence of independently arising tumours in *Runx1*-depleted mice, a phenotype that was intensified with

the additional loss of *Runx2*, is reflective of an enriched stem-like sub-population of MMECs within the mammary gland.

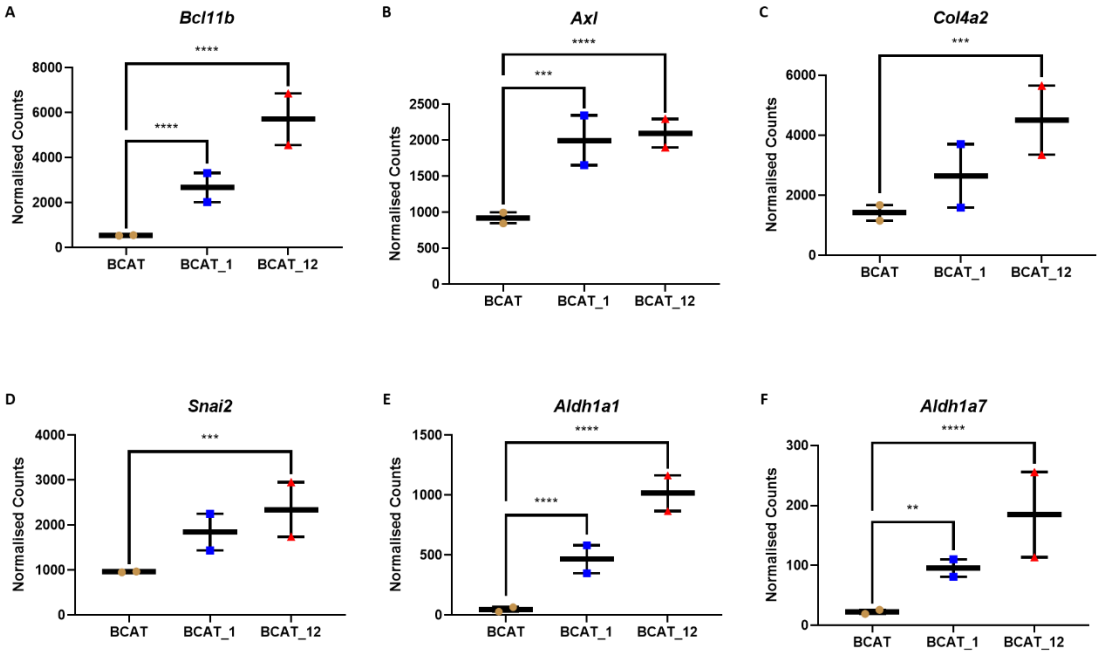


Figure 3.7 Upregulation of genes associated with normal mammary and breast cancer stem cells in BCAT_1 and BCAT_12 samples.

RNA-seq analysis of RFP-positive MMECs isolated from BCAT, BCAT_1, and BCAT_12 mice at 9 weeks of age. Normalised gene counts and adjusted p values were calculated for each of the genes analysed for the pilot RNA-seq experiment. Various genes, associated with stemness of the normal breast and breast cancer cells, were analysed and plotted in the above box and whisker plots [A-F]. Each of the plotted data points represents the normalised gene count obtained from a single sample. n=2 samples per cohort and each sample is a pool of RFP-positive MMECs isolated from 3 mice. The significance of each comparison with the BCAT group is also shown in each plot [** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$].

3.2.2 Increased sample-to-sample variation in extended RNA-seq experiment.

Following on from this promising pilot experiment, MMECs were isolated from additional experimental mice in order to generate test groups with larger sample numbers. This was essential to ensure adequate precision of the estimated changes in gene expression. A larger sample size was also necessary in order to increase the power of the study and the associated analyses, interpretations, and conclusions deduced from this experiment. The 3 sample groups analysed in the pilot were investigated in this extended RNA-seq experiment with sample sizes of $n \geq 4$. 2 of these were the samples analysed within the pilot experiment, discussed

above, and the rest were newly generated samples. An additional BCAT_2 test group was also introduced, however with a smaller sample size of n=2 due to a lack of available mice. As with samples used for the pilot experiment, each of the additional samples collected for the extended experiment was comprised of a combination of FACS-sorted (RFP+) MMECs isolated from multiple mice. The details of each sample, including RNA integrity numbers (RIN), are listed in Table 3.3. The samples were processed and analysed according to the methods described in 2.5 for RNA-seq.

Table 3.3 Additional samples, generated from pooled RFP-sorted (BCAT, BCAT_1, BCAT_2, and BCAT_12) MMECs, used for an extended RNA-seq analysis.

Additional Sample of Pooled RFP+ MMECs Individual Mice	**Total Number of Pooled RFP+ MMECs** Number of MMECs Collected Per Mouse	**RIN Value**
BCAT sample 3 — 6 mice	**~53,000**	**2.9**
YDG167.2c	~11,000	
YDG167.5e	~8,000	
YDG167.5h	~8,000	
YDG168.5g	~6,000	
YDG184.2g	~9,000	
YDG184.4f	11,237	
BCAT sample 4 — 9 mice	**~9000**	**2.6**
YDG167.3g	~750	
YDG167.3j	~630	
YDG167.4d	672	
YDG167.4h	339	
YDG168.2d	1,190	
YDG168.4f	1,011	
YDG170.2h	1,441	
YDG170.3h	~1,400	
YDG184.1g	1,618	
BCAT_1 sample 3 — 3 mice	**~54,000**	**2.9**
YDD118.1h	~29,000	
YDD119.8d	~14,000	
YDD119.8e	~11,000	
BCAT_1 sample 4 — 3 mice	**~107,000**	**8.5**
YDD116.5a	~7,000	
YDD116.7b	~95,000	
YDD117.2h	~5,000	
BCAT_1 sample 5 — 8 mice	**~65,000**	**2.7**
YDD116.7c	~29,000	
YDD117.6a	~19,000	
YDD119.7i	~8,000	
YDD116.1f	4,404	
YDD116.2h	1,221	
YDD116.6j	~700	
YDD118.1e	~1,500	
YDD118.4e	~1,000	

***continued on next page**

Table 3.3 (continued) Additional samples, generated from pooled RFP-sorted (BCAT, BCAT_1, BCAT_2, and BCAT_12) MMECs, used for an extended RNA-seq analysis.

Additional Sample of Pooled RFP+ MMECs Individual Mice	**Total Number of Pooled RFP+ MMECs** Number of MMECs Collected Per Mouse	**RIN Value**
BCAT_2 sample 1 — 3 mice	**~84,000**	**7.9**
YDG145.6d	35,698	
YDG150.2b	29,558	
YDG155.4b	18,575	
BCAT_2 sample 2 — 5 mice	**~37,000**	**3.1**
YDG161.1e	3,433	
YDG161.2d	27,111	
YDG165.5h	728	
YDG166.3e	4,048	
YDG182.3d	~1,500	
BCAT_12 sample 3 — 5 mice	**~85,000**	**2.8**
YDE100.3d	~34,000	
YDE101.6e	~17,000	
YDE104.2h	~10,000	
YDE105.3f	~11,000	
YDE108.3a	~13,000	
BCAT_12 sample 4 — 3 mice	**~79,000**	**7.2**
YDE101.6c	~67,000	
YDE105.1h	9,792	
YDE105.3g	~2,500	
BCAT_12 sample 5 — 3 mice	**~94,500**	**5.8**
YDE101.4i	5,021	
YDE104.3c	~89,000	
YDE105.6a	~640	

It is important to note that the majority of the additional samples collected for this extended experiment required a combination of MMECs extracted from a larger number of mice as compared to the pilot samples (≥4 mice per sample). This was due to a larger degree of variability in the number of sorted RFP+ cells obtained from the mammary glands of each experimental mouse. There was also a higher degree of sample-to-sample variability in the RIN values, with several of these being lower than the ideal RIN value of 7. It is important to keep these factors in mind while analysing these results, as this variation in the number of mice, whose MMECs were combined together to create each sample, could result in increased sample to sample variability, even within the same experimental group.

A PCA was initially conducted on sample data from this extended RNA-seq experiment and plotted in order to observe which samples are likely to be the most similar or dissimilar (Figure 3.8). Each of the pilot samples (S01 to S06) appeared to cluster into their own distinct experimental groups on the PCA plot. However, some of the additional samples (S07 to S16) did not cluster as closely together with the pilot samples for the same group. This was particularly true for the BCAT and BCAT_1 additional samples, which required a combination of MMECs from significantly more mice than the pilot samples. While there is a certain degree of clustering within each of the experimental groups, and each group does appear to cluster away from each other, this increased sample variation within the test groups could affect the statistical analyses of this data. This is a possibility, particularly given that the additional samples cluster away from their paired pilot samples on the PC1 axis (which accounts for most of the variation in the data) to a greater extent than the BCAT and BCAT_1 samples cluster away from each other on the PC2 axis. This would therefore indicate that the pilot and additional samples within one given test group (e.g. BCAT) are likely to be more different compared to each other than compared to the respective pilot and additional samples for a different test group (e.g. BCAT_1). Despite this increased variability in these sample groups, the BCAT_12 samples appeared to cluster rather closely to one another and distinctly away from samples of other groups. They were indicated to be most significantly distinct from samples within the BCAT group, and the BCAT_1 samples were plotted at a position intermediate to this.

These results are mostly in agreement with the PCA plot for the pilot experiment, and both appear to reflect phenotypic observations made in the ß-catenin-activated mouse models, whereby the loss of *Runx1* induced moderate changes that were significantly amplified with further loss of *Runx2*.

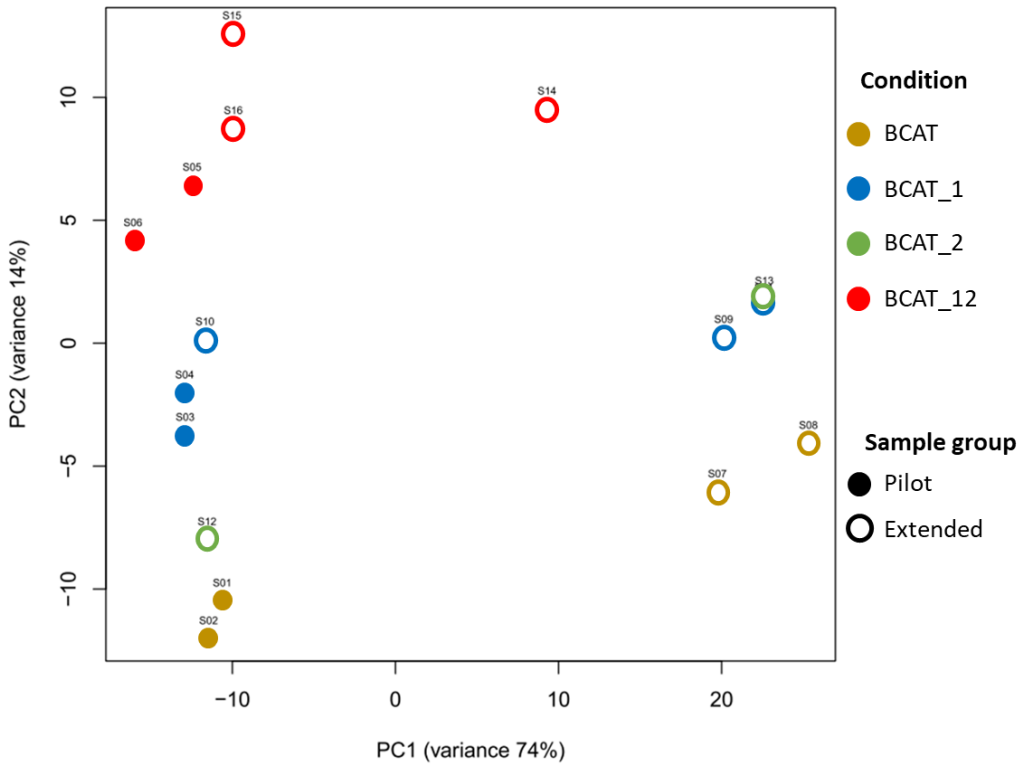


Figure 3.8 Clustering of experimental groups within extended analysis of RFP-sorted MMECs.

PCA plot for extended RNA-seq analysis of RFP-sorted *Blg-Cre;Catnb*$^{wt/lox(ex3)}$*;tdRFP* sample groups, isolated from 9 week old mice. n=4 BCAT samples; n=5 BCAT_1 samples; n=2 BCAT_2 samples; N=5 BCAT_12 samples. Each data point represents a pool of RFP-positive MMECs isolated from ≥3 individual mice, detailed in Table 3.1 (pilot) and Table 3.2 (extended/additional). Samples S01 to S06 (filled circles) were the same samples analysed in the pilot RNA-seq experiment (3.2.1), while S07 to S16 (hollow circles) were additional samples.

3.2.3 Significantly more genes uniquely altered in β-catenin-activated MMECs with simultaneous Runx1/Runx2 loss compared to *Runx1* loss alone.

As with the pilot experiment, gene expression data generated from the extended RNA-seq experiment was studied to compare the numbers of altered genes within each of the test groups compared to the wild-type control. During this investigation, it was revealed that the BCAT_2 mammary samples did not show any significant variation in gene expression compared to the controls for any of the analysed genes. While surprising, this supports observations relating to the phenotypes of these mice, which were virtually indistinguishable from that of their wild-type counterparts. Due to this finding, all subsequent analyses were conducted on the *Runx1*-depleted experimental groups and investigating the genes differentially expressed in each of these 2 groups compared to the wild-type group. The results of these initial gene expression analyses, comparing numbers of gene alterations in the BCAT_1 and BCAT_12 samples compared to the controls, were plotted in Venn diagrams (Figure 3.9). Similar to the results plotted for the pilot experiment (Figure 3.5), both test groups were shown to harbour significant gene alterations (both upregulation and downregulation) in a number of genes. There were also significantly more genes that were differentially expressed in the BCAT_12 MMECs, in which 1312 genes were altered, compared to those in the BCAT_1 group, where 88 genes were differentially expressed. Additionally, a large proportion of genes altered in the BCAT_1 group were shared with the BCAT-12 samples, with 75 of the 88 altered genes being shared, leaving only 13 unique to the BCAT_1 group. In contrast, the majority of gene alterations exhibited in BCAT_12 samples (1237 out of 1312 total genes) were unique to this group. These findings make sense within the context of the phenotypes observed within the mouse models, and indicate that the more extreme phenotype observed in BCAT_12 mice is reflective of the dramatically altered genomic landscape within the mammary epithelial cells of these mice.

While these results are in agreement with those shown in Figure 3.5, one notable difference was that there were significantly less genes found to be altered in this extended experiment compared to the pilot. Although it is normal for many comparisons to become less statistically significant with the addition of new samples, there were some considerations to take into account for this data.

Notably, the BCAT_1 group appeared to be disproportionately affected by this increase in sample size and exhibited a ~13 times reduction in the number of altered genes between experiments, while the difference in BCAT_12 samples was around ~2.2 times. The BCAT_1 samples showed significantly more sample-to-sample variation upon the inclusion of these additional samples, while the BCAT_12 samples remained relatively well-clustered (Figure 3.8), which could be causing this significant shift in the ratio of uniquely altered genes between these two groups.

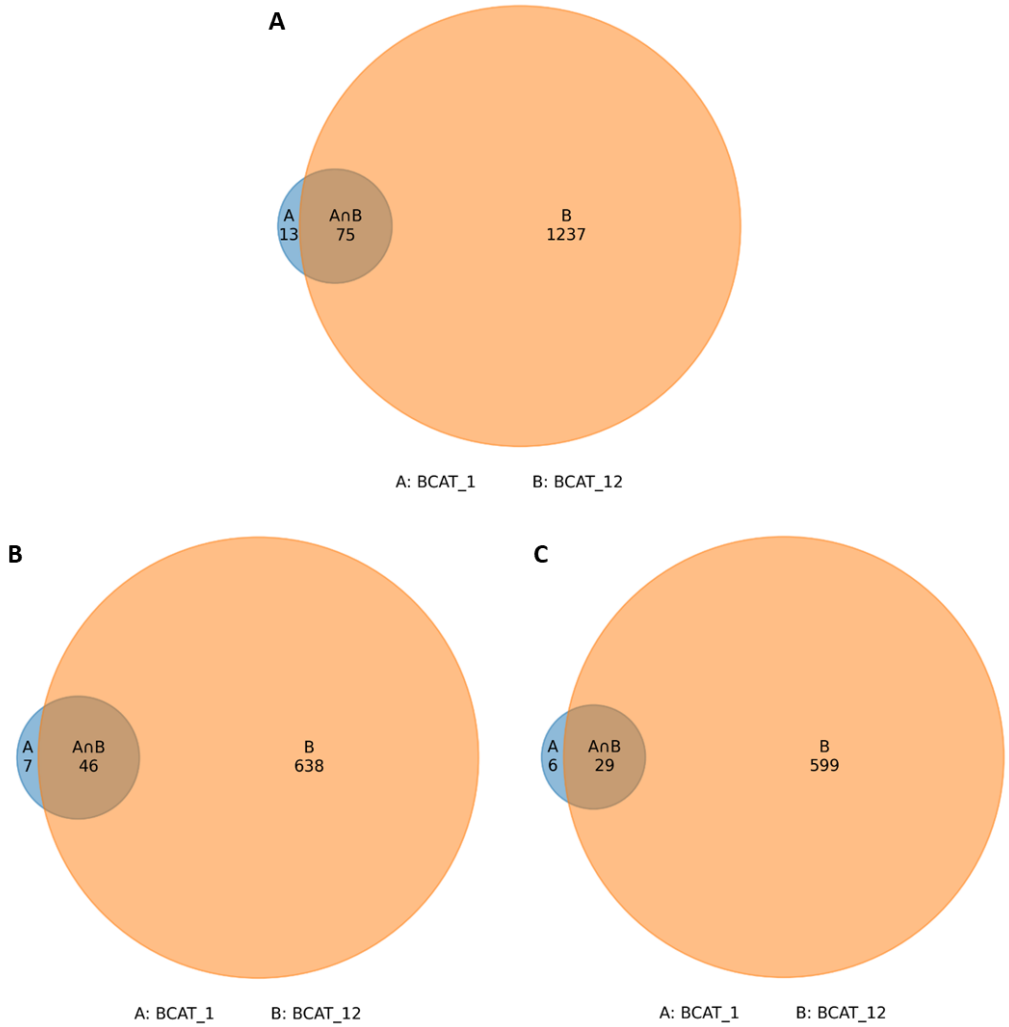


Figure 3.9 Venn diagrams of gene changes when Runx genes are deleted in ß-catenin-activated MMECs.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). Data obtained from the extended RNA-seq experiment was analysed for significantly altered genes (genes with Fold Change ≤−1.5 or ≥1.5, and Adjusted p value of <0.05) in each of the experimental groups (BCAT_1 and BCAT_12) compared with the wild-type (BCAT) control group, and plotted in Venn diagrams. A = genes uniquely altered in BCAT_1 samples. B = genes uniquely altered in BCAT_12 samples. A∩B = genes altered in both sample groups.

[A] Analysis of all significantly altered genes (up or down) in the BCAT_1 and BCAT_12 groups, summarised in a Venn diagram.

[B] Venn diagram summarising the analysis of significantly upregulated genes in the BCAT_1 and BCAT_12 experimental groups.

[C] Venn diagram of significantly downregulated genes in BCAT_1 and BCAT_12 test groups.

3.2.4 Various gene sets within the Hallmark collection are enriched in *Runx1*-depleted MMECs.

Gene Set Enrichment Analysis (GSEA) is a powerful analytical method that can be used to interpret complex gene expression data by comparing data at the level of gene sets (groups of genes that share biological functions and other common features) as opposed to focusing on single genes. This allows for a more informative, and relatively unbiased, analysis of data compared to analyses focusing on ranked lists of individual genes, which are often overwhelming in their size and complexity, or difficult to interpret due to it being more challenging to identify unifying biological themes in the data. The Molecular Signatures Database (MSigDB) offers over 32880 gene sets, divided into 9 major collections, for the analysis and interpretation of gene expression data. One such collection is known as the Hallmark collection of gene sets, which is an assortment of gene sets that summarise and represent specific well-defined biological functions. By identifying genes that overlap and display coordinate expression in other MSigDB gene sets, definitive gene lists that best represent specific biological functions were generated and organised into the Hallmark gene set collection.

The RNA-seq data, generated through the analysis of the completed experimental groups, was analysed using GSEA of the Hallmark collection. The Hallmark gene set collection (comprised of 50 total gene sets) were analysed for their enrichment in BCAT_1 and BCAT_12 sample groups, compared to the BCAT wild-type control group. The results of this investigation are presented in Figure 3.10, which shows the number of shared and unique enriched gene sets for the *Runx1*-depleted test groups. Similar to results for individually altered genes, the BCAT_12 samples showed enrichment of a greater number of Hallmark gene sets overall (29 altered in the BCAT_12 group compared to 17 in the BCAT_1 group), with the vast majority of these being positively enriched/upregulated (27 upregulated versus 2 downregulated). A large proportion of the gene sets that were altered in the BCAT_1 group were shared with the BCAT_12 group (12 shared and 5 unique), whereas a small majority of the gene sets showing differential expression in the BCAT_12 group were unique to this group (17 unique and 12 shared).

Further analysis of the specific gene sets from the Hallmark collection (Figure 3.11) that showed significant enrichment of various gene sets relating to pathways whose deregulation are implicated in the initiation and progression of various cancer types (including E2F, MYC, P53, Hedgehog, and WNT). There was also enrichment of several gene sets relating to Hanahan and Weinberg's hallmarks of cancer and to tumour aggression/progression (EMT, angiogenesis, oxidative phosphorylation, glycolysis, DNA repair). For the gene sets that were discovered to be positively enriched in both *Runx1*-deficient test groups, there were generally more significant or dramatic changes observed with the additional loss of *Runx2*.

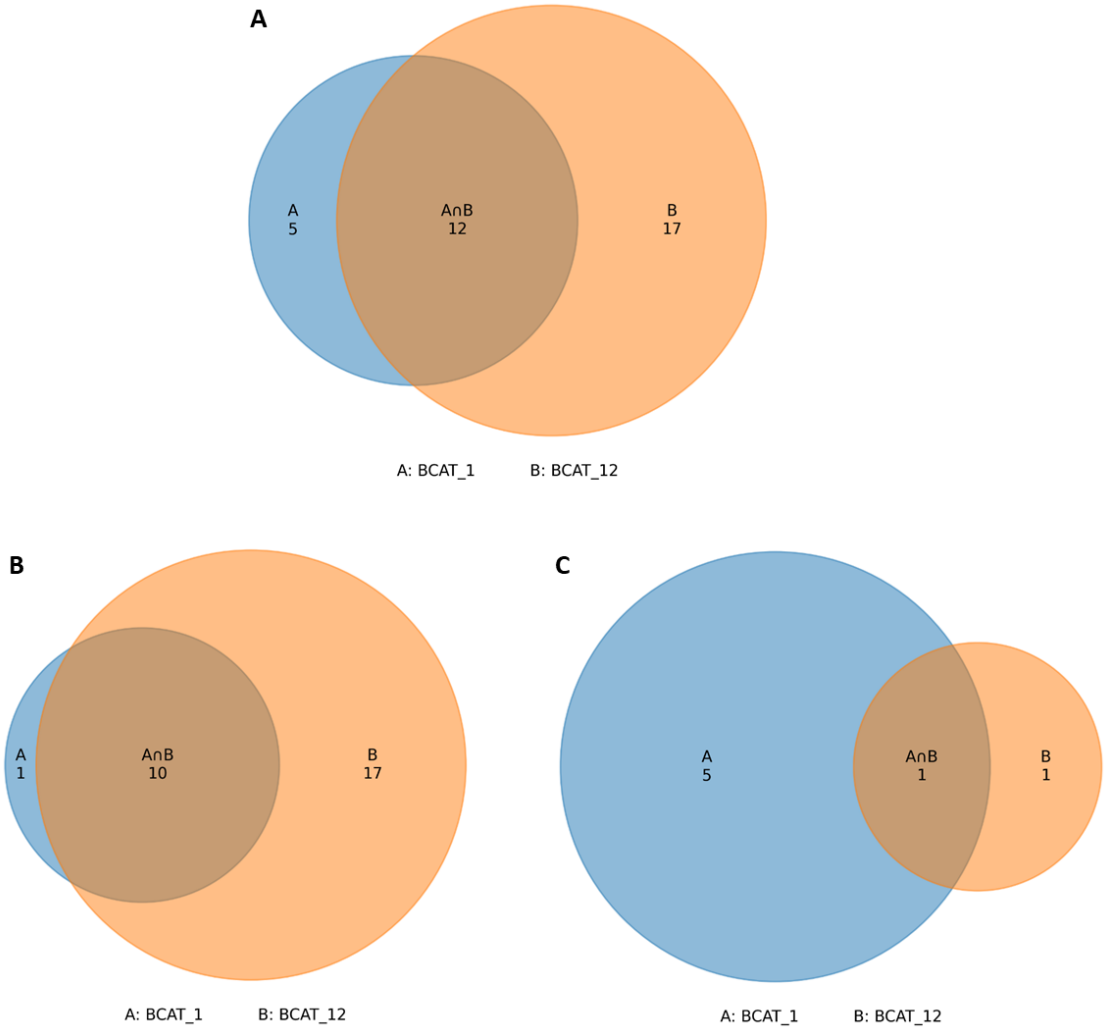


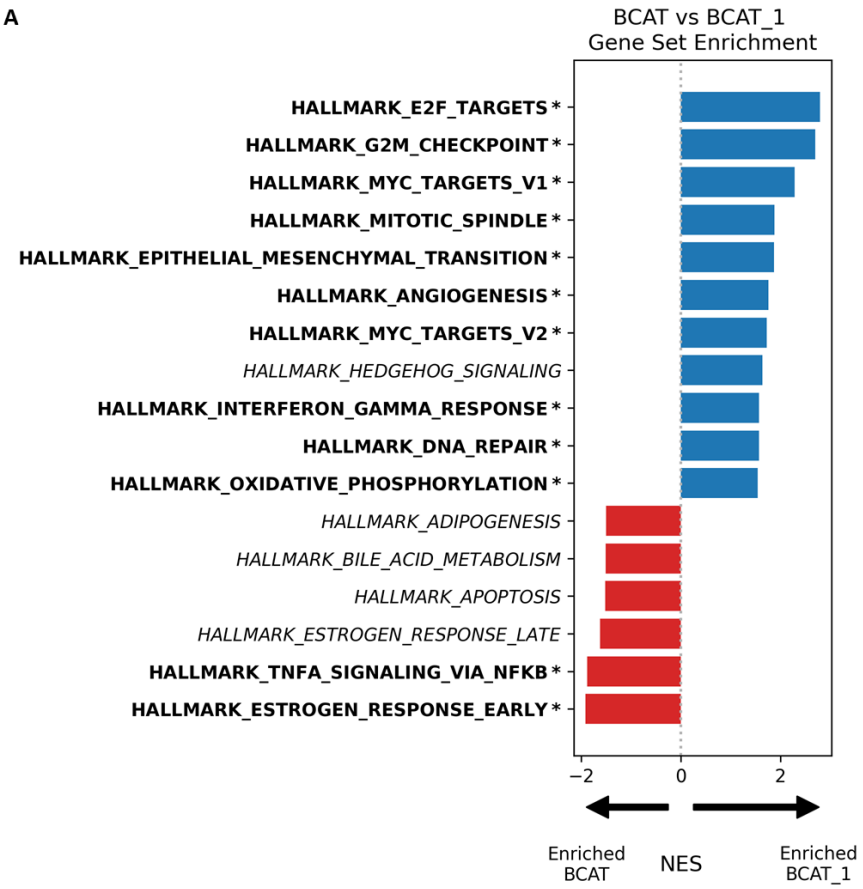
Figure 3.10 Venn diagrams summarising significantly altered Hallmark gene sets in BCAT_1 and BCAT_12 groups, and their overlap.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data was analysed for significantly altered Hallmark gene sets (NES ≤−1.5 or ≥1.5, FDR <0.05) in each of the experimental groups (BCAT_1 and BCAT_12) compared with the wild-type (BCAT) control group, and plotted in Venn diagrams. There are 50 total gene sets in the Hallmark gene set collection. A = gene sets uniquely altered in BCAT_1 samples. B = gene sets uniquely altered in BCAT_12 samples. A∩B = gene sets altered in both sample groups.

[A] Analysis of all significantly altered Hallmark gene sets (up or down) in the BCAT_1 and BCAT_12 groups, summarised in a Venn diagram.

[B] Summary of the significantly upregulated genes in the BCAT_1 and BCAT_12 experimental groups.

[C] Venn diagram of significantly downregulated genes in the BCAT_1 and BCAT_12 test groups.


A
BCAT vs BCAT_1
Gene Set Enrichment
HALLMARK_E2F_TARGETS*
HALLMARK_G2M_CHECKPOINT*
HALLMARK_MYC_TARGETS_V1*
HALLMARK_MITOTIC_SPINDLE*
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION*
HALLMARK_ANGIOGENESIS*
HALLMARK_MYC_TARGETS_V2*
HALLMARK_HEDGEHOG_SIGNALING
HALLMARK_INTERFERON_GAMMA_RESPONSE*
HALLMARK_DNA_REPAIR*
HALLMARK_OXIDATIVE_PHOSPHORYLATION*
HALLMARK_ADIPOGENESIS
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_APOPTOSIS
HALLMARK_ESTROGEN_RESPONSE_LATE
HALLMARK_TNFA_SIGNALING_VIA_NFKB*
HALLMARK_ESTROGEN_RESPONSE_EARLY*
−2
0
2
Enriched BCAT
NES
Enriched BCAT_1

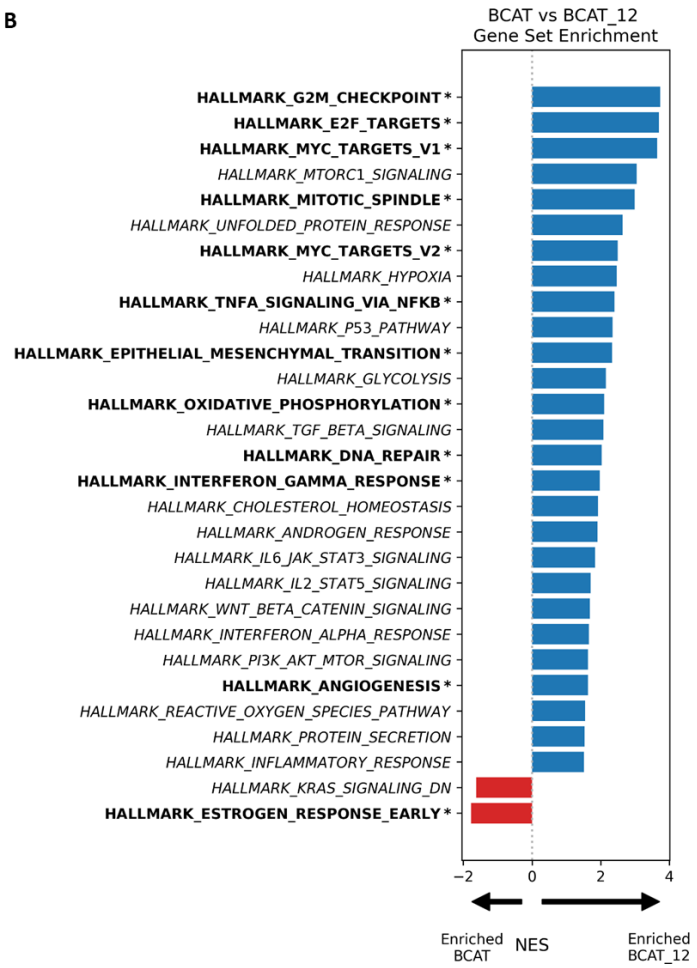

B
BCAT vs BCAT_12
Gene Set Enrichment
HALLMARK_G2M_CHECKPOINT*
HALLMARK_E2F_TARGETS*
HALLMARK_MYC_TARGETS_V1*
HALLMARK_MTORC1_SIGNALING
HALLMARK_MITOTIC_SPINDLE*
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_MYC_TARGETS_V2*
HALLMARK_HYPOXIA
HALLMARK_TNFA_SIGNALING_VIA_NFKB*
HALLMARK_P53_PATHWAY
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION*
HALLMARK_GLYCOLYSIS
HALLMARK_OXIDATIVE_PHOSPHORYLATION*
HALLMARK_TGF_BETA_SIGNALING
HALLMARK_DNA_REPAIR*
HALLMARK_INTERFERON_GAMMA_RESPONSE*
HALLMARK_CHOLESTEROL_HOMEOSTASIS
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_IL6_JAK_STAT3_SIGNALING
HALLMARK_IL2_STAT5_SIGNALING
HALLMARK_WNT_BETA_CATENIN_SIGNALING
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_PI3K_AKT_MTOR_SIGNALING
HALLMARK_ANGIOGENESIS*
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY
HALLMARK_PROTEIN_SECRETION
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_KRAS_SIGNALING_DN
HALLMARK_ESTROGEN_RESPONSE_EARLY*
−2
0
2
4
Enriched BCAT
NES
Enriched BCAT_12

Figure 3.11 Most significantly altered Hallmark gene sets in BCAT_1 and BCAT_12 mammary glands.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). BCAT_1 and BCAT_12 test groups were analysed for significantly altered Hallmark gene sets compared to their BCAT counterparts (NES ≤−1.5 or ≥1.5, FDR <0.05) and presented in the above bar charts. Gene sets significantly enriched in the ß-catenin-activated samples with *Runx1*-depletion, with positive NES values, are displayed in blue. Those enriched in the BCAT wild-type control group (i.e. those negatively enriched in the test group), with negative NES values, are shown in red. All gene sets significantly altered in the BCAT_1 group are summarised in [A]. All significantly altered groups from the BCAT_12 samples are summarised in [B]. Any gene sets alterations that are shared between the BCAT_1 and BCAT_12 groups are highlighted in bold text and marked with a "*" symbol.

3.2.5 Loss of functional *Runx1* in ß-catenin-activated glands associated with enrichment of a large number of Curated gene sets.

In addition to analysing the RNA-seq data at the level of the Hallmark collection of gene sets, it was also analysed using the MSigDB Curated collection of data sets. This collection is comprised of canonical pathways and experimental signatures curated from a multitude of different sources (including online pathway databases, the biomedical literature, and contributions from individual domain experts). Initial examinations of this data focused on the numbers of Curated gene sets found to be enriched in each of the BCAT_1 and BCAT_12 sample groups. The results of this analysis, including the numbers of gene sets either uniquely or commonly enriched within the BCAT_1 and BCAT_12 samples, are summarised in the Venn diagrams in Figure 3.12. In agreement with previous investigations into the differential expression of individual genes (Figure 3.9) and Hallmark gene sets (Figure 3.10), a much higher number of Curated gene sets was found to be enriched within the BCAT_12 sample group (1805 total gene sets altered) than was seen in the BCAT_1 group (685 total gene sets altered), with respect to their BCAT wild-type controls. The majority of alterations seen in the BCAT_1 group were shared with the BCAT_12 samples (with 593 of 685 gene sets shared and only 92 unique to the BCAT_1 group), while the majority of gene alterations exhibited in BCAT_12 samples (1212 out of 1805 total genes) were unique to this group.

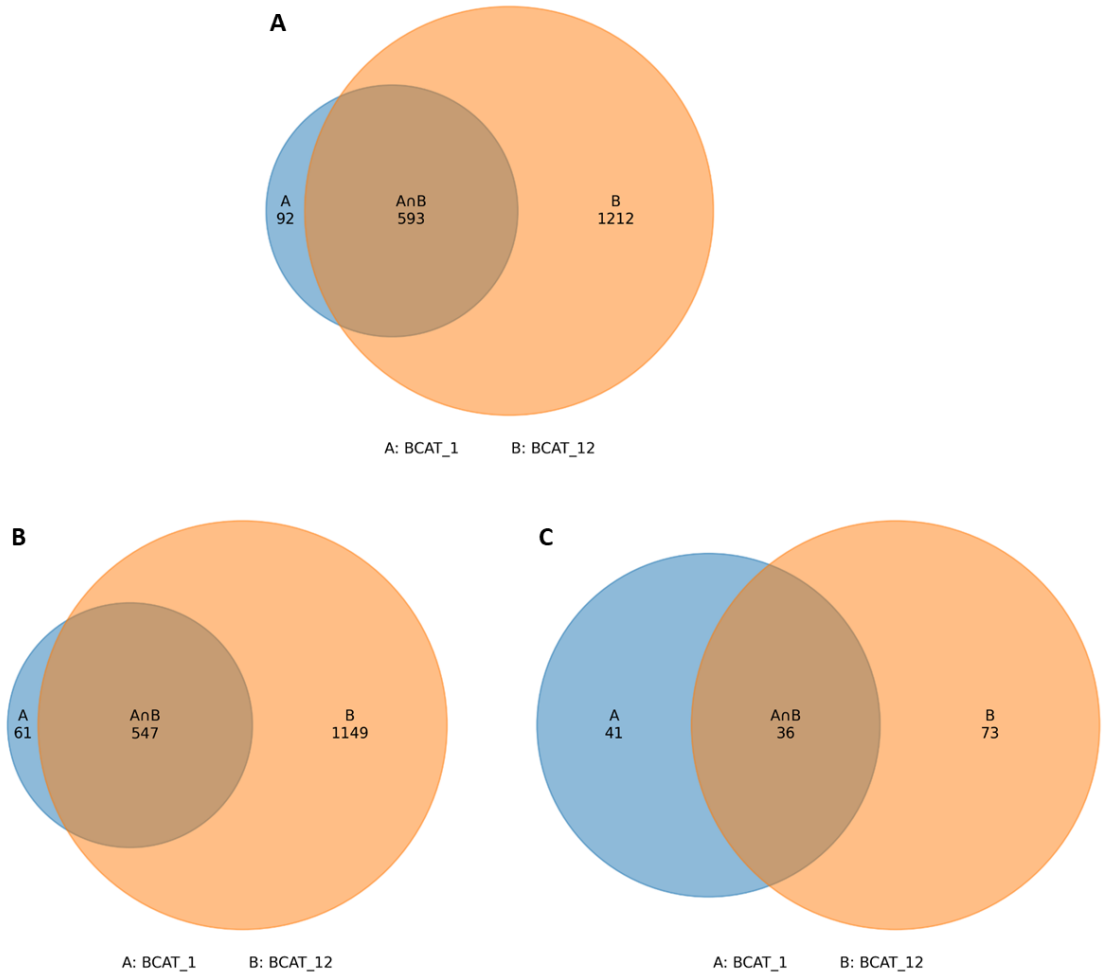


Figure 3.12 Venn diagrams summarising significantly altered Curated gene sets in BCAT_1 and BCAT_12 groups, and their overlap.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). Significantly altered Curated gene sets (NES ≤−1.5 or ≥1.5, FDR <0.05) from each of the RNA-seq test groups (BCAT_1 and BCAT_12) compared to the wild-type (BCAT) control, were plotted in Venn diagrams. There are 6366 total gene sets in the Curated gene set collection. A = gene sets uniquely altered in BCAT_1 samples. B = gene sets uniquely altered in BCAT_12 samples. A∩B = gene sets altered in both sample groups.

[A] Venn diagram summarising all significantly altered Curated gene sets (up or down) in the BCAT_1 and BCAT_12 groups.

[B] Summary of the significantly upregulated genes in the BCAT_1 and BCAT_12 experimental groups.

[C] Venn diagram of significantly downregulated genes in BCAT_1 and BCAT_12 test groups.

3.2.6 Enrichment of gene sets, associated with tumour cell aggression, in *Runx1*$^{fl/fl}$ and *Runx1*$^{fl/fl}$;*Runx2*$^{fl/fl}$ MMECs with activated ß-catenin.

The Curated collection is comprised of many gene sets sourced from various publications, in which the genetic landscape of these defined biological processes or signatures were investigated within a range of specific contexts. It is an immense collection of 6366 gene sets, some of which may share some common features or signatures. There was a substantial number of Curated gene sets that were enriched in MMECs with loss of functional *Runx1*, which were organised into groups using the aforementioned processes or signatures that are shared between gene sets. By doing this, it was possible to explore key biological features of the *Runx1*-depleted glands that were potentially contributing to the earlier emergence of more aggressive, stem-like tumours.

Firstly, the GSEA of Curated data sets was used to explore several key factors of breast cancer tumorigenesis that are required for breast cancer progression and are often associated a more aggressive phenotype (Feng, Spezia et al. 2018). Analyses of these distinct groups of biological features were summarised in NES and GSEA plots for each, presented in Figures 3.13 to 3.16. This data shows that there is enhancement of a basal-like molecular signature with loss of *Runx1* (Figure 3.13), a subtype which generally behaves more aggressively than others do, and a negative enrichment for gene sets relating to luminal cancers for which there are often more favourable prognoses. There was also enhanced EMT (Figure 3.14) that is essential for the migration and therefore invasion and metastasis (Figure 3.15) of tumour cells as well as being involved in therapeutic resistance (Figure 3.16). All of these aforementioned features of breast cancer progression and tumour aggression were shown to be positively enriched in MMECs upon depletion of *Runx1* and some even more drastically with the added loss of *Runx2*. Taken together, this data indicates that, even at an early point in their development, the mammary epithelial cells of mice with oncogenic Wnt activation and loss of *Runx1* function have a significantly higher propensity towards the development of tumours with more aggressive phenotype than their wild-type counterparts. This may facilitate both the earlier emergence and significantly faster progression of these tumours towards clinical end-point.

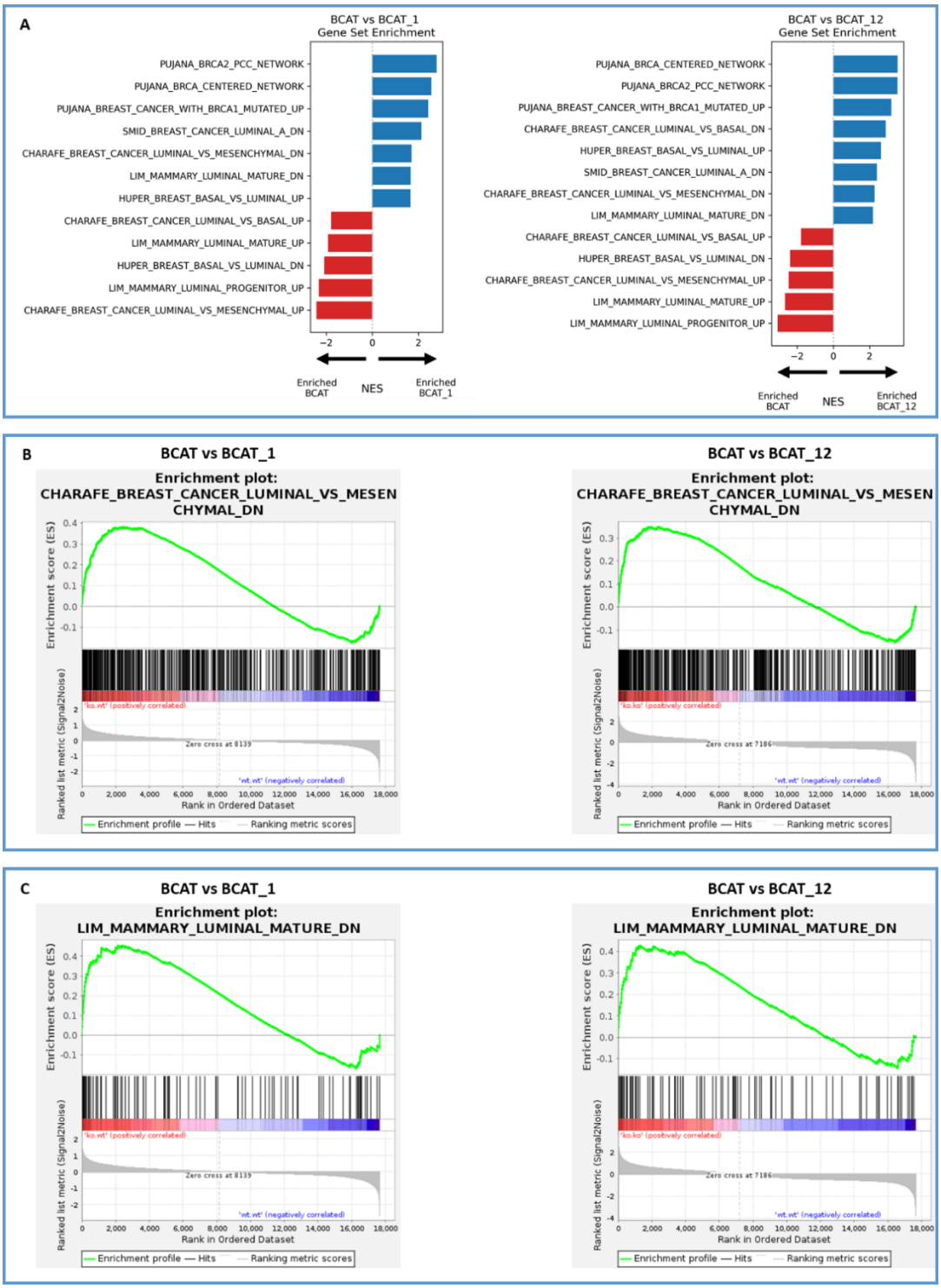


Figure 3.13 Significantly altered Curated gene sets relating to subtype in BCAT_1 and BCAT_12 samples reveal a less luminal and more basal-like gene signature.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data of *Runx1*-depleted test groups were analysed for significantly altered Curated gene sets relating to breast cancer subtypes and normal breast cell types (NES ≤−1.5 or ≥1.5, FDR <0.05) and presented in the above bar charts [A]. Gene sets significantly enriched in the *Runx1*-depleted samples, with positive NES values, are displayed in blue. Those enriched in the wild-type control group (i.e. those negatively enriched in the test group), with negative NES values, are shown in red. A selection of significantly enriched subtype-related gene sets from [A] were presented in enrichment plots [B] and [C].

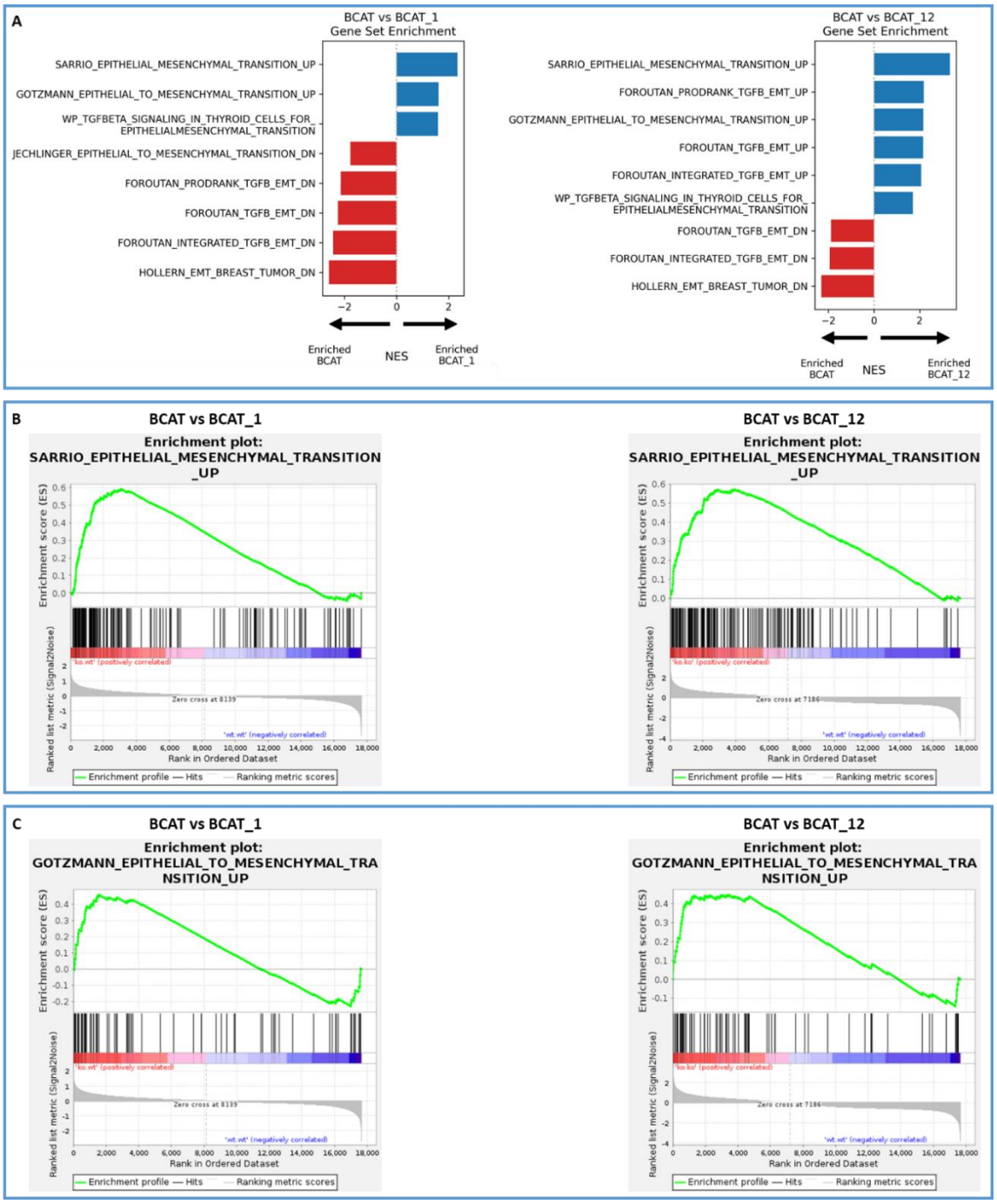


Figure 3.14 Enhancement of EMT-related Curated gene sets in ß-catenin-activated mammary glands with *Runx1* loss.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data of *Runx1*-depleted test groups were analysed for significantly altered Curated gene sets relating to EMT (NES ≤−1.5 or ≥1.5, FDR <0.05) and presented in the above bar charts [A]. Gene sets significantly enriched in the *Runx1*-depleted samples, with positive NES values, are displayed in blue. Those enriched in the wild-type control group (i.e. those negatively enriched in the test group), with negative NES values, are shown in red. A selection of significantly enriched EMT-related gene sets from [A] were presented in enrichment plots [B] and [C].

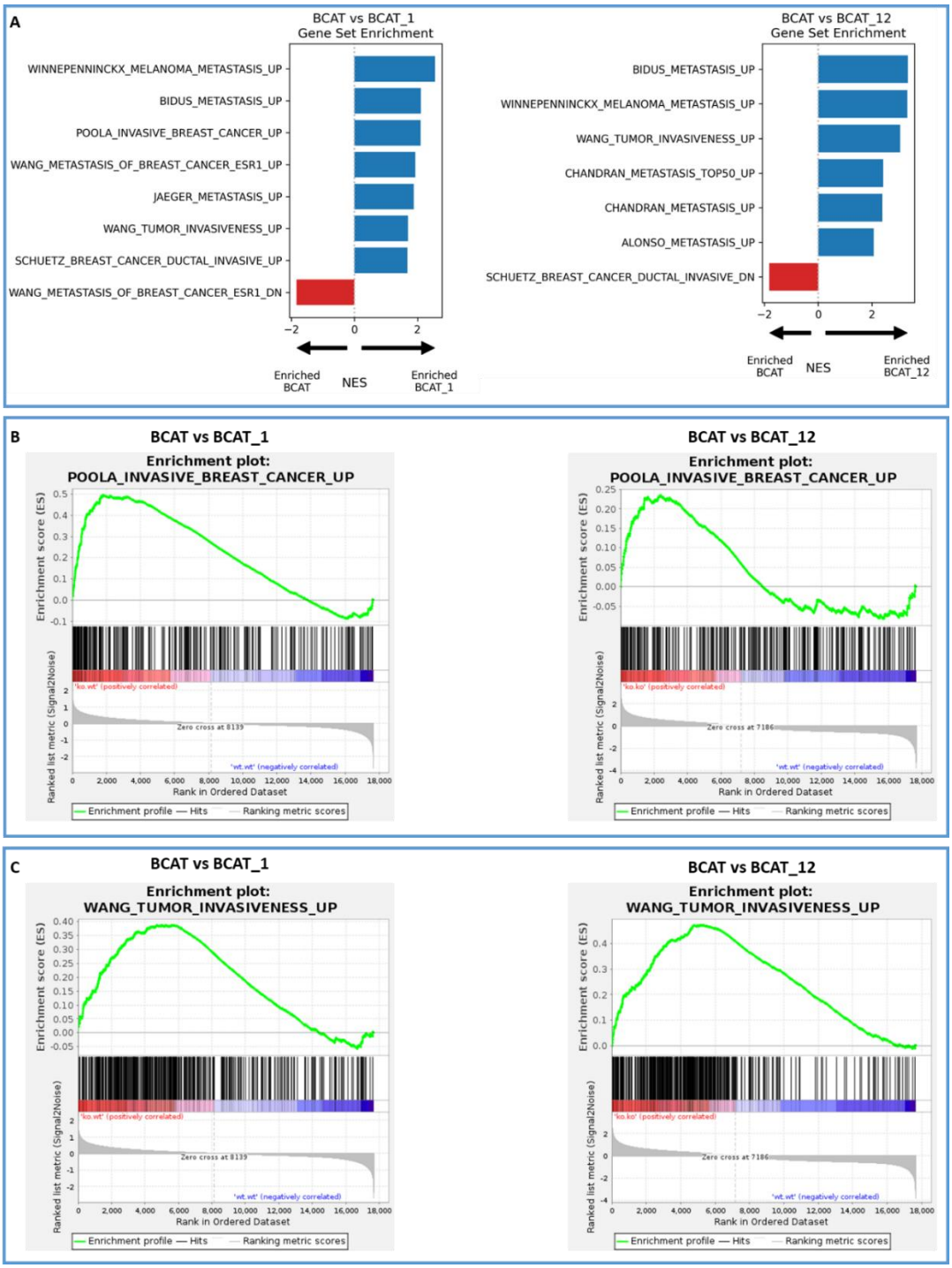


Figure 3.15 Enhancement of various Curated gene sets relating to tumour metastasis/invasiveness in BCAT_1 and BCAT_12 mammary glands.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data of *Runx1*-depleted test groups were analysed for significantly altered Curated gene sets relating to the metastasis/invasiveness of various tumour types (NES ≤−1.5 or ≥1.5, FDR <0.05) and presented in the above bar charts [A]. Gene sets significantly enriched in the *Runx1*-depleted samples, with positive NES values, are displayed in blue. Those enriched in the wild-type control group (i.e. those negatively enriched in the test group), with negative NES values, are shown in red. A selection of significantly enriched gene sets, associated with the metastasis/invasiveness of various tumour types, shown in [A] were presented in enrichment plots [B] and [C].

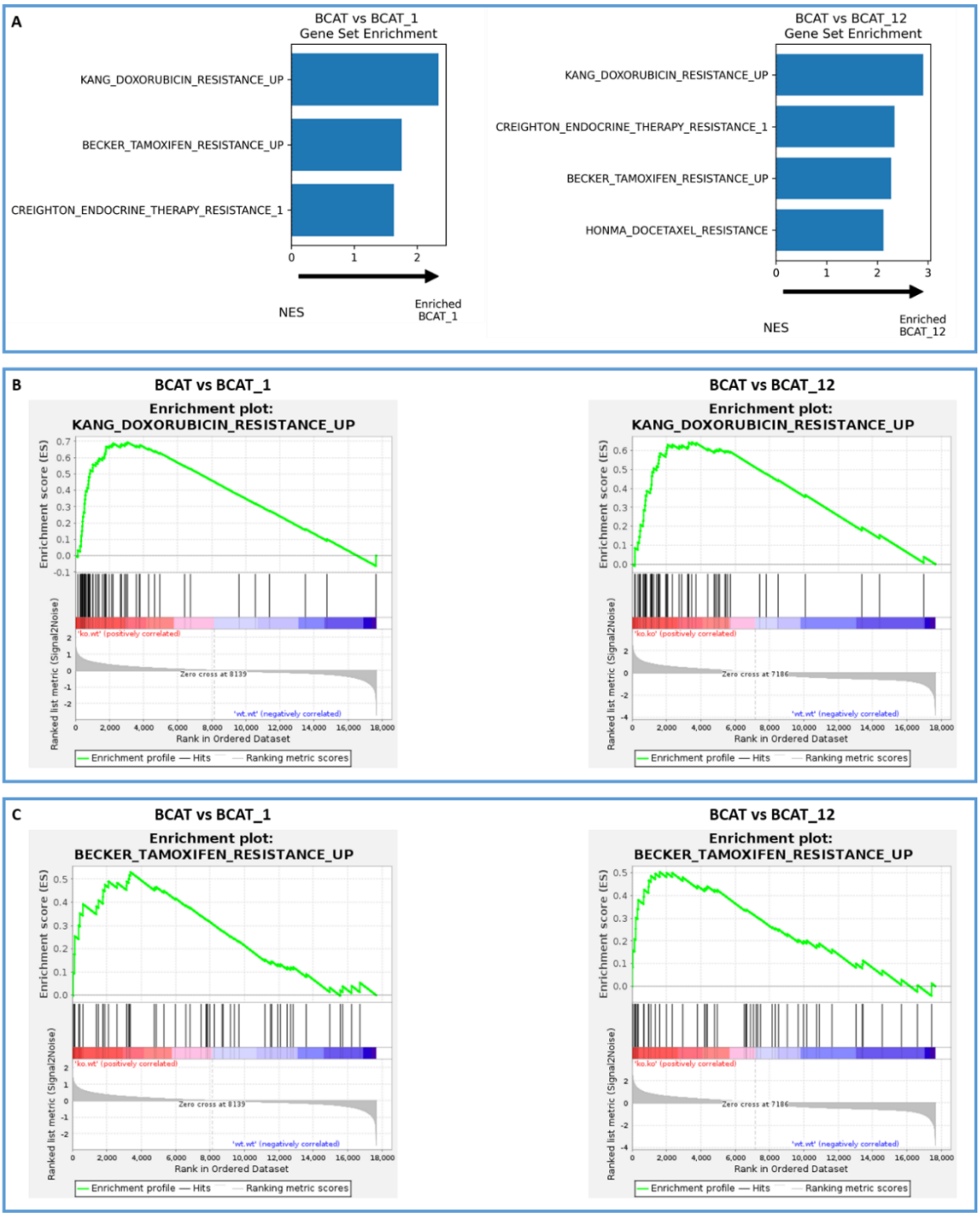


Figure 3.16 Enhancement of Curated gene sets associated with resistance to current breast cancer therapies in RFP+ MMECs isolated from BCAT_1 and BCAT_12 mice.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data of *Runx1*-depleted test groups were analysed for significantly altered Curated gene sets relating to resistance to therapies (NES ≤−1.5 or ≥1.5, FDR <0.05) and presented in the above bar charts [A]. All significantly altered gene sets were enriched in the *Runx1*-depleted samples, with positive NES values. A selection of significantly enriched gene sets associated with resistance to current therapies, displayed in [A], were also presented as enrichment plots in [B] and [C].

3.2.7 Indications of Wnt pathway enhancement in *Runx1*-depleted MMECs.

Results from the pilot RNA-seq experiment indicated there was remodelling of the Wnt pathway following the loss of functional *Runx1* in the mammary epithelium, in which there was constitutive activation of ß-catenin. These changes would likely lead to further dysregulation of the canonical Wnt pathway, as they involved enhancement of various pathway enhancers or targets and perturbation of some regulatory elements of the pathway (Figure 3.6). To investigate Wnt pathway modulation within the context of the expanded RNA-seq experiment, the data was first examined at the level of Wnt pathway-related gene sets within the MSigDB Curated collection. Results from this analysis are displayed in Figure 3.17, which shows NES plots for significantly altered Wnt pathway-related Curated data sets, for each of the *Runx1*-depleted sample groups (Figure 3.17A). Select GSEA plots for significantly enhanced gene sets are also shown (Figure 3.17B and 3.17C). All of the significantly enriched data sets concerning activation of the Wnt pathway were enhanced within each of the *Runx1*-deficient test groups in relation to the wild-type control group. There were more gene sets shown to be enhanced within the BCAT_12 samples and their NES values were generally greater, indicating that a larger number of genes were found to be enriched within the various gene sets analysed.

To investigate this enrichment of the Wnt pathway in more detail, individual genes involved in and targeted by the Wnt pathway were analysed and plotted for statistical comparisons of their normalised counts (Figure 3.18). Of note, the selection of MMEC-validated and GSEA Wnt pathway-related genes that were analysed in the pilot RNA-seq experiment (Figure 3.6) were also analysed in the extended RNA-seq samples. While several of the analysed genes were significantly altered, some of the genes that exhibited significantly altered levels of expression upon loss of *Runx1* within the pilot experiment were no longer affected to the same extent within the extended RNA-seq experiment. This reduced statistical significance can sometimes occur following the introduction of larger sample groups, for which there are generally more reliable statistics, however it is also possible that the data could have been artificially skewed by the large sample-to-sample variation, particularly in the additional control and BCAT_1 samples. Despite this, several genes (*Sox9*, *Wnt7a*, *Mmp9*, *Mmp12*) did seem to show the

same alteration trends, particularly within the BCAT_12 samples. Similar to indications from both the pilot experiment (Figure 3.6) and the GSEA analysis (Figure 3.17) that the Wnt pathway was generally more enhanced within the BCAT_12 mammary gland samples, a greater number of Wnt-related genes were found to be more significantly altered within these samples compared to BCAT_1 mammary samples. Notably, *Lgr4* and *Lgr6*, genes that have both been implicated in activating the Wnt/ß-catenin signalling pathway in normal and cancer processes ((Kong, Ou et al. 2020, Feng, Li et al. 2021, Ordaz-Ramos, Rosales-Gallegos et al. 2021), were specifically enhanced upon simultaneous loss of *Runx1* and *Runx2*, but not *Runx1* loss alone.

The GSEA and individual gene analysis together indicate an enhancement of the canonical Wnt pathway upon loss of *Runx1*, which is further enriched with simultaneous loss of *Runx2*. This progressive dysregulation of the Wnt/ß-catenin signalling pathway from the BCAT group through to the BCAT_12 samples may also partially explain their increasingly aggressive phenotypes within the mouse models.

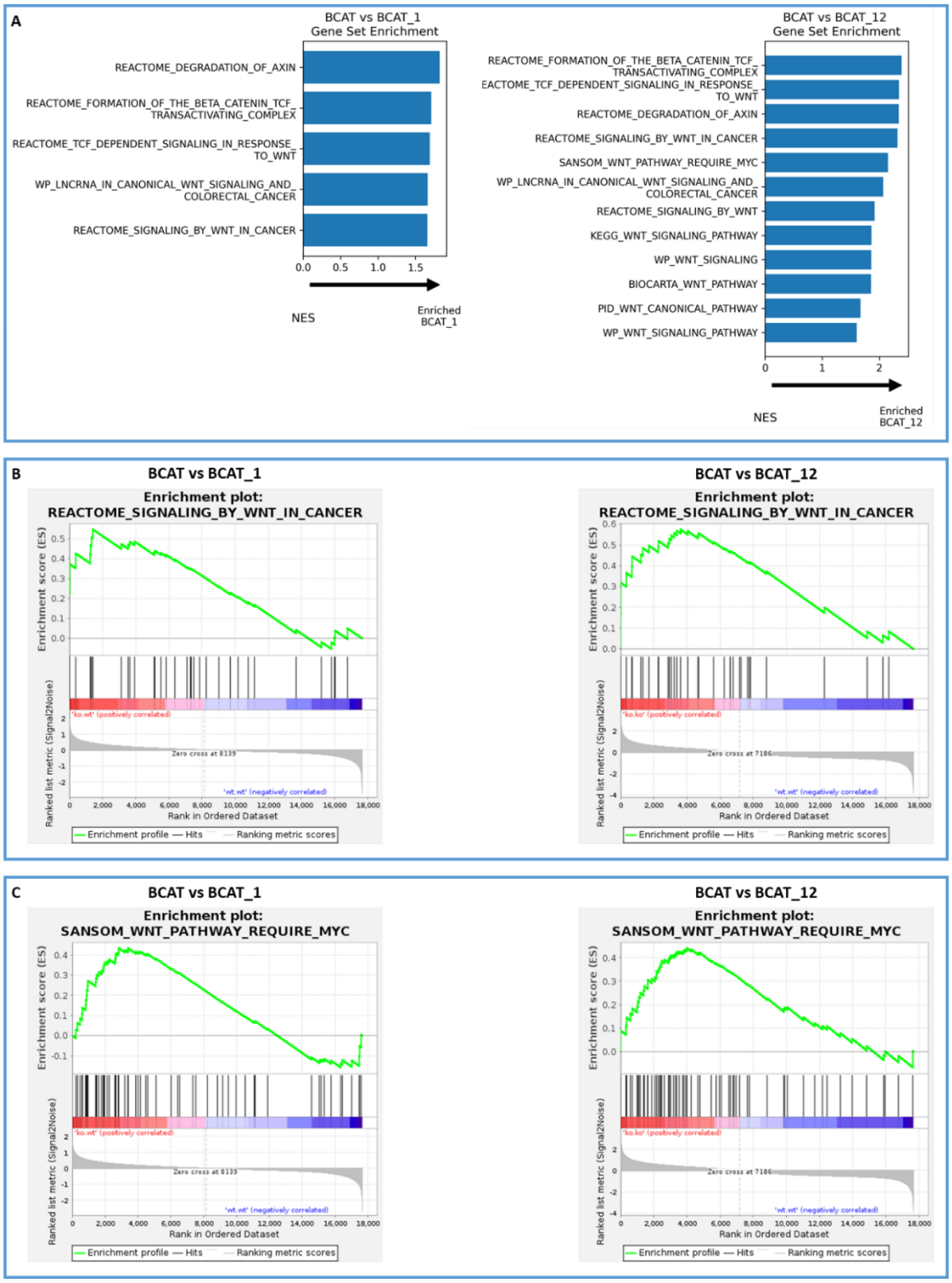


Figure 3.17 Enhancement of various Wnt pathway-related Curated gene sets in BCAT_1 and BCAT_12 mammary glands.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data of *Runx1*-depleted test groups were analysed for significantly altered Curated gene sets relating to the Wnt pathway (NES ≤−1.5 or ≥1.5, FDR <0.05) and presented in the above bar charts [A]. All significantly altered gene sets were enriched in the *Runx1*-depleted samples, with positive NES values. A selection of significantly enriched Wnt pathway-related gene sets from [A] were presented in enrichment plots [B] and [C].

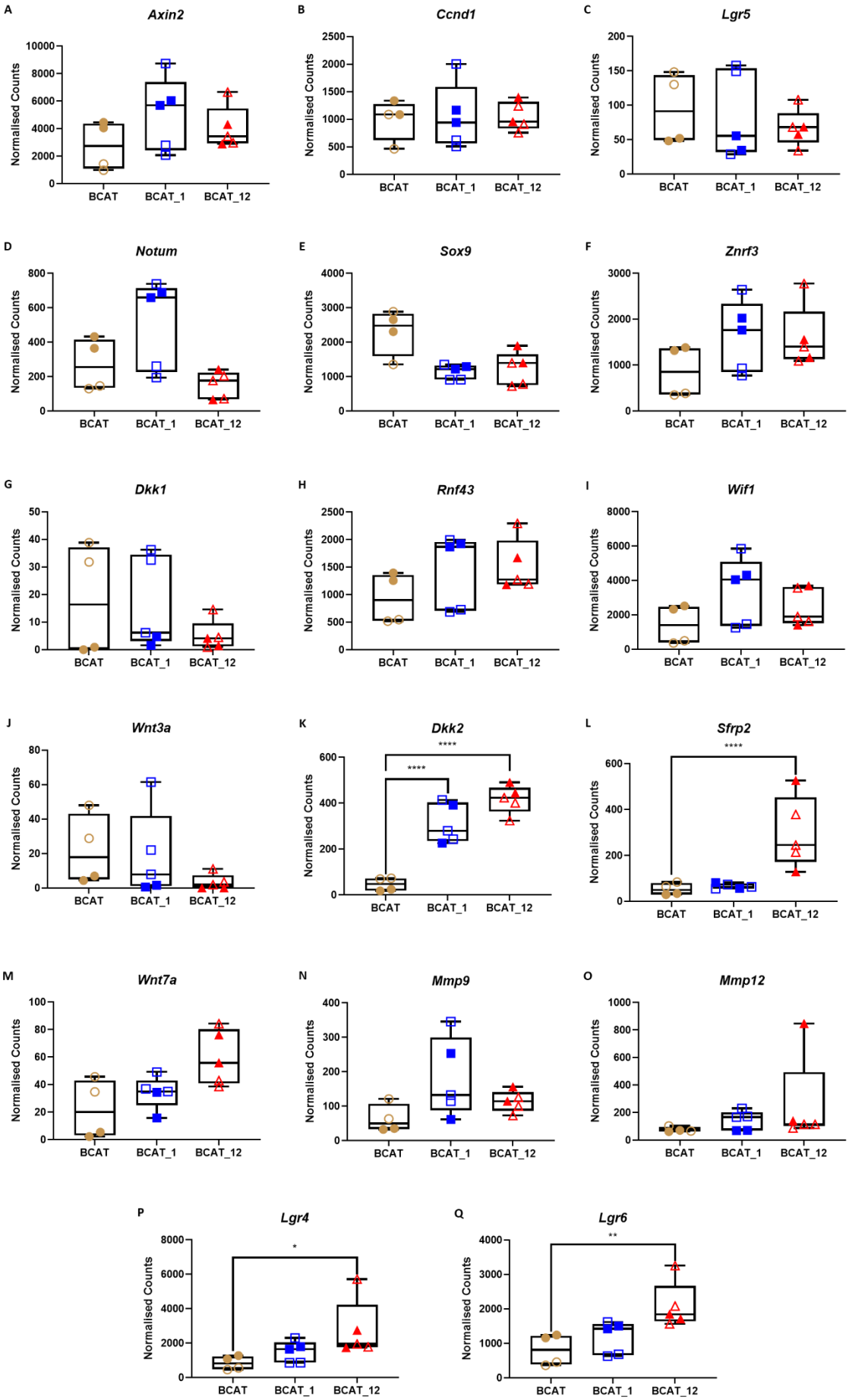

A
Axin2
B
Ccnd1
C
Lgr5
D
Notum
E
Sox9
F
Znrf3
G
Dkk1
H
Rnf43
I
Wif1
J
Wnt3a
K
Dkk2
L
Sfrp2
M
Wnt7a
N
Mmp9
O
Mmp12
P
Lgr4
Q
Lgr6
Normalised Counts
BCAT
BCAT_1
BCAT_12
****
*
**

Figure 3.18 RNA-seq analysis indicates an altered Wnt signalling upon loss of *Runx1*, particularly in *Lgr4* and *Lgr5*.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Normalised gene counts and adjusted p values were calculated for each of the genes analysed for the extended RNA-seq experiment. Various Wnt pathway-related genes were analysed and plotted in the above box and whisker plots, with each data point representative of the normalised gene counts from a single sample. Filled shapes represent the pilot samples and hollow shapes represent the additional samples. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). Significance of comparisons with the BCAT group are also shown in each plot [* $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$]. Wnt targets that were previously analysed in MMECs isolated from pregnant Blg-Cre;*Catnb*$^{wt/lox(Ex3)}$ mice (Table 3.2 and Appendix 1) are shown in [A-J], and additional Wnt pathway-associated genes are plotted in [K-Q].

3.2.8 Analyses of gene sets and individual genes reveal enhanced stem-like phenotype in *Runx1*-deficient MMECs.

Analysis of the pilot data importantly revealed enhanced stemness to be a potential molecular mechanism for the earlier emergence of independently arising mammary tumours in the *Runx1*-depleted mouse models with β-catenin activation (Figure 3.7). Data generated from the extended RNA-seq experiment were interrogated at the level of stem-/progenitor-related gene sets within the MSigDB Curated collection. NES plots (Figure 3.19A) and select GSEA plots (Figure 3.19B and 3.19C) for significantly altered stem cell-related Curated data sets are presented, which show positive enrichment for gene sets upregulated in stem cells and negative enrichment for those either downregulated in stem/progenitor cells or enhanced in mature cells. More stem cell-related gene sets were significantly enriched in the BCAT_12 sample group, indicating enrichment of a higher number of individual genes.

Alterations of individual genes were investigated by plotting and statistically comparing their normalised counts for each experimental group (Figure 3.20). While the significance of some comparisons was however lost for the *Aldh1a* genes, particularly within the BCAT_1 group, many of the genes whose expression was enhanced in *Runx1*-deleted MMECs within the pilot data were also significantly upregulated within the extended RNA-seq analysis. Among these was excitingly the stem cell maintenance gene *Bcl11b (Cai, Kalisky et al. 2017, Miller, Jin et al. 2018)*, a close paralogue to *Bcl11a*, which has links to stem cell function and increased tumour cell aggression within triple negative tumours (Khaled, Choon Lee et al. 2015, Seachrist, Hannigan et al. 2020, Wang, Xu et al. 2020). Notably, the significant elevation observed in *Snai2*, in both the BCAT_1 and BCAT_12 samples, was also maintained following the inclusion of additional samples. This is an EMT-associated transcription factor whose expression was found to be significantly upregulated in endocrine-resistant cells and in clinical samples its expression in metastases was associated with poorer progression-free survival for patients on endocrine treatment (Alves, Elias et al. 2018). Overall, this data reveals enhanced expression of various well-characterised stem cell-related genes. Similar to the pilot investigations, a larger number of the significantly altered stem cell-related genes were either uniquely or more highly significantly

upregulated in BCAT_12 MMECs compared to BCAT_1 mammary samples. An additional example of this identified from the extended RNA-seq data is *Fas*, a gene involved in the induction and maintenance of cancer stem cell activity through its activation of *STAT1* (Qadir, Ceppi et al. 2017), that was found to be significantly upregulated only in the samples with simultaneous loss of *Runx1* and *Runx2* function.

The results of this analysis support those generated in the pilot experiment and strengthen the proposed theory that the stem-like phenotype observed in tumours arising within *Runx1*-depleted mice, which was further exaggerated upon simultaneous depletion of *Runx2*, could be due to a step-wise enrichment of a population of MMECs with stem-like features.

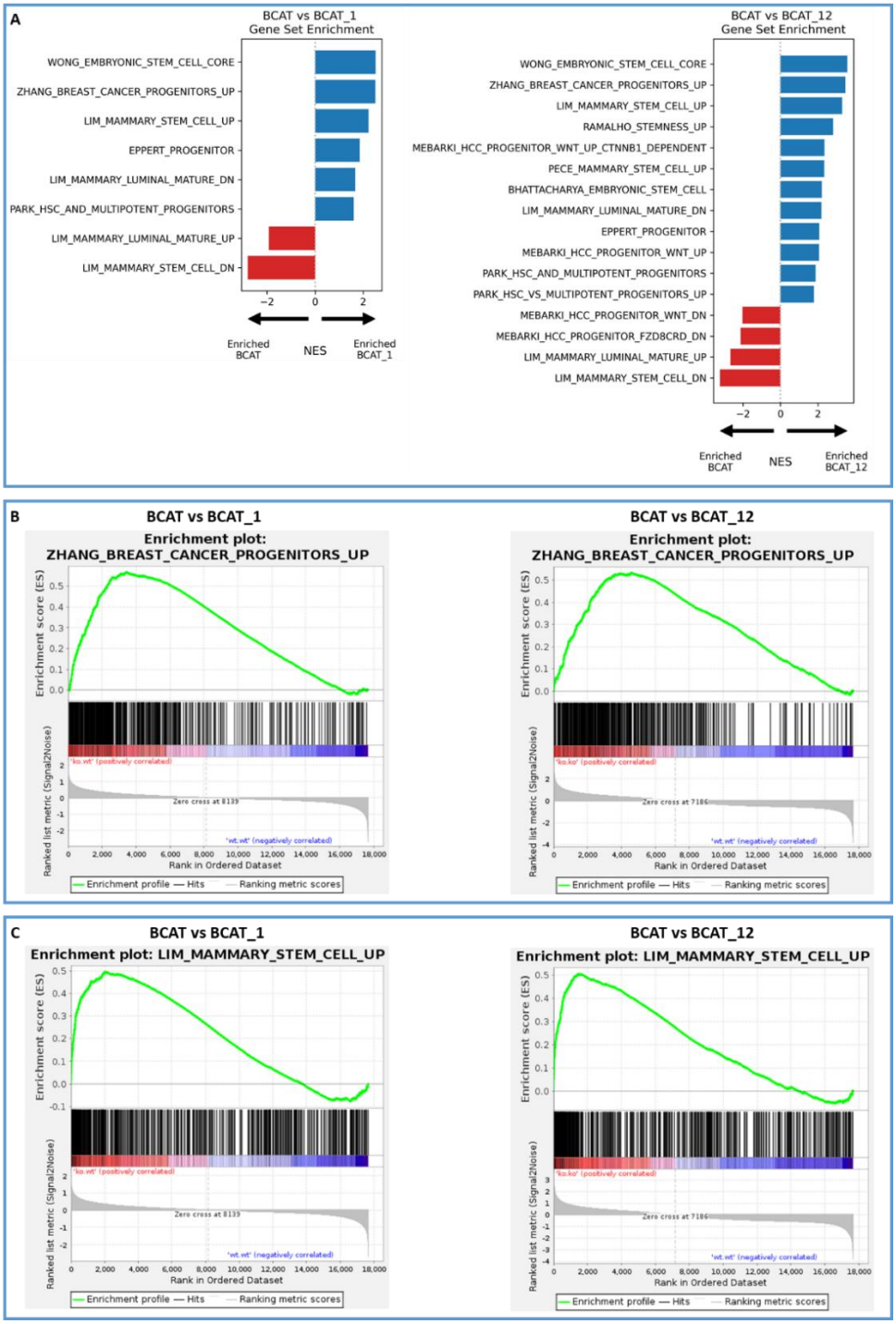


Figure 3.19 Enhancement of various stem/progenitor-related Curated gene sets in BCAT_1 and BCAT_12 mammary glands.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). RNA-seq data generated from *Runx1*-depleted test groups were analysed for significantly altered Curated gene sets associated with several types of normal and cancer associated stem/progenitor cells (NES $\leq$-1.5 or $\geq$1.5, FDR <0.05). Gene sets significantly enriched in the *Runx1*-depleted samples, with positive NES values, are displayed in blue. Those enriched in the wild-type control group (i.e. those negatively enriched in the test group), with negative NES values, are shown in red. A selection of significantly enriched gene sets, associated with the stemness of normal and cancer cells, from [A] were also presented as enrichment plots in [B] and [C].

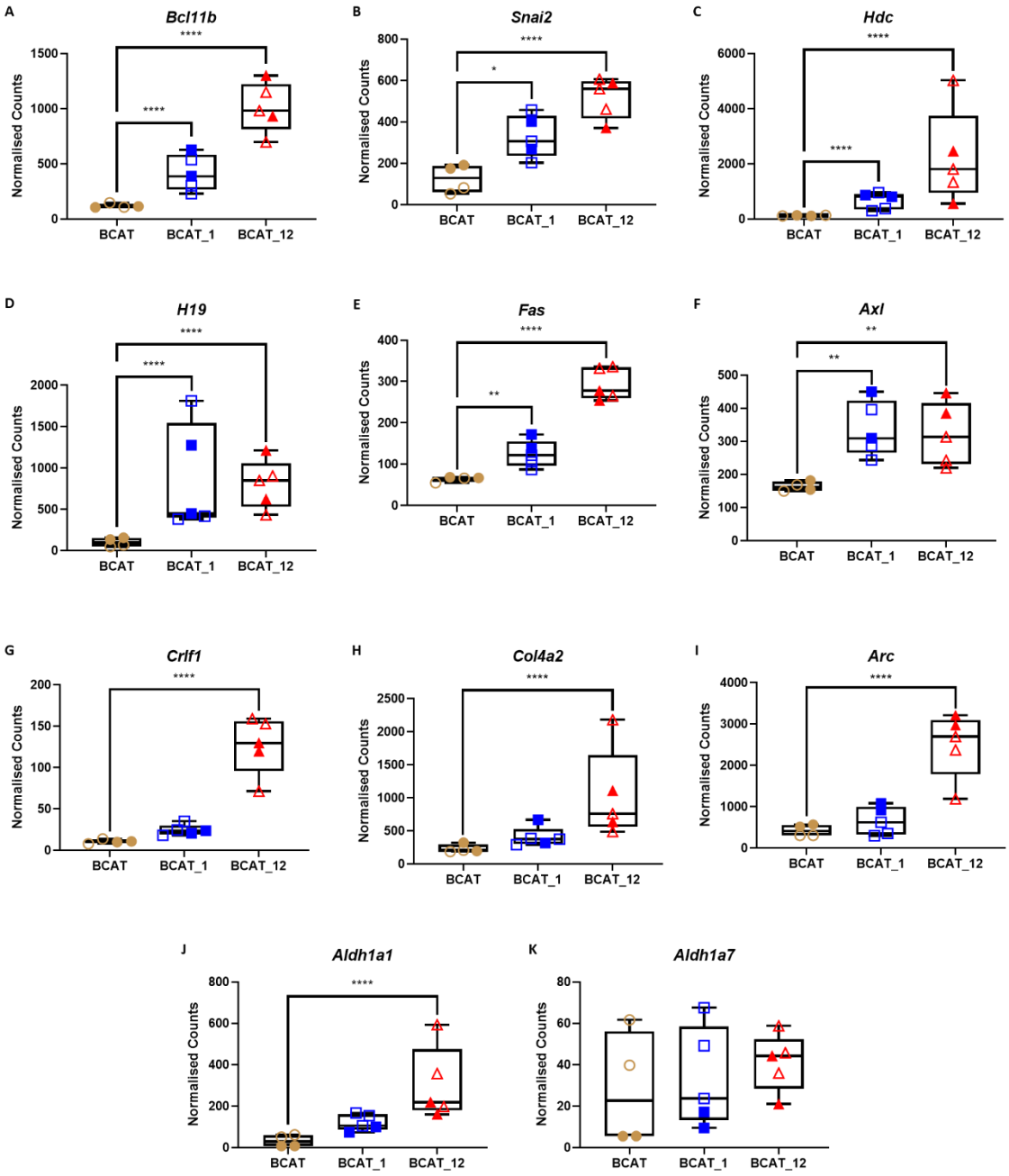


Figure 3.20 Genes, associated with normal and cancer stem cells upregulated in BCAT_1 and BCAT_12 mammary glands in an extended RNA-seq analysis.

Extended RNA-seq analysis of RFP-positive MMECs isolated from BCAT (n=4), BCAT_1 (n=5), and BCAT_12 (n=5) mice at 9 weeks of age. Normalised gene counts and adjusted p values were calculated for each of the genes analysed for the extended RNA-seq experiment. Various genes, associated with stemness of the normal breast and breast cancer cells, were analysed and plotted in the above box and whisker plots [A-J]. Each of the plotted data points represents the normalised gene count obtained from a single sample. Filled shapes represent the pilot samples and hollow shapes represent the additional samples. Each sample comprised of pooled RFP+ MMECs isolated from multiple mice (detailed in Table 3.1 and Table 3.3). The significance of each comparison with the BCAT wild-type control group is also shown in each plot [* $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$].

3.3 Discussion

The main purpose of the RNA-seq investigations, presented in this Chapter, was to investigate changes happening within the *Runx1*- and/or *Runx2*- depleted ß-catenin-activated mammary gland at the molecular level in order to discern the potential mechanisms for the phenotypic differences observed in the Blg-Cre;*Catnb*$^{wt/lox(Ex3)}$ mouse model, specifically at an early timepoint prior to frank tumorigenesis.

A previous RNA-seq experiment conducted in the Blyth lab was used to compare unsorted (bulk prep) mammary gland cell populations isolated from 9 week old experimental mice. While analyses of this data offered valuable insights into the changes happening within the total mammary gland tissue, it was not possible to discern from which cell type(s) each of the differential gene signatures was originating. This was particularly challenging while investigating the strongly enhanced pro-inflammatory/pro-tumorigenic immune signatures and ECM remodelling signatures, induced upon *Runx1* loss. There was uncertainty over which (if any) of these molecular changes was originating from the Blg-Cre-expressing MMECs (and therefore whether they were a direct result of *Runx1* loss) and which could be attributed to indirect changes happening in their microenvironment (i.e. immune cell infiltration). In addition, these strongly enhanced signatures, which likely originated from a combination of both MMECs and their microenvironment, appeared to be dominating much of the differential expression landscape of these samples. There was a strong possibility that any alterations happening specifically in the MMECs were effectively being drowned out by these more dominant molecular signatures. This consideration was particularly relevant when considering any stem cell-related gene expression data. Although the earlier emergence of multifocal and multicentric ß-catenin-driven tumours in BCAT_1 and BCAT_12 mice implied an acquisition of more stem-like behaviours in these mammary epithelial tumour cells, there did not appear to be a prominent enhancement of stem-associated signatures upon analysis of data. It is possible that, even if there was a strong upregulation of various stem cell-related genes within the MMECs, this signature could be significantly dampened by a lack of signature in the microenvironment.

To explore the molecular changes that were happening specifically within *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ MMECs, upon loss of *Runx1* and/or *Runx2* expression, mammary gland cell populations were FACS sorted using an RFP reporter system. RFP-sorted MMECs from ≥3 mice of the same genotype were combined to form each individual sample in order to obtain adequate RNA concentrations. Due to this requirement for increased numbers of mice for this experimental setup, a pilot RNA-seq experiment was conducted as a proof of concept. Analysis of the individual gene alterations exhibited in these cells revealed alterations in the expression of hundreds of genes in BCAT_1 MMECs compared to their wild-type controls, and differential expression of thousands more with the additional loss of *Runx2* function as seen in the BCAT_12 samples. It was noted during the analysis of this data that there was altered expression of many well-established Wnt pathway-associated genes. Closer inspections revealed that there was significant re-modelling of the Wnt pathway to promote its activation and limit its regulation upon loss of *Runx1*, and even more dramatically with simultaneous deletion of *Runx2*. Interestingly, various genes known to be involved in the stem-like behaviours of both normal and cancer cells in the mammary gland were significantly enhanced in a step-wise manner from BCAT to BCAT_1 to BCAT_12 samples.

Due to the informative results obtained in the pilot experiment, additional samples were collected for each experimental group and the pilot samples were run alongside these in a extended RNA-seq experiment. While analysing the PCA plot generated from this data (Figure 3.8), it was noted that there was significantly more sample-to-sample variation introduced with these additional samples, particularly in the BCAT and BCAT_12 groups, which was likely due to the larger number of mice that had been required to generate each sample. This was an important consideration during later investigations into the differential expression of individual genes, as it potentially contributed to the drastic reduction in the number of significantly altered genes, with a particular bias towards a reduction in the BCAT_1 mammary samples. It does, however, lend additional credence to the genes that were significantly altered despite this increased sample variation. Despite this, analysis of the numbers of genes with significantly altered expression levels showed that no genes were affected in the

BCAT_2 group compared to the control, almost 100 genes were altered in the BCAT_1 mammary samples, and over 1000 additional genes were uniquely up in the BCAT_12 MMECs. Each of these molecular profiles for these *Runx*-depleted samples is logical within the context of the phenotype observed for their respective mouse models. The larger the molecular distinction between the experimental cells and the wild-type control cells, the more significantly distinct the phenotype was between the two models.

In order to interpret the large volumes of data produced from this extended RNA-seq experiment, a GSEA approach was taken. Analysing the data at the level of gene sets, sourced from the MSigDB gene set collections, allowed for the identification of shared biological functions/features within the differentially expressed genes identified within each of the *Runx1*-depleted test groups. Similar to the trends seen with individual gene analysis, there were greater numbers of enriched gene sets within the BCAT_12 samples compared to BCAT_1 mammary samples. GSEA indicated a significant enrichment of gene sets with shared biological functions relating to the hallmarks of cancer and tumour progression/aggression. Notably, upon loss of *Runx1* there were significant enhancements of gene sets that are known to be involved in angiogenesis, DNA repair, EMT, metastasis and invasion, therapeutic resistance, Wnt pathway activation, and cancer stemness. Many of these alterations were even more significant with the additional loss of *Runx2*. These molecular changes exhibited with *Runx1* loss support already established knowledge relating to the function of the *RUNX1* gene in the human breast and in breast cancer, particularly those relating to its function in metastasis and EMT. For example, a study comparing gene expression profiles of various primary adenocarcinoma types with unmatched metastatic lesions revealed *RUNX1* downregulation to be part of a 17-gene expression signature that was associated with metastasis and poorer clinical outcome (Ramaswamy, Ross et al. 2003). Further to this, a previous publication from the Stein group demonstrated that loss of *RUNX1* was required for TGFß-induced EMT, a precondition essential for cancer metastasis (Hong, Messier et al. 2017).

A more in-depth analysis of the specific genes that were contributing to the enhanced Wnt signature revealed significantly (or trending towards significantly) enhanced expression of various Wnt pathway activators/enhancers and reduced expression of regulators. While some of the specific gene alterations seen in the pilot experiment lost their significance, the results shown throughout this chapter nonetheless support the theory that the Wnt pathway is undergoing remodelling or modulation upon loss of functional Runx genes. Analysis of individual genes relating to the stem cell-related signature within the *Runx1*-deficient samples identified several genes (*Aldh1a1*, *Snai2*, *Axl*) that have been identified as important predictors of disease progression and poorer overall survival for breast cancer patients.

While informative, the limitations of this specific analysis must also be considered and (if possible) addressed during any future investigations of this type. Following analysis of bulk mammary cell populations, the rationale for analysing RFP-sorted cells was so that the direct (MMEC-specific) molecular consequences of *Runx1* deletion could be investigated. As previously discussed, in order to obtain sufficient concentrations of RNA for this RNA-seq analysis, RFP-expressing MMECs isolated from several mice were combined together to form each individual sample. As exemplified in the PCA plots, increasing the numbers of mice per sample can often introduce more variability into each sample whilst also not increasing the overall power of the statistical analyses. One way of addressing this issue could be to use low input RNA sequencing technologies offered by companies such as Illumina and New England Biolabs (NEB). Alternatively, a single-cell RNA sequencing approach could be used for the bulk mammary gland cell populations. The advantage of this is that, rather than gene expression being averaged across all cell types and potentially diluting down some cell type-specific molecular signatures, gene expression is analysed at the level of individual cells and thus allows for a higher resolution analysis. Although this method also requires cell sorting, it generally requires less cells to be analysed per sample, compared to traditional bulk RNA-seq experiments. It would also facilitate an analysis of the effects of Runx alteration on various individual cell populations within the mammary gland tissue, therefore allowing for a more detailed examination of the immune pathways altered in the original bulk RNA-seq study.

An additional future consideration for this study is that, while the RNA extracted from these samples was adequate for RNA-seq analysis, the concentration and amount was not sufficient to perform confirmatory RT-qPCR analyses of differentially expressed genes. Additional samples (and therefore mice) would therefore be required to be able to perform this analysis which, in order to look at MMEC-specific gene changes, may also necessitate combining MMECs from several mice. Therefore, in order to take the 3Rs (Replacement, Reduction and Refinement) into account, it would be considered more appropriate to validate these gene alterations (in response to *Runx1* deletion) using *in vitro* investigations. This formed part of the rationale for the investigations conducted on HC11 mouse mammary epithelial cells, discussed in Chapters 4 and 5. In particular, one of the key messages from this chapter was the gene profile highlighting changes to stem cells. This necessitated further *in vitro* analysis to explore this aspect of RUNX1 in mediating stem like behaviour using a well-established HC11 murine mammary epithelial cell line (Danielson, Oborn et al. 1984, Ball, Friis et al. 1988). These results presented in Chapters 4 and 5 substantiate these findings and provide evidence that RUNX1 is functioning, at least in some aspects, to control mammary stemness and its loss may therefore facilitate earlier tumour onset.

Despite these discussed limitations, the RNA-seq experiments presented throughout this chapter provide a solid rationale for the distinct phenotypes seen in Blg-Cre; *Catnb*$^{wt/lox(Ex3)}$ cohort mice. While the BCAT_2 mice, which are phenotypically indistinct from their wild-type counterparts, showed no significant change in their molecular landscape, both *Runx1*-depleted cohorts exhibit significant gene expression changes that corresponded with the apparent intensity of their respective phenotypes. The various (related and distinct) molecular signatures and biological features, which were differentially regulated upon loss of *Runx1*, are likely all contributing towards the overall faster progressing and aggressive phenotypes identified in the mouse models. The wide range of biological functions, which were implicated in the dramatic phenotypes of the BCAT_1 and BCAT_12 mouse models, indicates that the tumour-suppressive functions of *Runx1* (and *Runx2*, together) are complex and multi-faceted, and thus the consequences of losing these functions are likely far-reaching and prominent.

4 Investigating the role of RUNX1 in the stem-like potential of the mammary gland through its overexpression

4.1 Introduction

4.1.1 The role of RUNX1 in stem cells

The heightened stem-like phenotype of *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ tumours with additional loss of *Runx1* function (or additional simultaneous loss of *Runx1* and *Runx2*), as indicated by the earlier and increased emergence of multifocal tumours and the RNA-Seq analyses presented in Chapter 3, prompted further exploration of the role of *RUNX1* in stem cell regulation in the normal mammary epithelium, and the associated protection from oncogenic insult. The purpose of this was to determine what could be happening mechanistically in the pre-tumour mammary epithelium of *Runx1*-depleted glands to promote this earlier emergence of independently arising tumours by the oncogenic activation of ß-catenin.

Runx1 has previously been linked to stem cell regulation in various tissues in the body (Hoi, Lee et al. 2010, Osorio, Lilja et al. 2011, Scheitz, Lee et al. 2012, Kim, Barron et al. 2014, Matsuo, Kimura et al. 2017). Mouse models have been key in determining a role for *Runx1* in definitive haematopoiesis and have shown that *Runx1* is arguably the most critical regulator in the formation of definitive haematopoietic stem cells (HSCs) (Okuda, van Deursen et al. 1996, Wang, Stacy et al. 1996, North, Gu et al. 1999, Nottingham, Jarratt et al. 2007). In one study, conditional knockout of *Runx1* resulted in increased populations of immature haematopoietic cells in a quiescent state, while *Runx1* overexpression in transplanted haematopoietic cells prevented the reconstitution of lethally irradiated bone marrow (Ichikawa, Goyama et al. 2008). The results from this study demonstrate that *Runx1* negatively regulates quiescent HSCs during haematopoiesis and the authors propose that expansion of HSCs, following loss of functional *Runx1*, may contribute to the development of haematological malignancies associated with disrupted *Runx1* function. A more recent study linked dysregulation of post-translational modifications (PTMs) of RUNX1, specifically defective methylation, with expansion of HSC populations that had lost their quiescence and possessed antiapoptotic properties (Matsumura, Nakamura-Ishizu et al. 2020). This increase in activated HSCs was seen to confer

survival advantage due to their resistance to apoptosis (induced by various stress conditions), indicated by the cells' increased capacity to reconstitute lethally irradiated bone marrow. As apoptosis acts as a mechanism to protect from cancer development initiated by damaged/abnormal cells, resistance to this programmed cell death is a hallmark of a preleukaemic clone. It was therefore postulated that, in the clinical context, this expansion of HSCs in populations with defective RUNX1 methylation might predispose affected individuals to leukaemia.

In a study relating to stem cells within the mammary gland, the Gupta lab devised a method, which they named perturbation-expression analysis of cell states (PEACS), that identified *RUNX1* as a regulator of mammary stem cell differentiation (Sokol, Sanduja et al. 2015). *RUNX1* silencing in MCF10A cells, using shRNAs, resulted in cells that were unable to produce mature lobules, and which had reduced capability to form ducts and ductal-lobular rudiments in 3D cultures. On the rare occasions that the collagen cultures produced spheres with ducts, they were significantly shorter and less branched than those produced by control cells. This phenotype was reversible upon rescue of *RUNX1* expression. *RUNX1* has also been studied within the context of breast cancer stem cells, and has been shown to suppress breast tumour growth, migration, and invasion by repressing the breast cancer stem cell phenotype (Hong, Fritz et al. 2018, Fritz, Hong et al. 2020).

4.1.2 Experimental Aims

While the aforementioned studies have been informative and have provided a substantial foundation on which to build, it would be useful to probe further into the possible mechanisms of mammary stem cell regulation by RUNX1. Work by the Stein lab provides an insight into the consequences of RUNX1 loss in already established tumours, initiated by other oncogenic influences. Their work focuses mainly on the relationship between RUNX1 and breast cancer stem cells, and cannot necessarily be used to inform that between RUNX1 and stem cells within the normal breast, particularly from a molecular or mechanistic perspective.

In addition to this, modelling loss of functional RUNX1 in already established cancers does not take into account the consequences of RUNX1 loss within the context of the normal mammary epithelium. The tumour suppressive functions of RUNX1 may not necessarily be fully addressed in terms of tumour initiation, and it does not consider that *RUNX1* mutations are driver mutations in their own right. Given the level of control that RUNX1 appears to be able to exert over the canonical Wnt pathway (as indicated in Chapter 2), it could be argued that *RUNX1* mutations could be an essential factor for the initiation of tumorigenesis in a significant number of mammary tumours.

It was therefore postulated that it would be useful to explore *RUNX1* alteration in a non-tumorigenic mammary epithelial cell environment, particularly within a non-transformed cell line capable of displaying stem cell properties, to determine by what mechanisms loss of RUNX1 function might drive tumour initiation in the mammary gland. This was explored by overexpressing RUNX1 in the HC11 mouse mammary epithelial cell line.

4.2 Results

4.2.1 RUNX1 Expression in HC11 Mouse Mammary Epithelial Cells

The HC11 cell line, an immortalised cell line originating from the normal mammary gland of a pregnant Balb/c mouse, is predominantly composed of luminal stem- or progenitor-like epithelial cells (Danielson, Oborn et al. 1984, Ball, Friis et al. 1988). This feature makes it a valuable model for exploring the stem-like properties of the mammary gland. The HC11 cell line was previously used by the Blyth lab during previous investigations that revealed a role for RUNX2 in mammary stem cell regulation (Ferrari, Riggio et al. 2015).

Initial Western Blot Analysis (Figure 4.1) showed that HC11 cells express intermediate levels of RUNX1, relative to several other cell types known to express the protein. Whilst the HC11 cells appear to express higher levels of RUNX1 compared to primary mouse mammary epithelial cells (MMECs), whose protein band was only visible following oversaturation of the other sample bands, there is

visibly less RUNX1 protein in these cells relative to levels expressed in a lymphoma cell line and primary mouse thymocytes

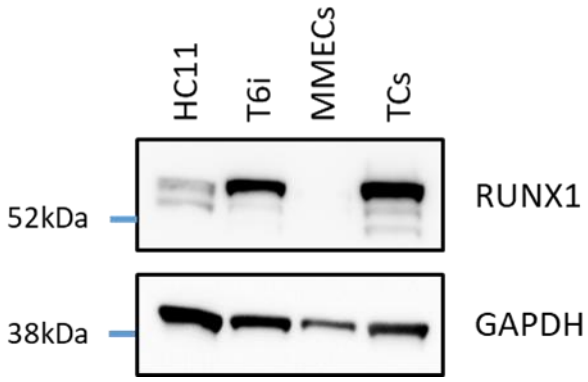


Figure 4.1 RUNX1 protein expression in various cell types.

Western Blot showing RUNX1 expression in HC11 cells compared with various cell types known to express relatively high levels of the protein. T6i = cell line derived from a mouse T-cell lymphoma. MMECs = mouse mammary epithelial cells. TCs = primary mouse thymocytes. GAPDH as a loading control. Amersham Full Range Rainbow Marker as a protein ladder, with positions and molecular weights of key markers shown.

4.2.2 RUNX1 overexpression reduces mammosphere- and colony-forming potential of HC11 cells without affecting 2D growth

Based on the RUNX1 expression data, it was reasoned that it should, in theory, be possible to induce either overexpression or knockout of *Runx1* and its associated protein. HC11 cells were transduced with retrovirus, produced from a pBABE-Puro vector that was modified to contain the mRunx1P1 construct (described in Materials and Methods), to induce overexpression of *Runx1* under the control of the P1 promoter. Retrovirus derived from an empty pBABE-Puro vector was also used to transduce HC11 cells, to create a control cell line. Following Puromycin selection, protein lysates extracted from transduced cells were analysed by Western Blot (Figure 4.2A) to determine whether this successfully induced overexpression of RUNX1 protein. Densitometry analysis using the Image Lab software revealed that transduction of HC11 cells with the pBABE-Puro-mRunx1P1 construct induced an almost ~2-fold increase in the relative density of RUNX1 protein compared to levels exhibited in the control (Empty Vector) line.

The RUNX1 overexpressing HC11 cells were initially compared with their control counterparts using 2D growth assays. For each cell line, cell counts were taken at

24 hour intervals and analysed on a CASY cell counter. As shown in Figure 4.2B there was no statistically significant difference in the growth of HC11 cells when RUNX1 was overexpressed under 2D conditions.

Although no differences in 2D growth were observed, follow-up experiments were conducted using mammosphere assays as a tool to quantify and compare stem cell/progenitor activity in HC11 Empty Vector and Runx1P1 cell lines. Stem cells are able to clonally proliferate to form spherical clusters in the relatively low-density, free-floating, and growth-limiting conditions of a mammosphere assay, where more differentiated cells would not survive (Shaw, Harrison et al. 2012). In contrast to 2D culture, the overexpression of RUNX1 in a mammosphere assay resulted in significantly fewer spheres, as shown in Figure 4.2C. A similar stem cell quantification assay, known as a 3D mammary colony-forming cell assay, was carried out to provide another independent method of analysis to support the results obtained from the mammosphere assays. 3D colony assays likewise involve seeding single cells at a low density in growth-limiting conditions, though conversely in a normal 24-well culture plate and suspended in a semi-solid Matrigel matrix. Each discrete 3D colony that is counted at the end of the assay originates from a single stem- or progenitor-like cell, thus allowing the stem-like behaviours of different cell lines to be relatively quantified and compared (Tornillo and Cabodi 2014). Statistical analyses showed significantly reduced colony-forming capabilities for the RUNX1 overexpressing HC11 cells (as shown in Figure 4.2D), thus demonstrating reduced stem cell activity in these cells.

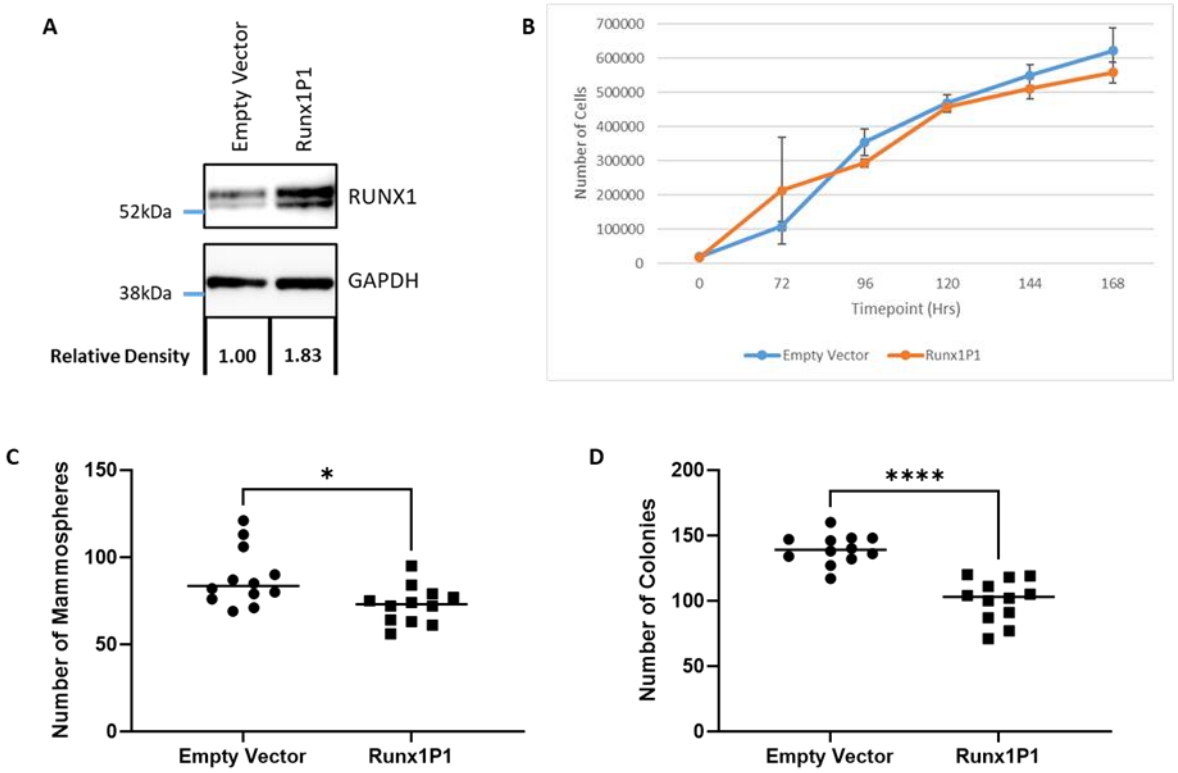


Figure 4.2 RUNX1 overexpression reduces "stemness" of HC11 mouse mammary epithelial cells, while having no effect on 2D growth.

[A] Western Blot of HC11 mouse mammary epithelial cells transduced with pBABE-Puro-Runx1P1 or pBABE-Puro Empty Vector as a control. Results from Image Lab software densitometry analysis reveals an almost 2-fold increase RUNX1 protein levels. GAPDH as a loading control and Amersham Full Range Rainbow Marker as a protein ladder, with key marker positions and molecular weights shown.

[B] 2D growth curve of Empty Vector and Runx1P1 cell lines. n=3 counts per cell line per day, plotted as averages (and population standard deviations). Results shown represent 1 of 4 experimental repeats.

[C] Primary mammosphere counts for Empty Vector and Runx1P1 cell lines, summarised in a scatter plot. Each data point represents the number of spheres counted in a single well. n=12 wells counted per cell line. Results shown represent 1 of 3 experimental repeats.

[D] Scatter plot summarising results from colony-forming assays comparing Empty Vector and Runx1P1 cells. Each data point represents the number of colonies counted in a single well. n=12 wells were counted per cell line at experimental endpoint. Results shown represent 1 of 2 experimental repeats.

Unpaired t-tests with Welch's correction performed for each [* $p < 0.05$; **** $p < 0.0001$].

4.2.3 Significantly larger mammospheres, with increased proliferation and reduced cell death, produced upon overexpressing RUNX1

Following manual counting, phase contrast images were taken of mammospheres using an Olympus CKX41 Inverted Microscope. It is apparent from the representative images shown (Figure 4.3A) that, although there are fewer spheres, the mammospheres derived from RUNX1-overexpressing HC11 cells are generally larger than those derived from control cells. In the ImageJ software, an image of a Stage Micrometer with a 1mm scale (subdivided into 100 divisions of 0.01mm) that was taken at x10 magnification was used to calibrate the scale bar, which allowed for the accurate quantification of the areas (μM^2) of the mammospheres captured. Results from this analysis confirmed that the Runx1P1 cells did indeed produce statistically larger mammospheres than their Empty Vector counterparts (Figure 4.3A).

To explore this change in mammosphere size resulting from increased RUNX1 expression, proliferation and apoptosis of cells within these structures was investigated. Sections taken from paraffin-embedded agarose plugs of Empty Vector and Runx1P1 cells were stained for Ki67 (Figure 4.3B), a widely used proliferation marker, and Caspase 3 (Figure 4.3C), a crucial mediator of programmed cell death, using IHC methods. Images were taken and analysed using a trained HALO software. Results from this analysis, shown in Figure 4.3B and Figure 4.3C indicate that the Runx1P1-derived mammospheres expressed significantly higher levels of Ki67 and reduced levels of Caspase 3 compared to those formed by Empty Vector cells. This indicates that there is a much higher proportion of cells within the Runx1P1-derived structures that are actively proliferating and a significantly lower percentage of cells undergoing apoptosis, which were likely contributing factors in their increased size. This rationale seems especially plausible with the knowledge of the essential role that RUNX1 plays in mammary stem/progenitor cell differentiation (Sokol, Sanduja et al. 2015).

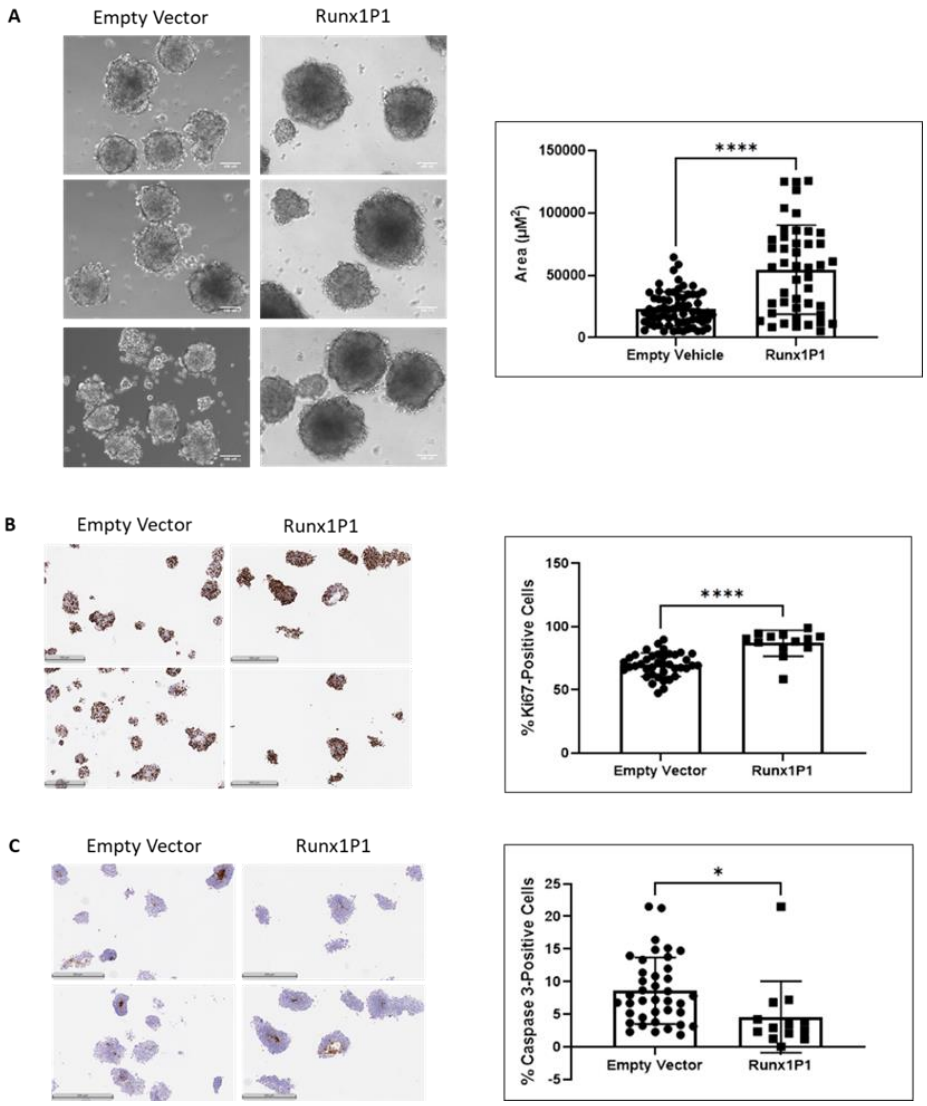


Figure 4.3 RUNX1 overexpression results in larger mammospheres, with increased proliferation and reduced cell death.

[A] Phase contrast images of mammospheres (left), derived from Empty Vector and Runx1P1 HC11 cells. Scale bar=100μM. Images are representative of 10 images taken for each cell line in this particular experimental setup. Mammosphere sizes (μM^2) were quantified using ImageJ software. A 1mm stage micrometer was used to calibrate the scale bar. Scatter plots comparing mammosphere sizes (μM^2) of Empty Vector and Runx1P1 cell lines (right). Each data point represents the size (μM^2) measured for a single mammosphere. Result shown represents 1 of 3 experimental repeats, each with n≥8 images taken per condition.

Representative images (left) and quantification (right) of Ki67 [B] and Caspase 3 [C] staining in sections taken from agarose plugs of HC11 mammospheres, analysed using HALO software. Representative images shown with 500μM scale bar. Scatter plots summarise results of the quantitative analysis, where each data point represents the percentage of total cells in a single mammosphere that stained positive. Results shown represent 1 of 2 experimental repeats.

Unpaired t-tests with Welch's correction conducted for each [* $p < 0.05$; **** $p < 0.0001$].

4.2.4 Stem cell promoting functions of recombinant WNT3A protein are diminished under RUNX1-overxpressing conditions

Whilst the previously described 3D growth assays demonstrate that RUNX1 may indeed play a role in mammary stem cell regulation within a normal population of mammary epithelial cells, and therefore potentially within the normal breast, it is not clear as to whether this function is maintained within the context of a mammary epithelium under oncogenic insult. It would be useful to determine whether RUNX1 overexpression is able to limit the stem cell enriching functions of WNT3A, a key activator of the canonical Wnt/ß-catenin pathway (He, Lu et al. 2015), which is often aberrantly activated in breast cancer (Zhan, Rindtorff et al. 2017). Through these means, it would therefore be possible to deduce whether the tumour suppressive role of Runx1 in the breast is at least partially mediated through its control over the stemness of the mammary epithelium, at least within the context of oncogenic Wnt signalling.

Mammospheres were chosen to model and compare stem cell activity in the Empty Vector and Runx1P1 HC11 cell populations, with and without the addition of recombinant WNT3A. These results support previous observations that RUNX1 overexpression limits mammosphere formation. The data also indicates that higher levels of RUNX1 expression can limit the stemness-promoting capabilities of recombinant WNT3A protein, as the significant enhancement of mammosphere formation in HC11 cells was dampened with RUNX1 overexpression (Figure 4.4). While there was a significant increase in mammospheres produced with WNT3A treatment of both cell lines, it was increased to a slightly lesser extent in the Runx1P1 cell line. Although this effect may not be immediately obvious when looking at the adjusted p values of vehicle versus WNT3A treatment for each cell line alone, both of which were $p < 0.0001$, it is useful to note that the fold increase in mammosphere counts with WNT3A treatment was reduced from 1.83 (Empty Vector) to 1.51 (Runx1P1). Based on this data, it is reasonable to conclude that RUNX1 may protect the mammary gland from oncogenic insult, at least in the context of abnormal Wnt/ß-catenin pathway activation and, at least in part, through its regulation of mammary epithelial stem/progenitor cell populations.

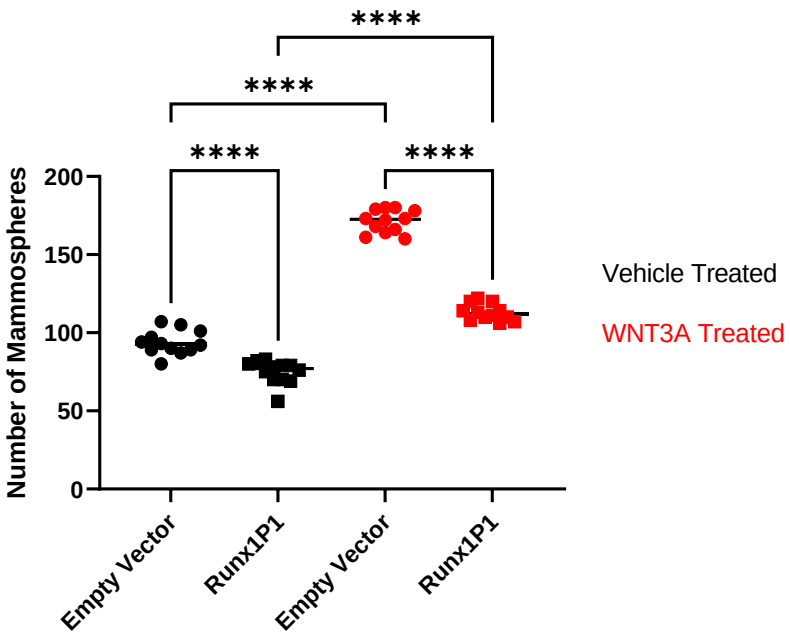


Figure 4.4 Overexpression of RUNX1 quenches the stemness-promoting capabilities of recombinant WNT3A protein.

Scatter plot summarising the quantification of mammospheres derived from Empty Vector and Runx1P1 HC11 cell lines, treated with 100ng/ml WNT3A or vehicle (distilled water). Each data point represents the number of spheres counted in a single well following 7 days of culture. Vehicle-treated HC11 cells in black and WNT3A treated cells in red. n=12 wells counted per condition. Results shown represent 1 of 3 experimental repeats. One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons, with the most pertinent comparisons shown [**** $p < 0.0001$].

4.2.5 Characterisation of RUNX1 overexpressing mammary epithelial cells using stem cell and Wnt-target gene panels

In order to evaluate which molecular markers were related to the stem cell-regulating functions of RUNX1 in mammary epithelial cells, quantitative PCR (qPCR) was used to analyse curated panels of stem cell and Wnt-target genes (detailed in 2.11.1). An appropriate stem cell panel was determined based on established stem cell players, some of which were differentially expressed upon *Runx1* depletion in the ß-catenin mouse model as demonstrated by RNA-Seq analysis in Chapter 3. The Wnt-target gene panel for this investigation was predominantly decided based on a definitive list of Wnt target genes relevant to colorectal cancer models (Sansom, Meniel et al. 2007), used to generate a gene set in the GSEA database (SANSOM_WNT_PATHWAY_REQUIRE_MYC). A selection of these genes were validated using qPCR analysis, by comparing their expression in MMECs isolated from pregnant *Blg-Cre;Catnb*$^{wt/wt}$ and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice. The results of this investigation are summarised in Chapter 3, Table 3.2, and box

plots of gene expression shown in Appendix 1. Due to the involvement of several Wnt pathway members in the control and maintenance of various types of stem cells, many of the genes in the Wnt panel may also be studied within the context of stemness. Analyses of 3 representative stem cell panel genes (*Aldh1a1*, *Aldh1a7*, and *Nanog*) and 2 Wnt target panel gene (*Lgr5* and *Wif1*) are shown in the below figures (Figures 4.5 to 4.9).

Aldh1a1 and its rodent-specific paralogue, *Aldh1a7*, are highly validated stem cell markers in both normal and malignant breast cells. *ALDH1A1* expression has previously been associated with tumour initiation, invasion and metastasis in breast cancer cells, in addition to poorer overall survival in breast cancer patients (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016). *Aldh1a1* and *Aldh1a7* were both upregulated in RFP+ mammary epithelial cells, isolated from *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ mammary glands with *Runx1* deletion compared with their *Runx* wild-type counterparts (Chapter 3).

Nanog, which is expressed at high levels in many breast cancers, has been associated with various stem-like properties in breast cancer cells, including their ability to: form spheres and 3D colonies; promote tumour growth and metastasis; and mediate drug resistance (Lu, Mazur et al. 2014, Jeter, Yang et al. 2015). Interestingly, one study found that the ability of *Nanog* to induce mammary tumorigenesis and promote metastasis was highly dependent on its co-expression with *Wnt1* (Lu, Mazur et al. 2014), indicating it may be involved in the Wnt signalling pathway (as a target or promoter).

Wif1 and *Lgr5* were initially chosen as Wnt target panel members based on their inclusion in the Sansom GSEA Wnt target gene set. *Lgr5* was validated as a relevant target of the Wnt/ß-catenin pathway in the context of the mouse mammary epithelium (Chapter 3, Table 3.2), while *Wif1* was not significantly altered in the Wnt activated mammary gland. *Wif1* is a negative regulator of the Wnt signalling pathway whose expression is positively correlated with Wnt pathway activation in the normal context. Its inactivation by epigenetic silencing is a common

occurrence in breast cancer, which allows for the oncogenic activation of the Wnt/ß-catenin signalling pathway (Ai, Tao et al. 2006). In addition to its role as a potentiator of the Wnt/ß-catenin signalling pathway, the *Lgr5* gene is a well-established molecular marker of stem cells in various organs and tissues (Nusse and Clevers 2017), including in the breast (Plaks, Brenot et al. 2013, Trejo, Luna et al. 2017). LGR5 was shown to be activated through the Wnt/ß-catenin signalling pathway in order to promote various aggressive behaviours in breast cancer cells that are associated with stem cells (including spheroid formation, cell mobility, invasion and metastasis). It was found to be overexpressed in primary breast cancer tissues, where its expression was additionally correlated with adverse clinical variables and poorer patient outcomes (Yang, Tang et al. 2015). In addition, the pilot RNA-Seq analysis in Chapter 3 identified *Lgr5* as differentially expressed in ß-catenin-activated mammary epithelial cells upon loss of functional *Runx1* and *Runx2* that, in addition to enhanced oncogenic activation of the Wnt pathway, displayed a more stem-like phenotype than their *Runx* wild-type counterparts. For the purpose of this study, this gene can therefore be considered both within the context of its role as a stem-cell regulator and Wnt pathway potentiator.

Analyses of the normalised relative expression values for each gene indicate that RUNX1 overexpression results in a reduction in the expression levels of a variety of stem cell and/or Wnt pathway genes (Figure 4.5 to 4.9). Part [A] of each figure gives an overview of the expression levels of each of the test genes within each of the experimental conditions. Parts [B] to [E] explore these test groups further by comparing the cell lines within each of the experimental conditions. RUNX1 overexpression in the HC11 cells resulted in notably reduced levels of all 5 genes, particularly in Vehicle [D] and WNT3A [E] treated mammospheres derived from this cell line compared to their Empty Vector counterparts. Although the change in *Lgr5* expression in the WNT3A treated mammospheres was not statistically significant, there is a clear trend towards reduced *Lgr5* levels upon RUNX1 overexpression in this condition. Gene expression changes within the 2D growth conditions, [B] and [C] were slightly variable from gene to gene. *Aldh1a1* (Figure 4.5) and *Wif1* (Figure 4.9) genes exhibited significantly decreased expression levels within the Runx1P1 cell line grown in both 2D conditions. Changes in

Aldh1a7 (Figure 4.6) and *Nanog* (Figure 4.7) expression only reached significance in 2D cultured Runx1P1 cells upon treatment with WNT3A, although there were clear trends towards reduced expression in the vehicle treated 2D Runx1P1 cultures compared with their Empty Vector counterparts. *Lgr5* was not significantly altered in the 2D context, nor was there a distinct bias towards either increased or reduced expression with RUNX1 overexpression.

Additional stem cell or Wnt pathway associated genes that were tested including *Trp63* (Appendix 2), *Ly6a* (Appendix 3), *Esr1* (Appendix 4), and *Notum* (Appendix 5) displayed varying degrees of depletion in the Runx1P1 cell line compared with the Empty Vector control cells. *Ly6a* was significantly reduced in RUNX1 overexpressing cells grown in all culture conditions. *Trp63* was significantly depleted in Runx1P1 cells cultured in all except the 2D vehicle treated condition, although there was still a clear trend towards its reduction in this context. *Esr1* expression was only significantly altered in mammospheres derived from Runx1P1 cells compared with their Empty Vector counterparts, and those grown in the 2D conditions did not display a bias towards either increased or decreased *Esr1* expression. It was only within the WNT3A treated conditions in which *Notum* expression exhibited clear tendency towards reduction upon RUNX1 overexpression, although this reduction only reached statistical significance within the mammospheres. Wnt pathway associated genes *Myc* (Appendix 6), *Sox9* (Appendix 7), and *Wnt3a* (Appendix 8), which also function as markers or mediators of stem cell activity in the normal and tumorigenic mammary gland (Zeng and Nusse 2010, Guo, Keckesova et al. 2012, Moumen, Chiche et al. 2012, Moumen, Chiche et al. 2013, Domenici, Aurrekoetxea-Rodríguez et al. 2019), did not appear to be altered in any way upon overexpression of RUNX1 within the HC11 cells.

Interestingly, not all of the analysed stem or Wnt pathway related genes were increased in mammospheres, which are enriched for cells with stem-like properties, compared with 2D cultured cells. Likewise, not all genes were responsive to recombinant WNT3A treatment, which would act as a mediator of stemness through its enhancement of the canonical Wnt/β-catenin pathway.

Representative genes *Aldh1a7* (Figure 4.6), *Lgr5* (Figure 4.8), and *Wif1* (Figure 4.9) display slightly increased levels of expression in Empty Vector HC11 cells grown within the 3D context compared with their 2D-cultured counterparts. These genes were not altered between the 2D and 3D growth conditions in the Runx1P1 cell line, however, due to the significantly dampened levels of expression for each gene within this cell line. For *Wif1*, this difference between the 2D and 3D cultures was only present when comparing those treated with WNT3A. Additionally, it was only within the 3D growth conditions that WNT3A treatment elevated *Wif1* expression levels above those seen with vehicle treatment. WNT3A treatment did not appear to have an effect on any of the other representative genes shown, and growing HC11 cells in 3D conditions as opposed to 2D appeared to have no effect on *Aldh1a1* and *Nanog* expression levels. Of the additional genes whose expression was tested but not shown in the figures below, *Notum*, which is a Wnt/ß-catenin target gene that in turn regulates Wnt signalling in a negative feedback loop (Giráldez, Copley et al. 2002, Flowers, Topczewska et al. 2012, Kakugawa, Langton et al. 2015), was increased when grown in 3D conditions compared to 2D cultures, in addition to being significantly increased upon treatment with WNT3A. Notably, the Wnt and stem cell related gene *Sox9* was significantly upregulated in 3D cultured HC11 cells compared with their 2D grown counterparts, although was not impacted by WNT3A.

Gene expression of *Runx2* was also evaluated and compared in each sample group, and the results summarised in Figure 4.10. The purpose of this was to rule out the possibility that the gene alterations contributing to a reduced mammosphere forming potential in Runx1P1 cells were caused by alterations in *Runx2* expression as opposed to RUNX1 upregulation. In a previous publication from the Blyth laboratory, it was revealed through *in vitro* and *in vivo* analyses that *Runx2* is involved in the stem/progenitor potential of the mammary epithelium and was identified as a mediator of Wnt signalling in mammary cultures enriched for stem cells (Ferrari, Riggio et al. 2015). There would therefore need to be a significant reduction in *Runx2* expression levels in each test condition in order to explain the reduced mammosphere forming potential, and depletion of stem cell and Wnt pathway related genes, in the Runx1P1 cell line. It can be clearly seen in Figure 4.10 that this is not the case, and in the 2D cultures there was even an increase

in *Runx2* expression levels that was statistically significant with vehicle treatment. This suggests that the overexpressed RUNX1 was perhaps outcompeting RUNX2 within this context in order to negatively regulate the stem-like potential of the cells and limit Wnt pathway activation.

These observations support results obtained from mammosphere experiments, which showed an overall reduction in stem cell activity and repression of Wnt-mediated stemness. Additionally, they give some indication as to which stem cell promoting and Wnt pathway mediating genes could potentially be regulated by *Runx1*, in order to prevent tumorigenesis in the context of the normal mammary epithelium. Differential expression of several stem and Wnt target genes in response to RUNX1 overexpression, particularly within stem cell enriched cultures (Figures 4.5-4.9, and Appendices 2-8), indicates that *Runx1* could be functioning upstream of these genes to regulate their expression, and therefore their function. This relationship suggests that the tumour suppressive role of *Runx1* could be at least partially mediated by exerting control over the proliferative capacity of stem cells within the mammary epithelial cells, thereby protecting these cells from oncogenic insult, at least within the context of a constitutively activated Wnt/ß-catenin pathway.

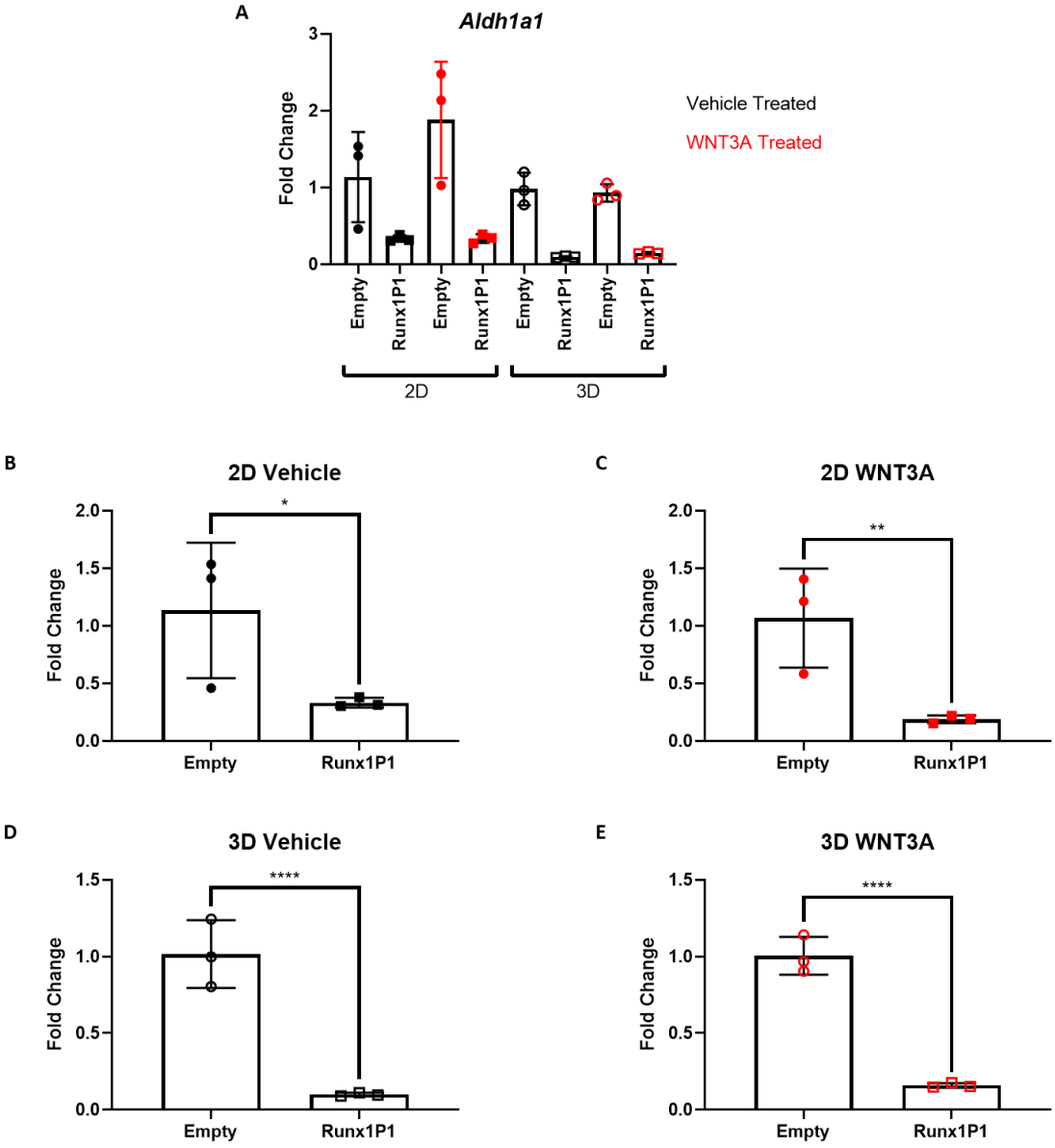


Figure 4.5 qPCR analysis reveals significantly reduced levels of stem cell marker gene, *Aldh1a1*, upon overexpression of RUNX1 in both the 2D and 3D contexts.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D (days of growth D0 and D7) or 3D (mammosphere) growth conditions with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Aldh1a1*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$].

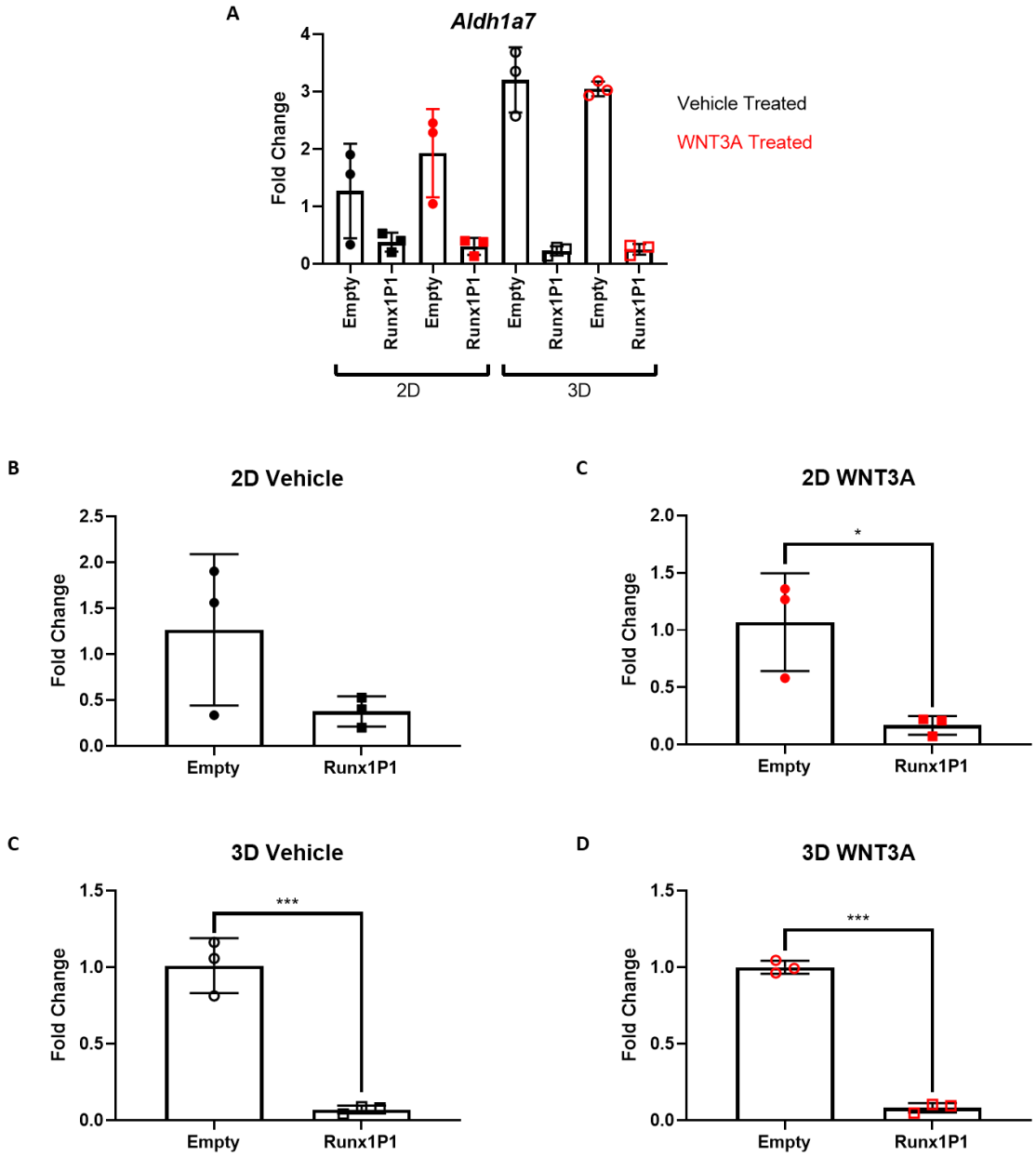


Figure 4.6 Expression of mouse-specific stem cell marker gene, *Aldh1a7*, is significantly reduced in Runx1P1-derived cultures.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Aldh1a7*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [* $p < 0.05$; *** $p < 0.001$].

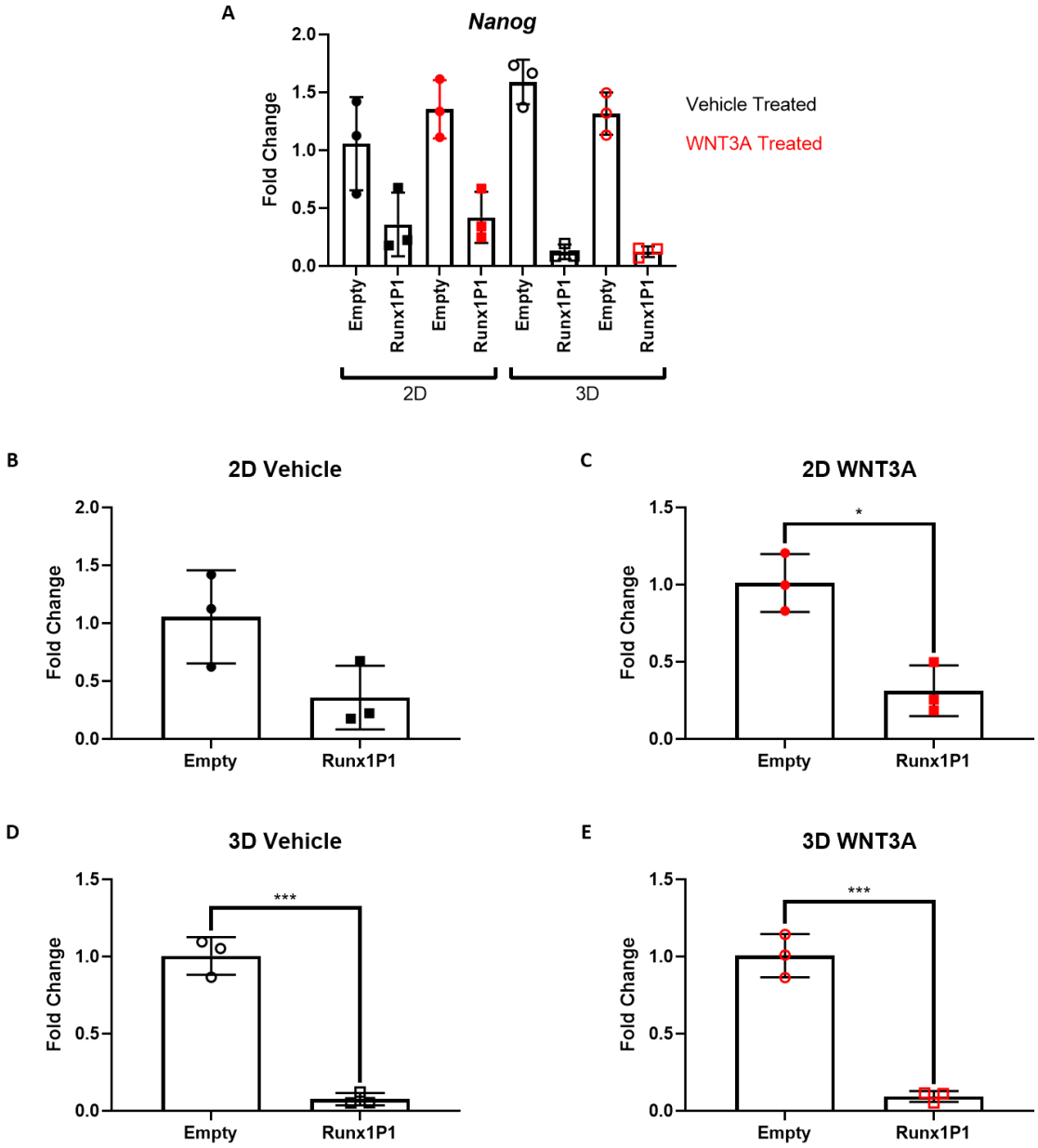


Figure 4.7 Expression of *Nanog*, a stem-related gene whose tumorigenic potential is dependent on *Wnt1* expression, is reduced in RUNX1-overexpressing cell cultures.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Nanog*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [* $p < 0.05$; *** $p < 0.001$].

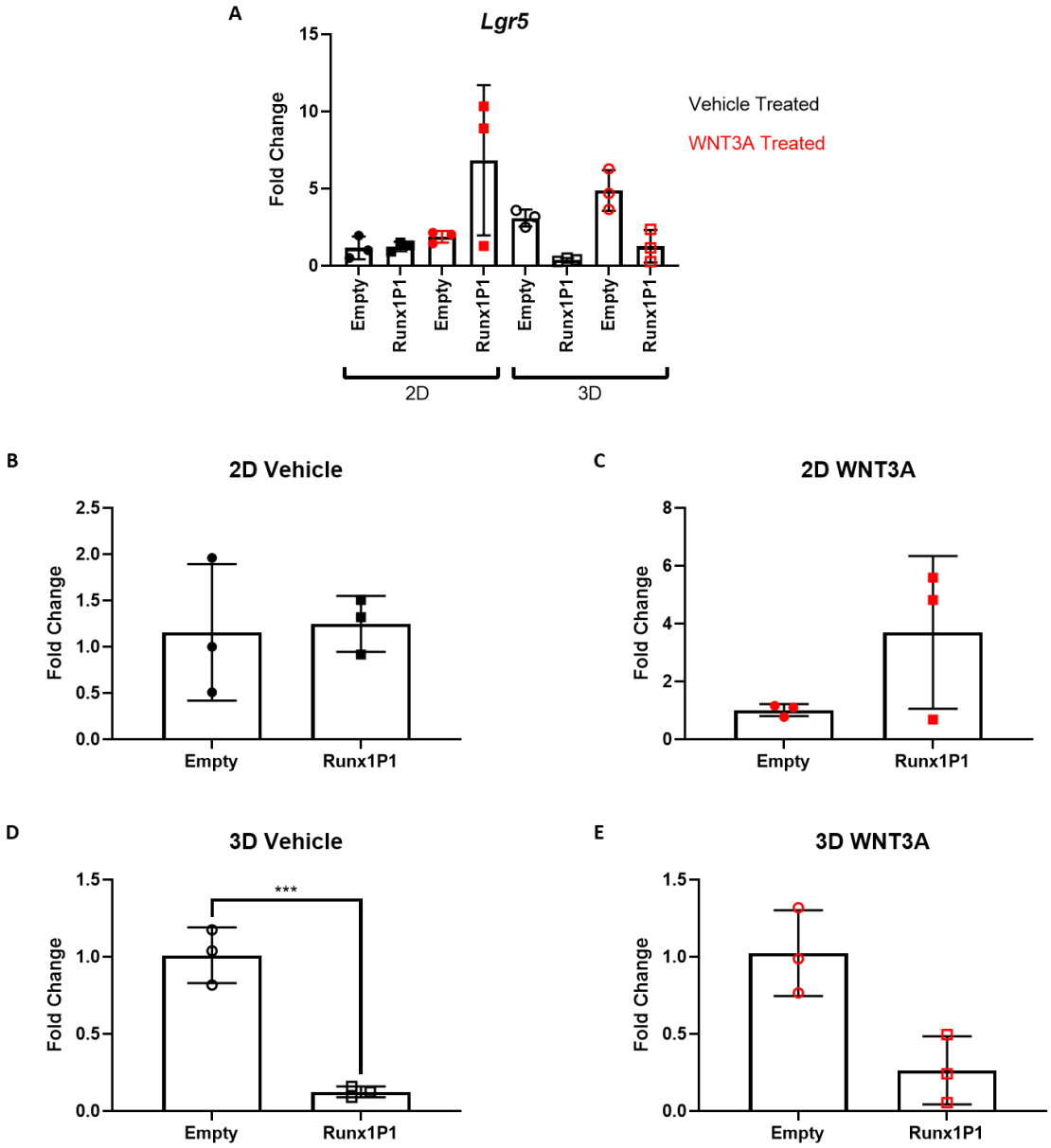


Figure 4.8 *Lgr5*, a potentiator of the Wnt/ß-Catenin pathway and a mammary stem cell marker, appears diminished with RUNX1 overexpression under 3D culture conditions.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Lgr5*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [*** $p < 0.001$].

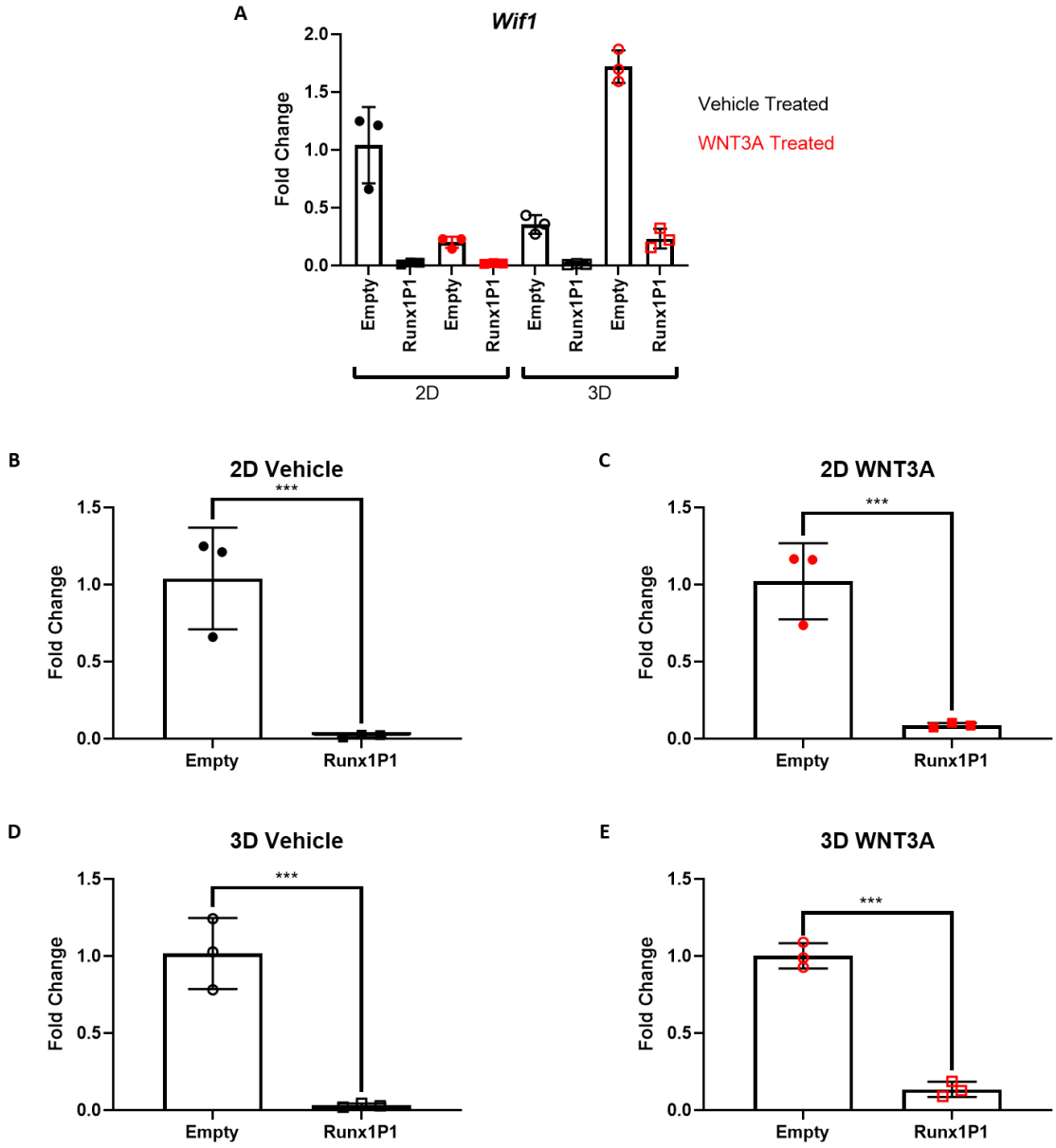


Figure 4.9 Reduced expression of *Wif1*, a negative regulator of the Wnt pathway, in RUNX1 overexpressing mammary epithelial cell cultures.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Wif1*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [*** $p < 0.001$].

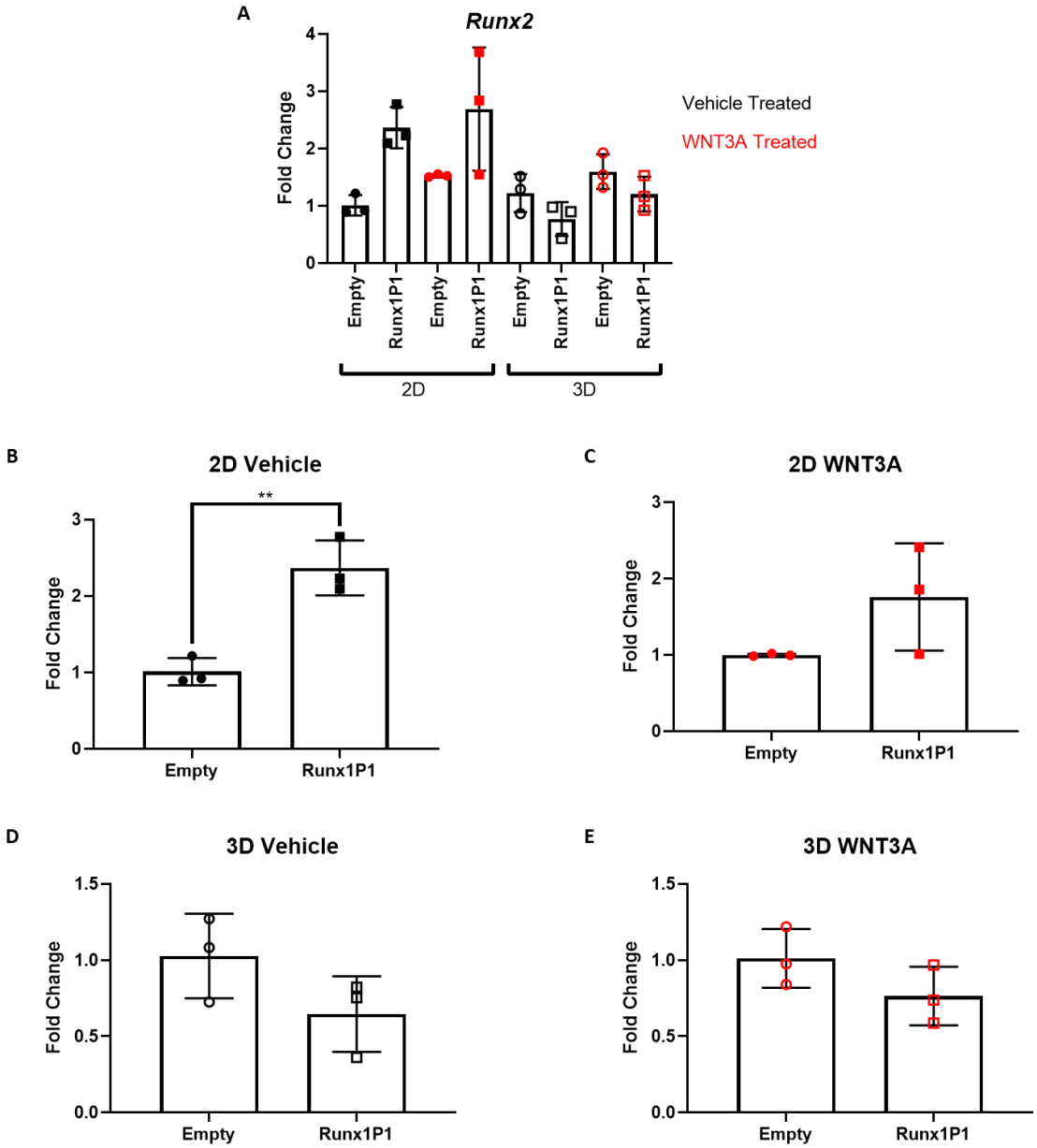


Figure 4.10 *Runx2* expression increased in 2D vehicle treated Runx1P1 cells, though is otherwise relatively unchanged.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Runx2*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [** $p < 0.01$].

4.3 Discussion

Throughout this investigation, the HC11 mouse mammary epithelial cell line was used as an *in vitro* model to study the role of the *Runx1* gene in non-tumorigenic mammary epithelial cell populations, particularly the stem/progenitor cell compartment. Due to the stem-/progenitor-like potential of the HC11 cells, and their ability to differentiate when grown at confluence in medium containing a combination of lactogenic hormones (dexamethasone, insulin, and prolactin), this cell line has been key for several investigations into mammary stem cell proliferation and differentiation (Slosberg, Klein et al. 1999, Pietersen, Evers et al. 2008, Williams, Helguero et al. 2009, Liu, Pawliwec et al. 2015, Sornapudi, Nayak et al. 2018, Niit, Geletu et al. 2019). Initial investigations using Western Blot analysis of various cell types revealed that the HC11 cell line expresses intermediate levels of RUNX1 relative to the other selected cell types that are known to express at least moderate levels of the protein. Based upon this expression analysis, it was reasonably assumed that the HC11 mammary epithelial cells could potentially be amenable to either overexpression or depletion of RUNX1. It was decided that initial attempts to alter RUNX1 expression would be in the form of RUNX1 overexpression, using a pBABE-Puro retroviral vector, as this vector was successfully used to overexpress RUNX2 in a previous Blyth lab publication (Ferrari, Riggio et al. 2015).

Mammosphere and 3D mammary colony-forming cell assays were then chosen as tools to aid with this investigation. These methods have been used in several other studies to examine the potentiation and self-renewal ability of mammary cell populations (Dontu, Abdallah et al. 2003, Stingl, Emerman et al. 2005, Cicalese, Bonizzi et al. 2009, Shaw, Harrison et al. 2012, Lombardo, de Giorgio et al. 2015).The mammosphere assay in particular has been shown to select for mammary stem/progenitor cells with a high capacity for self-renewal, as one study demonstrated that the percentage of cells with a bilineage differentiation potential increased significantly with each subsequent mammosphere generation, until almost all colonies (~98%) were derived from bipotent mammary stem/progenitor cells (Dontu, Abdallah et al. 2003). In another study, limiting dilutions of cell suspension from secondary mammospheres were injected into cleared fat pads of mice and successfully induced mammary outgrowth, which was

seen to a much lesser extent (primarily with lower cell dilutions) upon injecting primary mammary cells (Cicalese, Bonizzi et al. 2009). This indicates that the mammosphere assay can be used to enrich for populations of mammary stem/progenitor cells.

Mammosphere assays have also been used to study the stem-like properties and lactogenic differentiation of the HC11 cell line (Morrison and Cutler 2009, Thrasyvoulou, Vartholomatos et al. 2020). More relevantly, former Blyth lab PhD students used the HC11 mammary epithelial cell line as a tool to investigate the function of RUNX2 in the stem/progenitor cell compartment of the mammary epithelium (Ferrari, Riggio et al. 2015). Mammospheres that were derived from HC11 cells were shown to exhibit increased expression levels of the *Runx2* gene and its associated protein. Interestingly, while there were no differences observed in 2D growth upon overexpression of RUNX2, there was a significant increase in the mammosphere forming potential of HC11 cells. Results presented in this chapter show a strikingly similar context-specific change in cell behaviour in response to successful overexpression of RUNX1 in the HC11 cell line. Whilst RUNX1 overexpression was found to be inconsequential to the 2D growth of these cells, the mammosphere and colony forming potential was significantly altered, although it was reduced rather than enhanced. Phenotypic characterisation of mammospheres produced by Empty Vector and Runx1P1 HC11 cells revealed a significant increase in the size of mammospheres formed by cells that expressed heightened levels of RUNX1. Upon further inspection, these larger structures contained higher proportions of actively proliferating cells and lower percentages of cells undergoing programmed death. Collectively, the data presented here indicate that RUNX1 may function as a stem cell regulator in these non-tumorigenic mammary epithelial cells by initiating differentiation, leading to the reduced formation of mammospheres and colonies. The rate of differentiation within these Runx1P1-derived mammospheres were elevated, while the levels of cell death were reduced, which likely contributed to their larger sizes. Further experiments, looking into the differentiation status of the individual cells within each mammosphere, would need to be conducted to confirm this mechanism.

The regulation of the mammary stem cell population by *Runx1* was further explored within the context of the stemness-enhancing, cancer-associated Wnt signalling pathway. Treatment of HC11 cells with recombinant WNT3A protein, a key proponent of the canonical Wnt/ß-catenin pathway whose associated gene expression is frequently upregulated in breast cancer cell lines compared to normal mammary epithelial cells (Benhaj, Akcali et al. 2006, He, Lu et al. 2015), resulted in an enrichment of the stem cell compartment of these cells. However, there was a slight reduction in the enhancement of stem cell activity in these cells with RUNX1 overexpression, indicating that this protein is capable of limiting the stem-like phenotype of mammary epithelial cell populations both in the context of the normal mammary gland function and following oncogenic insult. Quantitative PCR analyses were conducted to explore the molecular mechanisms behind RUNX1-mediated stem cell regulation, with and without exogenous WNT3A. RUNX1 overexpression was seen to reduce the activity of various stem/progenitor cell-(*Aldh1a1*, *Aldh1a7*, *Nanog*, *Ly6a*, *Trp63*) and Wnt pathway-related (*Lgr5* and *Wif1*) genes, particularly in mammary stem cell-enriched cultures of HC11 cells. These results taken together indicate that the function of *Runx1*, within normal mammary gland homeostasis and as a tumour suppressor, are at least partially facilitated via its influence over the stem cell populations within the mammary epithelium. From these results, it could be inferred that RUNX1 is interacting with genes related to the stemness of the mammary epithelium, as well as directly/indirectly regulating several Wnt pathway targets, to maintain a controlled population of stem cells within the mammary gland. Loss of functional *Runx1* in mammary epithelial cells, as is often observed in ER-positive cancers, may lead to uncontrolled expansion of a stem-like population of mammary epithelial cells within the normal mammary gland, thus making it more susceptible to oncogenic insult.

The investigation presented in this chapter does have some additional factors that should be considered while interpreting the data. Firstly, it is possible that the tumour suppressive role of *Runx1* could be context specific, and that it may only work to suppress tumorigenesis in the setting of constitutively activated Wnt/ß-catenin, particularly when considering its role in limiting stemness to prevent tumour initiation. Investigations conducted by former PhD student, Alessandra

Riggio, helped to address concerns relating to the overall tumour suppressive capabilities of *Runx1* within the context of other independent models of mammary tumorigenesis. The *MMTV-PyMT* transgenic model was one such mouse line used during these studies, and was additionally considered within the context of stemness as described in Chapter 7. Another reflection from this work is that it cannot be said with absolute certainty that the differentially regulated stem cell markers, identified by qPCR, are functionally significant for the stemness of the mammary epithelium, or are simply markers whose expression coincides with the stem-like potential of the cells. To investigate this in future, it may be possible to rescue the expression of the specific stem/Wnt gene or its associated protein, using either a retroviral vector or recombinant protein respectively, in the RUNX1 overexpressing line and observe whether there is at least a partial rescue of the stem-like phenotype using mammosphere assays.

There are also limitations relating to the methodologies that should be discussed. For example, whilst *in vitro* studies offer many advantages over *in vivo* work (tighter control over the cell's environment, lower costs, shorter experiments, and higher throughput), they do have some drawbacks that should be considered. In order to facilitate the long-term culture of various cell lines, the primary cells used to make that cell line are genetically manipulated by the introduction of SV40 or HPV sequences, allowing the cells to evade normal cellular senescence. This genetic manipulation may alter their phenotype and behaviours, and serial passage could induce further significant variations and genetic drift over time, meaning that experiments using earlier passages could potentially yield rather different results compared to later passages. Within the context of the HC11 cells used, whilst the primary cells from which this cell line originated were considered normal and non-tumorigenic, the immortalised line may not accurately represent the mammary epithelial cells of a normal mammary gland. To circumvent this limitation during future investigations, it may be possible to extract normal primary MMECs from mice and alter RUNX1 expression *in vitro*.

There are also concerns relating specifically to the mammosphere in vitro assay. One caveat is that it is unclear as to whether the results from this assay truly

reflect the number of stem and progenitor cells present. In particular, it is not known whether the assay can identify quiescent stem cells in the mammary gland, or whether it measures cells that are simply most capable of adapting to the artificial/harsh conditions of the assay or which can act as proliferating mammary stem cells (Rota, Lazzarino et al. 2012). To support the findings from this experiment, it would therefore be useful if these results could be recapitulated in an *in vivo* context using a mammary reconstitution assay (Illa-Bochaca, Fernandez-Gonzalez et al. 2010, Rota, Lazzarino et al. 2012). This assay can be used to measure the stem cell frequency within an epithelial cell population by looking for the presence of epithelial outgrowths, over a period, in host mice that had serial dilutions of single-cell suspensions transplanted into their cleared fat pads. This assay was also used for the characterisation of a RUNX1-depleted HC11 cell line, discussed in Chapter 5. Finally, primary mammosphere formation was used as an indication of stem cell activity within the HC11 cells, however this does not give an indication of the capacity for ongoing self-renewal, which is a key feature of stem cells. Disaggregation of primary mammospheres and further passage of the cells as secondary, or even tertiary, mammospheres would allow for a more complete investigation into the stem-like properties of HC11 cells with altered expression of RUNX1, as addressed in Chapter 5.

Despite the limitations and considerations that should be taken into account in further similar studies, the data presented in this chapter align with previously published data indicating that RUNX1 is required for the differentiation of mammary stem cells in 3D culture (Sokol, Sanduja et al. 2015) and, in the context of breast cancer, suppresses the breast cancer stem cell phenotype and reduces tumour growth, migration, and invasion (Hong, Fritz et al. 2018, Fritz, Hong et al. 2020). It is also in agreement with findings that some of the developmental and regulatory functions of RUNX1 are enabled through interactions with several components of the Wnt pathway (Osorio, Lilja et al. 2011, Wu, Seay et al. 2012, Naillat, Yan et al. 2015, Luo, Zhang et al. 2019). One study of particular interest showed that RUNX1 could act as a mammary tumour suppressor against ER-positive breast cancer through its suppression of oestrogen-mediated inhibition of AXIN1, thus stabilising ß-catenin and preventing its deregulation (Chimge, Little et al. 2016). This chapter, along with Chapters 5 and 6, builds upon those

aforementioned studies as well as recent work from the Blyth lab (publication in progress) using *in vivo* models to study the tumour suppressor function of RUNX1 in breast cancer. The work presented in this thesis looks further into the potential molecular mechanisms by which loss of functional RUNX1 in the normal mammary gland may lead to oncogenesis, as opposed to focusing on the effects of RUNX1 depletion in an already established tumour. This will become clearer in the next chapter of this thesis, which aims to explore the effects of *Runx1* depletion in HC11 mouse mammary epithelial cells.

5 Deletion of RUNX1 enhances the progenitor-like potential of HC11 cells

5.1 Introduction

RUNX1 is mutated in ~6-9% of breast cancer cases studied in various targeted sequencing studies of primary breast tumours explored through cBioPortal (Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013). These mutations in the *Runx1* gene predominantly consist of loss of function mutations, particularly within the context of ER-positive breast cancers (as was discussed in Introduction) (Cerami, Gao et al. 2012, Gao, Aksoy et al. 2013). It is important that the models, used for investigating the function of *Runx1* in the normal mammary epithelium and its role in tumour suppression, recapitulate those findings in order for it to have clinical relevance. The previously described *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mouse model proved to be very informative during investigations into the role of *Runx1* as a tumour suppressor. Targeted loss of functional *Runx1* in the mammary epithelium resulted in earlier emergence and increased numbers of independently arising multifocal and multicentric tumours (Chapter 3). This stem-related phenotype, which was accompanied by enhanced expression of established stem cell markers, was even more dramatic with the additional loss of functional *Runx2* (Chapter 3). This potential role of RUNX1 in the stemness of the mammary epithelium was further investigated using an *in vitro* model that could be easily manipulated and studied.

5.1.1 Experimental Aims

Overexpressing RUNX1 in the HC11 cell line (Chapter 4) gave valuable insights into some of the functions of this protein, though they are restricted in what can be inferred about the implications of RUNX1 loss, as it cannot necessarily be assumed that the consequences of RUNX1 depletion will precisely mirror the effects seen by overexpressing the protein. This is partly because *Runx1* is involved in complex interactions with several intricate pathways, including the previously discussed canonical Wnt/ß-catenin pathway, and so there may be other compensatory mechanisms that are triggered in response to RUNX1 deletion to protect mammary epithelial cells from oncogenic transformation. For this reason, it was essential to consider knockout of the RUNX1 protein in HC11 cells.

Given that tumour emergence was remarkably accelerated with the additional loss of functional *Runx2* in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mouse model, another aim of this investigation was to explore the impact of simultaneous loss of RUNX1 and RUNX2 on mammary epithelial cells. The dramatic changes observed in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mice were rather surprising, given that *Runx2* loss alone resulted in a phenotype that was not discernible from wildtype controls (Chapter 3). In another *in vivo* study, embryonic mammary buds, dissected from *MMTV-PyMT* embryos (either *Runx2*$^{+/+}$ or *Runx2*$^{-/-}$), showed delays in tumour initiation and progression in the absence of *Runx2* following transplantation into *Rag1*$^{-/-}$ hosts (Owens, Rogers et al. 2014). In addition, a previous publication from the Blyth lab showed that elevated expression of RUNX2, rather than deletion, left the mammary gland more vulnerable to oncogenesis in a transgenic model (McDonald, Ferrari et al. 2014). *Runx2* has also shown strong associations with mammary cell compartments containing high proportions of stem/progenitor cells (Kendrick, Regan et al. 2008, Molyneux, Geyer et al. 2010, McDonald, Ferrari et al. 2014), including the terminal end buds (Kouros-Mehr and Werb 2006). Studies performed in the Blyth lab showed that RUNX2 is functionally important in mammary stem/progenitor cells (supporting data from Owens et al, demonstrating that *Runx2* overexpression perturbs HC11 differentiation (Owens, Rogers et al. 2014)) and potentially mediates Wnt signalling in mammary stem cells (Ferrari, Riggio et al. 2015). Considering this information, it could be hypothesised that loss of RUNX2 would cancel out the phenotype associated with RUNX1 depletion, rather than enhancing it. Regardless, the HC11 cell line offers a suitable model to investigate this fascinating observation in more depth.

5.2 Results

5.2.1 Stem-like potential of HC11 cell line significantly enhanced in three independent *Runx1* CRISPR clones

It was previously reasoned, based on the relatively intermediate RUNX1 protein expression levels shown in Chapter 4.2.1, that the HC11 cell line may be amenable to either RUNX1 overexpression or depletion. Results from a RUNX1 overexpressing cell line (Chapter 4) were informative in exploring some potential mechanisms behind the tumour suppressive function of RUNX1 in the mammary epithelium, particularly within the context of oncogenic Wnt/β-catenin signalling and in the

stemness of the mammary epithelium. Here, CRISPR-induced knockout of RUNX1 expression was carried out in order to confirm the role of this protein in the stem-like potential of HC11 mammary epithelial cells.

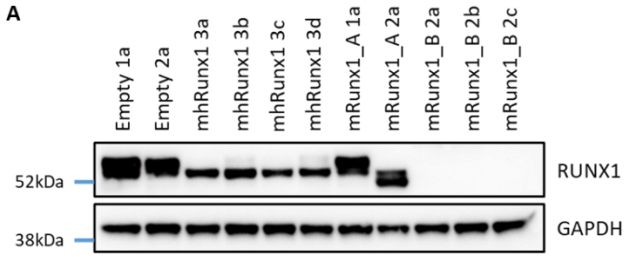
An SpCas9-bearing pX-459-Puro plasmid was modified to produce 3 individual constructs, each of which contained one of three *Runx1*-targeting guide RNA sequences to introduce loss-of-function deletions. HC11 cells were subsequently transfected using each of these three constructs, or using an empty vector as a control. The transfected cell lines were then subjected to antibiotic selection and single cell cloning in order to produce multiple cell lines, which were analysed by Western Blot. Protein expression was used as an indication of whether CRISPR-based targeting of *Runx1* expression was successful. In Figure 5.1A, below, it can be seen that, from the 8 total HC11 cell lines that were created, those generated from *mhRunx1* and m*Runx1*_A guide RNA (gRNA) sequences show incomplete depletion, as indicated by the still clearly detectable levels of protein that are comparable to those seen in the Empty control clones. However, in 3 independent clones, all of which were generated using the m*Runx1*_B gRNA sequence, successful sequence targeting and deletion is assumed based on the loss of detectable protein expression. These 3 clones will henceforth be referred to as Rx1_KO_A, Rx1_KO_B, and Rx1_KO_C.

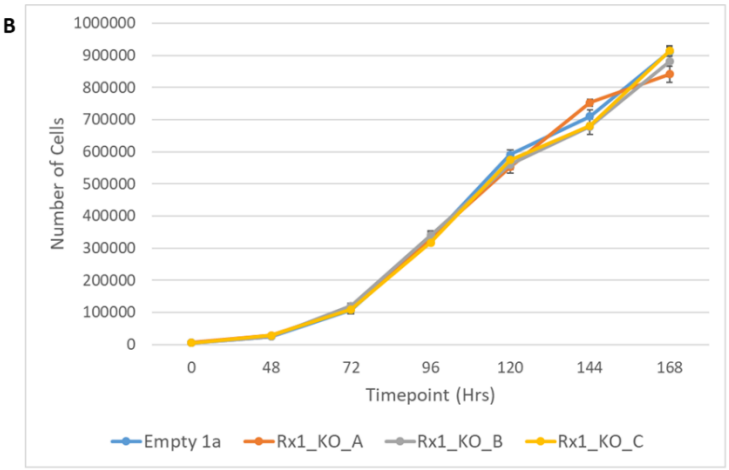
2D growth assays were used initially to compare each Runx1 knockout cell line with the Empty 1a control line. Cell counts were taken at 24 hour intervals for each cell line, set up in 2D growth conditions, using a CASY Cell Counter. The results of the 2D growth measurements, summarised in Figure 5.1B, demonstrate that depleting RUNX1 expression levels has no statistically significant effect on the growth of HC11 cells under 2D conditions. This data is in agreement with that of Chapter 4, in which RUNX1 overexpression did not significantly change the 2D growth of HC11 cells.

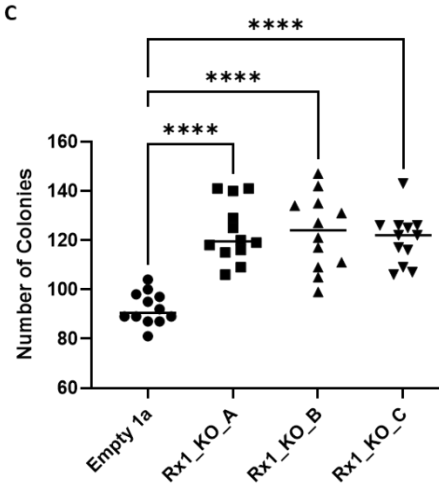
Due to the significant differences found in the colony and mammosphere forming potential of HC11 cells upon overexpressing RUNX1 (Chapter 4), a similar approach

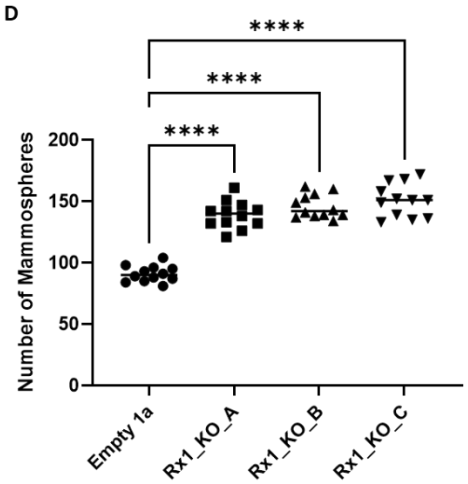
was adopted for the *Runx1* CRISPR cell lines. 3D colony formation assays were used to compare the stem-like potential of each of the RUNX1-depleted cell lines with their Empty vector control. In contrast with the findings that *Runx1* alteration does not significantly affect 2D mammary epithelial cell growth, loss of functional *Runx1* resulted in significantly higher 3D colony-forming potentials for each of the *Runx1* CRISPR clones, relative to the Empty 1a control clone, as shown in Figure 5.1C.

3D colony assays were supported with mammosphere forming assays to confirm whether loss of RUNX1 affects the stem-like potential of these cells. Significantly increased primary mammosphere formation was observed in each of the RUNX1-depleted clones, shown in Figure 5.1D. Further to evaluating primary formation efficiency, the effects of RUNX1 ablation on short-term self-renewal capacity of these mammospheres was investigated by further passaging these mammospheres in secondary mammosphere assays (Shaw, Harrison et al. 2012). This was also discovered to be significantly enhanced with loss of RUNX1 expression, as significantly more secondary mammosphere structures were counted for each of the *Runx1* CRISPR clones compared to their control (Figure 5.1E). Together, these results demonstrate that loss of RUNX1 expression in mammary epithelial cells leads to their increased stem-cell potential, as demonstrated by their increased ability to form and self-renew structures that have been previously linked to a stem-like capacity.


A
Empty 1a
Empty 2a
mhRunx1 3a
mhRunx1 3b
mhRunx1 3c
mhRunx1 3d
mRunx1_A 1a
mRunx1_A 2a
mRunx1_B 2a
mRunx1_B 2b
mRunx1_B 2c
52kDa
RUNX1
38kDa
GAPDH


B
Number of Cells
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
0
48
72
96
120
144
168
Timepoint (Hrs)
Empty 1a
Rx1_KO_A
Rx1_KO_B
Rx1_KO_C


C
****
****
****
Number of Colonies
160
140
120
100
80
60
Empty 1a
Rx1_KO_A
Rx1_KO_B
Rx1_KO_C


D
****
****
****
Number of Mammospheres
200
150
100
50
0
Empty 1a
Rx1_KO_A
Rx1_KO_B
Rx1_KO_C

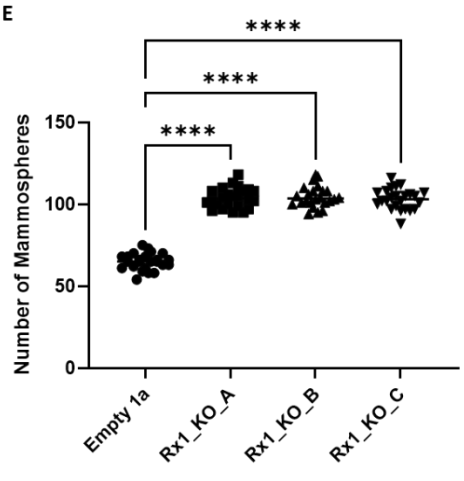

E
****
****
****
Number of Mammospheres
150
100
50
0
Empty 1a
Rx1_KO_A
Rx1_KO_B
Rx1_KO_C

Figure 5.1 Increased stem cell-like potential in three independent *Runx1* CRISPR cell lines.

[A] Western Blot of single cell clones obtained from HC11 cells, transfected with pX-459-Puro vector that was either Empty or contained one of three *Runx1*-targeting sequences. 2 Empty clones (1a and 2a), 4 mh*Runx1* clones (3a-d), 2 m*Runx1*_1 clones (1a and 2a), and 3 m*Runx1*_B clones (2a-c) were analysed. GAPDH as a loading control and Amersham Full Range Rainbow Marker as a protein ladder, with key marker positions and molecular weights shown.

[B] 2D growth curve of Empty 1a clone and 3 independent Rx1_KO clones. n=3 counts per cell line per day, plotted as averages (and population standard deviations). Results shown represent 1 of 3 experimental repeats.

[C] Scatter plot summarising results from colony-forming assays comparing HC11 Rx1_KO clones with Empty 1a control. Each data point represents the number of colonies counted in a single well. n=12 wells were counted per cell line at experimental endpoint of 7 days. Results are representative of 1 of 2 experimental repeats.

Primary [D] and secondary [E] mammosphere counts of Rx1_KO clones summarised in a scatter plot. Each data point represents the number of spheres counted in a single well. n=12 wells counted per cell line after 7 days of growth. Results shown represent 1 of 7 [D] and 4 [E] total experimental repeats, respectively.

One-way ANOVA performed for each, with Dunnett's multiple comparisons test [**** $p < 0.0001$].

5.2.2 Smaller and less proliferative mammospheres derived from HC11 *Runx1* CRISPR clones.

Representative images of HC11 mammospheres, generated from Empty 1a and *Runx1* CRISPR HC11 clones, are shown in Figure 5.2A. The mammospheres that originated from the CRISPR clones appeared to be smaller than those from the control cell line. ImageJ was used to quantify the area (μM^2) of each individual mammosphere in captured images. It is clear from the plotted results (Figure 5.2A) that the individual *Runx1* CRISPR clones exhibit reduced mammosphere sizes, to varying degrees of significance, in comparison to their associated control line.

Sections from paraffin embedded agarose plugs of control and *Runx1* CRISPR clone derived mammospheres were stained for Ki67 (Figure 5.2B) and Caspase 3 (Figure 5.2C) expression by IHC. Scans of the slides were analysed using HALO software, which was trained to detect the percentage of cells within each mammosphere that were stained positive. Mammospheres that were formed from HC11 cells with loss of RUNX1 expression were found to contain lower percentages of cells that were actively proliferating, as summarised in Figure 5.2B. There appeared to be a trend towards reduced apoptosis in all 3 clones, although only results for clone C reached statistical significance (Figure 5.2C). Despite the reduced levels of apoptosis in mammospheres originating from this clone, the more highly significant reduction in proliferation within these structures resulted in smaller mammospheres. This would also be true for mammospheres formed from clones A and B, in which mammosphere cell death was not significantly altered, whilst proliferation was significantly reduced, thereby contributing to their reduced size.

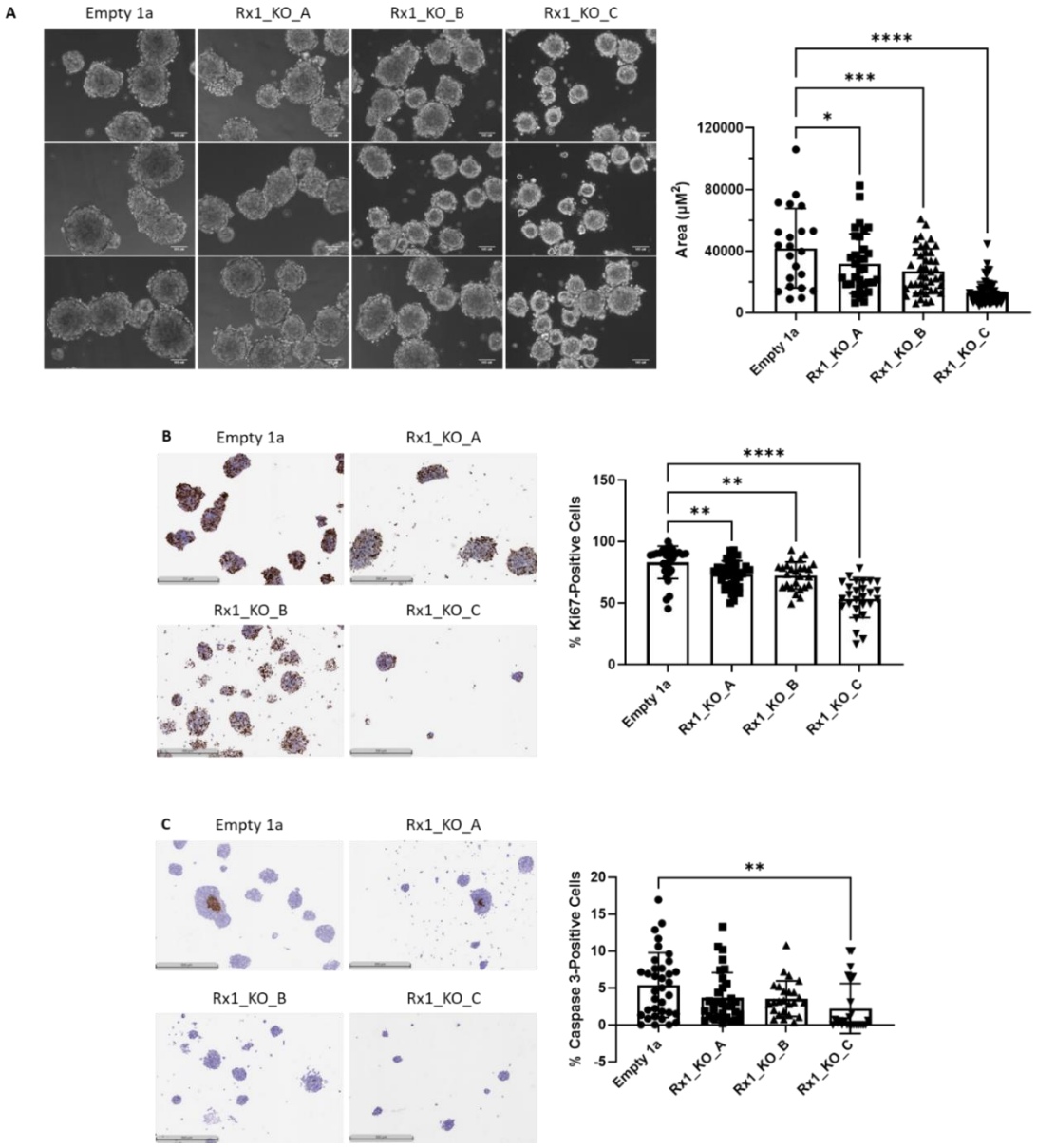


Figure 5.2 *Runx1* CRISPR clones produce smaller mammospheres composed of cells that are less proliferative.

[A] Phase contrast images (left) of mammospheres derived from *Runx1* CRISPR clones and the Empty 1a control. Scale bar=100μM. Images are representative of 6 images taken of each cell line for this particular experimental setup. Mammosphere sizes (μM^2) were quantified with ImageJ software. A 1mm stage micrometer was used to calibrate the scale bar. Scatter plots comparing mammosphere sizes (μM^2) of Empty 1a and *Runx1* CRISPR cell lines (right). Each data point represents the size of an individual mammosphere. Results shown representative of 1 of 3 experimental repeats each with n≥6 images taken per condition.

Representative images (left) and quantification (right) of Ki67 [B] and Caspase 3 [C] staining in sections taken from agarose plugs of HC11 mammospheres, analysed using HALO software. Scale bar=500μM. Scatter plots summarise results of the quantitative analysis, where each data point represents the percentage of total cells in a single mammosphere that stained positive. Results shown represent 1 of 2 experimental repeats.

One-way ANOVA used for each analysis, with Dunnett's Post-Hoc Test to correct for multiple comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$].

5.2.3 Enhanced expression of stem-like markers in RUNX1-deficient mammospheres.

Results thus far have pointed to a role for RUNX1 in the modulation of stemness in mammary epithelial cells. To investigate possible molecular mechanisms behind RUNX1-mediated control of the stem-like potential of the mammary gland, qPCR analysis was used to evaluate the relative expression of a panel of candidate stem cell-related genes in *Runx1* CRISPR clones. The aim of this analysis was to identify the stem cell-related genes that are differentially regulated in the Rx1_KO clones to identify which genes may be facilitating the expansion of a more stem-like subpopulation of mammary epithelial cells in response to the loss of functional RUNX1.

Relative gene expression analyses by qPCR revealed varying degrees of enhancement of a stem cell-related *Aldh1a1* gene and its rodent-specific paralogue, *Aldh1a7*, in each of the *Runx1* CRISPR clones (Figures 5.3 and 5.4). These alterations were generally more or only significant within the stem cell enriched populations of mammospheres that were analysed in parts [D] and [E] of each figure. In instances where statistical analyses did not reveal a particularly significant change in gene expression in every clone, for example within the 2D growth conditions, there was a trend towards increased expression that could be seen. Both of these genes were also shown to be significantly upregulated in *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ mice with the additional mammary-specific loss of functional *Runx1* (Chapter 3) that, in addition to them being highly validated markers of stem cells in both normal tissues and cancer cells (Ginestier, Hur et al. 2007, Khoury, Ademuyiwa et al. 2012, Tomita, Tanaka et al. 2016). Notably, these genes were both expressed at significantly higher levels in the stem cell enriched populations obtained from growth in the 3D conditions as compared with the HC11 cells grown in 2D conditions.

Additional mammary stem cell-related genes — including *Esr1*, *Sox9*, *Lgr5*, and *Runx2* (Figure 5.5A-D) — were also analysed for their expression by qPCR. *Esr1* expression was found to be overexpressed in 2 out of 3 Rx1_KO clones grown as primary and secondary mammospheres, while 1 clone was only altered in the

primary mammosphere condition. The *Sox9* gene, which is another key regulator of mammary stem cell activity and maintenance (Guo, Keckesova et al. 2012, Domenici, Aurrekoetxea-Rodríguez et al. 2019), was overexpressed in the same 2 clones specifically within the secondary mammosphere context. This was also true for *Lgr5* expression in secondary mammospheres, with 2 clones showing significant upregulation and, although it did not reach statistical significance clone Rx1_KO_A displayed a trend towards increased expression. Primary mammospheres derived from the Rx1_KO_C also demonstrated increased expression levels of this gene. *Lgr5* is an established marker of stem cells in the breast, and is also a known Wnt/ß-catenin pathway potentiator and target (Plaks, Brenot et al. 2013, Trejo, Luna et al. 2017). The expression of *Runx2* was also explored, and was found to be upregulated in the Rx1_KO_C clone, specifically within the secondary mammospheres derived from this clone. However, given that these results are inconsistent across the various experimental conditions and 3 individual clones, with only 1 out of 3 clones displaying significantly altered expression levels, it is unclear as to the true significance of this. It cannot necessarily be assumed that the differentially expressed genes that are unique to clone C are actually reflective of the molecular mechanisms of RUNX1, as it may be a clonal artefact or incidental alteration arising in this single clone as a result of genetic drift.

RT-qPCR analyses demonstrated that various stem cell marker genes were significantly, or generally trending towards, increased expression levels in RUNX1-depleted mammary epithelial cells, particularly in populations enriched for cells with a progenitor-like potential. This mirrors the qPCR results in Chapter 4 and supports observations from RNA-Seq experiments (Chapter 3) that revealed differential expression of established stem cell markers (including *Aldh1a1* and *Aldh1a7*) in mice which lacked functional *Runx1*, and in mice with simultaneous loss of *Runx1* and *Runx2* function.

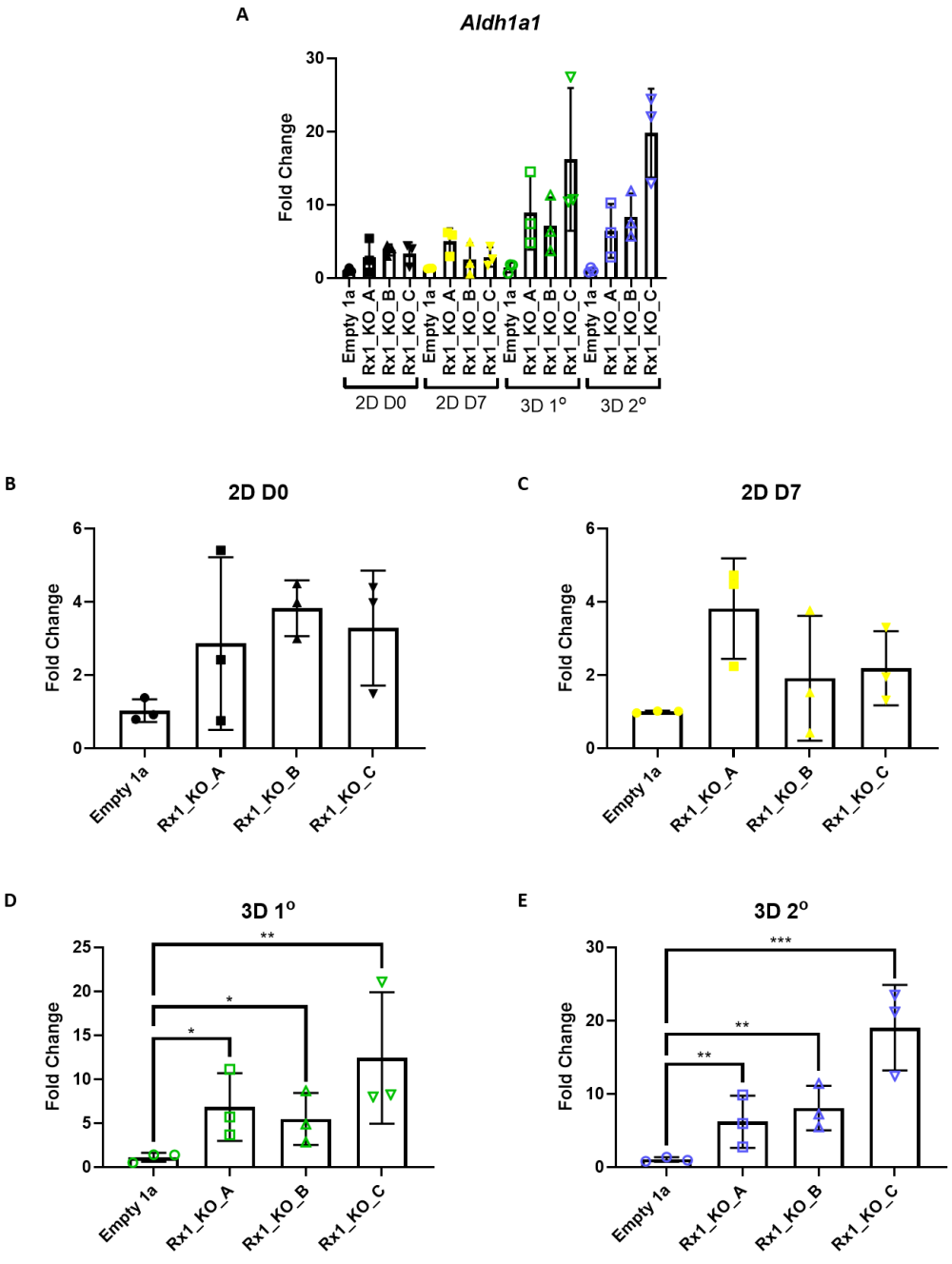


Figure 5.3 Increased expression of stem cell marker *Aldh1a1* in mammospheres derived from *Runx1* CRISPR clones.

Box and whisker plots summarising results from qPCR analysis of cell pellets qPCR analysis of 2D (days of growth D0 and D7) and 3D (primary/1° and secondary/2° mammospheres) HC11 cells, with and without *Runx1* deletion. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Aldh1a1*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$].

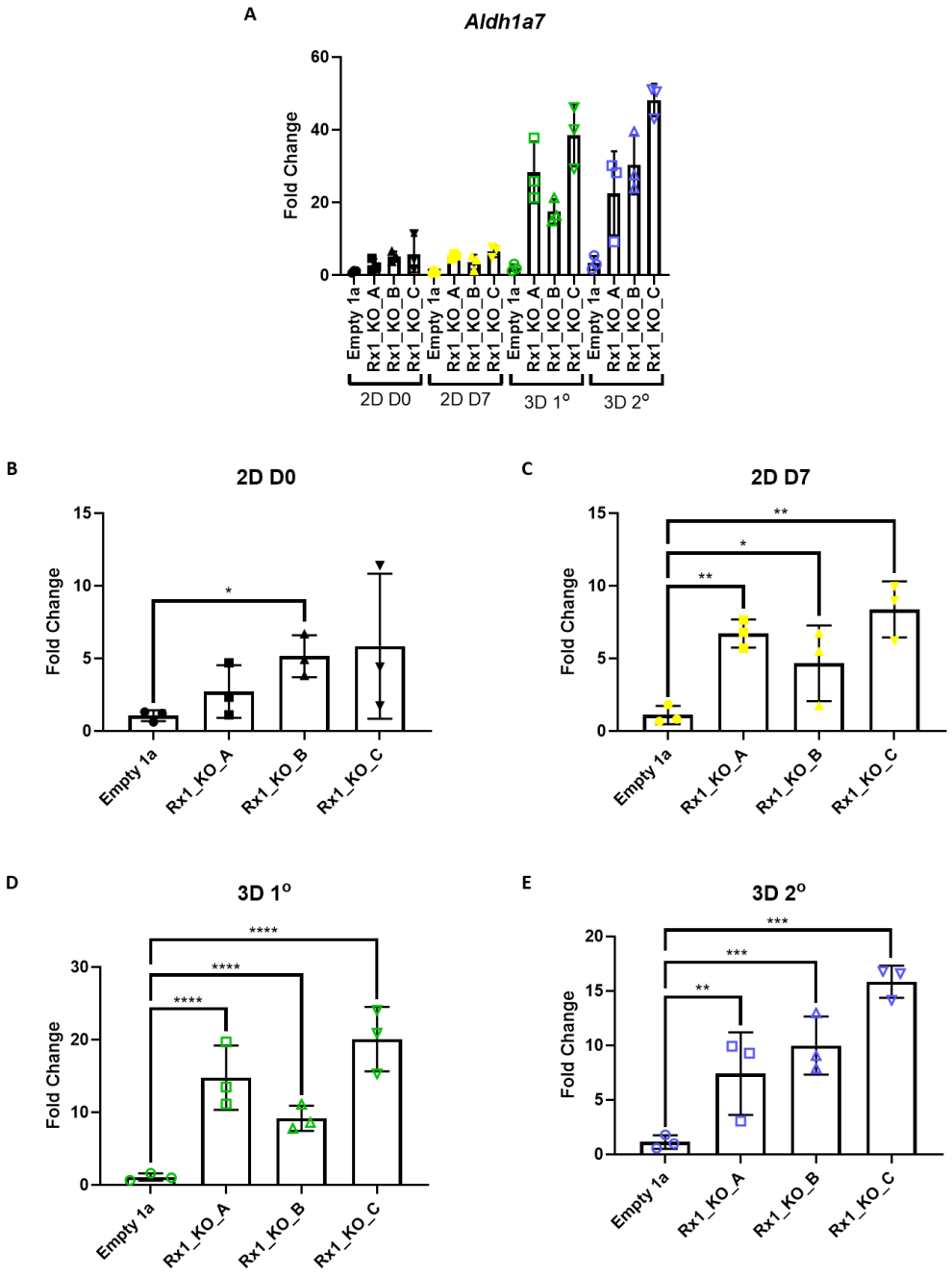


Figure 5.4 Rodent-specific *Aldh1a1* paralogue (*Aldh1a7*) expression is enhanced with RUNX1 depletion.

Box and whisker plots summarising results from qPCR analysis of cell pellets qPCR analysis of 2D (days of growth D0 and D7) and 3D (primary/1° and secondary/2° mammospheres) HC11 cells, with and without *Runx1* deletion. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Aldh1a7*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$].

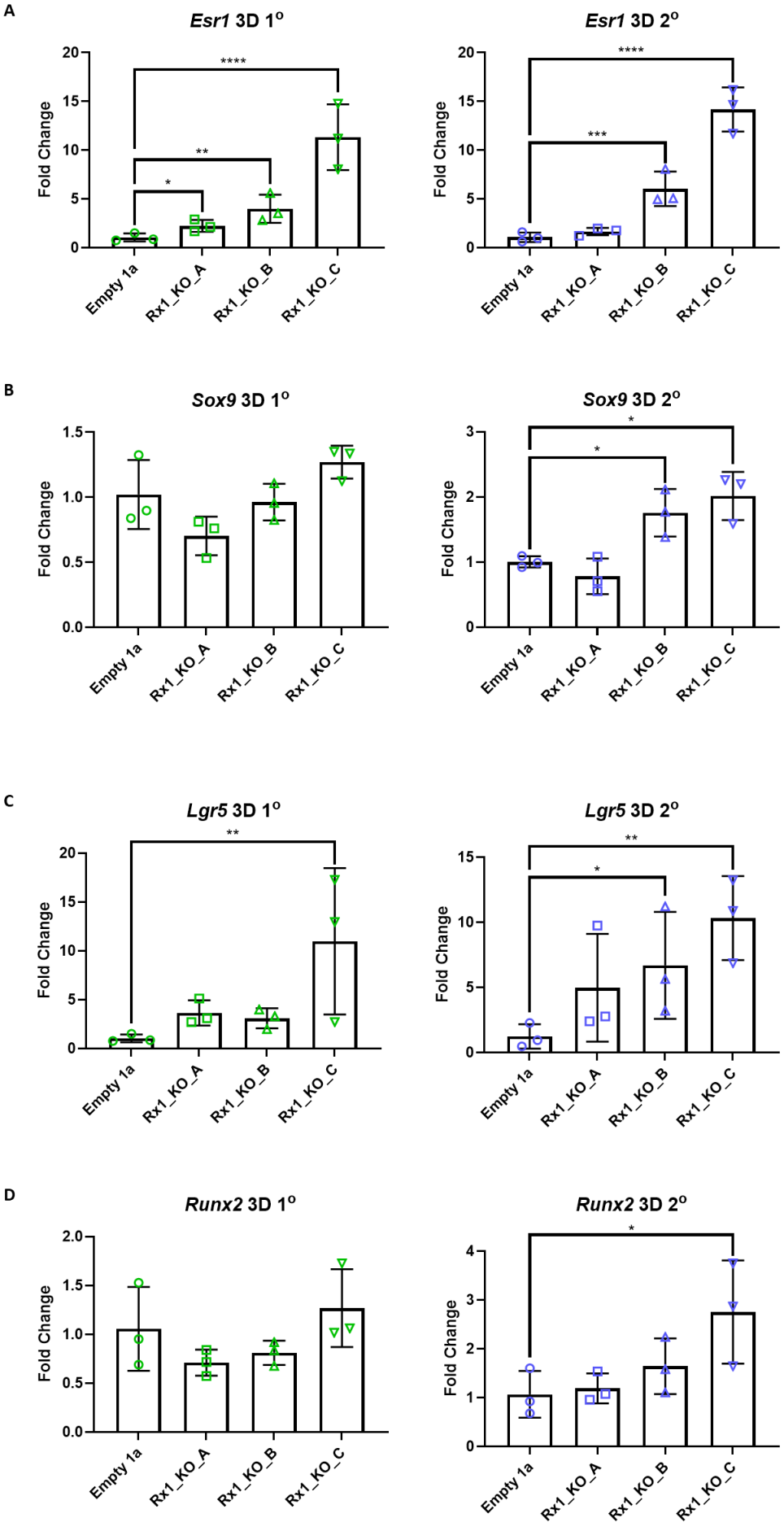

A
Esr1 3D 1°
Esr1 3D 2°
B
Sox9 3D 1°
Sox9 3D 2°
C
Lgr5 3D 1°
Lgr5 3D 2°
D
Runx2 3D 1°
Runx2 3D 2°
Fold Change
Empty 1a
Rx1_KO_A
Rx1_KO_B
Rx1_KO_C

Figure 5.5 Expression of stem cell panel genes in mammospheres derived from *Runx1*-CRISPR clones.

Box and whisker plots summarising results from qPCR analysis of cell pellets qPCR analysis of 3D-cultured (primary/1° and secondary/2° mammospheres) HC11 cells, with and without *Runx1* deletion. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for each test gene. Normalised relative expression shown for Esr1 [A], Sox9 [B], Lgr5 [C], and Runx2 [D]. Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$].

5.2.4 RUNX1 limits the stemness promoting capability of WNT3A protein.

A strong stem-like phenotype was observed in *Runx1*-deficient mice within the context of oncogenic ß-catenin signalling (as described in Chapter 3). To model the combined effects of *Runx1* loss with aberrant Wnt signalling on the stemness of the mammary epithelium, vehicle- and WNT3A-treated HC11 *Runx1*_KO cell lines were analysed for their mammosphere-forming potential, relative to the control line. In support of observations made for Figure 5.1C and D, loss of RUNX1 expression increased mammosphere formation in the HC11 cell line both with and without WNT3A treatment (Figure 5.7). Comparisons between Vehicle and WNT3A treatment responses for each cell line revealed that all cell lines exhibited greatly enhanced potential to form these stem cell-related 3D structures with WNT3A treatment, compared with Vehicle control (stats not displayed in figure, $p < 0.0001$ in all cases). Further to this, fold changes reveal further enhancement of WNT3A-mediated promoted stemness within the context of loss of functional RUNX1. WNT3A induced a fold increase of 1.77 in Empty 1a cells, whilst fold increases of 1.97, 1.99, and 1.95 were observed in each of the respective *Runx1* CRISPR clones. This data supports previous findings in Chapter 4, where overexpression of RUNX1 attenuated WNT3A-induced stemness. Taken together, these observations solidify the theory that, at least within the context of aberrant activation of the Wnt/ß-catenin pathway, the tumour suppressive capabilities of *Runx1* in the mammary epithelium could be partially mediated through its regulation of stem cell populations. Loss of functional *Runx1* could potentially leave the mammary epithelium more vulnerable to oncogenic insult through the unregulated expansion of a stem-like population.

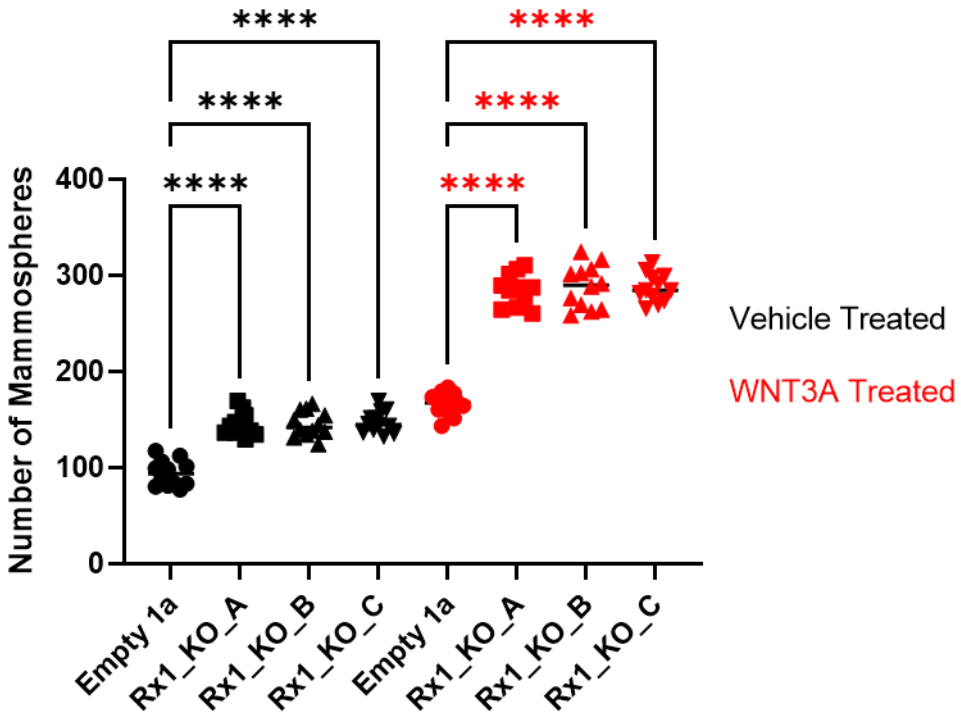


Figure 5.6 Possible additive or synergistic effect of RUNX1 deletion and WNT3A treatment on the progenitor-like potential of HC11 cell line.

Scatter plot summarising the quantification of mammospheres derived from Empty 1a cells and each Runx1 CRISPR cell line, treated with 100ng/ml WNT3A or vehicle (distilled water). Each data point represents the number of spheres counted in a single well following 7 days of culture. Vehicle-treated HC11 cells in black and WNT3A treated cells in red. n=12 wells counted per condition. Results shown represent 1 of 3 experimental repeats. One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons, with the most pertinent comparisons shown [**** $p < 0.0001$].

5.2.5 A potential role for RUNX1 in initiating stem cell quiescence in response to abnormal Wnt/ß-catenin pathway activation.

An MTS assay was used to investigate the effects of WNT3A treatment on HC11 cell lines, and explore further potential mechanisms for the increased production of mammospheres upon RUNX1 depletion. This is a sensitive and accurate colorimetric assay, which is used for the relative quantification of metabolically active cells within given populations, with a particular focus on cells' responses to growth factors and drug treatments. The MTS assay protocol is based on the conversion of the MTS tetrazolium compound into a coloured soluble formazan dye, whose presence can be quantified by measuring absorbance at ~490nm. The NAD(P)H-dependent cellular oxidoreductase enzyme, which is present in metabolically active cells, is responsible for this reduction of the MTS tetrazolium compound. The ability of cells to initiate this conversion is therefore used to

relatively assess their metabolic activity, which is often used as a surrogate for cell viability or proliferative capacity within cell populations.

Results from MTS assays, comparing metabolic activity of each *Runx1* CRISPR clone (and their Empty 1a control) in response to either Vehicle or WNT3A treatment, are summarised in Figure 5.7. Surprisingly, addition of recombinant WNT3A protein appeared to reduce the metabolic activity of HC11 cells relative to Vehicle controls. This decrease in the relative percentage of metabolically viable cells was more pronounced in the Empty 1a control clone, and appeared to be initiated earlier, compared with all 3 *Runx1* CRISPR cell lines.

This marked reduction in metabolic activity triggered by WNT3A, particularly within HC11 cells expressing normal levels of RUNX1, may be indicative of a protective mechanism of RUNX1. It could be possible that, in response to WNT3A treatment, RUNX1 may trigger cellular quiescence in an attempt to protect the HC11 mammary epithelial cells from transformation. If this theoretical RUNX1-dependent response to WNT3A was being triggered in cells with stem-like potential, this may provide further insight into the mechanisms behind RUNX1-mediated control of stem cell populations in the mammary epithelium in response to aberrant Wnt/ß-catenin signalling. This theory makes sense within the context of the WNT3A-treated mammosphere assays, in which higher RUNX1 expression levels were associated with dampened responses to the recombinant protein (Chapter 4 and Figure 5.7). As discussed previously, a limitation of this assay is its inability to detect quiescent stem cells. Therefore, if RUNX1 was initiating stem cell quiescence in response to abnormal Wnt/ß-catenin signalling, this would translate into limited formation of mammospheres, which was indeed observed, in both Chapter 4 and Figure 5.7, in the cells expressing higher levels of RUNX1 than their counterparts.

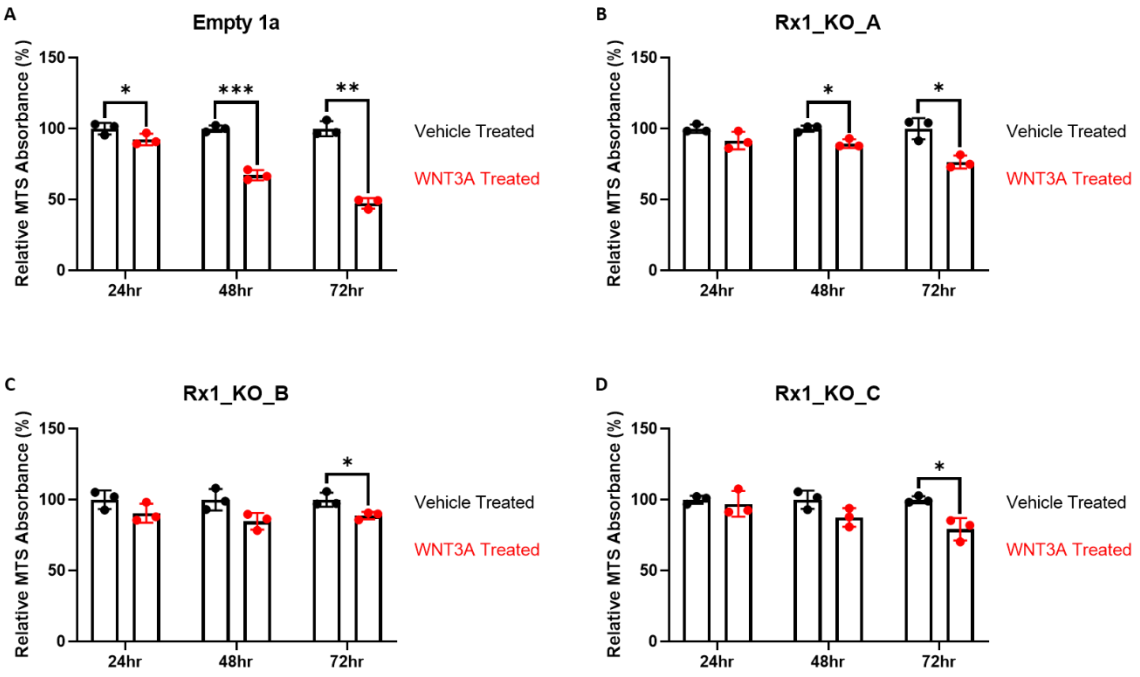


Figure 5.7 Loss of RUNX1 appears to protect HC11 cells from WNT3A-mediated attenuation of metabolic activity.

Box and whisker plots summarising results from MTS assays, exploring the effects of 100ng/ml WNT3A versus vehicle (distilled water) treatment on each HC11 cell line at 24hr, 48hr, and 72hr timepoints. Results shown for the Empty 1a clone [A] and each of the 3 *Runx1* CRISPR clones in [B] to [D]. For each cell line and timepoint, absorbance values were each normalised to the average absorbance value (from 3 technical repeats) of the vehicle control group (100%). Each data point represents the relative % metabolic viability in each technical repeat. Vehicle-treated HC11 cells in black and WNT3A treated cells in red. Paired t-tests were used to compare WNT3A treatment data with complementary Vehicle-treated control data at each timepoint within each cell line [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$].

5.2.6 Expression of putative Wnt targets elevated upon loss of functional RUNX1.

Results in Figures 5.6 and 5.7 indicate that loss of *Runx1* in the HC11 cell line results in an enhancement of Wnt-mediated enrichment of stemness, and possibly a reduction in stem cell quiescence in response to aberrant Wnt signaling. RNA-Seq analysis of mammary epithelial cells isolated from *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice revealed that, in addition to increased expression of several stem cell-related genes, various elements of the Wnt pathway were differentially expressed with the loss of normal *Runx1* expression. It is possible that *Runx1* may either directly or indirectly regulate various stemness-promoting elements of the canonical Wnt pathway, in addition to several non-Wnt stem-related gene targets, to protect the mammary gland from abnormal cell growth.

Analysing the expression of various canonical Wnt pathway target genes by qPCR revealed that the *Lgr5* gene, a Wnt pathway mediator and stem cell marker, was enhanced in mammospheres derived from of the Rx1_KO_C clone relative to the Empty 1a control (Figure 5.8). *Axin2* is also considered a classical Wnt target gene, which is expressed in Wnt-responsive stem cells of various organs, including within the mammary gland (Nusse and Clevers 2017). This gene was also validated as a relevant Wnt target in the breast, where its expression was significantly enhanced within MMECs extracted from pregnant *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice compared to *Blg-Cre;Catnb*$^{wt/wt}$ mice. *Axin2* is considered a reliable reporter of canonical Wnt pathway activation, as is clearly demonstrated in Figure 5.9 where its expression is greatly enhanced with the addition of WNT3A. It's role in the Wnt signalling pathway as a negative regulator, which engages in a negative feedback loop to moderate the duration and intensity of Wnt signalling, is also well established (Lustig, Jerchow et al. 2002, Bernkopf, Hadjihannas et al. 2015). The highly elevated expression of *Axin2* in response to WNT3A treatment may therefore explain why several other Wnt pathway potentiators (including *Lgr5*, Figure 5.8) did not display such heightened levels of expression with the addition of recombinant WNT3A. Significantly upregulated expression of *Axin2* was observed in at least 1 of the 3 *Runx1* CRISPR clones (Rx1_KO_C), grown in the 3D context, and a second clone (Rx1_KO_A) showed trends towards increased expression in the 3D WNT3A growth conditions (Figure 5.9).

Expression analyses of several other genes in the Wnt target panel— including representative genes *Rnf43*, *Znrf3,* and *Wif1* (Figure 5.10A-C) — revealed significant enhancements (to varying degrees) of several putative Wnt pathway targets in mammospheres derived from CRISPR clone C. However, this differential expression of Wnt pathway targets was inconsistent across experimental conditions and did not reach significance in either of the other *Runx1*-deficient clones (Figure 5.10). It is not feasible to determine the true significance of alterations that were exhibited in a minority of clones, as assumptions cannot necessarily be made as to whether these changes directly reflect the impact of RUNX1 alteration, or whether they might be coincidental mutations that occurred during the expansion of that single clone. The expression of Wnt/β-catenin target genes *Sox9* and *Myc* did not appear to be significantly altered in any of the cell

lines or conditions, although 3D-cultured Rx1_KO_C HC11 cells did appear to be at show at least slight elevations in expression levels of both genes.

Gene expression analyses of various putative Wnt targets demonstrated that some Wnt pathway genes were significantly, or generally trending towards, increased levels in Rx1_KO_C clone-derived mammospheres, and in some conditions displayed slightly (though not significant) increased levels in the other two clones. While the Rx1_KO_C clone presented significantly increased levels of several Wnt pathway targets, including *Lgr5* and *Axin2*, any trends towards alterations observed in the other two clones did not reach significance. While these results on their own cannot be used to conclusively determine a link between *Runx1* expression and Wnt pathway activation, due to the clone to clone variability, these results can be considered in combination with those from Runx1P1-overexpression analyses in Chapter 4. Together, these indicate that *Runx1* may directly or indirectly be involved in regulating the canonical Wnt pathway, including through the stemness-promoting components (such as *Lgr5*), to protect the mammary epithelium from hyperplasia that can be triggered by abnormal Wnt signalling.

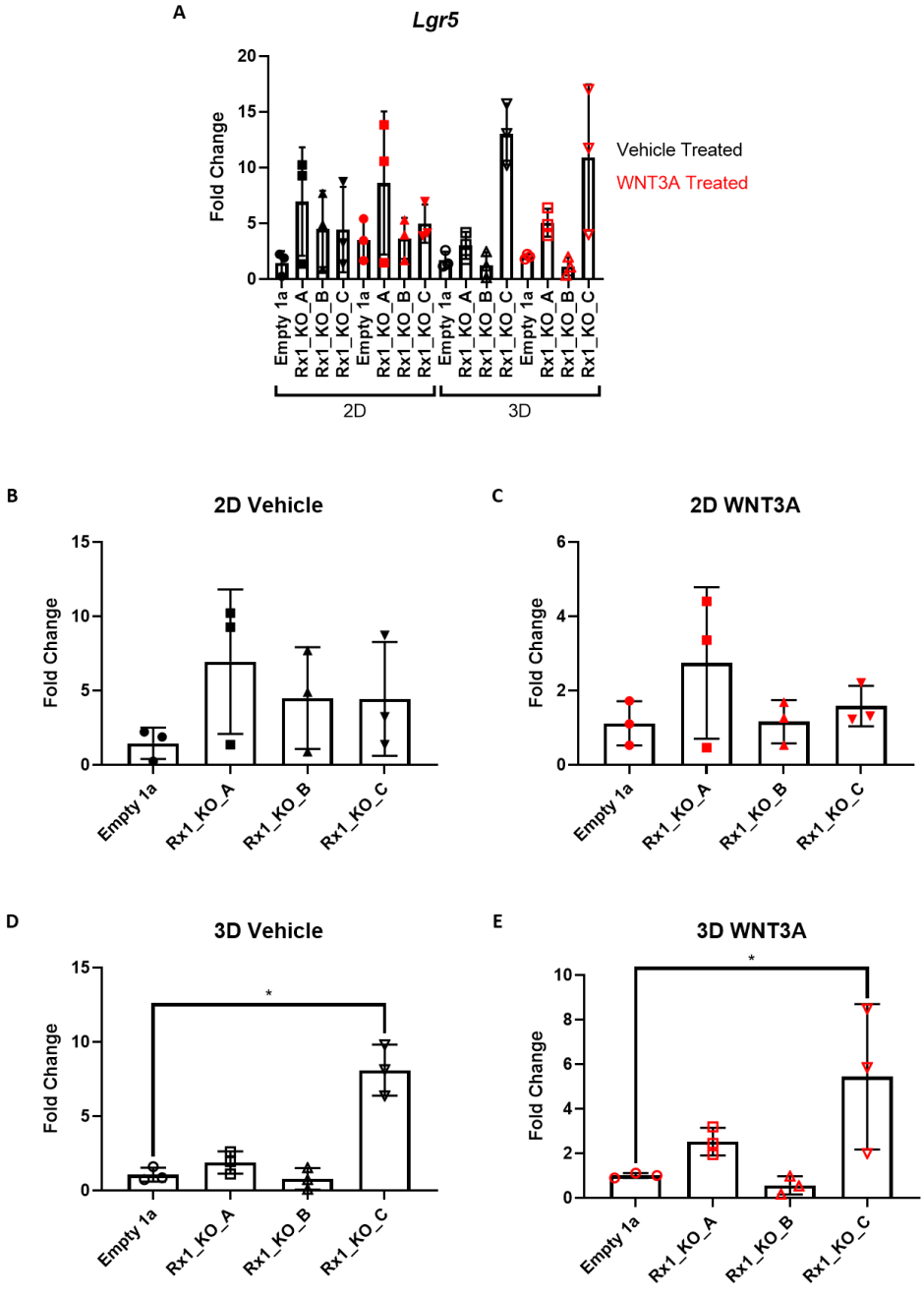


Figure 5.8 *Lgr5* gene expression is upregulated in HC11 Rx1_KO_C growth in 3D conditions.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty vector and Runx1-deleted HC11 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Lgr5*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons [* $p < 0.05$].

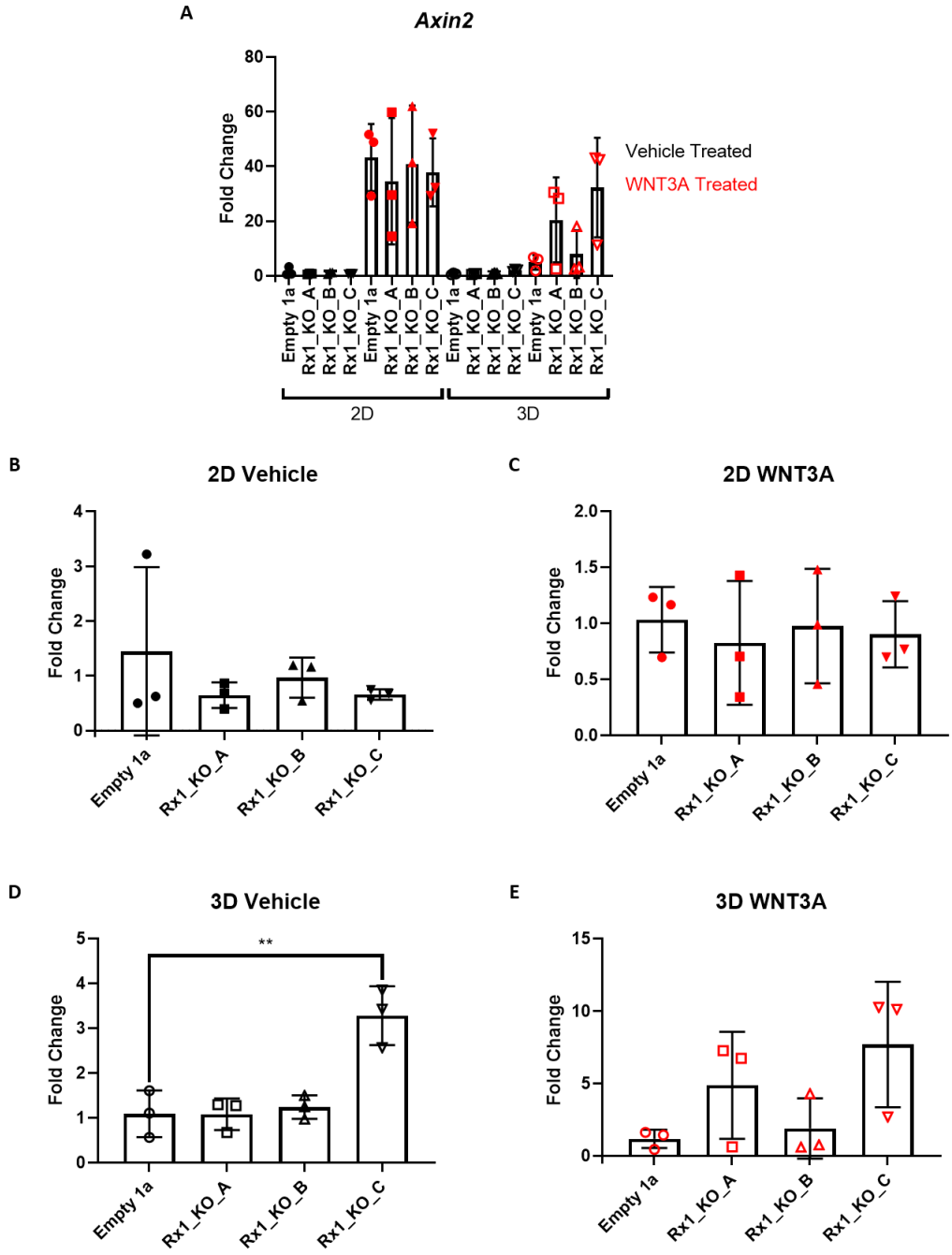


Figure 5.9 Canonical Wnt pathway target, *Axin2*, is altered in vehicle-treated mammospheres generated by the Rx1_KO_C clone.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty vector and Runx1-deleted HC11 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Axin2*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons [** $p < 0.01$].

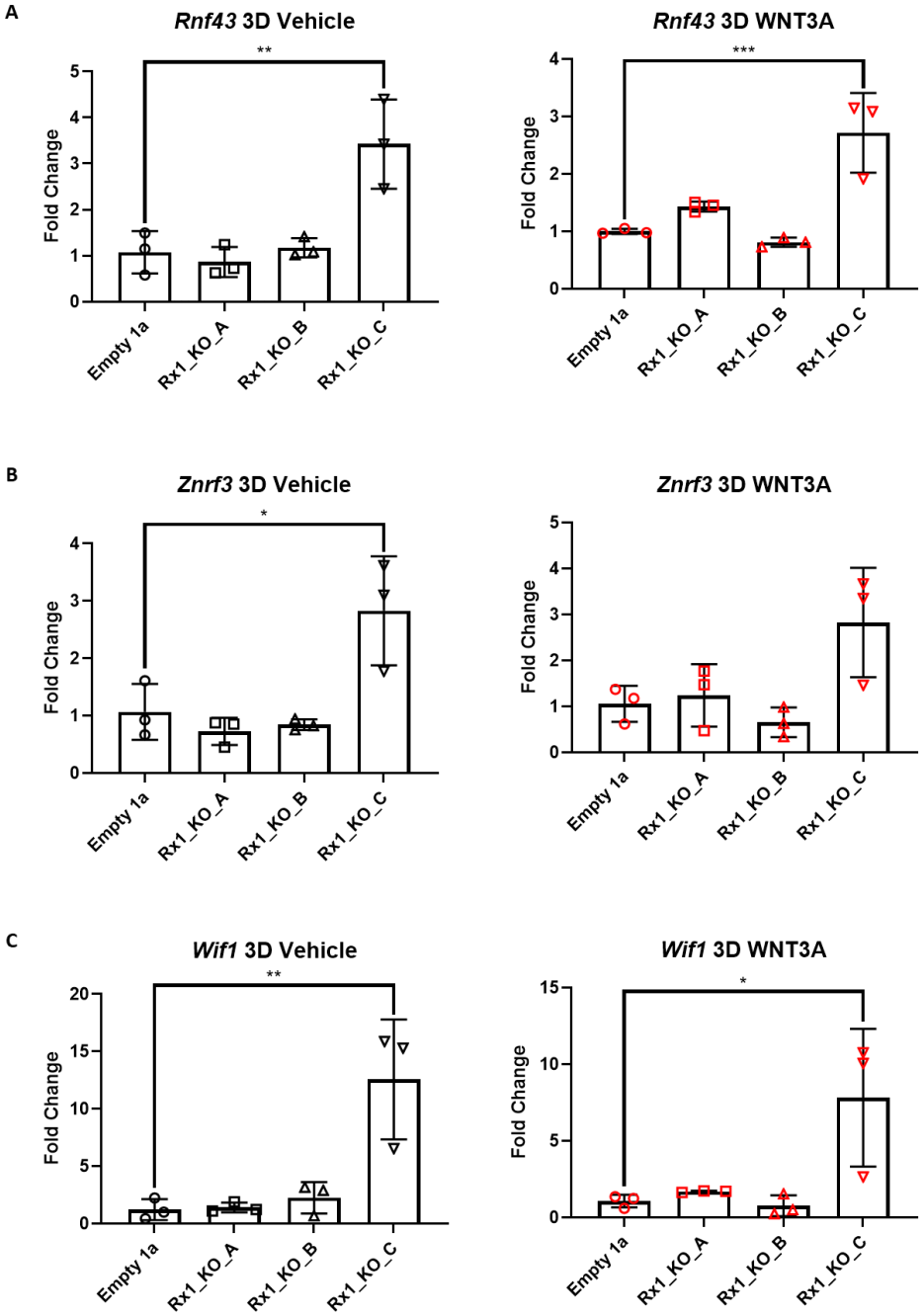


Figure 5.10 Expression levels of several Wnt pathway targets and mediators in *Runx1* CRISPR clone derived mammospheres.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty vector and Runx1-deleted HC11 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for each test gene: [A] *Rnf43*; [B] *Znrf3*; [C] *Wif1*. Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$].

5.2.7 Opposing roles for RUNX1 and RUNX2 in the regulation of mammary epithelial stem cells.

In vivo models of ß-catenin-mediated, mammary-specific oncogenesis exhibited significantly accelerated tumour onset with loss of *Runx1* in the mammary epithelium, in addition to increased incidences of independently arising multifocal and milticentric tumours, compared to wildtype controls (Chapter 3). This phenotype was dramatically enhanced with the additional loss of *Runx2*, despite *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx2*$^{fl/fl}$ mice having a phenotype that was indistinguishable from their wildtype counterparts. This was accompanied with more pronounced elevations in various stem cell-related and Wnt target genes in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mammary glands, compared to those with loss of *Runx1* alone (Chapter 3). To explore this fascinating, and apparently contradictory within the context of the role of *Runx2*, phenotype further the HC11 cell line was employed as a manipulatable system in which simultaneous loss of *Runx1* and *Runx2* in the mammary epithelium, and the consequences thereof, could be modelled and explored.

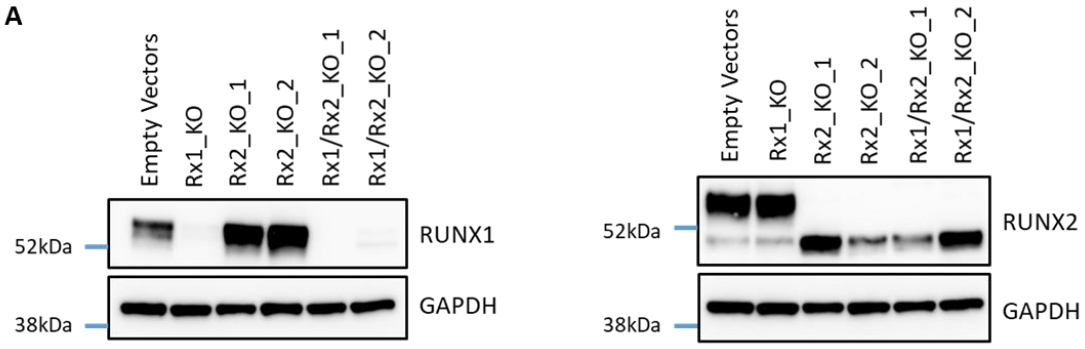
Deletion of each gene was achieved by sequential CRISPR Cas9-mediated gene alterations, with the *Runx2* gene being targeted first using the pX-459-Puro plasmid that was modified to contain a *Runx2*-targeting gRNA sequence. HC11 parental cells were transfected using the resulting construct, or an empty vector to produce a control line. Following antibiotic selection and single cell cloning, successful RUNX2 knockout was confirmed in 2 clones by Western Blot analysis and subsequent DNA sequencing (Appendices 11 and 12). DNA sequencing of these clones was necessary due to the presence of a second protein band, which was retained in the 2 successful CRISPR clones. The deletion of 30 and 39 bases in each of the respective clones, near the beginning of the RUNX2 coding region, would produce a truncated and likely non-functional RUNX2 variant, which would not be detectable by the RUNX2 antibody even if it were able to escape degradation. Given this information, and that the second band detected is of a lower molecular weight than predicted (55-62kDa), it can be safely assumed that this is a non-specific band that does not reflect RUNX2 expression. Both RUNX2-depleted clones, and an empty vector clone, were then transduced with lentiviruses, produced from lentiCRISPR-Neo vectors that were either empty (as a control) or

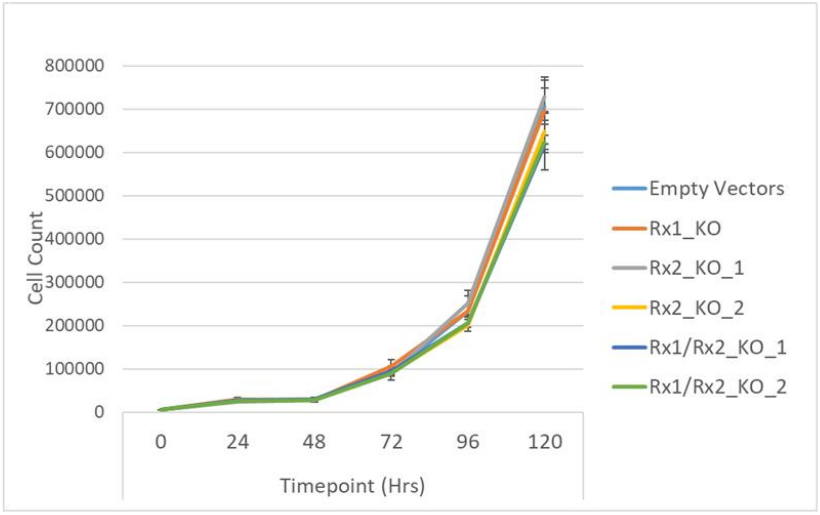
modified to hold the *Runx1* targeting gRNA sequence (which induced efficient depletion of RUNX1 in Figure 5.1). Consecutive CRISPR Cas9-mediated alterations in the *Runx1* and *Runx2* genes resulted in: 1 control clone (Empty Vectors); 1 RUNX1 single knockout clone (Rx1_KO), 2 RUNX2 single knockout clones (Rx2_KO_1 and Rx2_KO_2); and 2 RUNX1/RUNX2 double knockout clones (Rx1/Rx2_KO_1 and Rx1/Rx2_KO_2). Western Blot analyses of RUNX1 and RUNX2 protein expression, summarised in Figure 5.11A, demonstrates successful knockout in each clone.

RUNX1- and/or RUNX2-deficient HC11 cell lines were set up in 2D growth experiments, during which cell counts were taken every 24 hours using a CASY Cell Counter. The results of these daily cell counts were summarised in growth curves (Figure 5.11B), which do not reveal any statistically significant changes in 2D growth with loss of functional RUNX1 and/or RUNX2, compared to control cells. This supports data from 2D growth experiments examining HC11 cell growth in response to RUNX1 overexpression (Chapter 4, Figure 4.2B) or deletion (Figure 5.1B), or in response to RUNX2 overexpression (Ferrari, Riggio et al. 2015).

The impacts of these genetic alterations on the mammosphere forming efficiency, and thereby the stem-like behaviours, of HC11 mammary epithelial cells were evaluated. There was a statistically significant increase in the number of mammospheres produced by the RUNX1-deficient HC11 clone compared with the control line (Figure 5.11C). This is in agreement with findings from all 3 independent single CRISPR clones in Figure 5.1D and indicates a role for RUNX1 in the regulation of the progenitor-like activity of these cells. Contrastingly, loss of RUNX2 resulted in a reduced capacity to form these 3D structures in both independent single CRISPR clones, which supports evidence from a previous Blyth lab publication revealing a role for RUNX2 in mammary stem/progenitor cell function (Ferrari, Riggio et al. 2015). Interestingly, these opposing roles for RUNX1 and RUNX2 did not cancel each other out when observing the effects of simultaneous loss of these proteins on mammosphere forming potential as there was a significant increase in the number of mammospheres produced by both *Runx1*/*Runx2* double CRISPR clones. However, they were not able to fully overcome or compensate for the diminished progenitor-like activity caused by loss

of RUNX2 as they each still produced significantly less mammospheres than Rx1_KO cells (statistics not displayed on graph: $p < 0.001$ and $P < 0.0001$, respectively). This skew towards increased stemness, even with loss of RUNX2, indicates that RUNX1 could be playing a more significant role in the regulation of mammary stemness than RUNX2. It appears that the stem cell limiting functions of RUNX1 cannot be as effectively replaced, while progenitor-like cells can at least partially overcome their apparent reliance on RUNX2 expression. This indicates that compensatory mechanisms may be triggered to replace some of the functions of RUNX2, but only within the context of simultaneous loss of the RUNX1 and RUNX2 proteins.


A
Empty Vectors
Rx1_KO
Rx2_KO_1
Rx2_KO_2
Rx1/Rx2_KO_1
Rx1/Rx2_KO_2
52kDa
RUNX1
38kDa
GAPDH
Empty Vectors
Rx1_KO
Rx2_KO_1
Rx2_KO_2
Rx1/Rx2_KO_1
Rx1/Rx2_KO_2
52kDa
RUNX2
38kDa
GAPDH


B
800000
700000
600000
500000
400000
300000
200000
100000
0
Cell Count
0
24
48
72
96
120
Timepoint (Hrs)
Empty Vectors
Rx1_KO
Rx2_KO_1
Rx2_KO_2
Rx1/Rx2_KO_1
Rx1/Rx2_KO_2

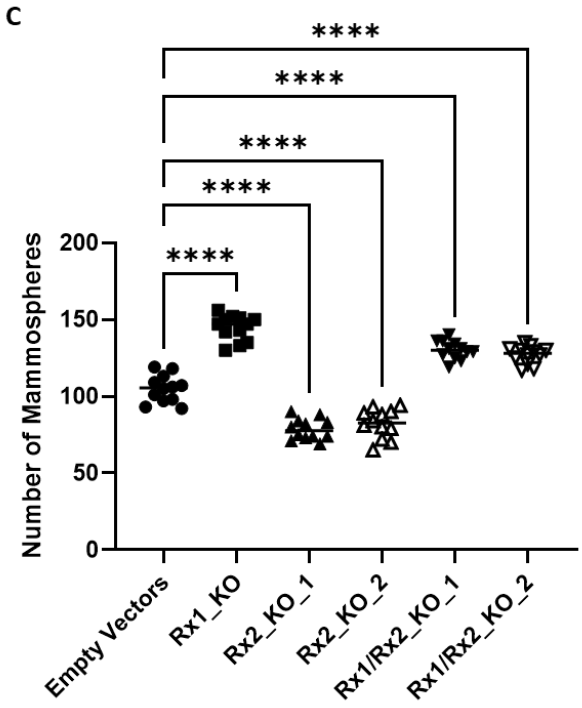

C
****
****
****
****
****
200
150
100
50
0
Number of Mammospheres
Empty Vectors
Rx1_KO
Rx2_KO_1
Rx2_KO_2
Rx1/Rx2_KO_1
Rx1/Rx2_KO_2

Figure 5.11 *Runx1* and *Runx2* have opposing functions in the stem-like potential of HC11 cells.

[A] Western Blot of single cell clones obtained from HC11 cells, first transfected with pX-459-Puro vector (Empty or containing *Runx2*-targeting sequence) and then transduced with lentivirus from lentiCRISPRv2-Neo vector (Empty or containing *Runx1*-targeting sequence). Successfully generated 1 control clone (Empty Vectors), 1 *Runx1* CRISPR clone (Rx1_KO), 2 *Runx2* CRISPR clones (Rx2_KO_1 and Rx2_KO_2), and 2 *Runx1*/*Runx2* double CRISPR clones (Rx1/Rx2_KO_1 and Rx1/Rx2_KO_2). GAPDH as a loading control and Amersham Full Range Rainbow Marker as a protein ladder, with key marker positions and molecular weights shown.

[B] 2D growth curve of HC11 Empty Vectors clone and clones with depletion of *Runx1* and/or *Runx2* expression. n=4 counts per cell line per day, plotted as averages (and population standard deviations). One-way ANOVA, with Dunnett's multiple comparisons test, performed for each timepoint. Results shown represent 1 of 2 experimental repeats.

[C] Primary mammosphere counts for HC11 *Runx1* and/or *Runx2* clones, summarised in a scatter plot. n=12 wells were counted per cell line after 7 days of growth. Each data point represents the number of spheres counted in a single well. Results shown represent 1 of 3 experimental repeats. One-way ANOVA, with Tukey's multiple comparisons test [**** $p < 0.0001$].

5.2.8 Loss of RUNX1 influences mammosphere size more significantly than RUNX2 depletion.

Shown in Figure 5.12A are representative images of mammospheres generated from each of the single and double RUNX1/RUNX2 CRISPR clones. Similar to Rx1_KO clones in Figure 5.2A, RUNX1 depletion in the HC11 cell line lead to the production of smaller mammospheres, as did the simultaneous disruption of RUNX1 and RUNX2 function. However, loss of RUNX2 alone did not appear to have as significant an impact on mammosphere size.

ImageJ was used to quantify individual mammosphere areas (μM^2) from the captured images, with results summarised in Figure 5.12B. Quantitative analysis confirmed this apparent reduction in cell size with the loss of RUNX1, both within the context of its singular knockout and with the additional loss of RUNX2. Mammospheres obtained from Rx2_KO cell lines, on the other hand, did not show consistent results between clones. Clone 1-derived mammospheres were statistically significantly smaller than those generated from control cells, however their size was not reduced to the same extent as those from Rx1_KO cells. Comparisons between clone 2 and Empty Vectors, however, did not show any statistically significant changes in size. Due to this inconsistency between clones, it cannot be determined which result is more representative of the true consequences of RUNX2 loss on mammosphere size, although it is clear that the loss of RUNX2 at least does not have as much of an impact as the loss of RUNX1. To probe further into this mammosphere size data, it was divided into quartiles to look at the spread of the values. Regardless of the cell line, the vast majority of mammospheres fell within the lowest quartile due to the large data range, however a larger and more visible percentage of mammospheres originating from the control line fell within quartiles 2, 3 and 4. While *Runx1* CRISPR cells produced mammospheres almost exclusively within the lowest quartile, RUNX2-deficient cells contrastingly produced spreads of data more similar to the control, with a significant percentage of mammospheres falling within quartile 2. *Runx1*/*Runx2* double CRISPR clones both produced mammospheres whose sizes were intermediately spread in relation to those originating from clones with single protein deletions of RUNX1 and RUNX2.

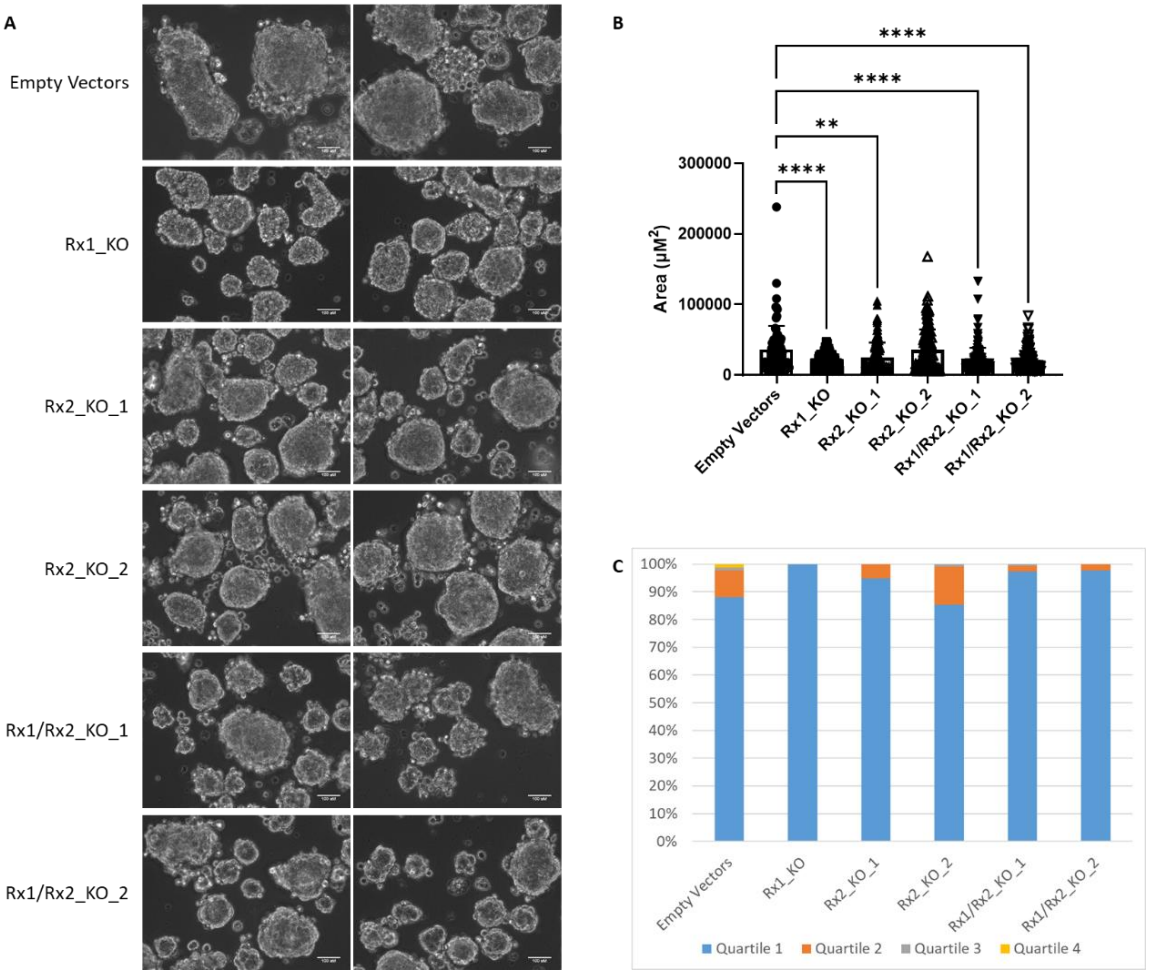


Figure 5.12 Loss of *Runx1* appears to have the most significant impact on mammosphere size.

[A] Phase contrast images of control and CRISPR clone-derived mammospheres. Images are representative of ≥18 images taken for each cell line in this particular experimental setup. Scale bar=100μM. 1mm stage micrometer used to calibrate the scale bar.

[B] Scatter plots comparing mammosphere sizes (μM^2) of each cell line, quantified using ImageJ software. Each data point represents the size (μM^2) measured for a single mammosphere. One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons [** $p < 0.01$; **** $p < 0.0001$].

[C] Mammosphere sizes obtained in part [B] were divided into quartiles, and the proportions of mammospheres (%) within each of the size quartiles were plotted as bar charts for each cell line.

All results shown are representative of 1 of 2 experimental repeats, each with n≥18 images taken per cell line.

5.2.9 Simultaneous loss of functional RUNX1 and RUNX2 dramatically enhances Wnt-dependent stemness.

While the loss of RUNX1 and RUNX2 together did still result in an increase in stemness compared to control cells (Figure 5.11C), these results did not completely recapitulate the very dramatic stem-like phenotype observed *in vivo*. While loss of *Runx1* alone in this model resulted in earlier emergence and increased incidence of independently arising tumours, accompanied by elevated expression of various stem cell-related genes in the mammary epithelium, this was significantly remarkably enhanced with the additional loss of *Runx2* (Chapter 3). Contrastingly, in HC11 cells mammosphere formation of double CRISPR clones was still restricted compared to Rx1_KO cells (Figure 5.11C), suggesting that the dampened stemness in response to RUNX2 loss could not be fully overcome. It is possible that the highly elevated stem-related phenotype in response to simultaneous loss of these genes in the mouse model was Wnt-dependent.

To model this combination of oncogenic Wnt activation with loss of both *Runx1* and *Runx2*, mammosphere formation was used to evaluate the stem-like potential of all CRISPR cell lines in the presence and absence of recombinant WNT3A protein. Results of these mammosphere assays are summarised in Figure 5.13, with data from vehicle-treated CRISPR cell lines supporting those in Figure 5.11C, which shows that although there is an enhancement of stem-like potential of HC11 cells with loss of both RUNX1 and RUNX2 together, it is significantly reduced when compared with RUNX1-depletion alone. Interestingly, however, it appears that this is not the case for WNT3A-treated cells, as the additional loss of RUNX2 in this context appears to enhance the stemness even further than RUNX1 loss alone. This is confirmed by looking at fold changes in mammosphere counts, within each cell line, in response to exogenous WNT3A expression. Addition of recombinant WNT3A induced a fold increase of 1.25 in Empty Vectors cells, which was increased to 1.31 in Rx1_KO cells, and further enhanced to 1.42 and 1.39 with the added loss of RUNX2 in each respective clone.

This data further supports the theory that the tumour-suppressive functions of *Runx1* may be partially mediated through regulating the stemness of the mammary

epithelium, particularly within the context of stemness-enhancing oncogenes such as β-catenin. It also offers some insight into the results obtained from mice with loss of functional *Runx1* and *Runx2*. Rx1/Rx2_KO cells could not fully overcome the restricted stem-like capacity, triggered by loss of RUNX2, in the absence of any oncogenic signal. Contrastingly, it appears that, within the context of aberrant activation of the Wnt signalling pathway, the stemness-promoting functions of RUNX2 can be compensated for, or in this case overcompensated, when RUNX1 is also lost. Interactions between RUNX2 and the Wnt pathway are well documented in various cell types, with β-catenin and RUNX2 capable of regulating one another and collaborating in the regulation of common downstream targets. In the absence of regulation by RUNX1, the aberrantly activated Wnt pathway may be able to compensate for RUNX2 loss by subverting the pathway towards alternative downstream targets, which may be able to replace or even enhance some of the activities performed by RUNX2 in the stem cell compartment.

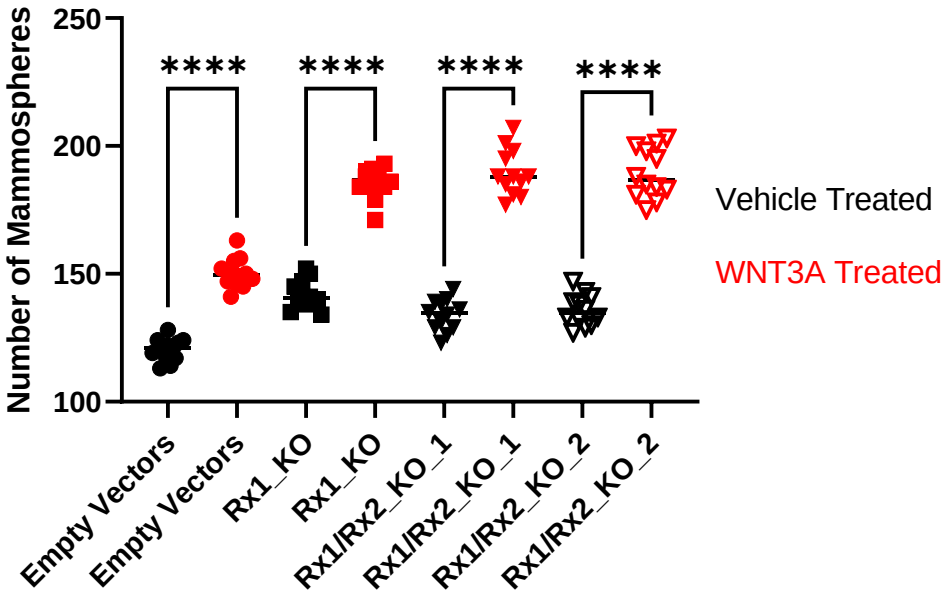


Figure 5.13 Enhancement of stem-like potential in WNT3A-treated HC11 cells with loss of *Runx1* and *Runx2*.

Scatter plot summarising the quantification of mammospheres derived from HC11 CRISPR and Empty clones, treated with 100ng/ml WNT3A or vehicle (distilled water). Each data point represents the number of spheres counted in a single well following 7 days of culture. Vehicle-treated HC11 cells in black and WNT3A treated cells in red. n=12 wells counted per condition. Results shown represent 1 of 2 experimental repeats. One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons [**** $p < 0.0001$].

5.2.10 Metabolic response to WNT3A treatment appears to be RUNX1-dependent.

The effects of exogenous WNT3A expression on each HC11 clonal cell line was explored further using an MTS assay. MTS data from 3 independent Rx1_KO clones hinted that one possible protective mechanism of RUNX1 against oncogenic insult (from an aberrantly activated Wnt pathway for example) could be through the initiation of stem cell quiescence, as there was a RUNX1-dependent reduction in metabolic activity in response to WNT3A addition (Figure 5.7). Given that the enhancement of progenitor-like activity in response to WNT3A is more pronounced in clones with the simultaneous loss of RUNX1 and RUNX2, it was important to also investigate potential changes in metabolic activity within this context.

Summarised in Figure 5.14 are MTS results comparing metabolic activity of each CRISPR clone in response to either Vehicle or WNT3A treatment. Results from Empty Vectors control cells and Rx1_KO cells support those from Figure 5.7, as the significant reduction in metabolic activity in the control line (in response to recombinant WNT3A) is not apparent in cells whose RUNX1 function is lost. This change in metabolic response with loss of RUNX1 was also observed with the additional loss of RUNX2. In one Rx1/Rx2_KO clone, there was even an enhancement of metabolic activity following 48hrs of WNT3A treatment, though this was not observed in other timepoints or in the other Rx1/Rx2_KO line. With loss of RUNX2 alone, however, the suppression of metabolic activity remains, indicating the reduction in metabolic rate is RUNX1-dependent.

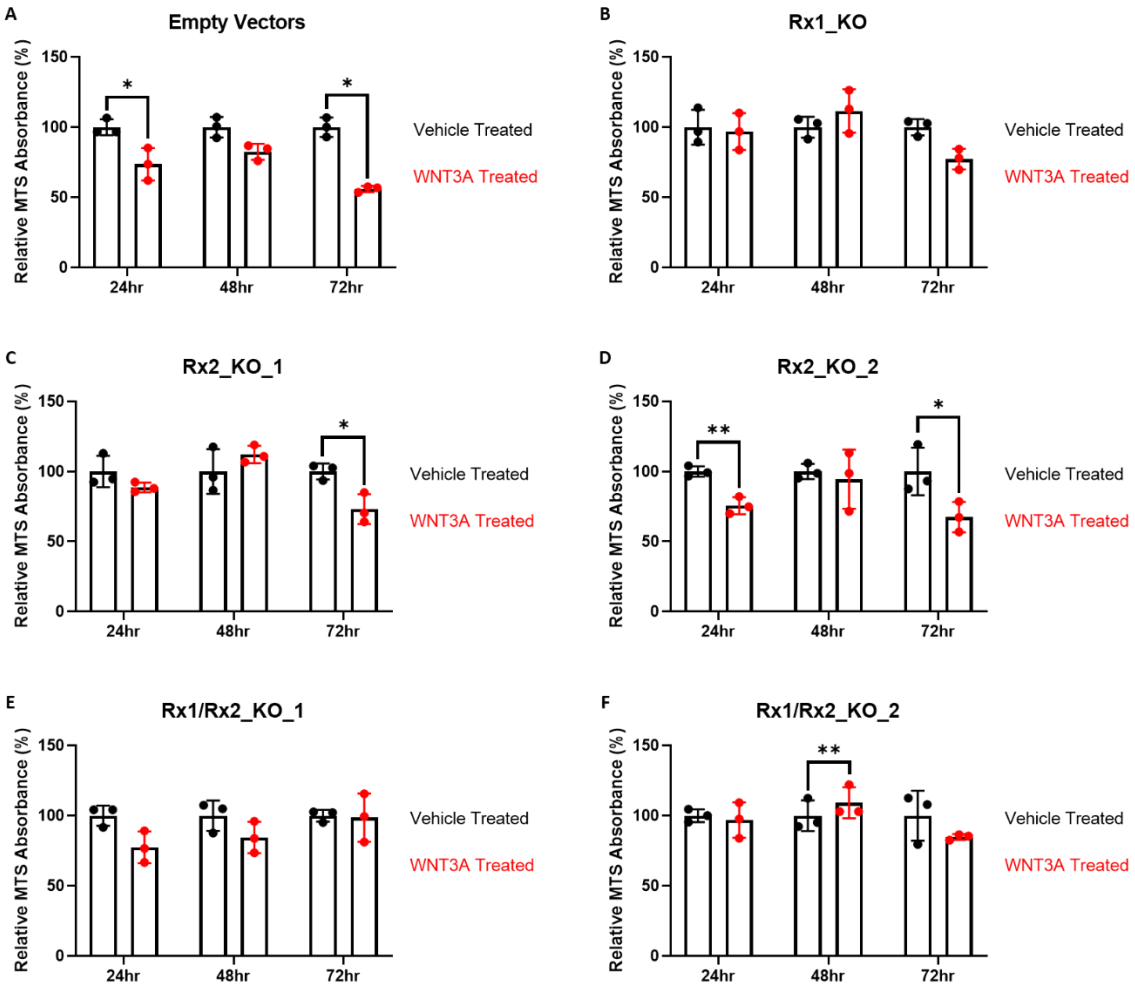


Figure 5.14 RUNX1, and not RUNX2, appears to be involved in the attenuation of metabolic activity in response to WNT3A addition.

Box and whisker plots summarising results from MTS assays, relatively comparing the overall metabolic viability in HC11 cells in the presence of 100ng/ml WNT3A compared to vehicle (distilled water) at 24hr, 48hr, and 72hr timepoints. Results shown for the Empty Vectors clone [A], the *Runx1* CRISPR clone, both *Runx2* CRISPR clones [C] and [D], and both *Runx1*/*Runx2* CRISPR clones [E] and [F]. For each cell line and timepoint, absorbance values were each normalised to the average absorbance value (from 3 technical repeats) of the vehicle control group (100%). Each data point represents the relative % metabolic viability in each technical repeat. Vehicle-treated HC11 cells in black and WNT3A treated cells in red. Paired t-tests were used to compare WNT3A treatment data with complementary Vehicle-treated control data at each timepoint within each cell line [* $p < 0.05$; ** $p < 0.01$].

5.3 Discussion

To more faithfully model the types *Runx1* alterations that are most commonly found in clinical samples of breast cancers (loss of function mutations, which have been functionally implicated as drivers of mammary tumorigenesis), it was important to explore the consequences of losing normal RUNX1 function in the mammary epithelium. Mammary-specific loss of *Runx1* function in a Wnt-driven model of mammary tumorigenesis resulted in a phenotype with potential links to stemness. The HC11 mouse mammary epithelial cell line was employed as a model to investigate this potential regulatory role for RUNX1 in mammary stemness, due to their stem-like potential and because they are demonstrably amenable to genetic manipulation. To explore any changes in the stem-like potential of HC11 cells in response to loss of functional RUNX1, mammosphere assays were used as an *in vitro* tool. The HC11 cell line has been used in several other investigations involving mammosphere assays in order to explore the cells' stem-like properties and capacity to differentiate (Morrison and Cutler 2009, Thrasyvoulou, Vartholomatos et al. 2020). Most notably, it used by the Blyth lab to explore the role of RUNX2 in the stem/progenitor potential of the mammary epithelium (Ferrari, Riggio et al. 2015).

Analyses of 3 independent single cell clones, in which successful knockout of RUNX1 was achieved using a CRISPR Cas9 gene editing tool, revealed significant enhancement of 3D colony and mammosphere forming potentials compared to control cells, whilst not significantly impacting their 2D growth. Additionally, self-renewal of mammospheres was demonstrated to be significantly enhanced with RUNX1 depletion, as indicated by the increased frequency of secondary spheres after passaging the primary 3D structures. Mammospheres formed by RUNX1-deficient cell lines were significantly smaller than those originating from the Empty Vector (Empty 1a) clone, which may be linked to the decrease in average percentage of actively proliferating cells. This further supports the theory from Chapter 4 that RUNX1 may be initiating differentiation in these cells in order to regulate their stem-like activity. Molecular characterisation of these clones by qPCR revealed relative enhancements of several validated stem cell related genes, particularly *Aldh1a1* and *Aldh1a7* that were notably also elevated in BCAT_1 and BCAT_12 mice. Not all tested genes showed consistent alterations in

each of the tested conditions or each of the 3 clones, with clone C often showing exaggerated enhancements of various stem markers. For genes whose alterations only reached significance in this single clone, it was not possible to assume that the observed alterations were meaningful or representative within the context of RUNX1 depletion. Nevertheless, this data infers that RUNX1 could be acting as a regulator of the stem-like potential of mammary epithelial cells through direct/indirect control of various stem/progenitor cell-related genes. Losing RUNX1 expression may release these stemness-promoting genes from tight regulation and result in uncontrolled expansion of the stem cell compartment, potentially leaving the mammary gland more vulnerable to oncogenic transformation.

The consequences of losing functional RUNX1 was also explored in the context of oncogenic Wnt pathway activation, as this is a stemness-promoting pathway whose aberrant activation is frequently associated with breast cancers (Xu, Zhang et al. 2020). It was also within the context of oncogenic Wnt/ß-catenin that the tumour suppressive function of RUNX1 was explored *in vivo* (Chapter 3), and so it was important to also model this *in vivo* to allow for analysis into the potential molecular mechanisms behind the striking phenotypes observed. The stemness-promoting functions of recombinant WNT3A were enhanced with loss of functional RUNX1, indicating that the tumour suppressive functions of this protein may be at least partially mediated through its control of the stem cell compartment of the mammary epithelium. One possible way in which RUNX1 may be controlling the stem cell compartment, to protect the mammary epithelium from Wnt-induced transformation, was revealed in MTS assays. There was a significant reduction in the metabolic activity of control HC11 cells treated with WNT3A, compared to those treated with Vehicle, however the metabolic rate was maintained in cells lacking functional RUNX1. This finding hints that, in order to protect the mammary gland from oncogenic insult (in this case aberrant activation of the canonical Wnt pathway), RUNX1 could potentially trigger cellular quiescence in mammary epithelial populations, such as the stem cell compartments. This may also partly explain the reduction in mammosphere forming potential as this assay cannot be used quantify quiescent stem cells. Expression of a panel of established Wnt targets was analysed by qPCR to explore potential molecular mechanisms for the

significantly enhanced WNT3A-dependent increase in stemness. As with stem cell markers, the results from clone C were often the most pronounced or exaggerated out of all 3 clones. Some Wnt targets, including the stem-related *Lgr5*, were shown to be significantly upregulated (or at least displaying trends towards upregulation) in RUNX1-depleted HC11 cells compared to their control. This indicates that RUNX1 could directly or indirectly be involved in the regulation of several components of the Wnt/ß-catenin pathway, and supports previous findings that RUNX1 is involved in the stabilisation of ß-catenin (Chimge, Little et al. 2016). Loss of RUNX1 in the mammary epithelium may lead to Wnt-dependent and –independent expansion of a subset of progenitor-like cells, which could leave the gland more vulnerable to abnormal growth and the development of tumours.

Depleting RUNX1 expression in the HC11 mammary epithelial cell line offered useful insights into the potential functions of this protein in the normal function of the mammary gland, as well as presenting some possible mechanisms behind its role as a tumour suppressor (particularly within the context of protecting the gland from Wnt-driven tumorigenesis). It was also important to investigate the intense stem-related phenotype observed in mice with mammary-specific loss of both *Runx1* and *Runx2*. This phenotype was unanticipated, especially given results from *Runx2* depletion alone (both in the Wnt-driven GEMM model in Chapter 3 and in a PyMT-driven transplantation model in a publication from Owens *et al* (Owens, Rogers et al. 2014)) in addition to the current knowledge of *Runx2*'s role in stem cell function and Wnt pathway mediation in the mammary epithelium. Successful depletion of RUNX1 and/or RUNX2 in the HC11 cell line was achieved by consecutive knockout experiments using CRISPR Cas9-based gene editing.

2D growth did not appear to be notably affected by either individual or simultaneous alteration of these proteins, however primary mammosphere formation was significantly impacted in each CRISPR clone. Results from the Rx1_KO clone are in agreement with those obtained from the 3 independent clones generated for single RUNX1 knockout experiments, the consequences of which were increased formation of mammospheres and a reduction in the average size of these 3D structures. RUNX2 loss of function, on the other hand, gave rise to a

reduced mammary forming potential in the HC11 cell line and there was a reduced (or negligible, depending upon the specific clone) impact on the average sphere size. Interestingly, while depletion of both proteins still lead to a statistically significant increase in the number of spheres, the magnitude of this change was not as large as with RUNX1 loss alone as both clones produced significantly less mammospheres than Rx1_KO cells.

Although the requirement for RUNX2 in the stemness of the mammary epithelium is at least partially overcome with the loss of RUNX1, the fact that the dramatic stem-associated phenotype seen in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mice was not fully recapitulated in the non-tumorigenic context indicates it may have been partly Wnt-dependent. To explore this theory, mammosphere forming potential was compared with and without recombinant WNT3A treatment for each CRISPR clone. The results presented from this assay more closely resemble those obtained *in vivo*, with WNT3A significantly enhancing the stem-like potential of Rx1/Rx2_KO cells to levels greater than that induced in Rx1_KO cells. This indicates that, in the absence of Wnt regulation by RUNX1, the stem cell promoting functions of RUNX2 can be replaced (potentially enhanced even) through the activation of alternative Wnt pathway mediators. An MTS assay was used to compare metabolic responses to WNT3A in each of the HC11 clones, and indicated that the reduction in metabolic activity (and the potential triggering of cellular quiescence that this result infers) in response to aberrant Wnt signalling is RUNX1-dependent. The control metabolic response to WNT3A was abrogated with the loss of RUNX1, in either single or double knockout clones, while it was maintained with RUNX2 loss alone. This indicates that RUNX1, and not RUNX2, is responsible for initiating this possibly protective mechanism in response to abnormal Wnt signalling.

Several methodological caveats, which were introduced in Chapter 4 Discussion, were addressed during investigations into the independent Rx1_KO clones. The generation of multiple clones for each genetic alteration to be tested proved to be valuable, as different clones sometimes produced results that were inconsistent with or unsupportive of one another. Had these observations been

made based on a single clone, certain results could have been under-representative or highly exaggerated, depending upon the specific clone selected. The availability of multiple independent clones provided an internal validation for the results, to allow them to be interpreted in a way that is more representative of the actual impacts of RUNX1 alteration. It was also noted that primary mammospheres alone are not sufficient to assess the self-renewal capacity of a population of cells, which is a key feature used to assess the stem-cell potential of a given cell population. For this reason, established cultures of primary spheres were passaged as single cell suspensions in secondary mammosphere assays in order to compare the short-term renewal capacity of each HC11 clone.

Although steps were taken to minimise the limitations of the studies conducted, it was not possible to eliminate all potential caveats within this chapter. For example, in Chapter 4 it was discussed that uncertainty as to whether the mammosphere assay captures the true relative stemness of mammary epithelial populations could be addressed using limiting dilution transplantation assays into cleared mammary fat pads. This method is thought to be the gold standard for the accurate determination of mammary stem cell frequency within a given population (Eirew, Stingl et al. 2008, Visvader 2009, Illa-Bochaca, Fernandez-Gonzalez et al. 2010). Mammary epithelial stem cells have been demonstrated to possess the ability to reconstitute fully functional mammary ductal branches due to their multi-lineage potential. It is therefore possible to evaluate the stem/progenitor cell activity within a given population of cells by transplanting increasingly diluted single-cell preparations into a de-epithelialized pubertal gland and evaluating their mammary repopulating efficiency (Eirew, Stingl et al. 2008, Visvader 2009, Illa-Bochaca, Fernandez-Gonzalez et al. 2010). Attempts were made to conduct such assays by transplanting limiting dilutions of HC11 (control and Rx1_KO) cells into the cleared fat pads of BALB/c mice. Unfortunately, none of the transplanted cell populations was capable of reconstituting the mammary gland, and remained as compacted balls of cells rather than growing out into branches. It is possible that the transplanted progenitor-like cells are failing to differentiate into ductal structures due to a lack of hormones required for their differentiation, causing them to stay in a de-differentiated state that is incapable of forming such complex and specialised

structures. For future implantation studies involving the HC11 cell line, recipient mice may need to either be administered the hormones necessary for these cells to undergo lactogenic differentiation (dexamethasone, insulin, prolactin (Merlo, Graus-Porta et al. 1996)), or perhaps a more appropriate method could be for the mice to undergo at least 2 rounds of parity to naturally stimulate lactogenic hormone production.

Regardless of limitations, the results presented in this chapter indicate a significant role for RUNX1 in regulating the stem-like potential of the mammary epithelium, both in a non-tumorigenic and tumorigenic context. While RUNX2 deletion has previously been shown to delay tumour initiation/progression *in vivo* (Owens, Rogers et al. 2014) and is important for mammary stem cell function (Owens, Rogers et al. 2014, Ferrari, Riggio et al. 2015), the loss of functional RUNX1 appears to be more essential in, or contribute more towards, promoting the stem-like functions of both the normal and tumorigenic mammary gland. The consequences of losing RUNX1, and particularly the mechanisms by which the loss of its function could lead to tumorigenesis within the normal mammary gland, are key concepts covered in this chapter. The data shown supports previous publications that indicate RUNX1 is involved in the suppression of the breast cancer stem cell phenotype, and thereby limits tumour growth, migration, and invasion (Hong, Fritz et al. 2018, Fritz, Hong et al. 2020), and prevents the deregulation of ß-catenin in ER-positive breast cancers (Chimge, Little et al. 2016). Additionally, this work offers further insights into some potential mechanisms behind the stem cell and Wnt pathway regulating functions of RUNX1 (which appear to overlap rather than being mutually exclusive) and by extension its tumour suppressive functions. Interestingly, RUNX1 was indicated to potentially to trigger quiescence of the stem cell compartment in response to aberrant Wnt activation — a potential tumour suppressive mechanism that has not previously been considered for this protein, especially within the context of the mammary gland. Flow cytometry can be employed as a method for cell cycle analysis in order to compare the percentages of cells within the different phases of the cell cycle in order to confirm whether RUNX1 is stimulating growth arrest in response to WNT3A.

6 Exploring possible mechanisms by which *Runx1* loss facilitates accelerated *MMTV-PyMT*-mediated tumour emergence

6.1 Introduction

The *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mouse model offered valuable insights into the tumour suppressive role of *Runx1* and some of its possible mechanisms, within the context of ß-catenin-driven tumorigenesis. While over 50% of breast cancer patients present with a disease in which the Wnt pathway components are amplified and overexpressed (Zhan, Rindtorff et al. 2017), it is important to remember that breast cancer is a very complex disease with a multitude of possible pathways to disease. The role of *RUNX1* within the context of other breast tumour-initiating factors should therefore also be explored in order to answer the question of whether *RUNX1* might act as a generic tumour suppressor in the mammary gland, or if it is specific within the context of constitutively activated ß-catenin.

The role of *Runx1* was therefore interrogated using one of the most commonly used transgenic models in the cancer research field — the mammary specific polyomavirus middle T antigen overexpression mouse model (*MMTV-PyMT*) (Lin, Jones et al. 2003, Attalla, Taifour et al. 2021). It is a well-established, reliable model within the field of breast cancer research and has frequently been used to explore a variety of breast cancer characteristics including: tumour initiation; cancer progression (both histologically and molecularly); and metastasis. This is primarily due to its relatively high penetrance and early onset of multifocal mammary tumours that are capable of metastasising to the lung, one of the most common metastatic sites in the human disease (Lin, Jones et al. 2003, Attalla, Taifour et al. 2021). Another essential feature of the *MMTV-PyMT* model is that it can more accurately recapitulate the progression of the human disease due to its advancement through various molecularly distinct stages of development. In addition, this model was chosen for this particular study as molecular and histological analyses of *MMTV-PyMT* tumours reveal that they most closely resemble the human luminal subtype of breast cancer (Lin, Jones et al. 2003, Attalla, Taifour et al. 2021). Because *RUNX1* loss-of-function mutations were documented predominantly in ER-positive tumours, this makes the *MMTV-PyMT*-

driven model appropriate for studying the tumour suppressive role of *RUNX1* in the mammary gland.

MMTV-PyMT mice, with mammary-specific conditional deletion of *Runx1*, were generated as described in 2.1.1.2. Given that *Cre*-dependent recombination of the *Runx1*$^{fl/fl}$ allele and PyMT-driven oncogenesis were introduced via two independent transgenes and are therefore not mutually dependent events, efforts were made to trace *MMTV-Cre* expression through the introduction of the red fluorescent protein *Gt(ROSA)26Sor*^{tm1Hjf} (*tdRFP*) transgenic reporter line (Luche, Weber et al. 2007).

Previous work in the Blyth laboratory (Riggio 2017) determined the age at which palpable lesions were first detected in the mammary glands of *MMTV-Cre;MMTV-PyMT;tdRFP* mice. Subsequently, lesions were routinely checked and measured until the tumour reached the clinical endpoint, which is the maximum size allowed on the project licence (15mm in this case). Through this close monitoring of mice, it was therefore possible to compare the individual mouse cohorts using 3 distinct parameters (Figure 6.1A): tumour onset (time from birth to tumour notice); tumour progression (time from tumour notice to end point); and overall survival (time from birth to end point). The work presented in Alessandra Riggio's PhD thesis (shown in Figure 6.1) demonstrated that, similar to the *Blg-Cre;Catnb*$^{wt/lox(ex3)}$ mouse model, tumour onset happened significantly earlier in mice with both heterozygous and homozygous loss of *Runx1* function, as compared with their wild-type counterparts (Figure 6.1B). However, one unique characteristic of the *Runx1*-deficient mice was the slower progression of their disease from initiation to clinical endpoint (Figure 6.1C). For these first two parameters, the changes were more significant or pronounced in mice with deletion of both copies of *Runx1* (*Runx1*$^{fl/fl}$) compared to those with deletion of a single copy (*Runx1*$^{wt/fl}$). Due to this delayed tumour growth, which balanced out the accelerated tumour emergence, there were no significant changes observed for the overall survival (time from birth to clinical endpoint) in either of the *Runx1*-depleted cohorts compared to the wild-type controls (Figure 6.1D).

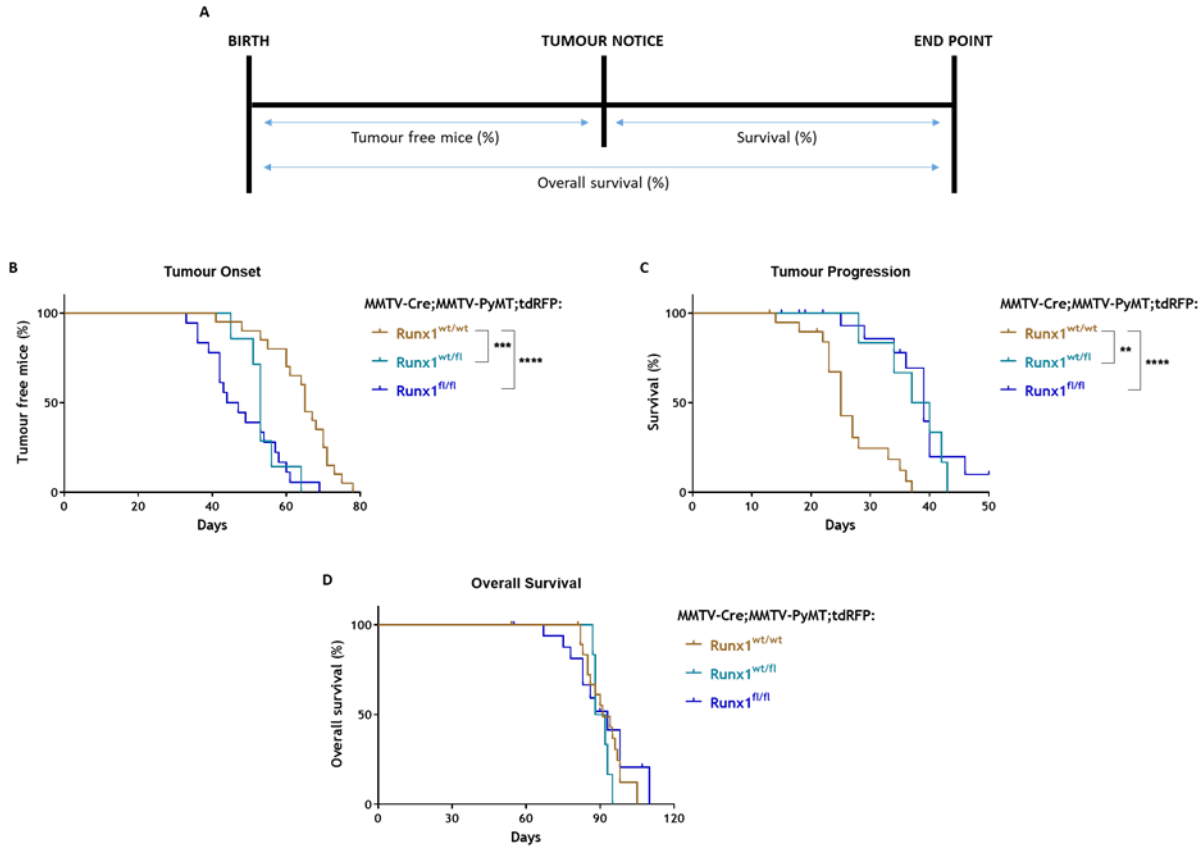


Figure 6.1 Deletion of *Runx1* accelerates *PyMT*-driven mammary tumorigenesis without affecting overall survival.

Data provided by Alessandra Riggio and shown here with permission

[A] A schematic representation of the three parameters that were used to assess the impact of partial/full loss of *Runx1* function on *PyMT*-driven tumorigenesis. Female *MMTV-Cre;MMTV-PyMT;tdRFP* mice, with and without the additional heterogeneous/homogeneous loss of *Runx1*, were monitored for tumour initiation by recording the first instances of any palpable tumours. Following tumour notice, size measurements were taken 2-3 times per week using callipers. Mice were sacrificed once a mammary tumour reached clinical end-point.

[B] Kaplan-Meier curve of time from birth to tumour notice observed in $Runx1^{wt/wt}$ (n=20), $Runx1^{wt/fl}$ (n=7), and $Runx1^{fl/fl}$ (n=18) mice.

[C] Kaplan-Meier curve of time from tumour notice to clinical end point for $Runx1^{wt/wt}$ (n=20), $Runx1^{wt/fl}$ (n=6), and $Runx1^{fl/fl}$ (n=18) mice.

[D] Kaplan-Meier curve of time from birth to clinical end point recorded for $Runx1^{wt/wt}$ (n=20), $Runx1^{wt/fl}$ (n=6), and $Runx1^{fl/fl}$ (n=18) mice.

Statistical analyses performed for each, using the Log-rank (Mantel-Cox) test [** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$].

6.1.1 Experimental Aims

These results provide compelling evidence for a tumour suppressive role for RUNX1 within a model of luminal breast cancer. This argument was even more convincing when presented along with data from the ß-catenin-driven model (described in Chapter 3), in which there was also a persuasive argument made for a tumour-suppressive function for RUNX1. The *MMTV-PyMT* model additionally indicates that, while loss of *Runx1* facilitates the earlier arrival of tumours, it also delays the progression of established tumours past this initiation phase. These results imply that there could be a stage-dependent role for *Runx1* in the early vs late stages of tumorigenesis, at least within this model. It is also important to note that there appeared to be an almost concentration-dependent effect to the phenotype observed with *Runx1* alteration, as loss of both copies significantly enhanced the early-emergence and delayed progression phenotypes of this model above those presented by mice with loss of a single copy. The results presented in the coming chapter follow on from this crucial work by probing further into the possible mechanisms of this apparent stage-dependent role of *Runx1*. Given the prominent phenotype observed with the simultaneous loss of *Runx1* and *Runx2*, even compared with loss of *Runx1* alone, within the context of Wnt-driven tumorigenesis (Chapters 3 and 5), it was also of great interest to determine whether this effect was also applicable to the *PyMT*-driven model.

6.2 Results

6.2.1 Examination of *MMTV-Cre* expression and its associated recombination of the *Runx1*$^{fl/fl}$ allele through RFP expression analysis.

In order to monitor *MMTV-Cre* expression and its associated recombination activity in the *MMTV-Cre;MMTV-PyMT;tdRFP* mice, a former PhD student in the lab (Alessandra Riggio) imaged end point mice for RFP expression using the IVIS Spectrum imaging system. The main purpose for this was to investigate a potential caveat within this model that, although PyMT-driven oncogenesis and *Cre*-mediated deletion of *Runx1* function are both under the control of the *MMTV* promoter, these events are not necessarily mutually dependent (and may even be occurring asynchronously or independently) due to them being introduced into the mouse genome via two independent transgenes. While almost all mice in the *MMTV-Cre;MMTV-PyMT;tdRFP* cohort exhibited intense levels of RFP fluorescence

(8/9 mice) in the majority of mammary glands — suggesting a relatively high degree of synchronicity between *MMTV-Cre* activation and PyMT-driven tumorigenesis — the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ mice presented with a much higher degree of variability in the detected RFP signal, with 5/14 mice appearing almost completely RFP-negative. It was hypothesised that this lack of RFP activity in the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ cohort could be as a result of the *MMTV-Cre* recombinase being active in only a small percentage of mammary epithelial cells, or due to depletion of *Runx1*$^{fl/fl}$ (RFP-expressing) cells during the progression of PyMT-driven tumours.

To explore both of these theories, RFP reporter fluorescence was used as a surrogate for both *MMTV-Cre* expression and associated recombination of the *Runx1*$^{fl/fl}$ allele, through the key stages of tumorigenesis. *MMTV-Cre;MMTV-PyMT;tdRFP* mice (henceforth referred to as MCMP) and *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ mice (henceforth referred to as MCMP_1), were euthanised at the date of notice (DON) or end point (EP) stages of tumour development and all glands (both tumour-bearing and non-tumour-bearing) were harvested. *MMTV-Cre;tdRPF* mice (henceforth referred to as MC) and *MMTV-Cre;tdRFP;Runx1*$^{fl/fl}$ mice (henceforth referred to as MC_1) were also analysed for RFP expression, as tumour-free controls. However, due to the absence of *MMTV-PyMT*-driven tumorigenesis in these mice, equivalent timepoints were required. The DON timepoint equivalent was chosen as 60 days old (60D) due to the average DON timepoint (for the MCMP and MCMP_1 cohorts combined) being 56.4 days. An EP equivalent of 90 days old (90D) was selected due to a combined average EP timepoint (for MCMP and MCMP_1 cohorts) of 90 days.

Results from the FACS analysis of RFP expression, summarised in Figure 6.2, demonstrate that RFP (*MMTV-Cre*) expression was not significantly altered with loss of functional *Runx1*, nor was it lost over time or with tumour progression. In non-tumour-bearing mice, which did not carry the *MMTV-PyMT* oncogene, excision of *Runx1* did not alter the proportion of RFP-expressing cells and neither was there a difference between 60 and 90 days of age in either the MC or MC_1 test groups (Figure 6.2A). There looked to be increased percentages of RFP-positive cells in

MMTV-PyMT-expressing animals compared with their *MMTV-PyMT*-naïve counterparts. There also appeared to be a reduction in the percentage of RFP-expressing cells in MCMP_1 animals at EP compared to MCMP_1 mice at DON, and compared to their MCMP wild-type controls, although none of these comparisons were statistically significant (Figure 6.2B).

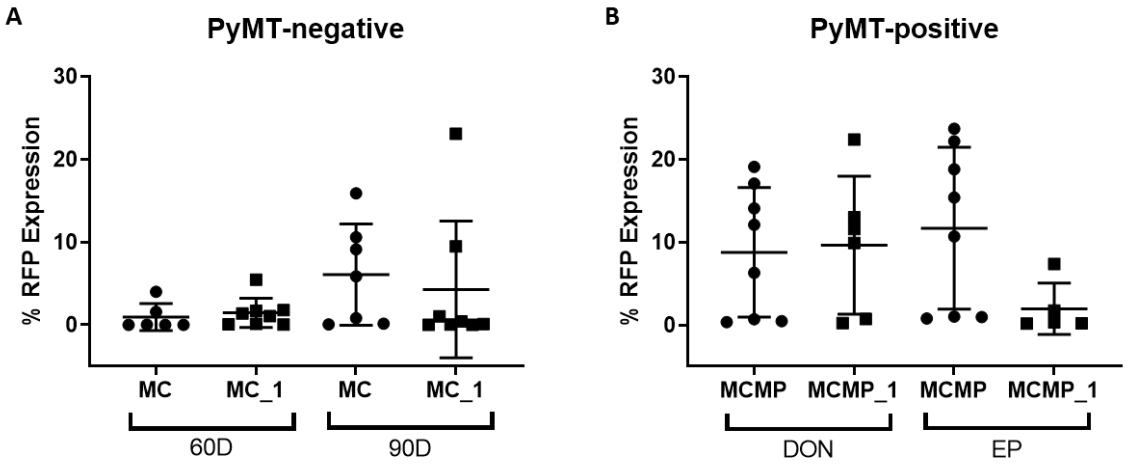


Figure 6.2 Initial analyses of mammary epithelial cells show that *Runx1* loss does not appear to have a statistically significant effect on RFP expression.

Box plots summarising FACS analysis of RFP expression in MMECs, isolated from *MMTV-Cre;MMTV-PyMT;tdRFP* cohorts (MCMP and MCMP_1) at date of notice (DON) and end-point (EP) [B] and from *PyMT*-naïve *MMTV-Cre;tdRFP* controls (MC and MC_1) at comparative 60 day (60D) and 90 day (90D) timepoints [A]. Each data point represents the percentage of live cells expressing RFP within a single sample. Each sample consists of the MMECs that were extracted from all mammary glands from a single mouse.

[A] 60D MC (n=6); 60D MC_1 (n=8); 90D MC (n=7); 90D MC_1 (n=8).

[B] DON MCMP (n=8); DON MCMP_1 (n=6); EP MCMP (n=8); EP MCMP_1 (n=5).

One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons, used for statistical comparisons.

6.2.2 Detailed *MMTV-Cre* expression analysis reveals *Runx1*$^{fl/fl}$ cells possess a selective advantage in early PyMT tumour development that is subsequently lost.

Following these initial analyses of RFP expression during the progression of the *MMTV-Cre;MMTV-PyMT;tdRFP* model of mammary tumorigenesis, it was considered that combining all of the glands from each mouse into a single sample might mask any subtle effects in individual glands. This could be particularly true for the DON timepoint for which there is a high proportion of non-tumour-bearing

glands compared to tumour-bearing glands. It was also thought that it could be valuable to examine RFP expression at an even earlier pre-neoplastic timepoint. In the initial analyses, *Runx1* depletion did not appear to significantly affect RFP expression at the DON timepoint, and there appeared to be a slight (though insignificant) decrease during the progression of MCMP_1 mice from DON to EP. Neither of these observations offers much explanation as to how loss of *Runx1* function results in earlier tumour emergence in this model. The fact that this significantly earlier tumour emergence is not coupled with an increase in the number of *MMTV-Cre*- and *MMTV-PyMT*-expressing cells at the DON timepoint indicates that the expansion of the tumour-initiating and RFP-expressing mammary epithelial cells in MCMP_1 mice may be happening at an even earlier timepoint in order to facilitate this earlier tumour incidence.

These theories, that RFP-expressing/*Runx1*-deleted cells may already be depleting in numbers by day of notice (DON), or their numbers may be drowned out at the DON stage by the high percentage of non-tumorigenic glands and untransformed mammary epithelial cells, are potential experimental caveats. This was addressed by adding an additional experimental timepoint at the pre-neoplastic (PN) stage. A 40 day equivalent timepoint was chosen for the PyMT-negative controls as this was the earliest possible post-pubertal timepoint occurring prior to the formation of any palpable lesions. Additionally, each of the glands was treated as an individual sample, as described in 2.1.5.2, which was analysed for RFP expression by FACS.

From the analysis in Figure 6.3A, it is clear that there are significantly higher levels of RFP expression in MC_1 mice at each of the 3 timepoints compared to those observed in MC mice. At the PN and DON timepoints, RFP expression levels are heightened to an even more significant extent within the context of *MMTV-PyMT*-driven oncogenesis in the individual mammary glands lacking *Runx1* expression, compared to their wild-type counterparts (Figure 6.3B). This indicates that there is possibly an expansion in the number of cells co-expressing the *MMTV-Cre* and *MMTV-PyMT* transgenes, and therefore an increase in the number of cells capable of initiating tumorigenesis. Looking at the RFP expression values for each

gland within a single mouse there is a visible amount of variation in RFP fluorescence from gland to gland. This indicates that the reason no significant differences were found in the analysis of combined glands was due to gland-to-gland variability and, in *MMTV-PyMT*-expressing mice, the tumorigenic glands with significant numbers of RFP-expressing cells being outnumbered by non-tumorigenic glands with relatively low levels of expression. While there is a significant enhancement of RFP fluorescence during the early stages of tumorigenesis in mice lacking *Runx1* expression, it is important to note that this is lost with progression of MCMP_1 tumours to endpoint, whereupon levels are relatively comparable to those expressed in the MCMP *Runx* wild-type control mouse glands (Figure 6.3B). Given that MCMP_1 mice display multifocal tumours across all glands and exhibit similar tumour burdens to MCMP mice, this reduction in RFP is unlikely to be due to a depletion of *MMTV-PyMT*-expressing cells, rather the *Runx1*-depleted cells are likely to be selected against for the progression of this model from DON to EP.

These results make sense within the context of the survival curves (Figure 6.1C) where loss of *Runx1* function resulted in an earlier emergence of tumours, but the progression of this model from DON to EP was delayed in comparison to their wild-type counterparts. It is possible that, while *Runx1* depletion can leave the mammary gland more susceptible to PyMT-driven tumorigenesis (potentially via the expansion of the tumour initiating population), a certain amount of *Runx1* expression could be required for the *PyMT*-driven tumours to progress past this initiation step. *Runx1*-proficient cells may therefore offer a selective advantage over those lacking *Runx1* within the developing tumour, resulting in the depletion of their associated RFP expression.

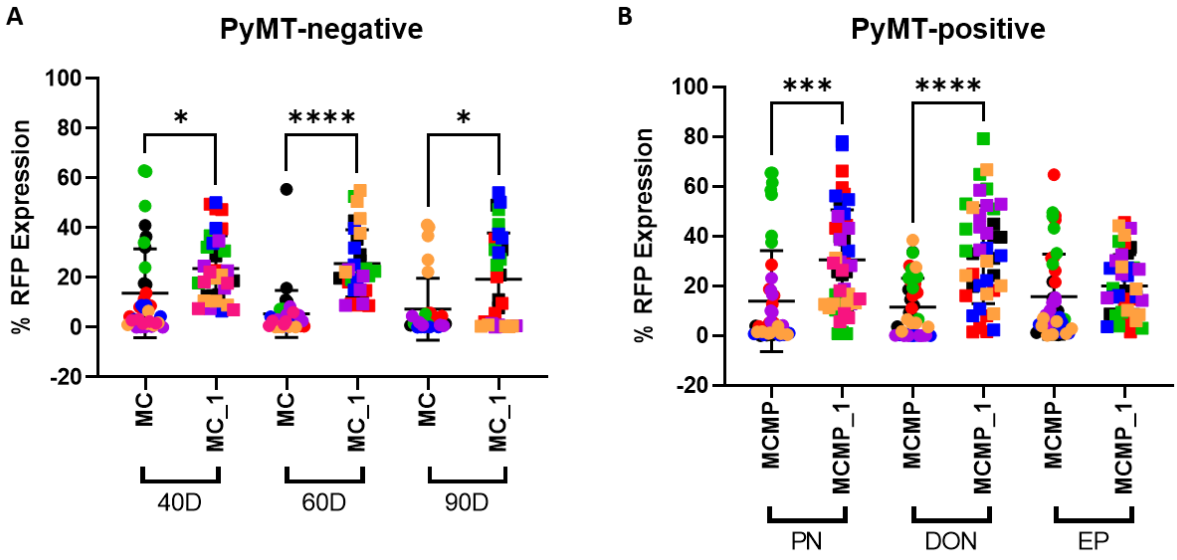


Figure 6.3 *Runx1* loss provides a selective advantage to tumour initiating cells within the mammary epithelium, but becomes disadvantageous in later tumour progression.

Box plots summarising FACS analysis of RFP expression in MMECs, isolated from *MMTV-Cre;MMTV-PyMT;tdRFP* cohorts (MCMP and MCMP_1) at pre-neoplastic (PN), date of notice (DON), and end-point (EP) [B] and from *PyMT*-naïve *MMTV-Cre;tdRFP* controls (MC and MC_1) at comparative 40 day (40D), 60 day (60D), and 90 day (90D) timepoints [A]. Each data point represents the percentage of live cells expressing RFP within a single sample. Each sample consists of the MMECs that were extracted from a single mammary gland. Each sample was given a colour to indicate the samples that originate from the same mouse within that cohort

[A] 40D MC (n=7); 40D MC_1 (n=7); 60D MC (n=7); 60D MC_1 (n=6); 90D MC (n=6); 90D MC_1 (n=6).

[B] PN MCMP (n=6); PN MCMP_1 (n=7); DON MCMP (n=6); DON MCMP_1 (n=6); EP MCMP (n=6); EP MCMP_1 (n=6).

One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons [* $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$].

6.2.3 *Runx1* loss results in expansion of cell populations with increased stem-like properties in emerging *PyMT*-driven tumours.

As stated above, the increase in RFP expression associated with the co-occurrence of *MMTV-Cre*-mediated *Runx1*$^{fl/fl}$ allele recombination and *MMTV-PyMT*-driven tumorigenesis followed by its significant depletion makes sense within the context of the progression/phenotype of this model. The results at PN and DON timepoints indicate that loss of *Runx1* could possibly result in the expansion of a sub-population of stem-like mammary epithelial cells capable of initiating tumour development, thereby resulting in earlier tumour incidence. Continuing with this

thinking, it is possible that the subsequent reduction in RFP expression therefore reflects either the exhaustion of or selection against these *Runx1*-depleted tumour-initiating cells with enhanced stem-like properties. Coupled with the fact that the progression of *Runx1*-deficient tumours is significantly delayed compared to their wild-type counterparts, it is possible that a certain level of *Runx1* expression is required for the efficient progression of *MMTV-PyMT* tumours from DON to EP, meaning *Runx1*-proficient tumour cells have a selective growth advantage. It is likely that MCMP_1 glands possess a lower percentage of tumour cells expressing functional *Runx1* when the initial tumour forms compared to the initial MCMP tumours. This may cause a reduction in the tumour growth rate, at least until the rarer *Runx1*-proficient cell population is able to acquire dominance within the tumour cell population.

To investigate the theory that an early-expanded stem-like population is exhausted or selected against during progression of the MCMP_1 mice, mammosphere/tumoursphere assays were employed as a tool to compare the stem-like capacities of the mammary epithelia of MCMP and MCMP_1 mice at DON and EP. MC and MC_1 control mice were also analysed at equivalent 60D and 90D timepoints to explore the effect of *Runx1* loss on the stem-like behaviour of the mammary epithelium in the absence of the *MMTV-PyMT* transgene. The results of these mammosphere/tumoursphere assays are summarised in Figure 6.4. It can be seen in Figure 6.4A that there is a significant increase in the mammosphere forming potential within the MC_1 mammary epithelium compared to MC controls at 60D, which is maintained at the 90D timepoint. Figure 6.4B shows an even greater enhancement in the mammosphere/tumoursphere forming potential of the MCMP_1 mammary epithelium, compared to that of the MCMP Runx wild-type controls, at tumour emergence. As the tumour progresses to the EP stages, however, it is apparent that this elevated capacity to form these stem cell-associated structures is diminished back to levels similar to their wild-type counterparts. Interestingly, the tumoursphere forming potential of both MCMP and MCMP_1 mammary epithelial cells at EP are both reduced as compared to MCMP-derived cell populations at DON.

The fact that this earlier emergence of tumours in the mammary epithelium of MCMP_1 mice coincides with this expansion of a stem-like population of cells, capable of forming 3D mammosphere/tumoursphere structures, supports the theory that loss of *Runx1* may result in the mammary epithelium becoming more vulnerable to oncogenic insult through the uncontrolled expansion of a stem-like population. This in turn implies that at least some of the tumour-suppressive functions of *Runx1* could be facilitated through its control over the stemness of the mammary epithelium. Following their early emergence, the delayed progression of these tumours is a surprising feature of this model, and appears to correspond with a statistically significant decrease in the tumoursphere forming potential within their cell populations. Coupled with the data in Figure 6.3, it is likely that (following tumour formation) there could be a certain requirement for *Runx1* expression to facilitate PyMT-driven tumour progression and, in order to meet this requirement, there is a selective pressure that favours the growth of (non-red and less stem-like) *Runx1*-proficient cells over the (red, tumour-initiating) *Runx1*-deleted cells. As there is likely to be lower proportions of *Runx1*-expressing cells in MCMP_1 DON tumours compared to their wild-type counterparts, these tumour cells will likely be delayed in their ability to progress tumour growth towards EP, which could explain the delayed progression of the MCMP_1 mouse model.

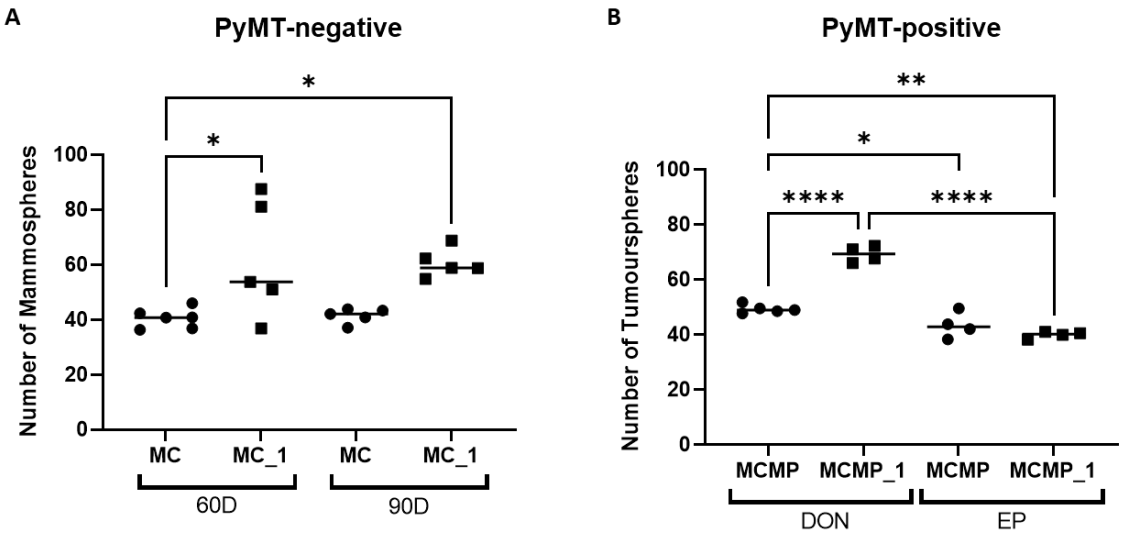


Figure 6.4 Enhanced mammosphere/tumoursphere forming potential of *Runx1*-deficient mammary epithelium at date of notice is lost upon progression to clinical end point.

Scatter plot summarising results from mammosphere/tumoursphere assays comparing *MMTV-Cre;MMTV-PyMT;tdRFP* cohorts (MCMP and MCMP_1) at date of tumour notice (DON) and endpoint (EP) [B] and comparing *MMTV-Cre;tdRFP* control cohorts (MP and MP_1) at equivalent 60 days (60D) and 90 days (90D) [A]. Each data point represents the average number of mammospheres/tumourspheres calculated (from an average of n=12 total counts per sample) for a single sample following 10 days of culture. Each sample constituted MMECs extracted from the combined glands of a single mouse. Results shown represent 1 of 2 experimental repeats.

[A] 60D MC (n=6); 60D MC_1 (n=5); 90D MC (n=5); 90D MC_1 (n=5).

[B] DON MCMP (n=5); DON MCMP_1 (n=4); EP MCMP (n=4); EP MCMP_1 (n=4).

One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons [* $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$].

6.2.4 Exploring potential molecular players in the stem-like potential of early *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1$^{fl/fl}$* tumours.

Similar to results observed in the *Blg-Cre;Catnb$^{wt/lox(Ex3)}$* and HC11 models, results in this chapter so far also appear to indicate a role for RUNX1 in the modulation of stemness within the context of *MMTV-PyMT*-driven tumorigenesis. The next step of this investigation was therefore to examine the possible molecular mechanisms by which RUNX1 is able to exert its control over the stem-like potential of the mammary epithelium in this context. qPCR analysis was selected as a method to evaluate the effects of *Runx1* loss on the expression of several established stem cell markers at the tumour initiation and clinical end stages of *MMTV-PyMT*-driven tumours.

The mammary epithelial cells that were collected from *MMTV-Cre;MMTV-PyMT;tdRFP* mice (with and without the additional loss of *Runx1*) and set up in the previously described mammosphere/tumoursphere assays were also sampled to analyse the relative expression levels of several stem-related genes by qPCR. The genes that were chosen for these analyses were those found to be the most significantly altered in *Runx1*-depleted HC11 cells — *Aldh1a1*, *Aldh1a7*, *Esr1*, and *Esr2*. There does not appear to be significant changes of expression in most of these genes, either with loss of functional *Runx1* or with tumour progression (Figure 6.5A-D). There did, however appear to be a slight (though not statistically significant) trend towards increased *Esr1* expression in MCMP_1 glands at tumour notice compared with their MCMP counterparts. Upon progression to end-point, there was a statistically significant reduction in the levels of *Esr1* in MCMP_1 glands (Figure 6.5C). This observation may offer some explanation as to why the *MMTV-PyMT*-driven tumours appeared to be more reliant on the expression of *Runx1* for their efficient progression past the initiation stage. Some aspects of the *PyMT* mouse model are thought to more accurately recapitulate the human breast cancer disease than some other models, such as their distinctive loss of ER and PR expression with disease progression (Lin, Jones et al. 2003, Attalla, Taifour et al. 2021). RUNX1 expression has previously been associated with poor patient prognosis in those diagnosed with ER-negative subtypes (Ferrari, Mohammed et al. 2014). Given that *ESR1* mRNA expression levels have been shown to closely correlate with ER protein expression (detected by IHC) in human breast cancer samples (Gong, Yan et al. 2007, Badve, Baehner et al. 2008), it would be reasonable to assume the same for the *PyMT* model. Taking all of this information together, this would indicate that *Runx1* loss may initially render the mammary gland more permissible to the earlier emergence of more ER-positive tumours (due to this increased ER expression in abnormal cells perhaps contributing to the faster expansion to form the initial tumour). Following tumour emergence, ER expression levels are significantly diminished during their progression, which may create an environment that is more admissible for the quick progression of *Runx1*-rich tumours than *Runx1*-diminished tumours.

While the diminishing expression of *Esr1* with *PyMT*-driven tumour progression provides some interesting insights into the possible reasons behind the stage-dependent role of *Runx1* within this model, the analyses of these select stem cell-related genes unfortunately do not appear to offer any potential candidate molecular markers or drivers for this stem-associated phenotype within the *Runx1*-depleted context.

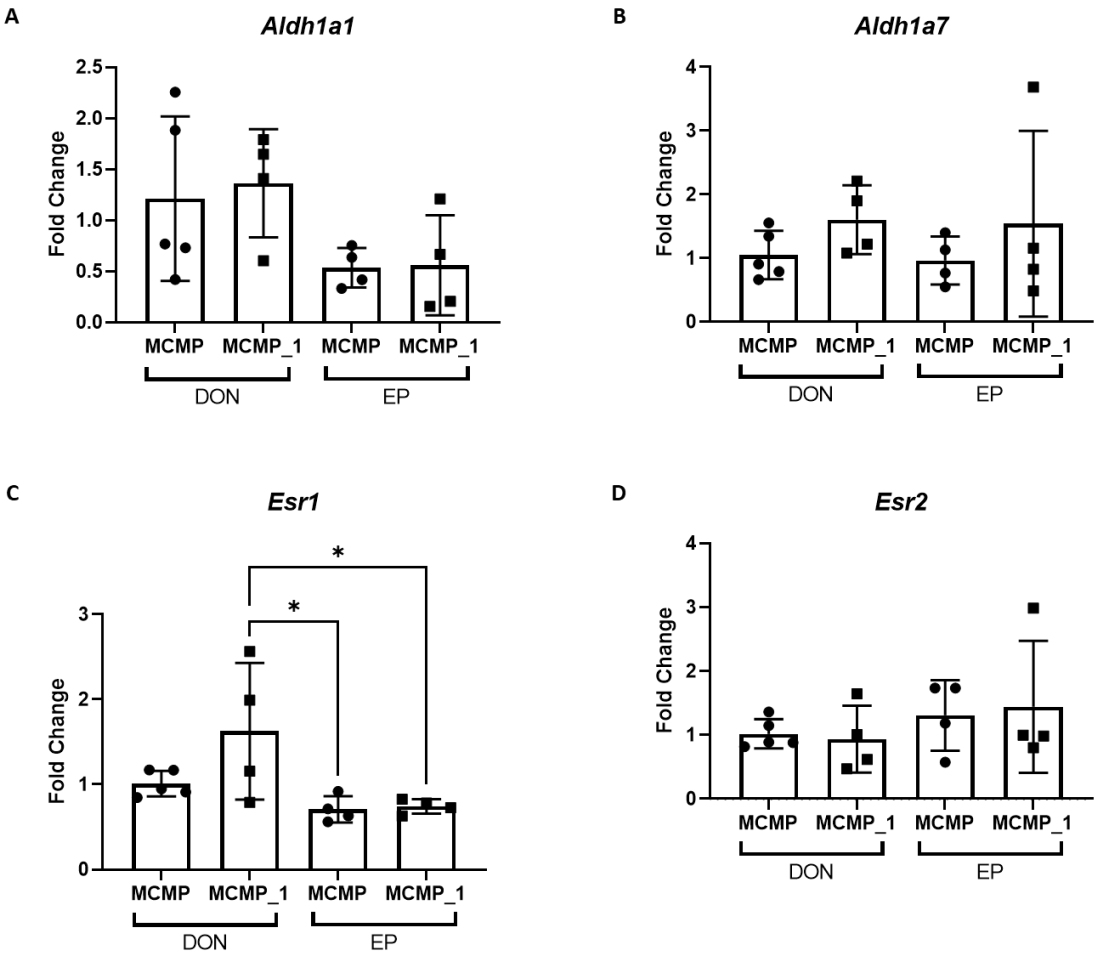


Figure 6.5 Stem cell gene expression is not significantly altered when *Runx1* is deleted in *MMTV-PyMT*-expressing glands.

Box and whisker plots summarising results from qPCR analysis of MMECs, extracted from *MMTV-Cre;MMTV-PyMT;tdRFP* cohort mice, with (MCMP_1) and without (MCMP) the additional deletion of *Runx1*, at tumour notice (DON) and clinical end-point (EP). *Gapdh* and *Actb* genes used as internal references to normalise the expression data for each test gene. Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual sample, where each sample consisted of MMECs extracted from the combined glands of a single mouse. DON MCMP (n=5); DON MCMP_1 (n=4); EP MCMP (n=4); EP MCMP_1 (n=4).

One-way ANOVA, with Tukey's Post-Hoc Test to correct for multiple comparisons [* $p < 0.05$].

6.2.5 Determining the impact of conditional *Runx2* loss on the survival of the *MMTV-Cre;MMTV-PyMT;tdRFP* model.

In the ß-catenin-driven model of oncogenesis, the accelerated tumour onset and stem-linked phenotype observed with mammary-specific loss of *Runx1* was dramatically enhanced with the additional knockout of *Runx2* (Chapter 3). It was wondered whether simultaneous loss of *Runx1* and *Runx2* might result in a similarly augmented phenotype within the context of *PyMT*-driven tumorigenesis. The $Runx2^{fl/fl}$ allele (Ferrari, Riggio et al. 2015) was therefore introduced into the *MMTV-Cre;MMTV-PyMT;tdRFP* colony to produce the *MMTV-Cre;MMTV-PyMT;tdRFP;*$Runx2^{fl/fl}$ and *MMTV-Cre;MMTV-PyMT;tdRFP;*$Runx1^{fl/fl}$;$Runx2^{fl/fl}$ cohorts. Similar to previous survival investigations (Chapter 3, Figure 6.1), this cohort of female animals were monitored for clinical onset of palpable tumours (tumour onset) which were then measured until clinical endpoint (tumour progression), as shown in the schematic in Figure 6.6A

In agreement with previous *MMTV-Cre;MMTV-PyMT;tdRFP;*$Runx1^{fl/fl}$ cohort data, deletion of both copies of *Runx1* resulted in the accelerated formation of palpable mammary gland lesions (median tumour-free survival of 41 days) compared to their wild-type counterparts (56 days median tumour-free survival). There was, however, no statistically significant difference in tumour onset between *Runx* wild-type and *Runx1*/*Runx2* double knockout mice (median tumour-free survival of 50 days), as shown in Figure 6.6B. Although tumour progression data (Figure 6.6C) did not show a statistically significant change in the tumour progression of *Runx1*-depleted mice (45 days median survival) compared to wild-type controls (42 days median survival), there did appear to be a slight trend towards a growth delay in the Kaplan-Meier curve. It is possible that this inconsistency with the results seen in Figure 6.1 could be due to some genetic drift in the *MMTV-Cre;MMTV-PyMT;tdRFP;*$Runx1^{wt/wt}$ and *MMTV-Cre;MMTV-PyMT;tdRFP;*$Runx1^{fl/fl}$ cohorts, which may have been caused by the multiple generations of back-crossing that were necessary for the generation of each cohort. Looking at the significantly earlier tumour initiation timepoints in both of these cohorts after the incorporation of the $Runx2^{fl/fl}$ mouse line (Figure 6.6B, median tumour-free survival of 56 and 41 days respectively) as compared with data collected from earlier analyses (Figure 6.1B, median tumour-free survival of 65 and 53 days

respectively), this certainly seems possible. There was a very significant delay in the progression of *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ tumours (56 days median survival). The difference in tumour initiation with loss of *Runx1* function was not reflected in the overall survival of these mice (87 days median overall survival), however there was a significant increase in the overall survival period with the additional loss of *Runx2* (104 days median overall survival) compared to wild-type mice (median overall survival of 88 days). Taking into account the knockout of *Runx2* alone, there was no significant impact on tumour emergence (Figure 6.6B, median tumour-free survival of 58 days) or the progression to end point (Figure 6.6C, although there was a slight trend towards delayed progression, with 47 days median survival). Despite this, the overall survival of these mice was significantly extended (Figure 6.6D, 105.5 days median overall survival) in comparison to the *Runx* wild-type cohort.

When looking at this survival data in isolation, it would likely be assumed that the loss of *Runx1* results in earlier emergence of tumours, while not significantly affecting tumour progression (although there was a slight, non-significant delay) and overall survival of the mice. It also appears that the additional loss of *Runx2* may at least partially compensate for this *Runx1*-depletion at the DON timepoint to overcome the stem-associated early-emergence phenotype. This implies that *Runx2* expression is essential to facilitate the expansion of this stem-like population, which is at least partially responsible for this accelerated initiation. Additionally, the loss of *Runx2* function seems to cause a significant delay in tumour growth to clinical end point. These findings are consistent with previous results, in which *Runx2* was shown to be functionally important in the mammary stem/progenitor cell population (Ferrari, Riggio et al. 2015). However, this does not entirely agree with other published works (Fritz, Hong et al. 2020) or data in Chapters 3 and 5 that indicate *Runx1* plays a more essential role than *Runx2* in the stemness and EMT of the normal and early tumorigenic mammary epithelium.

While it is possible that this inconsistency is due to the different contexts of the experiments, it is important to consider a potential caveat that was identified in the *Runx2*$^{fl/fl}$ model, which may provide some explanation for why this data does

not fully support that presented in previous chapters. Even in the absence of a *Cre* recombinase, the *Runx2*$^{fl/fl}$ mice exhibited a hypomorphic phenotype whereby their body weight was significantly lower than wild-type littermates (Figure 6.7A) throughout their development, bone development (particularly in the ribcage) was abnormal and lead to some breathing difficulties, and mammary gland development was delayed. The *Runx2* floxed allele was found to have retained the neomycin-resistance selection cassette (Neo cassette), which may have reduced gene expression and therefore resulted in this hypomorphic phenotype (Ferrari 2013).

Similar to observations made in previous investigations using this model, tumour burden and lung burden measurements (i.e. either the total mammary gland weight or lung weight, measured as a percentage of the body weight) at EP were unaffected by any combination of *Runx1* and/or *Runx2* deletion (Figure 6.7). While this may seem counterintuitive, particularly within the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ tumours given their significantly earlier emergence, this may be at least partially explained by the slower progression of these tumours to clinical end point. In addition, *PyMT* mice at clinical end point often bear multifocal tumours within the majority of the total mammary tissue, even without the context of *Runx1* loss. This could mean that it may not be possible for *Runx1* deletion to significantly increase the tumour burden at these end stages, as the maximal tumour burden is already reached with wild-type *Runx*.

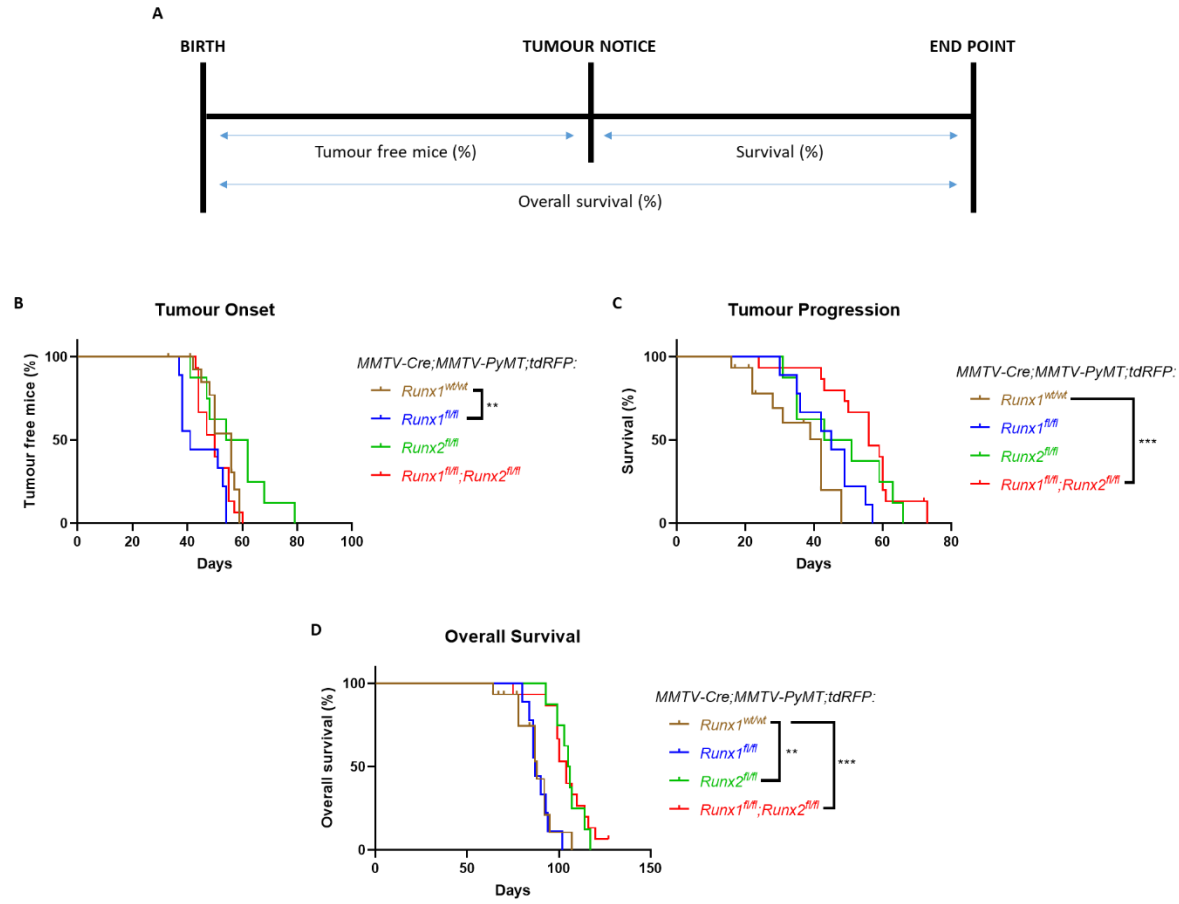


Figure 6.6 Simultaneous loss of *Runx1* and *Runx2* in an *MMTV-PyMT*-driven model of mammary tumorigenesis appears to extend overall survival due to delayed tumour progression.

[A] A schematic representation of the three parameters that were used to assess the impact of loss of *Runx1* and/or *Runx2* function on PyMT-driven tumorigenesis. Female *MMTV-Cre;MMTV-PyMT;tdRFP* mice, with and without the additional loss of *Runx1* and/or *Runx2*, were monitored for tumour initiation by recording the first instances of any palpable tumours. Following tumour notice, size measurements were taken 2-3 times per week using callipers. Mice were sacrificed once a mammary tumour reached clinical end-point.

[B] Kaplan-Meier curve of time from birth to tumour notice observed in $Runx^{wt/wt}$ (n=15), $Runx1^{fl/fl}$ (n=9), $Runx2^{fl/fl}$ (n=8), and $Runx1^{fl/fl}$;$Runx2^{fl/fl}$ (n=15) mice.

[C] Kaplan-Meier curve of time from tumour notice to clinical end point for $Runx^{wt/wt}$ (n=15), $Runx1^{fl/fl}$ (n=9), $Runx2^{fl/fl}$ (n=8), and $Runx1^{fl/fl}$;$Runx2^{fl/fl}$ (n=15) mice.

[D] Kaplan-Meier curve of time from birth to clinical end point recorded for $Runx^{wt/wt}$ (n=15), $Runx1^{fl/fl}$ (n=9), $Runx2^{fl/fl}$ (n=8), and $Runx1^{fl/fl}$;$Runx2^{fl/fl}$ (n=15) mice.

Statistical analyses performed for each, using the Log-rank (Mantel-Cox) test [** $p < 0.01$; *** $p < 0.001$].

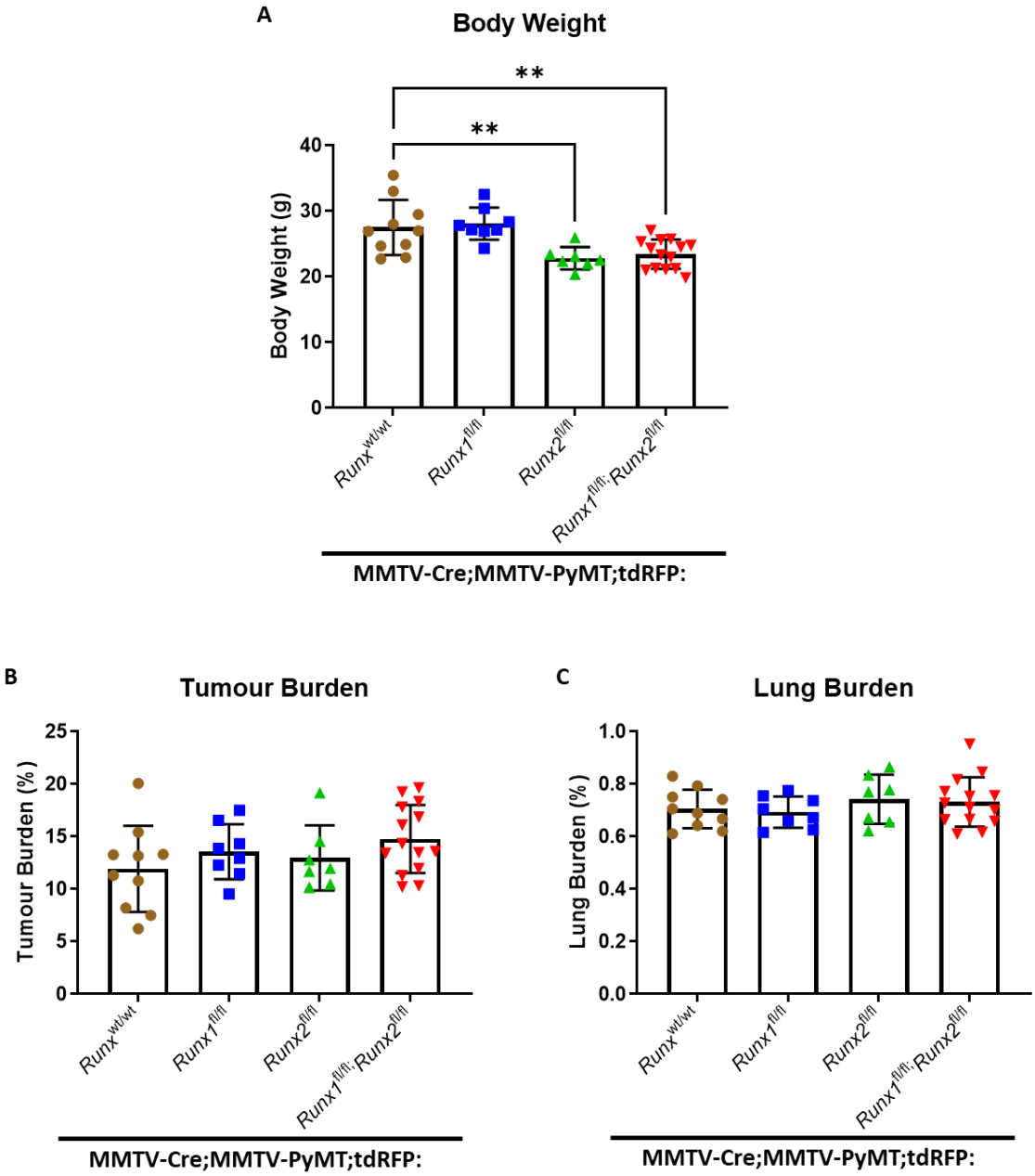


Figure 6.7 Body weight measurements and tumour/lung burdens calculated for *MMTV-Cre;MMTV-PyMT;tdRFP* cohort mice.

Box and whisker plots summarising body weights [A], tumour burdens (total mammary gland weight expressed as a percentage of body weight) [B], and lung burdens (lung weight expressed as a percentage of body weight) [C] for *MMTV-Cre;MMTV-PyMT;tdRFP* cohorts at clinical end-point – $Runx1^{wt/wt}$ (n=10), $Runx1^{fl/fl}$ (n=8), $Runx2^{fl/fl}$ (n=7), $Runx1^{fl/fl}$;$Runx2^{fl/fl}$ (n=14). Each data point represents a measurement or calculation taken for an individual cohort mouse. One-way ANOVA performed for each, with Dunnett's multiple comparisons test [** $p < 0.01$].

6.3 Discussion

Previous work in the Blyth laboratory, which explored the role of *Runx1* in *MMTV-PyMT*-driven tumorigenesis, demonstrated that the regulation of the stem-like behaviours of emerging tumour cell populations was not specific to the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ model. Both heterogeneous and homogeneous loss of *Runx1* resulted in earlier tumour emergence, however one unique observation from this model was their delayed progression to endpoint following tumour notice. *MMTV-Cre* activity could be examined through fluorescence imaging following incorporation of an RFP reporter. Within the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{wt/wt}$ cohort, intense levels of RFP expression were exhibited more consistently across mice within the majority of their mammary tissue, suggesting that *MMTV-Cre* expression coincided with *MMTV-PyMT*-mediated tumorigenesis. This was in stark contrast to the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ cohort, for which there were significantly reduced and more variable RFP expression patterns, both between mice and within the various glands of each mouse. Considering the tumour burdens of these mice were comparable to their wild-type counterparts, the cause of this variability was postulated to be due to either low *MMTV-Cre* activity in *Runx1*$^{fl/fl}$ glands or depletion of *Runx1*$^{fl/fl}$ mammary epithelial cells during tumour progression.

RFP expression was therefore analysed by FACS at key stages in tumour development to determine whether the reduced RFP expression at EP is due to *MMTV-Cre* activity being low throughout, or if the *Runx1*$^{fl/fl}$ cells are diminishing with disease progression. Initial combined mammary gland analyses at DON and EP stages of tumour development did not uncover any significant changes in RFP expression, by either *Runx1* depletion or disease progression, likely due to gland-to-gland variability. By analysing glands as individual samples, this allowed for the visualisation of this variability within the glands of each mouse, in addition to revealing a significantly expanded RFP-expressing population within *Runx1*$^{fl/fl}$ glands at the DON stage, and an additional PN stage, that was subsequently depleted upon progression of the model to EP.

By isolating mammary epithelial cells from DON and EP glands, this significant increase in the percentage of RFP-expressing cells in the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ group was shown to coincide with the expansion of a stem-like cell population. Likewise, the reduced RFP expression upon progression to EP was coupled with a depletion in these cells with stem-like characteristics. It was therefore theorised that, while *Runx1* depletion enabled the earlier emergence of mammary tumours through the expansion of a stem-like population of tumour-initiating cells, there may be some requirement for *Runx1* in order to efficiently advance past this initiation stage. This is reminiscent of the requirement for a certain level of RUNX1 activity for the growth of acute leukaemia (Goyama, Schibler et al. 2013, Choi, Illendula et al. 2017). This reliance on *Runx1* at later stages may cause a selective pressure against the RFP-expressing, tumour-initiating *Runx1*$^{fl/fl}$ cell population and resulting in their depletion.

While investigating possible molecular mechanisms behind these enhanced, and subsequently depleted, stem-like behaviours there was shown to be a significant reduction in the levels of *Esr1* from tumour notice to endpoint. This may at least partly explain the rather unique stage-dependent function of *Runx1* indicated within the *MMTV-PyMT*-driven model, as it was indicated this could be related to its subtype-specific functions. However, there were no significant changes found in the chosen panel of stem-associated genes upon *Runx1* loss. The small gene panel (consisting of *Aldh1a1*, *Aldh1a7*, *Esr1*, and *Esr2*) was chosen due to these genes being among the most significantly upregulated stem-associated genes following loss of *Runx1* function, in both the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ and HC11 models. It is important to consider the reasons why there might be a significant enhancement in mammosphere/tumoursphere formation in early tumours arising from *Runx1*$^{fl/fl}$ cells, but not an enhancement in the selected stem cell-associated genes. One of the possible reasons for this is that there could be alternative genes, outwith those within the chosen panel, which are being regulated by RUNX1 within the context of the PyMT mouse model. Another possibility is that, although the expression of *Aldh1a1*, *Aldh1a7*, *Ers1*, or *Ers2* is not significantly altered in the total mammary gland tissue with *Runx1* depletion, this does not definitively prove that there is no difference in expression of these, or other stem cell-associated, genes within these conditions. Because there is an almost binary element to the

formation of mammospheres (i.e. if a cell possesses stem-like abilities, it will form a sphere and if it does not, no sphere is formed), the final readout that is used to infer the stemness of a given cell population is not necessarily affected by the presence of non-stem-like cells. However, within the context of the qPCR assay, it is possible that the overall expression of each stem-associated gene could be drowned out by the amount of either non-transformed or non-stem-like cells in the analysed samples, particularly as these are likely to be the more dominant cell population within the total mammary gland tissues of a mouse at DON. It is therefore possible that cDNA isolated specifically from tumour tissue, or from purified populations of *PyMT*-transformed or stem-like cells (isolated by FACS sorting for RFP-expressing cells or collecting primary/secondary tumourspheres) at the DON timepoint would allow for a more accurate analysis of stem-associated genes.

The analyses of the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ cohort, presented throughout this chapter, gave useful insights into the functions of *Runx1* within a mouse model with some aspects that are thought to more accurately recapitulates the human disease. It is, however, also important to remember that the most pronounced phenotype observed in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ model occurred as a result of simultaneous loss of *Runx1* and *Runx2* expression (Chapter 3). The *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ genotype was therefore introduced as a new cohort for this model in order to explore the simultaneous loss of *Runx1* and *Runx2* within the context of *MMTV-PyMT*-driven oncogenesis. The data generated from observations made in the *MMTV-Cre;MMTV-PyMT;tdRFP;Runx1*$^{fl/fl}$ cohort were in agreement with results previously presented by Alessandra Riggio. However, due to a crucial experimental caveat in this model relating to the hypomorphic phenotype observed in *Runx2*$^{fl/fl}$ mice (caused by retention of the Neo cassette), it was not possible to confidently determine the impact of *Runx2* loss (either on its own or in addition to *Runx1* loss) on the *MMTV-PyMT*-driven model. While this hypomorphic phenotype was also present within the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ model, these mice are required to undergo multiple parities to induce sufficient levels of *Blg-Cre* (and therefore ß-catenin) expression, with the exception of the *Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ cohort due to its significant acceleration. It is overall a slower model of mammary tumorigenesis

than the *PyMT*-driven model, for which penetrance is significantly higher in virgin females and at a much younger age. It is therefore possible that, because tumour initiation occurs in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice at a stage significantly later than puberty, tumour initiation will likely not be hindered by the delayed development of the *Runx2*$^{fl/fl}$ mammary gland tissue in the way that it might in the *PyMT*-dependent model. This significantly later onset may allow enough time for *Runx2*$^{fl/fl}$ mammary glands to overcome this developmental delay and catch-up to their wild-type counterparts at a significantly earlier stage than ß-catenin-initiated tumorigenesis.

While the discussed experimental caveat meant it was not possible to make any definite conclusions about the implications of *Runx2* loss (with and without the additional depletion of *Runx1*) within the context of the *MMTV-PyMT* mouse model, it may be possible to address these issues in future experiments. From this *PyMT*-driven model of oncogenesis, endpoint tumours were extracted from each of the *Runx* genotypes, with an absence of the *MMTV-Cre*, and mammary epithelial cells were cultured for 5+ passages to generate stable cell lines. Using adenovirus-mediated *Cre* deletion of the floxed sequences within these primary cells, it may be possible to investigate the functions of *Runx1* and *Runx2* in established *MMTV-PyMT* tumours using various *in vitro* methods. It may even be possible to conduct *in vivo* investigations via syngeneic transplantation experiments into pure FVB mice to determine whether loss of *Runx1* and/or *Runx2* affects tumour growth (Lv, Dobrolecki et al. 2020). Furthermore, the limitations of the *Runx2*$^{fl/fl}$ model could be addressed using the more recent *Runx2*$^{fl/fl}$ cohort, generated by the Strathdee lab, in which the retained Neo cassette was removed from the targeted loci via CRISPR-Cas9-mediated deletion. This would allow for more reliable investigations into the impact of *Runx2* deletion at various stages of *MMTV-PyMT*-mediated tumorigenesis.

Despite it not being possible to reliably study the role of *Runx2* (either independently or within the context of *Runx1*) within this model during the timeframe of this project, the cell lines generated provide valuable resources to facilitate future investigations using a variety of experimental approaches.

Additionally, the work throughout this chapter not only indicates a stemness-related tumour-suppressive role for *Runx1* in a second independent (and well-established) mouse model for mammary tumorigenesis, but also provides evidence to support data from studies into the human disease that this tumour-suppressive function could be conditional based upon the expression of ER. This means that not only could the tumour-suppressive role of RUNX1 be dependent on the tumour subtype, but also on the stage of tumour progression.

7 Conclusions and Future Directions

The work presented throughout this thesis follows directly on from the exciting work conducted by former PhD students of the Blyth lab (Ferrari 2013, Riggio 2017, Riggio Unpublished), which provided the first *in vivo* confirmation (using genetically engineered and transgenic mouse models) of the putative tumour suppressor role for *RUNX1* in the context of breast tumorigenesis.

In two separate genetic models of breast cancer (*MMTV-Cre;MMTV-PyMT* and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$), mammary specific deletion of *Runx1* was shown to drive the earlier emergence of tumours in the mammary gland. Notably, within the ß-catenin-driven model, there appeared to be changes in the gross phenotypes presented in the resulting mammary tumours, with a larger number of independently arising tumours appearing in a larger proportion of mammary glands at the clinical end stage. This was supported with quantitative analysis, whereby the numbers of burdened glands and overall tumour burden (%) appeared to be slightly increased upon *Runx1* depletion. Another remarkable observation from this mouse model was made in mice with the simultaneous loss of *Runx1* and *Runx2*. While there was a lack of noteworthy phenotypic observations from the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx2*$^{fl/fl}$ mice, as compared to their *Runx* wild-type counterparts (*Blg-Cre;Catnb*$^{wt/lox(Ex3)}$), a particularly dramatic phenotypic change was observed in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mouse cohort. Not only did these mice develop mammary tumours at significantly earlier timepoints as compared with the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ wild-type control mice, tumours also (rather surprisingly) arose earlier than in their single *Runx1*-depleted (*Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$) counterparts. Strikingly, tumours emerged at a stage so significantly early in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mice as to negate the requirement (and ability) for these mice to undergo parity in order to potentiate the expression of *Blg-Cre* (Selbert, Bentley et al. 1998), and therefore ß-catenin, as demonstrated in characterisation experiments using the RFP reporter (1.6.2). The gross phenotype of tumours observed in these mice was even more dramatic than that seen in their *Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$ counterparts, with the vast majority of (or all) mammary glands presenting with many multifocal and multicentric tumours at clinical end stage. This was also strongly supported with the quantification of tumour-burdened glands and overall tumour burden (%),

which were statistically significantly higher in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mice when compared with their *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ controls. These results further substantiate previous studies (discussed in 1.5.3.1) indicating that *RUNX1* can play a tumour-suppressive role in the breast, and offer some insights into the dynamic between *RUNX1* and *RUNX2* in the breast. Intriguingly, the earlier emergence of higher numbers of independently arising tumours also indicated a possible role for *RUNX1* in the regulation of stem-like behaviours of the mammary gland cells and in cancer stem cells. This is in agreement with previous publications, discussed in 1.5.4, that demonstrate *RUNX1* suppresses stem-like behaviours in breast cancer cells (Chimge, Little et al. 2016, Hong, Fritz et al. 2018, Kulkarni, Tan et al. 2018, Fritz, Hong et al. 2020).

Although there is a large body of evidence, corroborated by this convincing *in vivo* work, that strongly supports an argument for the involvement of *RUNX1* in tumour suppression in the breast, there is relatively little known about the specific molecular mechanisms for this. For this reason, RNA-Seq was performed on mammary gland cell populations that were isolated from 9 week-old ß-catenin cohort mice. This post-pubertal timepoint, prior to the formation of palpable lesions, was chosen in order to investigate the molecular events that were happening at these early stages to facilitate the earlier emergence of independently arising tumours in the *Runx1*-deficient gland. The analyses of bulk mammary tissues isolated from these pre-tumour glands proved to be fruitful, and there were, rather interestingly and unexpectedly, enrichments in various pro-inflammatory/pro-tumorigenic immune signatures and ECM remodelling signatures in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ sample group. As each sample was formed from preparations of bulk mammary gland cells with various cell types, there was some uncertainty over whether some of the identified differentially expressed signatures were originating from the *Blg-Cre*-expressing (and therefore ß-catenin-expressing and *Runx1*/*Runx2*-depleted) MMECs or from their microenvironment.

An RFP reporter was therefore introduced into each of the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ cohorts to address this. This knock-in *Cre*-reporter enabled FACS-mediated sorting of mouse mammary epithelial cells (MMECs) for those expressing the *Blg-Cre* transgene, and thereby facilitated the isolation of mammary epithelial cells with simultaneous constitutive activation of ß-catenin alongside conditional knock-out of *Runx1* and/or *Runx2*. These isolated MMECs could then also be analysed using RNA-Seq technology in order to identify the molecular changes happening specifically during the early transformation of MMECs lacking functional *Runx1* and/or *Runx2*. The analysis of RFP-sorted MMECs was split into 2 experiments: a pilot experiment; and an extended experiment with additional samples. The pilot experiment served as a necessary proof-of-concept to determine whether the analysis would be technically feasible. Part of the reasoning behind this was due to there being a relatively small number of *Blg-Cre*-expressing cells within the combined mammary gland tissues of each mouse at this early developmental stage. This also meant that RFP-sorted MMECs needed to be combined together from several different mice to form each sample. While there were some differences and discrepancies between the pilot and extended analyses in terms of the sample quality, levels of sample-to-sample variability, and the resulting statistical interpretations (whereby significantly altered genes from the pilot experiment were rendered insignificant upon inclusion of additional samples for the expanded experiment), overall the results demonstrated that loss of functional *Runx1* leads to an enhancement of the canonical Wnt/ß-catenin pathway and various stem cell-related genes, which may be contributing to the observed stem-like phenotype. Given these enrichments were even more significant in the *Runx1*/*Runx2*-depleted MMECs, this could at least partially explain why the stem-related phenotype was much more pronounced within the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mouse cohort than the *Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$ mice.

One molecular alteration of particular note, identified in RNA-Seq data from isolated MMECs, was found in the *Bcl11b* gene, which was shown to be significantly upregulated in MMECs isolated from both *Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$ and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ cohort mice. This gene is involved in the maintenance of mammary stem cell populations in the mammary gland and drives

their self-renewal through inhibition of differentiation, and specifically commitment to a basal lineage, thus preserving their multipotency (Cai, Kalisky et al. 2017, Miller, Jin et al. 2018). *Bcl11b* is a close paralog of the *Bcl11a* gene, which has also been implicated in stem and progenitor functionality in TNBC. It is overexpressed in a large percentage of TNBC tumours, which was shown to promote tumour formation and drive cancer cell invasion and metastatic growth (Khaled, Choon Lee et al. 2015, Seachrist, Hannigan et al. 2020, Wang, Xu et al. 2020). Its knockdown or deletion reduced tumour formation in xenograft and genetic mouse models, and its inactivation in established tumours resulted in their regression (Khaled, Choon Lee et al. 2015). *Bcl11a* deletion in a genetically engineered mouse model resulted in depletion of the mammary stem cell fraction of the mammary gland, as demonstrated in flow cytometry analysis and confirmed with mammary colony formation assays and transplantation of limiting cell dilutions into cleared fat pads, thus demonstrating its role in stem cell maintenance (Khaled, Choon Lee et al. 2015).

Interestingly, the *Aldh1a1* gene and its rodent-specific paralogue, *Aldh1a7*, were also upregulated in the *Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$ and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ experimental groups compared to the wild-type controls. Previous publications revealed that *ALDH1A1* expression promotes stem/progenitor properties in normal and malignant breast cells (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016). It was correlated with tumour initiation within populations of breast cancer stem cells and promoted invasion and metastasis of breast cancer cells both *in vitro* and *in vivo* (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016). High mRNA levels of *ALDH1A1* in breast tumours were also correlated with poorer overall survival in breast cancer patients (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016).

Of the various Wnt pathway targets shown to be dysregulated in the *Runx1*/*Runx2*-depleted RFP-expressing MMECs, *Lgr5* was of particular interest due to its establishment as a bona fide molecular marker of stem cells in a number of actively self-renewing organs and tissues (Nusse and Clevers 2017) including the breast (Plaks, Brenot et al. 2013, Trejo, Luna et al. 2017). LGR5, a prototypic

member of the LGR-subfamily, is a potentiator of the Wnt/ß-catenin signalling pathway and was also confirmed to be a relevant canonical Wnt pathway target within the context of the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mouse model. LGR5 was previously shown to be overexpressed in primary breast cancer tissues compared to paired adjacent non-tumour tissues (Yang, Tang et al. 2015). Its expression in breast cancer tissues was also correlated with various clinical variables including: tumour size; TNM stage; recurrence and metastasis; and aggressive subtypes (Yang, Tang et al. 2015, Hou, Chen et al. 2018, Lee, Myung et al. 2021). Additionally, LGR5 expression was associated with poorer overall and disease-free survival (Yang, Tang et al. 2015, Hou, Chen et al. 2018). Through the use of *in vitro* and *in vivo* techniques, it was shown that LGR5 was able to promote breast tumour cell migration, invasion, metastasis, and EMT, which are behaviours generally associated with stemness and aggression of cancers (Yang, Tang et al. 2015). Additionally, LGR5 was found to be upregulated in spheroid cells (generated from 2 independent breast cancer cell lines) compared to both surrounding single cells, which failed to form spheroids, and compared to their 2D cultured counterparts (Yang, Tang et al. 2015). Silencing LGR5 expression resulted in a reduced capacity to form spheroids, while overexpression caused the formation of increased numbers of larger spheroids (Yang, Tang et al. 2015). These aggressive and stem-like behaviours were promoted by LGR5 through its activation of the Wnt/ß-catenin signalling pathway (Yang, Tang et al. 2015). The *Lgr5* data generated from the RNA-Seq experiment, coupled with this information from the literature, indicate a role for this gene in potentiating the stem-like behaviours of the early-transforming mammary epithelium, potentially through further Wnt pathway enhancement, in the absence of functional *Runx1* and *Runx2*.

While it was not possible to validate these RNA-Seq findings by RT-qPCR analysis of the RFP-sorted MMECs (due to insufficient cDNA quantities), the immortalised HC11 mouse mammary epithelial cell line offered a valuable and easily manipulatable *in vitro* tool for investigating the differentially regulated genes and for elucidating the role of RUNX1 in the stem-like behaviours of mammary epithelial cells. This cell line proved to be particularly useful in previous investigations conducted in the Blyth lab that focused on RUNX2 and its function in the mammary stem/progenitor populations (Ferrari, Riggio et al. 2015).

CRISPR/Cas9-mediated deletion of RUNX1 in this cell line mediated the discovery that loss of RUNX1 drives mammary cell stemness in mammosphere assays and colony forming assays. Loss of RUNX1 function also appeared to enhance the stemness-promoting effects of WNT3A-treatment in the 3D context. Conversely, ectopic expression of RUNX1 in these cells decreased mammosphere and colony formation capabilities, and dampened the stemness-promoting abilities of WNT3A. Interestingly, there was no effect on 2D growth with either RUNX1 loss or upregulation. Size quantification and IHC analyses of mammospheres formed from *Runx1*-overexpressing and –depleted HC11 cell lines indicated there was also a positive correlation between RUNX1 expression and both mammosphere area (measured in μM^2) and relative proliferation levels (as determined by % Ki67 staining), while there was an inverse correlation with cell death (using % Caspase-3 staining).

RT-qPCR analyses were used to evaluate relative expression levels of panels of Wnt target genes and candidate stem cell-related genes in RNA extracted from *Runx1*-altered HC11 cell lines grown in 2D and 3D conditions, with and without WNT3A treatment. This enabled investigations into the possible molecular mechanisms behind RUNX1-mediated control of the canonical Wnt/ß-catenin pathway and the stem-like potential of the mammary gland. The results from these analyses supported observations from the mammosphere assays in addition to the data obtained from RNA-Seq analysis of RFP-sorted MMEC samples, isolated from *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ cohort mice. Expression of several stem cell marker genes, including the *Aldh1a1* and *Aldh1a7* genes that were significantly upregulated in RFP-sorted MMECs isolated from *Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$ and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mouse mammary glands, was increased in *Runx1*-depleted HC11 cells and downregulated upon *Runx1* overexpression. The canonical Wnt pathway target and stem cell marker, *Lgr5*, showed similar expression changes with *Runx1* alteration, as did other genes within the panel of Wnt target genes. This data, together with the mammosphere results, indicate that *Runx1* limits the stem-like behaviours and the expression of various stem cell-related genes in the HC11 cell line. It was not clear, however, which (if any) of these stemness-related genes were directly contributing to the stem-like behaviours of the HC11 cells within this context, or if they were acting as markers

for cells with an increased stem-like potential. Functional studies would need to be conducted for these genes to determine this (either with CRISPR-mediated deletion or pharmacological suppression, and subsequent rescue). The RT-qPCR data also indicates that *Runx1* functions as a regulator of the Wnt/ß-catenin pathway, and is supportive of observations from HC11 cell line-derived, WNT3A-treated mammospheres, and a previous key study supporting a tumour suppressor role for *RUNX1* in ER-positive breast cancer (Chimge, Little et al. 2016).

To delve further into the fascinating phenotype observed in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$*;Runx1*$^{fl/fl}$*;Runx2*$^{fl/fl}$ mouse cohort, a series of HC11 cell lines was generated with singular knockout of the *Runx1* or *Runx2* genes, or with simultaneous *Runx1*/*Runx2* knockout. Results from this analysis again revealed no change in 2D growth with *Runx1* and/or *Runx2* manipulation. Mammosphere assays, however, supported results from single *Runx1* knockout experiments, and previous results from the Blyth lab showing that *Runx2* potentiates the mammosphere-forming potential of the HC11 cell line (Ferrari, Riggio et al. 2015). *Runx1*/*Runx2* knockout HC11 cells exhibited a significant increase in their mammosphere forming potential compared to control cells, albeit a relatively modest increase when compared with the *Runx1* deletion alone, indicating they were able to compensate for *Runx2* loss through simultaneous *Runx1* depletion. This data demonstrates opposing functions for *Runx1* and *Runx2* in the stem-like behaviours of the mammary gland and indicates that there is a greater reliance on *Runx1* expression for the regulation of stem-like behaviours of these mammary cells than there is on *Runx2* to potentiate their stemness. In addition to this, the additive enhancement of the stem-like potential in HC11 cells, in response to *Runx1* loss and WNT3A treatment, was further enriched with additional loss of functional *Runx2*. While loss of *Runx2* was previously shown to impair the activation of several Wnt target genes following recombinant WNT3A treatment (Ferrari, Riggio et al. 2015), the results from these WNT3A-treated mammospheres indicate the stemness-promoting functions of RUNX2 can be compensated for, or in this case overcompensated through alternate activation routes, when RUNX1 expression is also diminished.

The MTS assay was used to assess the impact of WNT3A treatment on HC11 cell behaviour (whereby the relative proportion of metabolically active cells was used as a surrogate for cell viability) relative to their vehicle-treated controls, and thereby explore further potential mechanisms for the increased production of mammospheres upon RUNX1 and/or RUNX2 depletion. Surprisingly, the proportion of metabolically active Empty Vector control HC11 cells was reduced, relative to Vehicle controls, with the addition of recombinant WNT3A protein. This reduction in metabolic activity (and by extension, reduction in cell viability) appeared to be RUNX1-dependent, as this WNT3A-triggered reduction was not observed in cells whose RUNX1 function is lost, either with or without the additional loss of RUNX2. In cells with RUNX2 loss alone, however, the suppression of metabolic activity in response to exogenous WNT3A treatment was maintained. Given the previously established role for RUNX1 as a Wnt/β-catenin pathway suppressor in the context of ER-positive breast cancer (Chimge, Little et al. 2016), supported by mammosphere and RT-qPCR results from this work, it was theorised that this metabolic suppression may be protective mechanism initiated by RUNX1 in response to excess levels of WNT3A. Taking this rationale slightly further, this gradual reduction in metabolic activity over the course of the 72 hour-long assay could indicate that RUNX1 is perhaps triggering cellular quiescence in an attempt to protect the HC11 mammary epithelial cells from oncogenic transformation. It would be beneficial to confirm whether cellular quiescence is indeed being initiated, and this could be facilitated through IHC analysis of bromodeoxyuridine(BrdU)-stained cell plugs from WNT3A treated HC11 cell lines (control and Rx1_KO).

Transplantation of limiting dilutions of HC11 (control and Rx1_KO) cells into cleared mammary fat pads of BALB/c mice were also attempted in order to address the limitations of the 3D *in vitro* methods used to evaluate their relative stem/progenitor-like capabilities. Attempts at limiting dilution transplantation assays into cleared mammary fat pads proved unsuccessful, however, as none of the transplanted cell populations was capable of reconstituting the mammary gland, and remained as compacted balls of cells rather than growing out into branches. Due to the fact that transplantation assays are generally considered the gold standard for determining the relative stem/progenitor-like potential of given

populations of mammary epithelial cells, it would be a potentially fruitful avenue for future investigation. As discussed in 5.3, it was likely that, in the absence of the necessary hormones, these HC11 cells failed to differentiate into the diverse cell lineages required to form the specialised structures of the mammary ductal branches, and instead remained in a de-differentiated state. To further this work, the most appropriate method for stimulating the natural production of lactogenic hormones, necessary for these cells to undergo differentiation, within the mammary glands would likely be for the mice to undergo at least 2 rounds of parity.

The previously described promising results, showing a role for RUNX1 in the regulation of stem cells in the normal mammary gland and within the context of oncogenic Wnt/β-catenin-mediated tumorigenesis, also prompted similar investigations within the context of the *MMTV-PyMT*-driven model of mammary tumorigenesis. This avenue of exploration was further rationalised following indications, using flow cytometry analysis, of an expanded population of RFP-positive (*MMTV-Cre*- and *MMTV-PyMT*-expressing) cells, capable of tumour initiation, within *Runx1*-deficient mammary glands during early transformation of the tissue, which was depleted upon progression to the clinical end point. It was believed that the expansion of a stem-like population of cells with tumour-initiating capacity, in the absence of functional *Runx1*, could be at least partially facilitating the earlier emergence of tumours in the mammary glands of *MMTV-Cre;MMTV-PyMT;Runx1*$^{fl/fl}$ mice. Their exhaustion or depletion at later stages in tumour development, perhaps as these *Runx1*-lacking cells may not offer a selective advantage with the gradually reducing expression of *Esr1*, may explain why these tumours are much slower to progress than their *Runx1*-proficient counterparts.

Tumoursphere assays of mammary epithelial cells isolated from mice at tumour notice (DON) and clinical endpoint (EP) stages appeared to confirm these suspicions, as 3D cultures derived from *MMTV-Cre;MMTV-PyMT;Runx1*$^{fl/fl}$ glands showed an enhancement in their tumoursphere-forming potential compared to their *MMTV-Cre;MMTV-PyMT* counterparts at the DON timepoint, which was not

evident at EP. Following these analyses up with RT-qPCR investigations did not reveal significant changes in *Aldh1a1* and *Aldh1a7*, the stem markers that were shown to be significantly upregulated with loss of *Runx1* function in the context of the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ model. While it is possible that these genes are indeed unaffected within this specific model, it is also conceivable that the overall expression of each gene, associated with rarer stem-like cell populations, could be drowned out by gene signatures/expression levels from the dominant non-transformed or non-stem-like cell populations in the analysed samples. To rule this out in future investigations, it would be necessary to isolate pure populations of PyMT-transformed or stem-like cells, by FACS sorting for RFP-expressing cells or collecting primary/secondary tumourspheres, at the DON timepoint to facilitate a more accurate RT-qPCR analysis of stem-associated genes. Analysis of a larger panel of stem-associated genes using these means could yield results that may be key to understanding this earlier emergence and slower-growing phenotype of *Runx1*-depleted glands undergoing oncogenic transformation by the *MMTV-PyMT* transgene.

To determine whether the addition of conditional *Runx2* knockout in this *MMTV-Cre;MMTV-PyMT* mouse model recapitulated the remarkable phenotype observed in the *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ model, the *Runx2*$^{fl/fl}$ line was crossed onto each *MMTV-Cre;MMTV-PyMT* cohort. Unfortunately, the hypomorphic phenotype observed in *Runx2*$^{fl/fl}$ mice (caused by retention of the Neo cassette), coupled with the speed of this model, resulted in this study yielding inconclusive results. It would be extremely beneficial to explore the implications of *Runx2* loss (with and without the additional depletion of *Runx1*) within the context of the *MMTV-PyMT* mouse model. This would help to determine if the dramatic acceleration of tumour emergence and progression was dependent on Wnt/ß-catenin-mediated tumorigenesis, or whether this role may also be evident in other independent models of mammary oncogenesis. To address the limitations of the *Runx2*$^{fl/fl}$ model, the Strathdee lab at the CRUK Beatson Institute generated a *Runx2*$^{fl/fl}$ mouse line in which the retained Neo cassette was removed from the targeted loci via CRISPR-*Cas9*-mediated deletion, which would be a valuable research for future *in vivo* investigations. For example, crossing this *Runx2*$^{fl/fl}$ line and the *Runx1*$^{fl/fl}$ line (Growney, Shigematsu et al. 2005) in various combinations with the *MMTV-*

Cre;MMTV-PyMT mouse model (or another independent model) could allow for a more thorough analysis of the consequences of losing functional *Runx1* and/or *Runx2* within a faster-progressing (and metastatic) model of mammary tumorigenesis, under the control of an independent oncogene. This would enable an investigation into whether the dramatic stem-like phenotype and associated molecular signatures, seen in *Runx1*-depleted mice within the ß-catenin-driven model, were unique to this context and therefore solely based on the regulatory roles that *Runx1* and *Runx2* play in Wnt-driven oncogenesis and stemness, or whether these phenotypic and molecular observations were representative of their generic roles in breast cancer and breast cancer stem cell populations. It may also be beneficial to consider additional strategies that are less breeding-intensive. For example, primary cell lines were derived from MMECs, extracted from endpoint tumours of mice from each *MMTV-PyMT* cohort (although without *MMTV-Cre* expression), as described in 6.3. Adenovirus-mediated *Cre* deletion of the floxed sequences within these primary cells may facilitate future *in vitro* (2D growth, tumourspheres, mammary colony formation, and RT-qPCR) and *in vivo* (orthotopic transplantation into cleared fat pads of syngeneic, pure FVB, mice) investigations into the functions of *Runx1* and *Runx2* in established *MMTV-PyMT* tumours.

The body of work presented in this thesis offers the first GEMM- or transgenic mouse model-based evidence of *Runx1* restricting tumour development at least partially due to its control over the stem-like properties of the breast. The data shown also aligns with previous publications investigating associations between *RUNX1* and the stem-like behaviours of several cell types (Kim, Barron et al. 2014, Yzaguirre, de Bruijn et al. 2017), including those contained within the normal and tumorigenic breast tissues. With respect to its role in the normal mammary stem cell populations, a 2015 publication indicated that RUNX1 expression is essential for the differentiation of mammary stem cells and subsequent mammary morphogenesis in 3D culture models (Sokol, Sanduja et al. 2015). Work from the Stein group also demonstrate that RUNX1 plays a role in the suppression of the breast cancer stem cell phenotype, tumour growth, migration, and invasion using human cell lines (Hong, Fritz et al. 2018, Fritz, Hong et al. 2020). The work presented here not only supports, but also builds on these analyses by exploring

the stem-related role of RUNX1 in a wider variety of stages and contexts of mammary tumorigenesis, within normal mammary epithelial cells, and delves further into the potential molecular mechanisms behind this regulatory role using various combinations of molecular and analytical techniques. Together, the work relating to the role of RUNX1 in mammary stemness indicates that it is an essential factor for the regulation of the stem-like compartment of epithelial cells contained within the mammary gland. It also demonstrates that if RUNX1 function is lost (due to loss-of-function mutations in the gene) this may lead to an expansion of this stem-like sup-population of mammary cells within the gland, which represents the population of cells that are more susceptible to oncogenic transformation and represent the mammary cell population with tumour-initiating capabilities. This expansion of stem cell populations may be facilitated through a reduced capacity to negatively regulate various stem cell markers (including *ALDH1*) that have previously been implicated in prognostic outcomes for breast cancer patients (Ginestier, Hur et al. 2007, Tomita, Tanaka et al. 2016). In the context of aberrant Wnt pathway activation, this increased stemness may also be facilitated through the loss of control over the activation of various Wnt pathway mediators and targets that not only promote oncogenic Wnt pathway activation but also possess stem cell-promoting functions. Excitingly, a novel role for RUNX1 in the metabolic suppression of mammary epithelial cells, and possibly in initiating cellular quiescence, in response to oncogenic insult was potentially identified as a result of the work carried out during this project. The research presented throughout this thesis adds to the exciting and ever-growing body of research focusing on this pathway as a promising therapeutic target, though there is still more work required to extend our understanding of the complex and context/stage-dependent functions of *RUNX1* before this can be translated into a clinical context.

Bibliography

(2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015." Lancet **388**(10053): 1459-1544.

(2022). "The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the Global Burden of Disease Study 2019." Lancet **400**(10352): 563-591.

Aherne, G. W., E. McDonald and P. Workman (2002). "Finding the needle in the haystack: why high-throughput screening is good for your health." Breast cancer research : BCR **4**(4): 148-154.

Ahmad, A. S., N. Ormiston-Smith and P. D. Sasieni (2015). "Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960." Br J Cancer **112**(5): 943-947.

Ai, L., Q. Tao, S. Zhong, C. R. Fields, W. J. Kim, M. W. Lee, Y. Cui, K. D. Brown and K. D. Robertson (2006). "Inactivation of Wnt inhibitory factor-1 (WIF1) expression by epigenetic silencing is a common event in breast cancer." Carcinogenesis **27**(7): 1341-1348.

Ali, S. and R. C. Coombes (2002). "Endocrine-responsive breast cancer and strategies for combating resistance." Nat Rev Cancer **2**(2): 101-112.

Alves, C. L., D. Elias, M. B. Lyng, M. Bak and H. J. Ditzel (2018). "SNAI2 upregulation is associated with an aggressive phenotype in fulvestrant-resistant breast cancer cells and is an indicator of poor response to endocrine therapy in estrogen receptor-positive metastatic breast cancer." Breast Cancer Res **20**(1): 60.

Andrechek, E. R., W. R. Hardy, P. M. Siegel, M. A. Rudnicki, R. D. Cardiff and W. J. Muller (2000). "Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis." Proc Natl Acad Sci U S A **97**(7): 3444-3449.

Attalla, S., T. Taifour, T. Bui and W. Muller (2021). "Insights from transgenic mouse models of PyMT-induced breast cancer: recapitulating human breast cancer progression in vivo." Oncogene **40**(3): 475-491.

Bäckström, S., M. Wolf-Watz, C. Grundström, T. Härd, T. Grundström and U. H. Sauer (2002). "The RUNX1 Runt domain at 1.25A resolution: a structural switch and specifically bound chloride ions modulate DNA binding." J Mol Biol **322**(2): 259-272.

Badve, S. S., F. L. Baehner, R. P. Gray, B. H. Childs, T. Maddala, M. L. Liu, S. C. Rowley, S. Shak, E. A. Perez, L. J. Shulman, S. Martino, N. E. Davidson, G. W. Sledge, L. J. Goldstein and J. A. Sparano (2008). "Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory." J Clin Oncol **26**(15): 2473-2481.

Ball, R. K., R. R. Friis, C. A. Schoenenberger, W. Doppler and B. Groner (1988). "Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line." Embo j **7**(7): 2089-2095.

Banerji, S., K. Cibulskis, C. Rangel-Escareno, K. K. Brown, S. L. Carter, A. M. Frederick, M. S. Lawrence, A. Y. Sivachenko, C. Sougnez, L. Zou, M. L. Cortes, J. C. Fernandez-Lopez, S. Peng, K. G. Ardlie, D. Auclair, V. Bautista-Piña, F. Duke, J. Francis, J. Jung, A. Maffuz-Aziz, R. C. Onofrio, M. Parkin, N. H. Pho, V. Quintanar-Jurado, A. H. Ramos, R. Rebollar-Vega, S. Rodriguez-Cuevas, S. L. Romero-Cordoba, S. E. Schumacher, N. Stransky, K. M. Thompson, L. Uribe-Figueroa, J. Baselga, R. Beroukhim, K. Polyak, D. C.

Sgroi, A. L. Richardson, G. Jimenez-Sanchez, E. S. Lander, S. B. Gabriel, L. A. Garraway, T. R. Golub, J. Melendez-Zajgla, A. Toker, G. Getz, A. Hidalgo-Miranda and M. Meyerson (2012). "Sequence analysis of mutations and translocations across breast cancer subtypes." Nature **486**(7403): 405-409.

Barker, N., M. van de Wetering and H. Clevers (2008). "The intestinal stem cell." Genes Dev **22**(14): 1856-1864.

Barker, N., J. H. van Es, J. Kuipers, P. Kujala, M. van den Born, M. Cozijnsen, A. Haegebarth, J. Korving, H. Begthel, P. J. Peters and H. Clevers (2007). "Identification of stem cells in small intestine and colon by marker gene Lgr5." Nature **449**(7165): 1003-1007.

Ben-Ami, O., D. Friedman, D. Leshkowitz, D. Goldenberg, K. Orlovsky, N. Pencovich, J. Lotem, A. Tanay and Y. Groner (2013). "Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1." Cell Rep **4**(6): 1131-1143.

Benhaj, K., K. C. Akcali and M. Ozturk (2006). "Redundant expression of canonical Wnt ligands in human breast cancer cell lines." Oncol Rep **15**(3): 701-707.

Bernkopf, D. B., M. V. Hadjihannas and J. Behrens (2015). "Negative-feedback regulation of the Wnt pathway by conductin/axin2 involves insensitivity to upstream signalling." J Cell Sci **128**(1): 33-39.

Bhat-Nakshatri, P., H. Appaiah, C. Ballas, P. Pick-Franke, R. Goulet, Jr., S. Badve, E. F. Srour and H. Nakshatri (2010). "SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype." BMC Cancer **10**: 411.

Bray F, M. B. (2006). "Predicting the future burden of cancer." Nat Rev Cancer **6**(1): 63-74.

Brenner, O., D. Levanon, V. Negreanu, O. Golubkov, O. Fainaru, E. Woolf and Y. Groner (2004). "Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia." Proceedings of the National Academy of Sciences of the United States of America **101**(45): 16016-16021.

Briest, S. and V. Stearns (2011). 17 - Neoadjuvant Therapy. Early Diagnosis and Treatment of Cancer Series: Breast Cancer. L. Jacobs and C. A. Finlayson. Saint Louis, W.B. Saunders**:** 261-279.

Browne, G., H. Taipaleenmäki, N. M. Bishop, S. C. Madasu, L. M. Shaw, A. J. van Wijnen, J. L. Stein, G. S. Stein and J. B. Lian (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**(10): 2522-2532.

Burton, R. and R. Bell (2013). "The global challenge of reducing breast cancer mortality." The oncologist **18**(11): 1200-1202.

Butti, R., S. Das, V. P. Gunasekaran, A. S. Yadav, D. Kumar and G. C. Kundu (2018). "Receptor tyrosine kinases (RTKs) in breast cancer: signaling, therapeutic implications and challenges." Mol Cancer **17**(1): 34.

Cai, S., T. Kalisky, D. Sahoo, P. Dalerba, W. Feng, Y. Lin, D. Qian, A. Kong, J. Yu, F. Wang, E. Y. Chen, F. A. Scheeren, A. H. Kuo, S. S. Sikandar, S. Hisamori, L. J. van Weele, D. Heiser, S. Sim, J. Lam, S. Quake and M. F. Clarke (2017). "A Quiescent Bcl11b High Stem Cell Population Is Required for Maintenance of the Mammary Gland." Cell Stem Cell **20**(2): 247-260.e245.

Callahan, R. and S. Hurvitz (2011). "Human epidermal growth factor receptor-2-positive breast cancer: Current management of early, advanced, and recurrent disease." Current opinion in obstetrics & gynecology **23**(1): 37-43.

Cancer Research UK. "Breast cancer statistics." September 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero.

Cancer Research UK. "Cancer incidence statistics." September 2022, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence.

Carvalho, R., A. N. A. Milne, M. Polak, W. E. Corver, G. J. A. Offerhaus and M. A. J. Weterman (2005). "Exclusion of RUNX3 as a tumour-suppressor gene in early-onset gastric carcinomas." Oncogene **24**(56): 8252-8258.

cBioportal for Cancer Genomics. August 2022, from https://www.cbioportal.org/.

Cerami, E., J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander and N. Schultz (2012). "The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data." Cancer Discov **2**(5): 401-404.

Chen, H., J. Wu, Z. Zhang, Y. Tang, X. Li, S. Liu, S. Cao and X. Li (2018). "Association Between BRCA Status and Triple-Negative Breast Cancer: A Meta-Analysis." Frontiers in pharmacology **9**: 909-909.

Chew, H. K. (2001). "Adjuvant therapy for breast cancer: who should get what?" The Western journal of medicine **174**(4): 284-287.

Chimge, N.-O., G. H. Little, S. K. Baniwal, H. Adisetiyo, Y. Xie, T. Zhang, A. O'Laughlin, Z. Y. Liu, P. Ulrich, A. Martin, P. Mhawech-Fauceglia, M. J. Ellis, D. Tripathy, S. Groshen, C. Liang, Z. Li, D. E. Schones and B. Frenkel (2016). "RUNX1 prevents oestrogen-mediated AXIN1 suppression and β-catenin activation in ER-positive breast cancer." Nature communications **7**: 10751-10751.

Chimge, N. O., G. H. Little, S. K. Baniwal, H. Adisetiyo, Y. Xie, T. Zhang, A. O'Laughlin, Z. Y. Liu, P. Ulrich, A. Martin, P. Mhawech-Fauceglia, M. J. Ellis, D. Tripathy, S. Groshen, C. Liang, Z. Li, D. E. Schones and B. Frenkel (2016). "RUNX1 prevents oestrogen-mediated AXIN1 suppression and β-catenin activation in ER-positive breast cancer." Nat Commun **7**: 10751.

Choi, A., A. Illendula, J. A. Pulikkan, J. E. Roderick, J. Tesell, J. Yu, N. Hermance, L. J. Zhu, L. H. Castilla, J. H. Bushweller and M. A. Kelliher (2017). "RUNX1 is required for oncogenic Myb and Myc enhancer activity in T-cell acute lymphoblastic leukemia." Blood **130**(15): 1722-1733.

Cicalese, A., G. Bonizzi, C. E. Pasi, M. Faretta, S. Ronzoni, B. Giulini, C. Brisken, S. Minucci, P. P. Di Fiore and P. G. Pelicci (2009). "The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells." Cell **138**(6): 1083-1095.

Colacino, J. A., E. Azizi, M. D. Brooks, R. Harouaka, S. Fouladdel, S. P. McDermott, M. Lee, D. Hill, J. Madden, J. Boerner, M. L. Cote, M. A. Sartor, L. S. Rozek and M. S. Wicha (2018). "Heterogeneity of Human Breast Stem and Progenitor Cells as Revealed by Transcriptional Profiling." Stem Cell Reports **10**(5): 1596-1609.

Cornen, S., A. Guille, J. Adélaïde, L. Addou-Klouche, P. Finetti, M.-R. Saade, M. Manai, N. Carbuccia, I. Bekhouche, A. Letessier, S. Raynaud, E. Charafe-Jauffret, J. Jacquemier, S. Spicuglia, H. de The, P. Viens, F. Bertucci, D. Birnbaum and M. Chaffanet (2014). "Candidate luminal B breast cancer genes identified by genome, gene expression and DNA methylation profiling." PloS one **9**(1): e81843-e81843.

Daniel, C. W., B. D. Aidells, D. Medina and L. J. Faulkin, Jr. (1975). "Unlimited division potential of precancerous mouse mammary cells after spontaneous or carcinogen-induced transformation." Fed Proc **34**(1): 64-67.

Danielson, K. G., C. J. Oborn, E. M. Durban, J. S. Butel and D. Medina (1984). "Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro." Proc Natl Acad Sci U S A **81**(12): 3756-3760.

De Braekeleer, E., N. Douet-Guilbert, F. Morel, M. J. Le Bris, C. Férec and M. De Braekeleer (2011). "RUNX1 translocations and fusion genes in malignant hemopathies." Future Oncol **7**(1): 77-91.

De Braekeleer, E., C. Férec and M. De Braekeleer (2009). "RUNX1 translocations in malignant hemopathies." Anticancer Res **29**(4): 1031-1037.

Deome, K. B., L. J. Faulkin, Jr., H. A. Bern and P. B. Blair (1959). "Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice." Cancer Res **19**(5): 515-520.

Djuretic, I. M., D. Levanon, V. Negreanu, Y. Groner, A. Rao and K. M. Ansel (2007). "Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells." Nat Immunol **8**(2): 145-153.

Doğukan, R., R. Uçak, F. M. Doğukan, C. Tanık, B. Çitgez and F. Kabukcuoğlu (2019). "Correlation between the Expression of PD-L1 and Clinicopathological Parameters in Triple Negative Breast Cancer Patients." European journal of breast health **15**(4): 235-241.

Domenici, G., I. Aurrekoetxea-Rodríguez, B. M. Simões, M. Rábano, S. Y. Lee, J. S. Millán, V. Comaills, E. Oliemuller, J. A. López-Ruiz, I. Zabalza, B. A. Howard, R. M. Kypta and M. D. Vivanco (2019). "A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells." Oncogene **38**(17): 3151-3169.

Dontu, G., W. M. Abdallah, J. M. Foley, K. W. Jackson, M. F. Clarke, M. J. Kawamura and M. S. Wicha (2003). "In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells." Genes Dev **17**(10): 1253-1270.

Dulak, A. M., S. E. Schumacher, J. van Lieshout, Y. Imamura, C. Fox, B. Shim, A. H. Ramos, G. Saksena, S. C. Baca, J. Baselga, J. Tabernero, J. Barretina, P. C. Enzinger, G. Corso, F. Roviello, L. Lin, S. Bandla, J. D. Luketich, A. Pennathur, M. Meyerson, S. Ogino, R. A. Shivdasani, D. G. Beer, T. E. Godfrey, R. Beroukhim and A. J. Bass (2012). "Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis." Cancer Res **72**(17): 4383-4393.

Eirew, P., J. Stingl, A. Raouf, G. Turashvili, S. Aparicio, J. T. Emerman and C. J. Eaves (2008). "A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability." Nat Med **14**(12): 1384-1389.

Ella, E., Y. Harel, M. Abraham, H. Wald, O. Benny, A. Karsch-Bluman, D. Vincent, D. Laurent, G. Amir, U. Izhar, O. M. Shapira, D. Yoon, H. S. Lee, D. J. Sugarbaker, B. Burt, A. Peled and O. Wald (2018). "Matrix metalloproteinase 12 promotes tumor propagation in the lung." J Thorac Cardiovasc Surg **155**(5): 2164-2175.e2161.

Ellis, M. J., L. Ding, D. Shen, J. Luo, V. J. Suman, J. W. Wallis, B. A. Van Tine, J. Hoog, R. J. Goiffon, T. C. Goldstein, S. Ng, L. Lin, R. Crowder, J. Snider, K. Ballman, J. Weber, K. Chen, D. C. Koboldt, C. Kandoth, W. S. Schierding, J. F. McMichael, C. A. Miller, C. Lu, C. C. Harris, M. D. McLellan, M. C. Wendl, K. DeSchryver, D. C. Allred, L. Esserman, G. Unzeitig, J. Margenthaler, G. V. Babiera, P. K. Marcom, J. M. Guenther, M. Leitch, K. Hunt, J. Olson, Y. Tao, C. A. Maher, L. L. Fulton, R. S. Fulton, M. Harrison, B. Oberkfell, F. Du, R. Demeter, T. L. Vickery, A. Elhammali, H. Piwnica-Worms, S. McDonald, M. Watson, D. J. Dooling, D. Ota, L.-W. Chang, R. Bose, T. J. Ley, D. Piwnica-Worms, J. M. Stuart, R. K. Wilson and E. R. Mardis (2012). "Whole-genome analysis informs breast cancer response to aromatase inhibition." Nature **486**(7403): 353-360.

Engel, C., K. Rhiem, E. Hahnen, S. Loibl, K. E. Weber, S. Seiler, S. Zachariae, J. Hauke, B. Wappenschmidt, A. Waha, B. Blümcke, M. Kiechle, A. Meindl, D. Niederacher, C. R. Bartram, D. Speiser, B. Schlegelberger, N. Arnold, P. Wieacker, E. Leinert, A. Gehrig, S. Briest, K. Kast, O. Riess, G. Emons, B. H. F. Weber, J. Engel and R. K. Schmutzler (2018). "Prevalence of pathogenic BRCA1/2 germline mutations among 802 women with unilateral triple-negative breast cancer without family cancer history." BMC Cancer **18**(1): 265.

Engelsen, A. S. T., K. Wnuk-Lipinska, S. Bougnaud, F. A. Pelissier Vatter, C. Tiron, R. Villadsen, M. Miyano, M. L. Lotsberg, N. Madeleine, P. Panahandeh, S. Dhakal, T. Z. Tan, S. D. m. Peters, S. Grøndal, S. M. Aziz, S. Nord, L. Herfindal, M. R. Stampfer, T. Sørlie, R. A. Brekken, O. Straume, N. Halberg, G. Gausdal, J. P. Thiery, L. A. Akslen, O. W. Petersen, M. A. LaBarge and J. B. Lorens (2020). "AXL Is a Driver of Stemness in Normal Mammary Gland and Breast Cancer." iScience **23**(11): 101649-101649.

Farnie, G. and R. B. Clarke (2007). "Mammary stem cells and breast cancer--role of Notch signalling." Stem Cell Rev **3**(2): 169-175.

Feng, Q., S. Li, H. M. Ma, W. T. Yang and P. S. Zheng (2021). "LGR6 activates the Wnt/ß-catenin signaling pathway and forms a ß-catenin/TCF7L2/LGR6 feedback loop in LGR6(high) cervical cancer stem cells." Oncogene **40**(42): 6103-6114.

Feng, Y., M. Spezia, S. Huang, C. Yuan, Z. Zeng, L. Zhang, X. Ji, W. Liu, B. Huang, W. Luo, B. Liu, Y. Lei, S. Du, A. Vuppalapati, H. H. Luu, R. C. Haydon, T.-C. He and G. Ren (2018). "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis." Genes & diseases **5**(2): 77-106.

Ferlay, J., M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor and F. Bray (2021). "Cancer statistics for the year 2020: An overview." Int J Cancer.

Ferlay J, L. M., Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2020). "Global Cancer Observatory." September 2022, from https://gco.iarc.fr/.

Ferrari, N. (2013). "Investigating RUNX transcription factors in mammary gland development and breast cancer." PhD Thesis, University of Glasgow.

Ferrari, N., Z. M. Mohammed, C. Nixon, S. M. Mason, E. Mallon, D. C. McMillan, J. S. Morris, E. R. Cameron, J. Edwards and K. Blyth (2014). "Expression of RUNX1 correlates with poor patient prognosis in triple negative breast cancer." PLoS One **9**(6): e100759.

Ferrari, N., A. I. Riggio, S. Mason, L. McDonald, A. King, T. Higgins, I. Rosewell, J. C. Neil, M. J. Smalley, O. J. Sansom, J. Morris, E. R. Cameron and K. Blyth (2015). "Runx2 contributes to the regenerative potential of the mammary epithelium." Sci Rep **5**: 15658.

Fijneman, R. J., R. A. Anderson, E. Richards, J. Liu, M. Tijssen, G. A. Meijer, J. Anderson, A. Rod, M. G. O'Sullivan, P. M. Scott and R. T. Cormier (2012). "Runx1 is a tumor suppressor gene in the mouse gastrointestinal tract." Cancer Sci **103**(3): 593-599.

Flowers, G. P., J. M. Topczewska and J. Topczewski (2012). "A zebrafish Notum homolog specifically blocks the Wnt/ß-catenin signaling pathway." Development **139**(13): 2416-2425.

French, R. and G. Tornillo (2019). "Heterogeneity of Mammary Stem Cells." Adv Exp Med Biol **1169**: 119-140.

Fritz, A. J., D. Hong, J. Boyd, J. Kost, K. H. Finstaad, M. P. Fitzgerald, S. Hanna, A. H. Abuarqoub, M. Malik, J. Bushweller, C. Tye, P. Ghule, J. Gordon, S. Frietze, S. K. Zaidi, J. B. Lian, J. L. Stein and G. S. Stein (2020). "RUNX1 and RUNX2 transcription factors

function in opposing roles to regulate breast cancer stem cells." J Cell Physiol **235**(10): 7261-7272.

Gao, J., B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander and N. Schultz (2013). "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal." Sci Signal **6**(269): pl1.

Gashaw, I., P. Ellinghaus, A. Sommer and K. Asadullah (2011). "What makes a good drug target?" Drug Discov Today **16**(23-24): 1037-1043.

Ginestier, C., M. H. Hur, E. Charafe-Jauffret, F. Monville, J. Dutcher, M. Brown, J. Jacquemier, P. Viens, C. G. Kleer, S. Liu, A. Schott, D. Hayes, D. Birnbaum, M. S. Wicha and G. Dontu (2007). "ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome." Cell Stem Cell **1**(5): 555-567.

Giráldez, A. J., R. R. Copley and S. M. Cohen (2002). "HSPG modification by the secreted enzyme Notum shapes the Wingless morphogen gradient." Dev Cell **2**(5): 667-676.

Gjerdrum, C., C. Tiron, T. Høiby, I. Stefansson, H. Haugen, T. Sandal, K. Collett, S. Li, E. McCormack, B. T. Gjertsen, D. R. Micklem, L. A. Akslen, C. Glackin and J. B. Lorens (2010). "Axl is an essential epithelial-to-mesenchymal transition-induced regulator of breast cancer metastasis and patient survival." Proc Natl Acad Sci U S A **107**(3): 1124-1129.

Gong, Y., K. Yan, F. Lin, K. Anderson, C. Sotiriou, F. Andre, F. A. Holmes, V. Valero, D. Booser, J. E. Pippen, Jr., S. Vukelja, H. Gomez, J. Mejia, L. J. Barajas, K. R. Hess, N. Sneige, G. N. Hortobagyi, L. Pusztai and W. F. Symmans (2007). "Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study." Lancet Oncol **8**(3): 203-211.

Goyama, S., J. Schibler, L. Cunningham, Y. Zhang, Y. Rao, N. Nishimoto, M. Nakagawa, A. Olsson, M. Wunderlich, K. A. Link, B. Mizukawa, H. L. Grimes, M. Kurokawa, P. P. Liu, G. Huang and J. C. Mulloy (2013). "Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells." J Clin Invest **123**(9): 3876-3888.

Goyama, S., J. Schibler, L. Cunningham, Y. Zhang, Y. Rao, N. Nishimoto, M. Nakagawa, A. Olsson, M. Wunderlich, K. A. Link, B. Mizukawa, H. L. Grimes, M. Kurokawa, P. P. Liu, G. Huang and J. C. Mulloy (2013). "Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells." The Journal of clinical investigation **123**(9): 3876-3888.

Greenman, C., P. Stephens, R. Smith, G. L. Dalgliesh, C. Hunter, G. Bignell, H. Davies, J. Teague, A. Butler, C. Stevens, S. Edkins, S. O'Meara, I. Vastrik, E. E. Schmidt, T. Avis, S. Barthorpe, G. Bhamra, G. Buck, B. Choudhury, J. Clements, J. Cole, E. Dicks, S. Forbes, K. Gray, K. Halliday, R. Harrison, K. Hills, J. Hinton, A. Jenkinson, D. Jones, A. Menzies, T. Mironenko, J. Perry, K. Raine, D. Richardson, R. Shepherd, A. Small, C. Tofts, J. Varian, T. Webb, S. West, S. Widaa, A. Yates, D. P. Cahill, D. N. Louis, P. Goldstraw, A. G. Nicholson, F. Brasseur, L. Looijenga, B. L. Weber, Y. E. Chiew, A. DeFazio, M. F. Greaves, A. R. Green, P. Campbell, E. Birney, D. F. Easton, G. Chenevix-Trench, M. H. Tan, S. K. Khoo, B. T. Teh, S. T. Yuen, S. Y. Leung, R. Wooster, P. A. Futreal and M. R. Stratton (2007). "Patterns of somatic mutation in human cancer genomes." Nature **446**(7132): 153-158.

Gross, J. M. and D. Yee (2002). "How does the estrogen receptor work?" Breast Cancer Res **4**(2): 62-64.

Growney, J. D., H. Shigematsu, Z. Li, B. H. Lee, J. Adelsperger, R. Rowan, D. P. Curley, J. L. Kutok, K. Akashi, I. R. Williams, N. A. Speck and D. G. Gilliland (2005). "Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype." Blood **106**(2): 494-504.

Gu, T. L., T. L. Goetz, B. J. Graves and N. A. Speck (2000). "Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1)." Mol Cell Biol **20**(1): 91-103.

Guo, W., Z. Keckesova, J. L. Donaher, T. Shibue, V. Tischler, F. Reinhardt, S. Itzkovitz, A. Noske, U. Zürrer-Härdi, G. Bell, W. L. Tam, S. A. Mani, A. van Oudenaarden and R. A. Weinberg (2012). "Slug and Sox9 cooperatively determine the mammary stem cell state." Cell **148**(5): 1015-1028.

Guy, C. T., R. D. Cardiff and W. J. Muller (1992). "Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease." Mol Cell Biol **12**(3): 954-961.

Haber, D. A. and J. Settleman (2007). "Cancer: drivers and passengers." Nature **446**(7132): 145-146.

Hall, B., A. Limaye and A. B. Kulkarni (2009). "Overview: generation of gene knockout mice." Curr Protoc Cell Biol **Chapter 19**: Unit 19.12 19.12.11-17.

Hanahan, D. (2022). "Hallmarks of Cancer: New Dimensions." Cancer Discov **12**(1): 31-46.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.

Hansen, A. R. and P. L. Bedard (2013). "Clinical application of high-throughput genomic technologies for treatment selection in breast cancer." Breast Cancer Res **15**(5): R97.

Haque, M. M. and K. V. Desai (2019). "Pathways to Endocrine Therapy Resistance in Breast Cancer." Frontiers in endocrinology **10**: 573-573.

Harada, N., Y. Tamai, T. Ishikawa, B. Sauer, K. Takaku, M. Oshima and M. M. Taketo (1999). "Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene." Embo j **18**(21): 5931-5942.

He, S., Y. Lu, X. Liu, X. Huang, E. T. Keller, C. N. Qian and J. Zhang (2015). "Wnt3a: functions and implications in cancer." Chin J Cancer **34**(12): 554-562.

Herschkowitz, J. I., K. Simin, V. J. Weigman, I. Mikaelian, J. Usary, Z. Hu, K. E. Rasmussen, L. P. Jones, S. Assefnia, S. Chandrasekharan, M. G. Backlund, Y. Yin, A. I. Khramtsov, R. Bastein, J. Quackenbush, R. I. Glazer, P. H. Brown, J. E. Green, L. Kopelovich, P. A. Furth, J. P. Palazzo, O. I. Olopade, P. S. Bernard, G. A. Churchill, T. Van Dyke and C. M. Perou (2007). "Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors." Genome Biol **8**(5): R76.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke and R. A. Young (2013). "Super-enhancers in the control of cell identity and disease." Cell **155**(4): 934-947.

Hoi, C. S., S. E. Lee, S. Y. Lu, D. J. McDermitt, K. M. Osorio, C. M. Piskun, R. M. Peters, R. Paus and T. Tumbar (2010). "Runx1 directly promotes proliferation of hair follicle stem cells and epithelial tumor formation in mouse skin." Mol Cell Biol **30**(10): 2518-2536.

Hong, D., A. J. Fritz, K. H. Finstad, M. P. Fitzgerald, A. Weinheimer, A. L. Viens, J. Ramsey, J. L. Stein, J. B. Lian and G. S. Stein (2018). "Suppression of Breast Cancer Stem Cells and Tumor Growth by the RUNX1 Transcription Factor." Mol Cancer Res **16**(12): 1952-1964.

Hong, D., T. L. Messier, C. E. Tye, J. R. Dobson, A. J. Fritz, K. R. Sikora, G. Browne, J. L. Stein, J. B. Lian and G. S. Stein (2017). "Runx1 stabilizes the mammary epithelial cell phenotype and prevents epithelial to mesenchymal transition." Oncotarget **8**(11): 17610-17627.

Hou, M. F., P. M. Chen and P. Y. Chu (2018). "LGR5 overexpression confers poor relapse-free survival in breast cancer patients." BMC Cancer **18**(1): 219.

Hou, Y., H. Nitta, L. Wei, P. M. Banks, B. Portier, A. V. Parwani and Z. Li (2017). "HER2 intratumoral heterogeneity is independently associated with incomplete response to anti-HER2 neoadjuvant chemotherapy in HER2-positive breast carcinoma." Breast Cancer Res Treat **166**(2): 447-457.

Housman, G., S. Byler, S. Heerboth, K. Lapinska, M. Longacre, N. Snyder and S. Sarkar (2014). "Drug resistance in cancer: an overview." Cancers **6**(3): 1769-1792.

Huang, S. P., Y. H. Lan, T. L. Lu, J. B. Pao, T. Y. Chang, H. Z. Lee, W. H. Yang, C. J. Hsieh, L. M. Chen, L. C. Huang, W. C. Ting and B. Y. Bao (2011). "Clinical significance of runt-related transcription factor 1 polymorphism in prostate cancer." BJU Int **107**(3): 486-492.

Hwang-Verslues, W. W., K. J. Chang, E. Y. Lee and W. H. Lee (2008). "Breast cancer stem cells and tumor suppressor genes." J Formos Med Assoc **107**(10): 751-766.

Ichikawa, M., S. Goyama, T. Asai, M. Kawazu, M. Nakagawa, M. Takeshita, S. Chiba, S. Ogawa and M. Kurokawa (2008). "AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis." J Immunol **180**(7): 4402-4408.

Ikeda, T., H. Jinno, A. Matsu, S. Masamura and M. Kitajima (2002). "The role of neoadjuvant chemotherapy for breast cancer treatment." Breast Cancer **9**(1): 8-14.

Illa-Bochaca, I., R. Fernandez-Gonzalez, D. N. Shelton, B. E. Welm, C. Ortiz-de-Solorzano and M. H. Barcellos-Hoff (2010). "Limiting-dilution transplantation assays in mammary stem cell studies." Methods Mol Biol **621**: 29-47.

Ingraham, C. A., G. C. Park, H. P. Makarenkova and K. L. Crossin (2011). "Matrix metalloproteinase (MMP)-9 induced by Wnt signaling increases the proliferation and migration of embryonic neural stem cells at low O2 levels." The Journal of biological chemistry **286**(20): 17649-17657.

Inoue, K., S. Ozaki, K. Ito, T. Iseda, S. Kawaguchi, M. Ogawa, S. C. Bae, N. Yamashita, S. Itohara, N. Kudo and Y. Ito (2003). "Runx3 is essential for the target-specific axon pathfinding of trkc-expressing dorsal root ganglion neurons." Blood Cells Mol Dis **30**(2): 157-160.

Institute of Medicine Committee on Cancer Control in, L.-a. M.-I. C. (2007). The National Academies Collection: Reports funded by National Institutes of Health. Cancer Control Opportunities in Low- and Middle-Income Countries. F. A. Sloan and H. Gelband. Washington (DC), National Academies Press (US)

Iqbal, N. and N. Iqbal (2014). "Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications." Molecular biology international **2014**: 852748-852748.

Ito, K., K. i. Inoue, S. C. Bae and Y. Ito (2009). "Runx3 expression in gastrointestinal tract epithelium: resolving the controversy." Oncogene **28**(10): 1379-1384.

Ito, Y. (2008). "RUNX genes in development and cancer: regulation of viral gene expression and the discovery of RUNX family genes." Adv Cancer Res **99**: 33-76.

Jeon, J., S. Nim, J. Teyra, A. Datti, J. L. Wrana, S. S. Sidhu, J. Moffat and P. M. Kim (2014). "A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening." Genome medicine **6**(7): 57-57.

Jeter, C. R., T. Yang, J. Wang, H. P. Chao and D. G. Tang (2015). "Concise Review: NANOG in Cancer Stem Cells and Tumor Development: An Update and Outstanding Questions." Stem Cells **33**(8): 2381-2390.

Ji, X., Y. Lu, H. Tian, X. Meng, M. Wei and W. C. Cho (2019). "Chemoresistance mechanisms of breast cancer and their countermeasures." Biomedicine & Pharmacotherapy **114**: 108800.

Jiang, H. and H. Li (2021). "Prognostic values of tumoral MMP2 and MMP9 overexpression in breast cancer: a systematic review and meta-analysis." BMC Cancer **21**(1): 149.

Johnston, S. R., G. Saccani-Jotti, I. E. Smith, J. Salter, J. Newby, M. Coppen, S. R. Ebbs and M. Dowsett (1995). "Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer." Cancer Res **55**(15): 3331-3338.

Jordan, V. C. and B. W. O'Malley (2007). "Selective estrogen-receptor modulators and antihormonal resistance in breast cancer." J Clin Oncol **25**(36): 5815-5824.

Joseph, C., M. Alsaleem, N. Orah, P. L. Narasimha, I. M. Miligy, S. Kurozumi, I. O. Ellis, N. P. Mongan, A. R. Green and E. A. Rakha (2020). "Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival." Breast cancer research and treatment **182**(2): 267-282.

Kadota, M., H. H. Yang, B. Gomez, M. Sato, R. J. Clifford, D. Meerzaman, B. K. Dunn, L. M. Wakefield and M. P. Lee (2010). "Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines." PloS one **5**(2): e9201-e9201.

Kakugawa, S., P. F. Langton, M. Zebisch, S. Howell, T. H. Chang, Y. Liu, T. Feizi, G. Bineva, N. O'Reilly, A. P. Snijders, E. Y. Jones and J. P. Vincent (2015). "Notum deacylates Wnt proteins to suppress signalling activity." Nature **519**(7542): 187-192.

Kalager, M., T. Haldorsen, M. Bretthauer, G. Hoff, S. O. Thoresen and H.-O. Adami (2009). "Improved breast cancer survival following introduction of an organized mammography screening program among both screened and unscreened women: a population-based cohort study." Breast cancer research : BCR **11**(4): R44-R44.

Karayiannakis, A. J., L. Nakopoulou, H. Gakiopoulou, A. Keramopoulos, P. S. Davaris and M. Pignatelli (2001). "Expression patterns of beta-catenin in in situ and invasive breast cancer." Eur J Surg Oncol **27**(1): 31-36.

Karn, T., L. Pusztai, U. Holtrich, T. Iwamoto, C. Y. Shiang, M. Schmidt, V. Müller, C. Solbach, R. Gaetje, L. Hanker, A. Ahr, C. Liedtke, E. Ruckhäberle, M. Kaufmann and A. Rody (2011). "Homogeneous datasets of triple negative breast cancers enable the identification of novel prognostic and predictive signatures." PloS one **6**(12): e28403-e28403.

Kendrick, H., J. L. Regan, F. A. Magnay, A. Grigoriadis, C. Mitsopoulos, M. Zvelebil and M. J. Smalley (2008). "Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate." BMC Genomics **9**: 591.

Kesson, E. M., G. M. Allardice, W. D. George, H. J. Burns and D. S. Morrison (2012). "Effects of multidisciplinary team working on breast cancer survival: retrospective, comparative, interventional cohort study of 13 722 women." Bmj **344**: e2718.

Khaled, W. T., S. Choon Lee, J. Stingl, X. Chen, H. Raza Ali, O. M. Rueda, F. Hadi, J. Wang, Y. Yu, S. F. Chin, M. Stratton, A. Futreal, N. A. Jenkins, S. Aparicio, N. G. Copeland, C. J. Watson, C. Caldas and P. Liu (2015). "BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells." Nat Commun **6**: 5987.

Khoury, T., F. O. Ademuyiwa, R. Chandrasekhar, M. Jabbour, A. Deleo, S. Ferrone, Y. Wang and X. Wang (2012). "Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to neoadjuvant chemotherapy." Mod Pathol **25**(3): 388-397.

Kim, W., D. A. Barron, R. San Martin, K. S. Chan, L. L. Tran, F. Yang, S. J. Ressler and D. R. Rowley (2014). "RUNX1 is essential for mesenchymal stem cell proliferation and myofibroblast differentiation." Proc Natl Acad Sci U S A **111**(46): 16389-16394.

Klupp, F., L. Neumann, C. Kahlert, J. Diers, N. Halama, C. Franz, T. Schmidt, M. Koch, J. Weitz, M. Schneider and A. Ulrich (2016). "Serum MMP7, MMP10 and MMP12 level as negative prognostic markers in colon cancer patients." BMC Cancer **16**: 494.

Koboldt, D. C., R. S. Fulton, M. D. McLellan, H. Schmidt, J. Kalicki-Veizer, J. F. McMichael, L. L. Fulton, D. J. Dooling, L. Ding, E. R. Mardis, R. K. Wilson, A. Ally, M. Balasundaram, Y. S. N. Butterfield, R. Carlsen, C. Carter, A. Chu, E. Chuah, H.-J. E. Chun, R. J. N. Coope, N. Dhalla, R. Guin, C. Hirst, M. Hirst, R. A. Holt, D. Lee, H. I. Li, M. Mayo, R. A. Moore, A. J. Mungall, E. Pleasance, A. Gordon Robertson, J. E. Schein, A. Shafiei, P. Sipahimalani, J. R. Slobodan, D. Stoll, A. Tam, N. Thiessen, R. J. Varhol, N. Wye, T. Zeng, Y. Zhao, I. Birol, S. J. M. Jones, M. A. Marra, A. D. Cherniack, G. Saksena, R. C. Onofrio, N. H. Pho, S. L. Carter, S. E. Schumacher, B. Tabak, B. Hernandez, J. Gentry, H. Nguyen, A. Crenshaw, K. Ardlie, R. Beroukhim, W. Winckler, G. Getz, S. B. Gabriel, M. Meyerson, L. Chin, P. J. Park, R. Kucherlapati, K. A. Hoadley, J. Todd Auman, C. Fan, Y. J. Turman, Y. Shi, L. Li, M. D. Topal, X. He, H.-H. Chao, A. Prat, G. O. Silva, M. D. Iglesia, W. Zhao, J. Usary, J. S. Berg, M. Adams, J. Booker, J. Wu, A. Gulabani, T. Bodenheimer, A. P. Hoyle, J. V. Simons, M. G. Soloway, L. E. Mose, S. R. Jefferys, S. Balu, J. S. Parker, D. Neil Hayes, C. M. Perou, S. Malik, S. Mahurkar, H. Shen, D. J. Weisenberger, T. Triche Jr, P. H. Lai, M. S. Bootwalla, D. T. Maglinte, B. P. Berman, D. J. Van Den Berg, S. B. Baylin, P. W. Laird, C. J. Creighton, L. A. Donehower, G. Getz, M. Noble, D. Voet, G. Saksena, N. Gehlenborg, D. DiCara, J. Zhang, H. Zhang, C.-J. Wu, S. Yingchun Liu, M. S. Lawrence, L. Zou, A. Sivachenko, P. Lin, P. Stojanov, R. Jing, J. Cho, R. Sinha, R. W. Park, M.-D. Nazaire, J. Robinson, H. Thorvaldsdottir, J. Mesirov, P. J. Park, L. Chin, S. Reynolds, R. B. Kreisberg, B. Bernard, R. Bressler, T. Erkkila, J. Lin, V. Thorsson, W. Zhang, I. Shmulevich, G. Ciriello, N. Weinhold, N. Schultz, J. Gao, E. Cerami, B. Gross, A. Jacobsen, R. Sinha, B. Arman Aksoy, Y. Antipin, B. Reva, R. Shen, B. S. Taylor, M. Ladanyi, C. Sander, P. Anur, P. T. Spellman, Y. Lu, W. Liu, R. R. G. Verhaak, G. B. Mills, R. Akbani, N. Zhang, B. M. Broom, T. D. Casasent, C. Wakefield, A. K. Unruh, K. Baggerly, K. Coombes, J. N. Weinstein, D. Haussler, C. C. Benz, J. M. Stuart, S. C. Benz, J. Zhu, C. C. Szeto, G. K. Scott, C. Yau, E. O. Paull, D. Carlin, C. Wong, A. Sokolov, J. Thusberg, S. Mooney, S. Ng, T. C. Goldstein, K. Ellrott, M. Grifford, C. Wilks, S. Ma, B. Craft, C. Yan, Y. Hu, D. Meerzaman, J. M. Gastier-Foster, J. Bowen, N. C. Ramirez, A. D. Black, R. E. Pyatt, P. White, E. J. Zmuda, J. Frick, T. M. Lichtenberg, R. Brookens, M. M. George, M. A. Gerken, H. A. Harper, K. M. Leraas, L. J. Wise, T. R. Tabler, C. McAllister, T. Barr, M. Hart-Kothari, K. Tarvin, C. Saller, G. Sandusky, C. Mitchell, M. V. Iacocca, J. Brown, B. Rabeno, C. Czerwinski, N. Petrelli, O. Dolzhansky, M. Abramov, O. Voronina, O. Potapova, J. R. Marks, W. M. Suchorska, D. Murawa, W. Kycler, M. Ibbs, K. Korski, A. Spychała, P. Murawa, J. J. Brzeziński, H. Perz, R. Łaźniak, M. Teresiak, H. Tatka, E. Leporowska, M. Bogusz-Czerniewicz, J. Malicki, A. Mackiewicz, M. Wiznerowicz, X. Van Le, B. Kohl, N. Viet Tien, R. Thorp, N. Van Bang, H. Sussman, B. Duc Phu, R. Hajek, N. Phi Hung, T. Viet The Phuong, H. Quyet Thang, K. Zaki Khan, R. Penny, D. Mallery, E. Curley, C. Shelton, P. Yena, J. N. Ingle, F. J. Couch, W. L. Lingle, T. A. King, A. Maria Gonzalez-Angulo, G. B. Mills, M. D. Dyer, S. Liu, X. Meng, M. Patangan, N. The Cancer Genome Atlas, L.

Genome sequencing centres: Washington University in St, B. C. C. A. Genome characterization centres, I. Broad, Brigham, H. Women's, S. Harvard Medical, C. H. University of North Carolina, H. University of Southern California/Johns, M. Genome data analysis: Baylor College of, B. Institute for Systems, C. Memorial Sloan-Kettering Cancer, H. Oregon, U. Science, M. D. A. C. C. The University of Texas, S. C. B. I. University of California, Nci, R. Biospecimen core resource: Nationwide Children's Hospital Biospecimen Core, A.-I. Tissue source sites, Christiana, Cureline, C. Duke University Medical, C. The Greater Poland Cancer, Ilsbio, C. International Genomics, C. Mayo, Mskcc and M. D. A. C. Center (2012). "Comprehensive molecular portraits of human breast tumours." Nature **490**(7418): 61-70.

Kohler, B. A., R. L. Sherman, N. Howlader, A. Jemal, A. B. Ryerson, K. A. Henry, F. P. Boscoe, K. A. Cronin, A. Lake, A.-M. Noone, S. J. Henley, C. R. Eheman, R. N. Anderson and L. Penberthy (2015). "Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State." Journal of the National Cancer Institute **107**(6): djv048-djv048.

Komori, T., H. Yagi, S. Nomura, A. Yamaguchi, K. Sasaki, K. Deguchi, Y. Shimizu, R. T. Bronson, Y. H. Gao, M. Inada, M. Sato, R. Okamoto, Y. Kitamura, S. Yoshiki and T. Kishimoto (1997). "Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts." Cell **89**(5): 755-764.

Kong, Y., X. Ou, X. Li, Y. Zeng, G. Gao, N. Lyu and P. Liu (2020). "LGR6 Promotes Tumor Proliferation and Metastasis through Wnt/β-Catenin Signaling in Triple-Negative Breast Cancer." Mol Ther Oncolytics **18**: 351-359.

Kordon, E. C. and G. H. Smith (1998). "An entire functional mammary gland may comprise the progeny from a single cell." Development **125**(10): 1921-1930.

Kouros-Mehr, H. and Z. Werb (2006). "Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis." Dev Dyn **235**(12): 3404-3412.

Kramer, K., J. Wu and D. L. Crowe (2016). "Tumor suppressor control of the cancer stem cell niche." Oncogene **35**(32): 4165-4178.

Kulkarni, M., T. Z. Tan, N. B. Syed Sulaiman, J. M. Lamar, P. Bansal, J. Cui, Y. Qiao and Y. Ito (2018). "RUNX1 and RUNX3 protect against YAP-mediated EMT, stem-ness and shorter survival outcomes in breast cancer." Oncotarget **9**(18): 14175-14192.

Kurokawa, M., T. Tanaka, K. Tanaka, N. Hirano, S. Ogawa, K. Mitani, Y. Yazaki and H. Hirai (1996). "A conserved cysteine residue in the runt homology domain of AML1 is required for the DNA binding ability and the transforming activity on fibroblasts." J Biol Chem **271**(28): 16870-16876.

Lamb, R., M. P. Ablett, K. Spence, G. Landberg, A. H. Sims and R. B. Clarke (2013). "Wnt pathway activity in breast cancer sub-types and stem-like cells." PLoS One **8**(7): e67811.

Lee, H. J., J. K. Myung, H. S. Kim, D. H. Lee, H. S. Go, J. H. Choi, H. M. Koh, S. J. Lee and B. Jang (2021). "Expression of LGR5 in mammary myoepithelial cells and in triple-negative breast cancers." Sci Rep **11**(1): 17750.

Lee, M. A., J. H. Park, S. Y. Rhyu, S. T. Oh, W. K. Kang and H. N. Kim (2014). "Wnt3a expression is associated with MMP-9 expression in primary tumor and metastatic site in recurrent or stage IV colorectal cancer." BMC Cancer **14**: 125.

Levanon, D., Y. Bernstein, V. Negreanu, K. R. Bone, A. Pozner, R. Eilam, J. Lotem, O. Brenner and Y. Groner (2011). "Absence of Runx3 expression in normal gastrointestinal epithelium calls into question its tumour suppressor function." EMBO molecular medicine **3**(10): 593-604.

Levanon, D., D. Bettoun, C. Harris-Cerruti, E. Woolf, V. Negreanu, R. Eilam, Y. Bernstein, D. Goldenberg, C. Xiao, M. Fliegauf, E. Kremer, F. Otto, O. Brenner, A. Lev-Tov and Y. Groner (2002). "The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons." The EMBO journal **21**(13): 3454-3463.

Levanon, D., O. Brenner, V. Negreanu, D. Bettoun, E. Woolf, R. Eilam, J. Lotem, U. Gat, F. Otto, N. Speck and Y. Groner (2001). "Spatial and temporal expression pattern of Runx3 (Aml2) and Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis." Mechanisms of Development **109**(2): 413-417.

Levanon, D., O. Brenner, F. Otto and Y. Groner (2003). "Runx3 knockouts and stomach cancer." EMBO reports **4**(6): 560-564.

Levanon, D., V. Negreanu, J. Lotem, K. R. Bone, O. Brenner, D. Leshkowitz and Y. Groner (2014). "Transcription factor Runx3 regulates interleukin-15-dependent natural killer cell activation." Molecular and cellular biology **34**(6): 1158-1169.

Li, M., L. Zhou, S. Li, L. Fang, L. Yang, X. Wu, C. Yang, Y. Bao, S. Lan, Z. Tong, S. Zheng, B. Tang, E. Zeng, S. Xie, C. Chen and T. Hong (2021). "MMP12 is a potential therapeutic target for Adamantinomatous craniopharyngioma: Conclusions from bioinformatics analysis and in vitro experiments." Oncol Lett **22**(1): 536.

Li, Q., Q. Lai, C. He, Y. Fang, Q. Yan, Y. Zhang, X. Wang, C. Gu, Y. Wang, L. Ye, L. Han, X. Lin, J. Chen, J. Cai, A. Li and S. Liu (2019). "RUNX1 promotes tumour metastasis by activating the Wnt/β-catenin signalling pathway and EMT in colorectal cancer." J Exp Clin Cancer Res **38**(1): 334.

Li, Q. L., K. Ito, C. Sakakura, H. Fukamachi, K. Inoue, X. Z. Chi, K. Y. Lee, S. Nomura, C. W. Lee, S. B. Han, H. M. Kim, W. J. Kim, H. Yamamoto, N. Yamashita, T. Yano, T. Ikeda, S. Itohara, J. Inazawa, T. Abe, A. Hagiwara, H. Yamagishi, A. Ooe, A. Kaneda, T. Sugimura, T. Ushijima, S. C. Bae and Y. Ito (2002). "Causal relationship between the loss of RUNX3 expression and gastric cancer." Cell **109**(1): 113-124.

Li, S., S. Li, Y. Sun and L. Li (2014). "The expression of β-catenin in different subtypes of breast cancer and its clinical significance." Tumour Biol **35**(8): 7693-7698.

Lim, E., D. Wu, B. Pal, T. Bouras, M. L. Asselin-Labat, F. Vaillant, H. Yagita, G. J. Lindeman, G. K. Smyth and J. E. Visvader (2010). "Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways." Breast Cancer Res **12**(2): R21.

Lin, E. Y., J. G. Jones, P. Li, L. Zhu, K. D. Whitney, W. J. Muller and J. W. Pollard (2003). "Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases." Am J Pathol **163**(5): 2113-2126.

Lin, S. Y., W. Xia, J. C. Wang, K. Y. Kwong, B. Spohn, Y. Wen, R. G. Pestell and M. C. Hung (2000). "Beta-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression." Proc Natl Acad Sci U S A **97**(8): 4262-4266.

Liu, F., A. Pawliwec, Z. Feng, Z. Yasruel, J. J. Lebrun and S. Ali (2015). "Prolactin/Jak2 directs apical/basal polarization and luminal linage maturation of mammary epithelial cells through regulation of the Erk1/2 pathway." Stem Cell Res **15**(2): 376-383.

Liu, Y. N., W. W. Lee, C. Y. Wang, T. H. Chao, Y. Chen and J. H. Chen (2005). "Regulatory mechanisms controlling human E-cadherin gene expression." Oncogene **24**(56): 8277-8290.

Lombardo, Y., A. de Giorgio, C. R. Coombes, J. Stebbing and L. Castellano (2015). "Mammosphere formation assay from human breast cancer tissues and cell lines." J Vis Exp(97).

Lu, X., S. J. Mazur, T. Lin, E. Appella and Y. Xu (2014). "The pluripotency factor nanog promotes breast cancer tumorigenesis and metastasis." Oncogene **33**(20): 2655-2664.

Luche, H., O. Weber, T. Nageswara Rao, C. Blum and H. J. Fehling (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**(1): 43-53.

Luo, Y., Y. Zhang, G. Miao, Y. Zhang, Y. Liu and Y. Huang (2019). "Runx1 regulates osteogenic differentiation of BMSCs by inhibiting adipogenesis through Wnt/β-catenin pathway." Arch Oral Biol **97**: 176-184.

Lustig, B., B. Jerchow, M. Sachs, S. Weiler, T. Pietsch, U. Karsten, M. van de Wetering, H. Clevers, P. M. Schlag, W. Birchmeier and J. Behrens (2002). "Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors." Mol Cell Biol **22**(4): 1184-1193.

Lv, X., L. E. Dobrolecki, Y. Ding, J. M. Rosen, M. T. Lewis and X. Chen (2020). "Orthotopic Transplantation of Breast Tumors as Preclinical Models for Breast Cancer." J Vis Exp(159).

Lyons, T. G. (2019). "Targeted Therapies for Triple-Negative Breast Cancer." Current Treatment Options in Oncology **20**(11): 82.

Mandelblatt, J., N. van Ravesteyn, C. Schechter, Y. Chang, A.-T. Huang, A. M. Near, H. de Koning and A. Jemal (2013). "Which strategies reduce breast cancer mortality most? Collaborative modeling of optimal screening, treatment, and obesity prevention." Cancer **119**(14): 2541-2548.

Matsumura, T., A. Nakamura-Ishizu, S. Muddineni, D. Q. Tan, C. Q. Wang, K. Tokunaga, R. Tirado-Magallanes, S. Sian, T. Benoukraf, T. Okuda, N. Asou, M. Matsuoka, M. Osato and T. Suda (2020). "Hematopoietic stem cells acquire survival advantage by loss of RUNX1 methylation identified in familial leukemia." Blood **136**(17): 1919-1932.

Matsuo, J., S. Kimura, A. Yamamura, C. P. Koh, M. Z. Hossain, D. L. Heng, K. Kohu, D. C. Voon, H. Hiai, M. Unno, J. B. So, F. Zhu, S. Srivastava, M. Teh, K. G. Yeoh, M. Osato and Y. Ito (2017). "Identification of Stem Cells in the Epithelium of the Stomach Corpus and Antrum of Mice." Gastroenterology **152**(1): 218-231.e214.

McCann, K. E., S. A. Hurvitz and N. McAndrew (2019). "Advances in Targeted Therapies for Triple-Negative Breast Cancer." Drugs **79**(11): 1217-1230.

McDonald, L., N. Ferrari, A. Terry, M. Bell, Z. M. Mohammed, C. Orange, A. Jenkins, W. J. Muller, B. A. Gusterson, J. C. Neil, J. Edwards, J. S. Morris, E. R. Cameron and K. Blyth (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**(5): 525-534.

McKillop, A. J., J. Edwards, E. Johnson, S. Mason, E. R. Cameron and K. Blyth (2017). "Investigating RUNX1 and RUNX2 in prostate cancer." Journal of Clinical Oncology **35**(6_suppl): 232-232.

McNally, S. and T. Stein (2017). "Overview of Mammary Gland Development: A Comparison of Mouse and Human." Methods Mol Biol **1501**: 1-17.

Meisel, J. L., V. A. Venur, M. Gnant and L. Carey (2018). "Evolution of Targeted Therapy in Breast Cancer: Where Precision Medicine Began." Am Soc Clin Oncol Educ Book **38**: 78-86.

Merid, S. K., D. Goranskaya and A. Alexeyenko (2014). "Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis." BMC bioinformatics **15**(1): 308-308.

Merlo, G. R., D. Graus-Porta, N. Cella, B. M. Marte, D. Taverna and N. E. Hynes (1996). "Growth, differentiation and survival of HC11 mammary epithelial cells: diverse effects of receptor tyrosine kinase-activating peptide growth factors." Eur J Cell Biol **70**(2): 97-105.

Mevel, R., J. E. Draper, A. L. M. Lie, V. Kouskoff and G. Lacaud (2019). "RUNX transcription factors: orchestrators of development." Development **146**(17).

Miller, D. H., D. X. Jin, E. S. Sokol, J. R. Cabrera, D. A. Superville, R. A. Gorelov, C. Kuperwasser and P. B. Gupta (2018). "BCL11B Drives Human Mammary Stem Cell Self-Renewal In Vitro by Inhibiting Basal Differentiation." Stem Cell Reports **10**(3): 1131-1145.

Mittendorf, E. A., A. V. Philips, F. Meric-Bernstam, N. Qiao, Y. Wu, S. Harrington, X. Su, Y. Wang, A. M. Gonzalez-Angulo, A. Akcakanat, A. Chawla, M. Curran, P. Hwu, P. Sharma, J. K. Litton, J. J. Molldrem and G. Alatrash (2014). "PD-L1 expression in triple-negative breast cancer." Cancer Immunol Res **2**(4): 361-370.

Miyagawa, K., C. Sakakura, S. Nakashima, T. Yoshikawa, S. Kin, Y. Nakase, K. Ito, H. Yamagishi, H. Ida, S. Yazumi, T. Chiba, Y. Ito and A. Hagiwara (2006). "Down-regulation of RUNX1, RUNX3 and CBFbeta in hepatocellular carcinomas in an early stage of hepatocarcinogenesis." Anticancer Res **26**(5b): 3633-3643.

Molyneux, G., F. C. Geyer, F. A. Magnay, A. McCarthy, H. Kendrick, R. Natrajan, A. Mackay, A. Grigoriadis, A. Tutt, A. Ashworth, J. S. Reis-Filho and M. J. Smalley (2010). "BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells." Cell Stem Cell **7**(3): 403-417.

Morrison, B. and M. L. Cutler (2009). "Mouse Mammary Epithelial Cells form Mammospheres During Lactogenic Differentiation." J Vis Exp(32).

Moumen, M., A. Chiche, C. Decraene, V. Petit, A. Gandarillas, M. A. Deugnier, M. A. Glukhova and M. M. Faraldo (2013). "Myc is required for β-catenin-mediated mammary stem cell amplification and tumorigenesis." Mol Cancer **12**(1): 132.

Moumen, M., A. Chiche, M. A. Deugnier, V. Petit, A. Gandarillas, M. A. Glukhova and M. M. Faraldo (2012). "The proto-oncogene Myc is essential for mammary stem cell function." Stem Cells **30**(6): 1246-1254.

Mundlos, S., F. Otto, C. Mundlos, J. B. Mulliken, A. S. Aylsworth, S. Albright, D. Lindhout, W. G. Cole, W. Henn, J. H. M. Knoll, M. J. Owen, R. Mertelsmann, B. U. Zabel and B. R. Olsen (1997). "Mutations Involving the Transcription Factor CBFA1 Cause Cleidocranial Dysplasia." Cell **89**(5): 773-779.

Musgrove, E. A. and R. L. Sutherland (2009). "Biological determinants of endocrine resistance in breast cancer." Nature Reviews Cancer **9**(9): 631-643.

Nahta, R., L. X. Yuan, B. Zhang, R. Kobayashi and F. J. Esteva (2005). "Insulin-like growth factor-I receptor/human epidermal growth factor receptor 2 heterodimerization contributes to trastuzumab resistance of breast cancer cells." Cancer Res **65**(23): 11118-11128.

Naillat, F., W. Yan, R. Karjalainen, A. Liakhovitskaia, A. Samoylenko, Q. Xu, Z. Sun, B. Shen, A. Medvinsky, S. Quaggin and S. J. Vainio (2015). "Identification of the genes regulated by Wnt-4, a critical signal for commitment of the ovary." Exp Cell Res **332**(2): 163-178.

Nelson, R. (October 2019). "Increased Cancer Incidence May Reflect 'Overdiagnosis'." September 2020, from https://www.medscape.com/viewarticle/919503.

Newman, L. A., J. S. Reis-Filho, M. Morrow, L. A. Carey and T. A. King (2015). "The 2014 Society of Surgical Oncology Susan G. Komen for the Cure Symposium: triple-negative breast cancer." Ann Surg Oncol **22**(3): 874-882.

Ng, K. T., X. Qi, K. L. Kong, B. Y. Cheung, C. M. Lo, R. T. Poon, S. T. Fan and K. Man (2011). "Overexpression of matrix metalloproteinase-12 (MMP-12) correlates with poor prognosis of hepatocellular carcinoma." Eur J Cancer **47**(15): 2299-2305.

Niini, T., J. Kanerva, K. Vettenranta, U. M. Saarinen-Pihkala and S. Knuutila (2000). "AML1 gene amplification: a novel finding in childhood acute lymphoblastic leukemia." Haematologica **85**(4): 362-366.

Niit, M., M. Geletu, Z. Taha, R. Arulanandam, J. Cass, V. Hoskin, B. Elliott, P. Gunning and L. Raptis (2019). "Regulation of Differentiation of HC11 Mouse Breast Epithelial Cells by the Signal Transducer and Activator of Transcription-3." Anticancer Res **39**(6): 2749-2756.

Nik-Zainal, S., H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L. B. Alexandrov, S. Martin, D. C. Wedge, P. Van Loo, Y. S. Ju, M. Smid, A. B. Brinkman, S. Morganella, M. R. Aure, O. C. Lingjærde, A. Langerød, M. Ringnér, S.-M. Ahn, S. Boyault, J. E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J. A. Foekens, M. Gerstung, G. K. J. Hooijer, S. J. Jang, D. R. Jones, H.-Y. Kim, T. A. King, S. Krishnamurthy, H. J. Lee, J.-Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporté, X. Pivot, C. A. Purdie, K. Raine, K. Ramakrishnan, F. G. Rodríguez-González, G. Romieu, A. M. Sieuwerts, P. T. Simpson, R. Shepherd, L. Stebbings, O. A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G. G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. v. t. Veer, A. Tutt, S. Knappskog, B. K. T. Tan, J. Jonkers, Å. Borg, N. T. Ueno, C. Sotiriou, A. Viari, P. A. Futreal, P. J. Campbell, P. N. Span, S. Van Laere, S. R. Lakhani, J. E. Eyfjord, A. M. Thompson, E. Birney, H. G. Stunnenberg, M. J. van de Vijver, J. W. M. Martens, A.-L. Børresen-Dale, A. L. Richardson, G. Kong, G. Thomas and M. R. Stratton (2016). "Landscape of somatic mutations in 560 breast cancer whole-genome sequences." Nature **534**(7605): 47-54.

North, T., T. L. Gu, T. Stacy, Q. Wang, L. Howard, M. Binder, M. Marín-Padilla and N. A. Speck (1999). "Cbfa2 is required for the formation of intra-aortic hematopoietic clusters." Development **126**(11): 2563-2575.

Nottingham, W. T., A. Jarratt, M. Burgess, C. L. Speck, J. F. Cheng, S. Prabhakar, E. M. Rubin, P. S. Li, J. Sloane-Stanley, A. S. J. Kong and M. F. de Bruijn (2007). "Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer." Blood **110**(13): 4188-4197.

Nusse, R. and H. Clevers (2017). "Wnt/ß-Catenin Signaling, Disease, and Emerging Therapeutic Modalities." Cell **169**(6): 985-999.

Okada, H., T. Watanabe, M. Niki, H. Takano, N. Chiba, N. Yanai, K. Tani, H. Hibino, S. Asano, M. L. Mucenski, Y. Ito, T. Noda and M. Satake (1998). "AML1(-/-) embryos do not express certain hematopoiesis-related gene transcripts including those of the PU.1 gene." Oncogene **17**(18): 2287-2293.

Okuda, T., J. van Deursen, S. W. Hiebert, G. Grosveld and J. R. Downing (1996). "AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis." Cell **84**(2): 321-330.

Onitilo, A. A., J. M. Engel, R. T. Greenlee and B. N. Mukesh (2009). "Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival." Clin Med Res **7**(1-2): 4-13.

Ordaz-Ramos, A., V. H. Rosales-Gallegos, J. Melendez-Zajgla, V. Maldonado and K. Vazquez-Santillan (2021). "The Role of LGR4 (GPR48) in Normal and Cancer Processes." Int J Mol Sci **22**(9).

Osato, M. (2004). "Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia." Oncogene **23**(24): 4284-4296.

Osorio, K. M., K. C. Lilja and T. Tumbar (2011). "Runx1 modulates adult hair follicle stem cell emergence and maintenance from distinct embryonic skin compartments." J Cell Biol **193**(1): 235-250.

Otto, F., A. P. Thornell, T. Crompton, A. Denzel, K. C. Gilmour, I. R. Rosewell, G. W. Stamp, R. S. Beddington, S. Mundlos, B. R. Olsen, P. B. Selby and M. J. Owen (1997). "Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development." Cell **89**(5): 765-771.

Owens, T. W., R. L. Rogers, S. Best, A. Ledger, A. M. Mooney, A. Ferguson, P. Shore, A. Swarbrick, C. J. Ormandy, P. T. Simpson, J. S. Carroll, J. Visvader and M. J. Naylor (2014). "Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer." Cancer Res **74**(18): 5277-5286.

Pan, H., R. Gray, J. Braybrooke, C. Davies, C. Taylor, P. McGale, R. Peto, K. I. Pritchard, J. Bergh, M. Dowsett, D. F. Hayes and Ebctcg (2017). "20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years." The New England journal of medicine **377**(19): 1836-1846.

Park, S., Y. Han, Y. Liu, A. T. Toriola, L. L. Peterson, G. A. Colditz, S. I. Kim, Y. U. Cho, B.-W. Park and Y. Park (2019). "Adjuvant chemotherapy and survival among patients 70 years of age and younger with node-negative breast cancer and the 21-gene recurrence score of 26-30." Breast cancer research : BCR **21**(1): 110-110.

Peng, Y. J., Q. Y. Chen, D. J. Fu, Z. M. Liu, T. T. Mao, J. Li and W. T. She (2017). "A novel gene mutation of Runx2 in cleidocranial dysplasia." J Huazhong Univ Sci Technolog Med Sci **37**(5): 772-776.

Pereira, B., S.-F. Chin, O. M. Rueda, H.-K. M. Vollan, E. Provenzano, H. A. Bardwell, M. Pugh, L. Jones, R. Russell, S.-J. Sammut, D. W. Y. Tsui, B. Liu, S.-J. Dawson, J. Abraham, H. Northen, J. F. Peden, A. Mukherjee, G. Turashvili, A. R. Green, S. McKinney, A. Oloumi, S. Shah, N. Rosenfeld, L. Murphy, D. R. Bentley, I. O. Ellis, A. Purushotham, S. E. Pinder, A.-L. Børresen-Dale, H. M. Earl, P. D. Pharoah, M. T. Ross, S. Aparicio and C. Caldas (2016). "The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes." Nature Communications **7**(1): 11479.

Pietersen, A. M., B. Evers, A. A. Prasad, E. Tanger, P. Cornelissen-Steijger, J. Jonkers and M. van Lohuizen (2008). "Bmi1 regulates stem cells and proliferation and differentiation of committed cells in mammary epithelium." Curr Biol **18**(14): 1094-1099.

Plaks, V., A. Brenot, D. A. Lawson, J. R. Linnemann, E. C. Van Kappel, K. C. Wong, F. de Sauvage, O. D. Klein and Z. Werb (2013). "Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis." Cell Rep **3**(1): 70-78.

Pohlmann, P. R., I. A. Mayer and R. Mernaugh (2009). "Resistance to Trastuzumab in Breast Cancer." Clinical cancer research : an official journal of the American Association for Cancer Research **15**(24): 7479-7491.

Pon, J. R. and M. A. Marra (2015). "Driver and passenger mutations in cancer." Annu Rev Pathol **10**: 25-50.

Potten, C. S., R. Gandara, Y. R. Mahida, M. Loeffler and N. A. Wright (2009). "The stem cells of small intestinal crypts: where are they?" Cell Prolif **42**(6): 731-750.

Public Health England (2020). Cancer statistics: availability and location - March 2020 update.

Qadir, A. S., P. Ceppi, S. Brockway, C. Law, L. Mu, N. N. Khodarev, J. Kim, J. C. Zhao, W. Putzbach, A. E. Murmann, Z. Chen, W. Chen, X. Liu, A. R. Salomon, H. Liu, R. R. Weichselbaum, J. Yu and M. E. Peter (2017). "CD95/Fas Increases Stemness in Cancer Cells by Inducing a STAT1-Dependent Type I Interferon Response." Cell Rep **18**(10): 2373-2386.

Rakha, E. A., S. E. Pinder, J. M. Bartlett, M. Ibrahim, J. Starczynski, P. J. Carder, E. Provenzano, A. Hanby, S. Hales, A. H. Lee and I. O. Ellis (2015). "Updated UK Recommendations for HER2 assessment in breast cancer." J Clin Pathol **68**(2): 93-99.

Ramaswamy, S., K. N. Ross, E. S. Lander and T. R. Golub (2003). "A molecular signature of metastasis in primary solid tumors." Nat Genet **33**(1): 49-54.

Raveh, E., S. Cohen, D. Levanon, Y. Groner and U. Gat (2005). "Runx3 is involved in hair shape determination." Dev Dyn **233**(4): 1478-1487.

Rexer, B. N. and C. L. Arteaga (2012). "Intrinsic and acquired resistance to HER2-targeted therapies in HER2 gene-amplified breast cancer: mechanisms and clinical implications." Critical reviews in oncogenesis **17**(1): 1-16.

Riggio, A. I. (2017). "The role of Runx1 in genetic models of breast cancer." PhD Thesis, University of Glasgow.

Riggio, A. I., et al. (Unpublished).

Ring, A. and M. Dowsett (2004). "Mechanisms of tamoxifen resistance." Endocr Relat Cancer **11**(4): 643-658.

Rios, A. C., N. Y. Fu, G. J. Lindeman and J. E. Visvader (2014). "In situ identification of bipotent stem cells in the mammary gland." Nature **506**(7488): 322-327.

Ritchie, H., Spooner, F., and Roser, M. (2018). "Causes of death." Our World in Data.

Robinson, G. W. and L. Hennighausen (2011). "MMTV-Cre transgenes can adversely affect lactation: considerations for conditional gene deletion in mammary tissue." Anal Biochem **412**(1): 92-95.

Rody, A., T. Karn, C. Liedtke, L. Pusztai, E. Ruckhaeberle, L. Hanker, R. Gaetje, C. Solbach, A. Ahr, D. Metzler, M. Schmidt, V. Müller, U. Holtrich and M. Kaufmann (2011). "A clinically relevant gene signature in triple negative and basal-like breast cancer." Breast cancer research : BCR **13**(5): R97-R97.

Rooney, N., S. M. Mason, L. McDonald, J. H. M. Däbritz, K. J. Campbell, A. Hedley, S. Howard, D. Athineos, C. Nixon, W. Clark, J. D. G. Leach, O. J. Sansom, J. Edwards, E. R. Cameron and K. Blyth (2020). "RUNX1 Is a Driver of Renal Cell Carcinoma Correlating with Clinical Outcome." Cancer Res **80**(11): 2325-2339.

Rossi, L., D. Stevens, J.-Y. Pierga, F. Lerebours, F. Reyal, M. Robain, B. Asselain and R. Rouzier (2015). "Impact of Adjuvant Chemotherapy on Breast Cancer Survival: A Real-World Population." PloS one **10**(7): e0132853-e0132853.

Rota, L. M., D. A. Lazzarino, A. N. Ziegler, D. LeRoith and T. L. Wood (2012). "Determining mammosphere-forming potential: application of the limiting dilution analysis." J Mammary Gland Biol Neoplasia **17**(2): 119-123.

Rudas, M., M. Filipits, S. Taucher, T. Stranzl, G. G. Steger, R. Jakesz, R. Pirker and G. Pohl (2003). "Expression of MRP1, LRP and Pgp in breast carcinoma patients treated with preoperative chemotherapy." Breast Cancer Res Treat **81**(2): 149-157.

Sakakura, C., A. Hagiwara, K. Miyagawa, S. Nakashima, T. Yoshikawa, S. Kin, Y. Nakase, K. Ito, H. Yamagishi, S. Yazumi, T. Chiba and Y. Ito (2005). "Frequent downregulation of the runt domain transcription factors RUNX1, RUNX3 and their cofactor CBFB in gastric cancer." Int J Cancer **113**(2): 221-228.

Sansom, O. J., V. S. Meniel, V. Muncan, T. J. Phesse, J. A. Wilkins, K. R. Reed, J. K. Vass, D. Athineos, H. Clevers and A. R. Clarke (2007). "Myc deletion rescues Apc deficiency in the small intestine." Nature **446**(7136): 676-679.

Scheitz, C. J., T. S. Lee, D. J. McDermitt and T. Tumbar (2012). "Defining a tissue stem cell-driven Runx1/Stat3 signalling axis in epithelial cancer." Embo j **31**(21): 4124-4139.

Seachrist, D. D., M. M. Hannigan, N. N. Ingles, B. M. Webb, K. L. Weber-Bonk, P. Yu, G. Bebek, S. Singh, S. T. Sizemore, V. Varadan, D. D. Licatalosi and R. A. Keri (2020). "The transcriptional repressor BCL11A promotes breast cancer metastasis." J Biol Chem **295**(33): 11707-11719.

Selbert, S., D. J. Bentley, D. W. Melton, D. Rannie, P. Lourenço, C. J. Watson and A. R. Clarke (1998). "Efficient BLG-Cre mediated gene deletion in the mammary gland." Transgenic Res **7**(5): 387-396.

Shackleton, M., F. Vaillant, K. J. Simpson, J. Stingl, G. K. Smyth, M. L. Asselin-Labat, L. Wu, G. J. Lindeman and J. E. Visvader (2006). "Generation of a functional mammary gland from a single stem cell." Nature **439**(7072): 84-88.

Sharma, P., J. R. Klemp, B. F. Kimler, J. D. Mahnken, L. J. Geier, Q. J. Khan, M. Elia, C. S. Connor, M. K. McGinness, J. M. Mammen, J. L. Wagner, C. Ward, L. Ranallo, C. J. Knight, S. R. Stecklein, R. A. Jensen, C. J. Fabian and A. K. Godwin (2014). "Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing." Breast Cancer Res Treat **145**(3): 707-714.

Shaw, F. L., H. Harrison, K. Spence, M. P. Ablett, B. M. Simões, G. Farnie and R. B. Clarke (2012). "A detailed mammosphere assay protocol for the quantification of breast stem cell activity." J Mammary Gland Biol Neoplasia **17**(2): 111-117.

Shehata, M., A. Teschendorff, G. Sharp, N. Novcic, I. A. Russell, S. Avril, M. Prater, P. Eirew, C. Caldas, C. J. Watson and J. Stingl (2012). "Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland." Breast Cancer Res **14**(5): R134.

Simon, M., V. L. Grandage, D. C. Linch and A. Khwaja (2005). "Constitutive activation of the Wnt/beta-catenin signalling pathway in acute myeloid leukaemia." Oncogene **24**(14): 2410-2420.

Sinha, A., V. B. Fan, A. B. Ramakrishnan, N. Engelhardt, J. Kennell and K. M. Cadigan (2021). "Repression of Wnt/β-catenin signaling by SOX9 and Mastermind-like transcriptional coactivator 2." Sci Adv **7**(8).

Sjöblom, T., S. Jones, L. D. Wood, D. W. Parsons, J. Lin, T. D. Barber, D. Mandelker, R. J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. D. Markowitz, J. Willis, D. Dawson, J. K. Willson, A. F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. H. Park, K. E. Bachman, N. Papadopoulos, B. Vogelstein, K. W. Kinzler and V. E. Velculescu (2006). "The consensus coding sequences of human breast and colorectal cancers." Science **314**(5797): 268-274.

Slattery, M. L., A. Lundgreen, J. S. Herrick, B. J. Caan, J. D. Potter and R. K. Wolff (2011). "Associations between genetic variation in RUNX1, RUNX2, RUNX3, MAPK1 and eIF4E and riskof colon and rectal cancer: additional support for a TGF-ß-signaling pathway." Carcinogenesis **32**(3): 318-326.

Slosberg, E. D., M. G. Klein, Y. Yao, E. K. Han, I. Schieren and I. B. Weinstein (1999). "The alpha isoform of protein kinase C mediates phorbol ester-induced growth inhibition and p21cip1 induction in HC11 mammary epithelial cells." Oncogene **18**(48): 6658-6666.

Smalley, M. J., J. Titley and M. J. O'Hare (1998). "Clonal characterization of mouse mammary luminal epithelial and myoepithelial cells separated by fluorescence-activated cell sorting." In Vitro Cell Dev Biol Anim **34**(9): 711-721.

Smetana, K., Jr., L. Lacina, P. Szabo, B. Dvořánková, P. Brož and A. Šedo (2016). "Ageing as an Important Risk Factor for Cancer." Anticancer Res **36**(10): 5009-5017.

Snippert, H. J., L. G. van der Flier, T. Sato, J. H. van Es, M. van den Born, C. Kroon-Veenboer, N. Barker, A. M. Klein, J. van Rheenen, B. D. Simons and H. Clevers (2010). "Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells." Cell **143**(1): 134-144.

Sokol, E. S., S. Sanduja, D. X. Jin, D. H. Miller, R. A. Mathis and P. B. Gupta (2015). "Perturbation-expression analysis identifies RUNX1 as a regulator of human mammary stem cell differentiation." PLoS Comput Biol **11**(4): e1004161.

Song, D., M. Cui, G. Zhao, Z. Fan, K. Nolan, Y. Yang, P. Lee, F. Ye and D. Y. Zhang (2014). "Pathway-based analysis of breast cancer." Am J Transl Res **6**(3): 302-311.

Sornapudi, T. R., R. Nayak, P. K. Guthikonda, A. K. Pasupulati, S. Kethavath, V. Uppada, S. Mondal, S. Yellaboina and S. Kurukuti (2018). "Comprehensive profiling of transcriptional networks specific for lactogenic differentiation of HC11 mammary epithelial stem-like cells." Sci Rep **8**(1): 11777.

Sternlicht, M. D. (2006). "Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis." Breast Cancer Res **8**(1): 201.

Stingl, J., C. J. Eaves, U. Kuusk and J. T. Emerman (1998). "Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast." Differentiation **63**(4): 201-213.

Stingl, J., C. J. Eaves, I. Zandieh and J. T. Emerman (2001). "Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue." Breast Cancer Res Treat **67**(2): 93-109.

Stingl, J., P. Eirew, I. Ricketson, M. Shackleton, F. Vaillant, D. Choi, H. I. Li and C. J. Eaves (2006). "Purification and unique properties of mammary epithelial stem cells." Nature **439**(7079): 993-997.

Stingl, J., J. T. Emerman and C. J. Eaves (2005). "Enzymatic dissociation and culture of normal human mammary tissue to detect progenitor activity." Methods Mol Biol **290**: 249-263.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray (2021). "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." CA Cancer J Clin **71**(3): 209-249.

Szostakowska, M., A. Trębińska-Stryjewska, E. A. Grzybowska and A. Fabisiewicz (2019). "Resistance to endocrine therapy in breast cancer: molecular mechanisms and future goals." Breast cancer research and treatment **173**(3): 489-497.

Tabár, L., P. B. Dean, T. H. Chen, A. M. Yen, S. L. Chen, J. C. Fann, S. Y. Chiu, M. M. Ku, W. Y. Wu, C. Y. Hsu, Y. C. Chen, K. Beckmann, R. A. Smith and S. W. Duffy (2019).

"The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening." Cancer **125**(4): 515-523.

Tang, Y. Y., B. E. Crute, J. J. Kelley, X. Huang, J. Yan, J. Shi, K. L. Hartman, T. M. Laue, N. A. Speck and J. H. Bushweller (2000). "Biophysical characterization of interactions between the core binding factor alpha and beta subunits and DNA." FEBS Lett **470**(2): 167-172.

Tang, Y. Y., J. Shi, L. Zhang, A. Davis, J. Bravo, A. J. Warren, N. A. Speck and J. H. Bushweller (2000). "Energetic and functional contribution of residues in the core binding factor beta (CBFbeta) subunit to heterodimerization with CBFalpha." J Biol Chem **275**(50): 39579-39588.

Taniuchi, I., M. Osato, T. Egawa, M. J. Sunshine, S. C. Bae, T. Komori, Y. Ito and D. R. Littman (2002). "Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development." Cell **111**(5): 621-633.

Thrasyvoulou, S., G. Vartholomatos, G. Markopoulos, D. Noutsopoulos, S. Mantziou, F. Gkartziou, P. Papageorgis, A. Charchanti, P. Kouklis, A. I. Constantinou and T. Tzavaras (2020). "VL30 retrotransposition is associated with induced EMT, CSC generation and tumorigenesis in HC11 mouse mammary stem-like epithelial cells." Oncol Rep **44**(1): 126-138.

Tolaney, S. M., H. Guo, S. Pernas, W. T. Barry, D. A. Dillon, L. Ritterhouse, B. P. Schneider, F. Shen, K. Fuhrman, M. Baltay, C. T. Dang, D. A. Yardley, B. Moy, P. K. Marcom, K. S. Albain, H. S. Rugo, M. J. Ellis, I. Shapira, A. C. Wolff, L. A. Carey, B. Overmoyer, A. H. Partridge, C. A. Hudis, I. E. Krop, H. J. Burstein and E. P. Winer (2019). "Seven-Year Follow-Up Analysis of Adjuvant Paclitaxel and Trastuzumab Trial for Node-Negative, Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer." J Clin Oncol **37**(22): 1868-1875.

Tomita, H., K. Tanaka, T. Tanaka and A. Hara (2016). "Aldehyde dehydrogenase 1A1 in stem cells and cancer." Oncotarget **7**(10): 11018-11032.

Topol, L., W. Chen, H. Song, T. F. Day and Y. Yang (2009). "Sox9 inhibits Wnt signaling by promoting beta-catenin phosphorylation in the nucleus." J Biol Chem **284**(5): 3323-3333.

Tornillo, G. and S. Cabodi (2014). "3D Mammary Colony-Forming Cell Assay." Bio-protocol **4**(7): e1087.

Trejo, C. L., G. Luna, C. Dravis, B. T. Spike and G. M. Wahl (2017). "Lgr5 is a marker for fetal mammary stem cells, but is not essential for stem cell activity or tumorigenesis." NPJ Breast Cancer **3**: 16.

Trimble, E. L., R. S. Ungerleider, J. A. Abrams, R. S. Kaplan, E. G. Feigal, M. A. Smith, C. L. Carter and M. A. Friedman (1993). "Neoadjuvant therapy in cancer treatment." Cancer **72**(11 Suppl): 3515-3524.

Valkenburg, K. C., C. R. Graveel, C. R. Zylstra-Diegel, Z. Zhong and B. O. Williams (2011). "Wnt/ß-catenin Signaling in Normal and Cancer Stem Cells." Cancers (Basel) **3**(2): 2050-2079.

van Bragt, M. P. A., X. Hu, Y. Xie and Z. Li (2014). "RUNX1, a transcription factor mutated in breast cancer, controls the fate of ER-positive mammary luminal cells." eLife **3**: e03881-e03881.

Van Keymeulen, A., A. S. Rocha, M. Ousset, B. Beck, G. Bouvencourt, J. Rock, N. Sharma, S. Dekoninck and C. Blanpain (2011). "Distinct stem cells contribute to mammary gland development and maintenance." Nature **479**(7372): 189-193.

Vassilopoulos, A., R. H. Wang, C. Petrovas, D. Ambrozak, R. Koup and C. X. Deng (2008). "Identification and characterization of cancer initiating cells from BRCA1 related mammary tumors using markers for normal mammary stem cells." Int J Biol Sci **4**(3): 133-142.

Vernieri, C., M. Milano, M. Brambilla, A. Mennitto, C. Maggi, M. S. Cona, M. Prisciandaro, C. Fabbroni, L. Celio, G. Mariani, G. V. Bianchi, G. Capri and F. de Braud (2019). "Resistance mechanisms to anti-HER2 therapies in HER2-positive breast cancer: Current knowledge, new research directions and therapeutic perspectives." Critical Reviews in Oncology/Hematology **139**: 53-66.

Visvader, J. E. (2009). "Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis." Genes Dev **23**(22): 2563-2577.

Visvader, J. E. and J. Stingl (2014). "Mammary stem cells and the differentiation hierarchy: current status and perspectives." Genes Dev **28**(11): 1143-1158.

Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr. and K. W. Kinzler (2013). "Cancer genome landscapes." Science **339**(6127): 1546-1558.

Wagner, K. U., R. J. Wall, L. St-Onge, P. Gruss, A. Wynshaw-Boris, L. Garrett, M. Li, P. A. Furth and L. Hennighausen (1997). "Cre-mediated gene deletion in the mammary gland." Nucleic Acids Res **25**(21): 4323-4330.

Wan, L., X. Lu, S. Yuan, Y. Wei, F. Guo, M. Shen, M. Yuan, R. Chakrabarti, Y. Hua, H. A. Smith, M. A. Blanco, M. Chekmareva, H. Wu, R. T. Bronson, B. G. Haffty, Y. Xing and Y. Kang (2014). "MTDH-SND1 interaction is crucial for expansion and activity of tumor-initiating cells in diverse oncogene- and carcinogen-induced mammary tumors." Cancer Cell **26**(1): 92-105.

Wang, L., J. S. Brugge and K. A. Janes (2011). "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 of the United States of America **108**(40): E803-E812.

Wang, Q., T. Stacy, M. Binder, M. Marin-Padilla, A. H. Sharpe and N. A. Speck (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**(8): 3444-3449.

Wang, X., Y. Xu, K. Xu, Y. Chen, X. Xiao and X. Guan (2020). "BCL11A confers cell invasion and migration in androgen receptor-positive triple-negative breast cancer." Oncol Lett **19**(4): 2916-2924.

Wang, X., H. Zhang and X. Chen (2019). "Drug resistance and combating drug resistance in cancer." Cancer Drug Resistance **2**(2): 141-160.

Welm, B. E., S. B. Tepera, T. Venezia, T. A. Graubert, J. M. Rosen and M. A. Goodell (2002). "Sca-1(pos) cells in the mouse mammary gland represent an enriched progenitor cell population." Dev Biol **245**(1): 42-56.

Williams, C., L. Helguero, K. Edvardsson, L. A. Haldosén and J. A. Gustafsson (2009). "Gene expression in murine mammary epithelial stem cell-like cells shows similarities to human breast cancer gene expression." Breast Cancer Res **11**(3): R26.

Wisdom, A. J., Y. M. Mowery, C. S. Hong, J. E. Himes, B. Y. Nabet, X. Qin, D. Zhang, L. Chen, H. Fradin, R. Patel, A. M. Bassil, E. S. Muise, D. A. King, E. S. Xu, D. J. Carpenter, C. L. Kent, K. S. Smythe, N. T. Williams, L. Luo, Y. Ma, A. A. Alizadeh, K. Owzar, M. Diehn, T. Bradley and D. G. Kirsch (2020). "Single cell analysis reveals distinct immune landscapes in transplant and primary sarcomas that determine response or resistance to immunotherapy." Nat Commun **11**(1): 6410.

Wood, L. D., D. W. Parsons, S. Jones, J. Lin, T. Sjöblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezso, V. Ustyanksky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P. A. Wilson, J. S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J. K. Willson, S. Sukumar, K. Polyak, B. H. Park, C. L. Pethiyagoda, P. V. Pant, D. G. Ballinger, A. B. Sparks, J. Hartigan, D. R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S. D. Markowitz, G. Parmigiani, K. W. Kinzler, V. E. Velculescu and B. Vogelstein (2007). "The genomic landscapes of human breast and colorectal cancers." Science **318**(5853): 1108-1113.

Woolf, E., C. Xiao, O. Fainaru, J. Lotem, D. Rosen, V. Negreanu, Y. Bernstein, D. Goldenberg, O. Brenner, G. Berke, D. Levanon and Y. Groner (2003). "Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis." Proc Natl Acad Sci U S A **100**(13): 7731-7736.

Wotton, S., M. Stewart, K. Blyth, F. Vaillant, A. Kilbey, J. C. Neil and E. R. Cameron (2002). "Proviral insertion indicates a dominant oncogenic role for Runx1/AML-1 in T-cell lymphoma." Cancer Res **62**(24): 7181-7185.

Wu, B., S. P. Crampton and C. C. W. Hughes (2007). "Wnt signaling induces matrix metalloproteinase expression and regulates T cell transmigration." Immunity **26**(2): 227-239.

Wu, J. Q., M. Seay, V. P. Schulz, M. Hariharan, D. Tuck, J. Lian, J. Du, M. Shi, Z. Ye, M. Gerstein, M. P. Snyder and S. Weissman (2012). "Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line." PLoS Genet **8**(3): e1002565.

Xu, L., W. Lin, L. Wen and G. Li (2019). "Lgr5 in cancer biology: functional identification of Lgr5 in cancer progression and potential opportunities for novel therapy." Stem Cell Res Ther **10**(1): 219.

Xu, W. a., Q. Chen, C. Liu, J. Chen, F. Xiong and B. Wu (2017). "A novel, complex RUNX2 gene mutation causes cleidocranial dysplasia." BMC medical genetics **18**(1): 13-13.

Xu, X., M. Zhang, F. Xu and S. Jiang (2020). "Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities." Mol Cancer **19**(1): 165.

Yang, L., H. Tang, Y. Kong, X. Xie, J. Chen, C. Song, X. Liu, F. Ye, N. Li, N. Wang and X. Xie (2015). "LGR5 Promotes Breast Cancer Progression and Maintains Stem-Like Cells Through Activation of Wnt/β-Catenin Signaling." Stem Cells **33**(10): 2913-2924.

Yang, X., H. Wang and B. Jiao (2017). "Mammary gland stem cells and their application in breast cancer." Oncotarget **8**(6): 10675-10691.

Yersal, O. and S. Barutca (2014). "Biological subtypes of breast cancer: Prognostic and therapeutic implications." World J Clin Oncol **5**(3): 412-424.

Yin, P., W. Wang, Z. Zhang, Y. Bai, J. Gao and C. Zhao (2018). "Wnt signaling in human and mouse breast cancer: Focusing on Wnt ligands, receptors and antagonists." Cancer science **109**(11): 3368-3375.

Yokomizo, T., M. Ogawa, M. Osato, T. Kanno, H. Yoshida, T. Fujimoto, S. Fraser, S. Nishikawa, H. Okada, M. Satake, T. Noda, S.-I. Nishikawa and Y. Ito (2001). "Requirement of Runx1/AML1/PEBP2αB for the generation of haematopoietic cells from endothelial cells." Genes to Cells **6**(1): 13-23.

Yoshida, C. A., H. Yamamoto, T. Fujita, T. Furuichi, K. Ito, K.-i. Inoue, K. Yamana, A. Zanma, K. Takada, Y. Ito and T. Komori (2004). "Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog." Genes & development **18**(8): 952-963.

You, W. and M. Henneberg (2018). "Cancer incidence increasing globally: The role of relaxed natural selection." Evol Appl **11**(2): 140-152.

Yuan, T., Y. Wang, L. Pao, S. M. Anderson and H. Gu (2011). "Lactation defect in a widely used MMTV-Cre transgenic line of mice." PLoS One **6**(4): e19233.

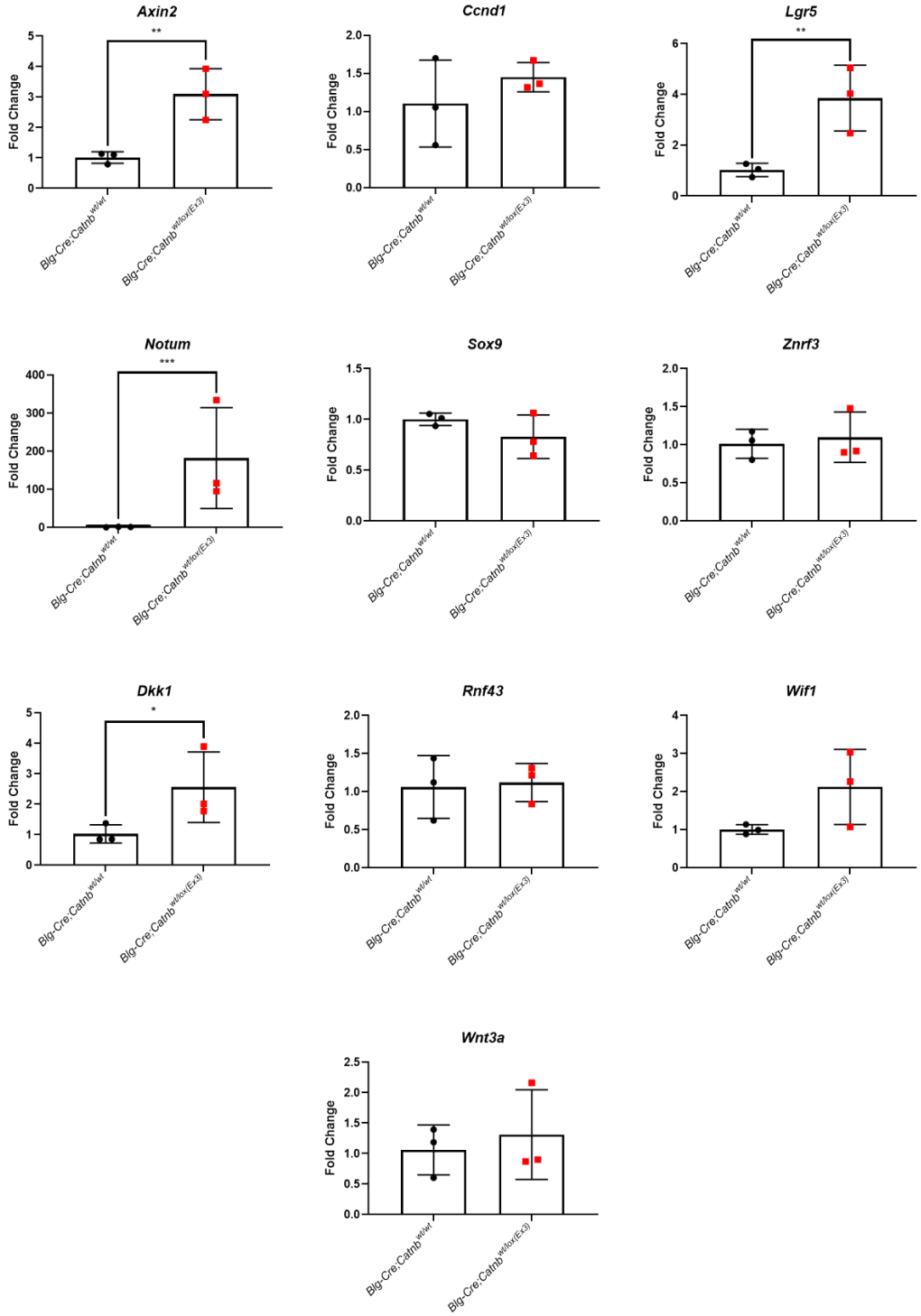
Yzaguirre, A. D., M. F. de Bruijn and N. A. Speck (2017). "The Role of Runx1 in Embryonic Blood Cell Formation." Adv Exp Med Biol **962**: 47-64.

Zaha, D. C. (2014). "Significance of immunohistochemistry in breast cancer." World journal of clinical oncology **5**(3): 382-392.

Zeng, Y. A. and R. Nusse (2010). "Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture." Cell Stem Cell **6**(6): 568-577.

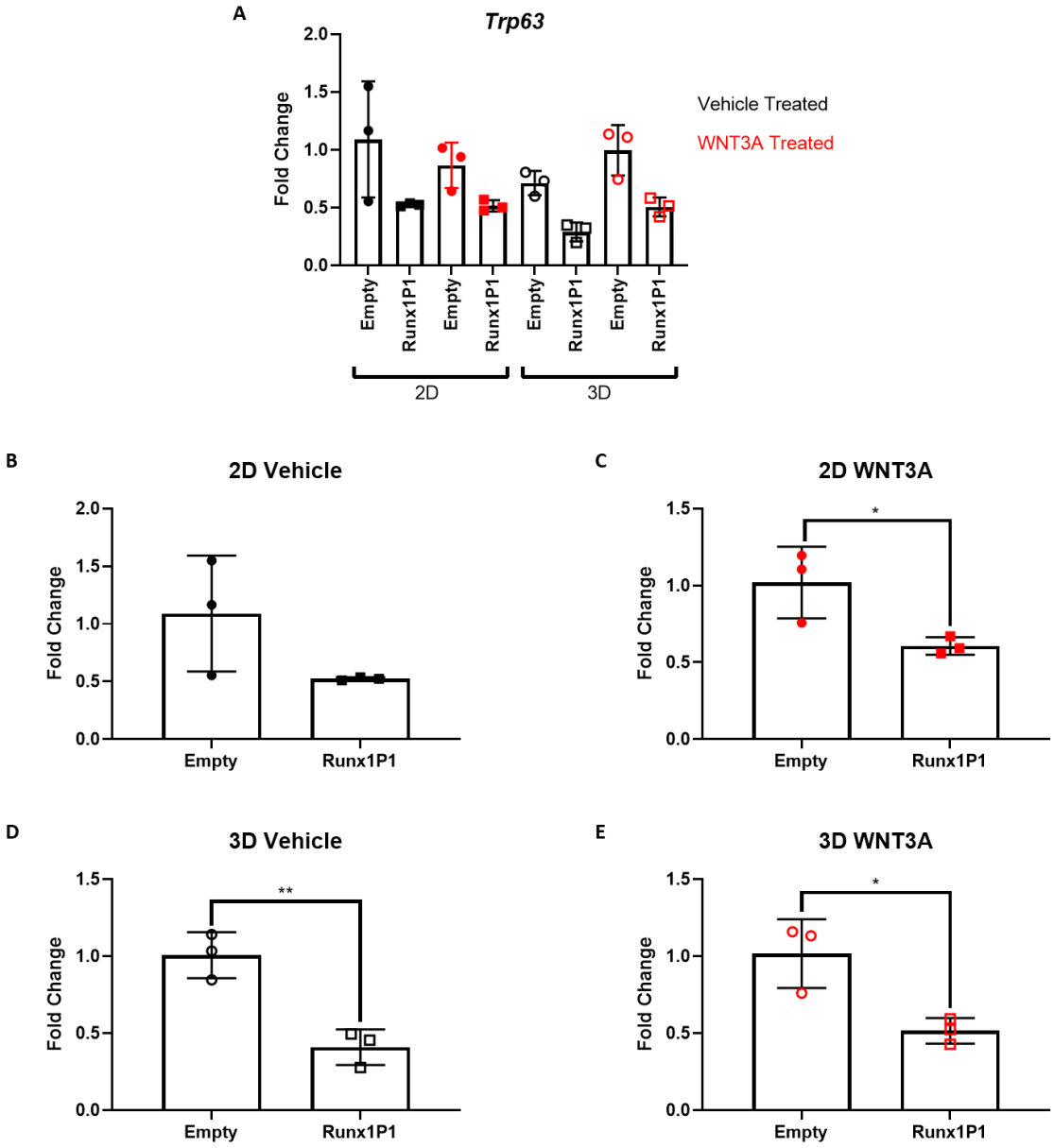
Zhan, T., N. Rindtorff and M. Boutros (2017). "Wnt signaling in cancer." Oncogene **36**(11): 1461-1473.

Appendix



Appendix 1 qPCR validation of Wnt targets in *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice.

Box and whisker plots summarising results from qPCR analysis of MMECs, isolated from pregnant *Blg-Cre;Catnb*$^{wt/wt}$ and *Blg-Cre;Catnb*$^{wt/lox(Ex3)}$ mice. *Actb* used as internal reference to normalise the expression data for each test gene shown. Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual sample, where each sample consisted of MMECs extracted from the combined glands of a single mouse. Unpaired t-tests with Welch's correction performed for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$].

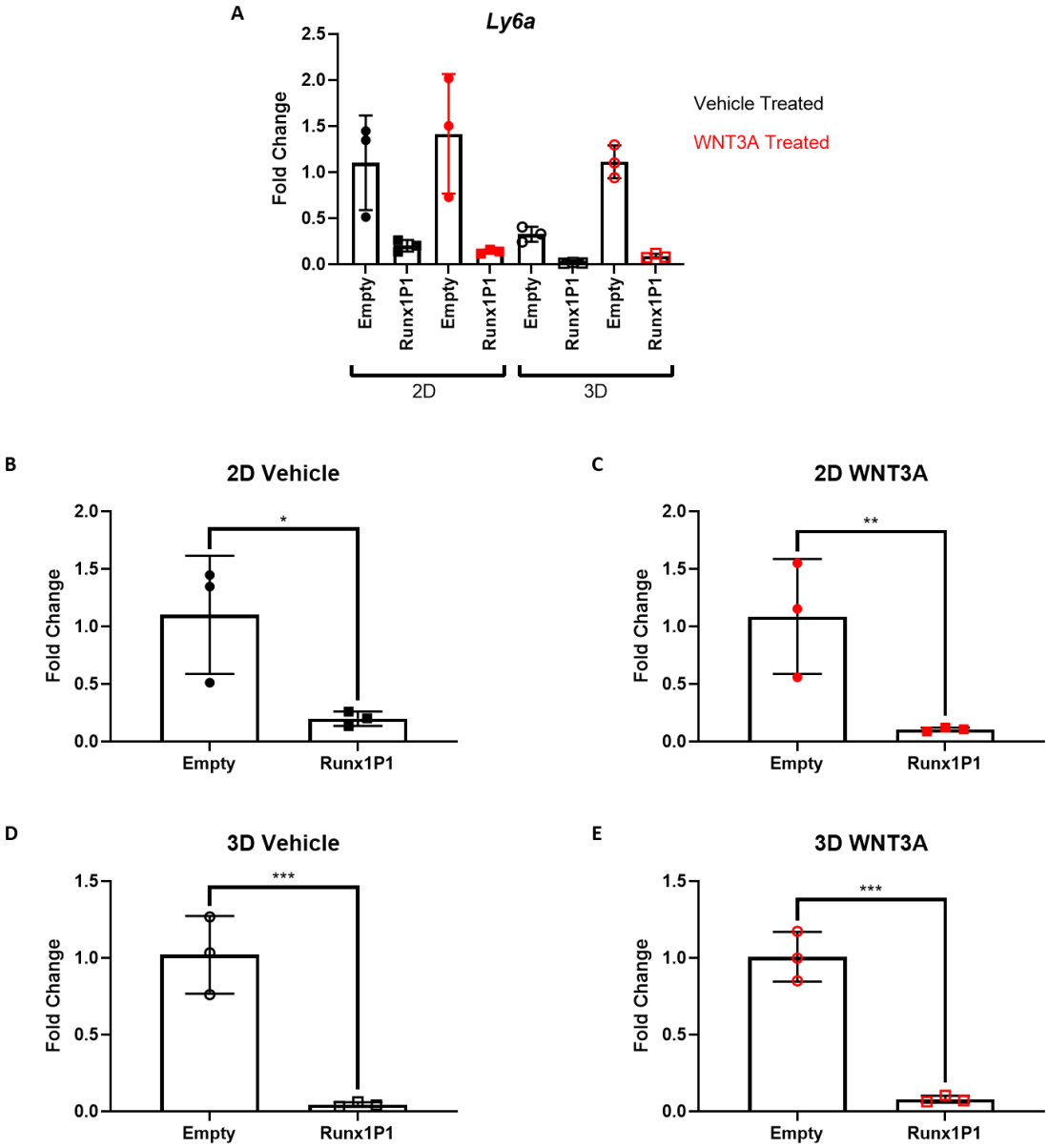


Appendix 2 qPCR analysis comparing *Trp63* in RUNX1 overexpressing and control HC11 cells.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Trp63*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [* $p < 0.05$; ** $p < 0.01$].

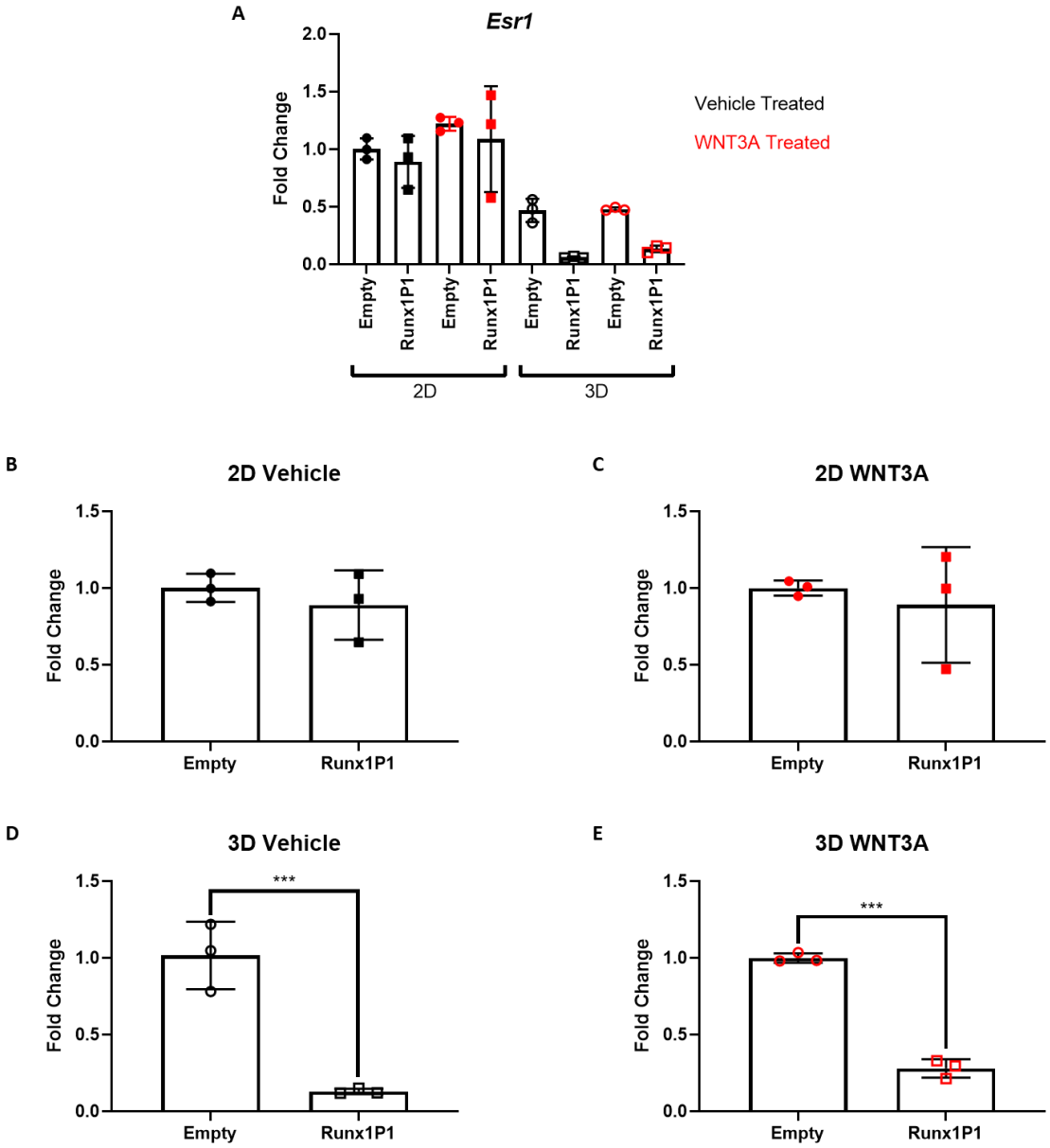


Appendix 3 Analysing the effects of RUNX1 overexpression on *Ly6a* expression by qPCR.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Ly6a*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$].

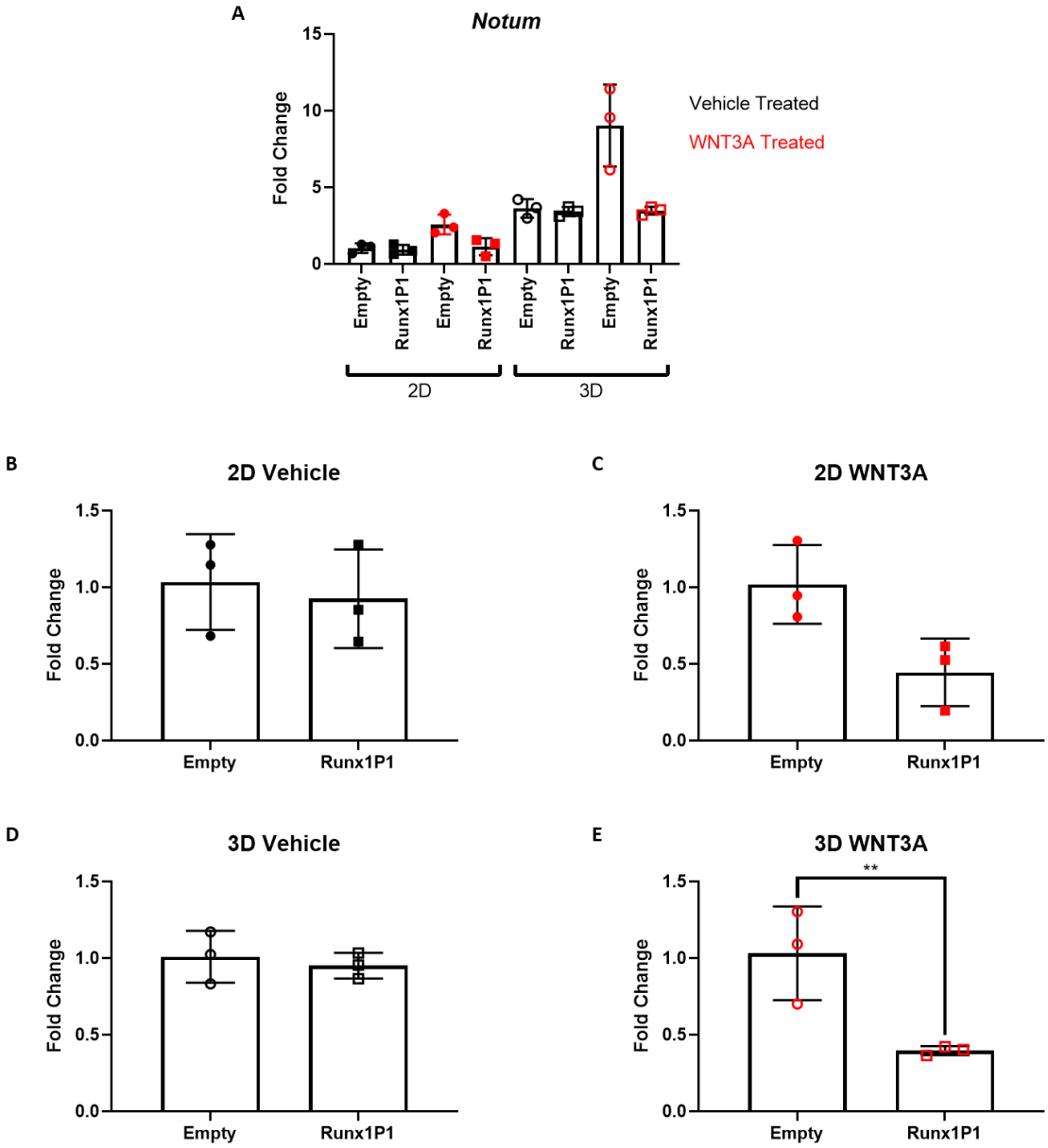


Appendix 4 qPCR analysis of *Esr1* expression in control versus Runx1P1 cells.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Esr1*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons [*** $p < 0.001$].

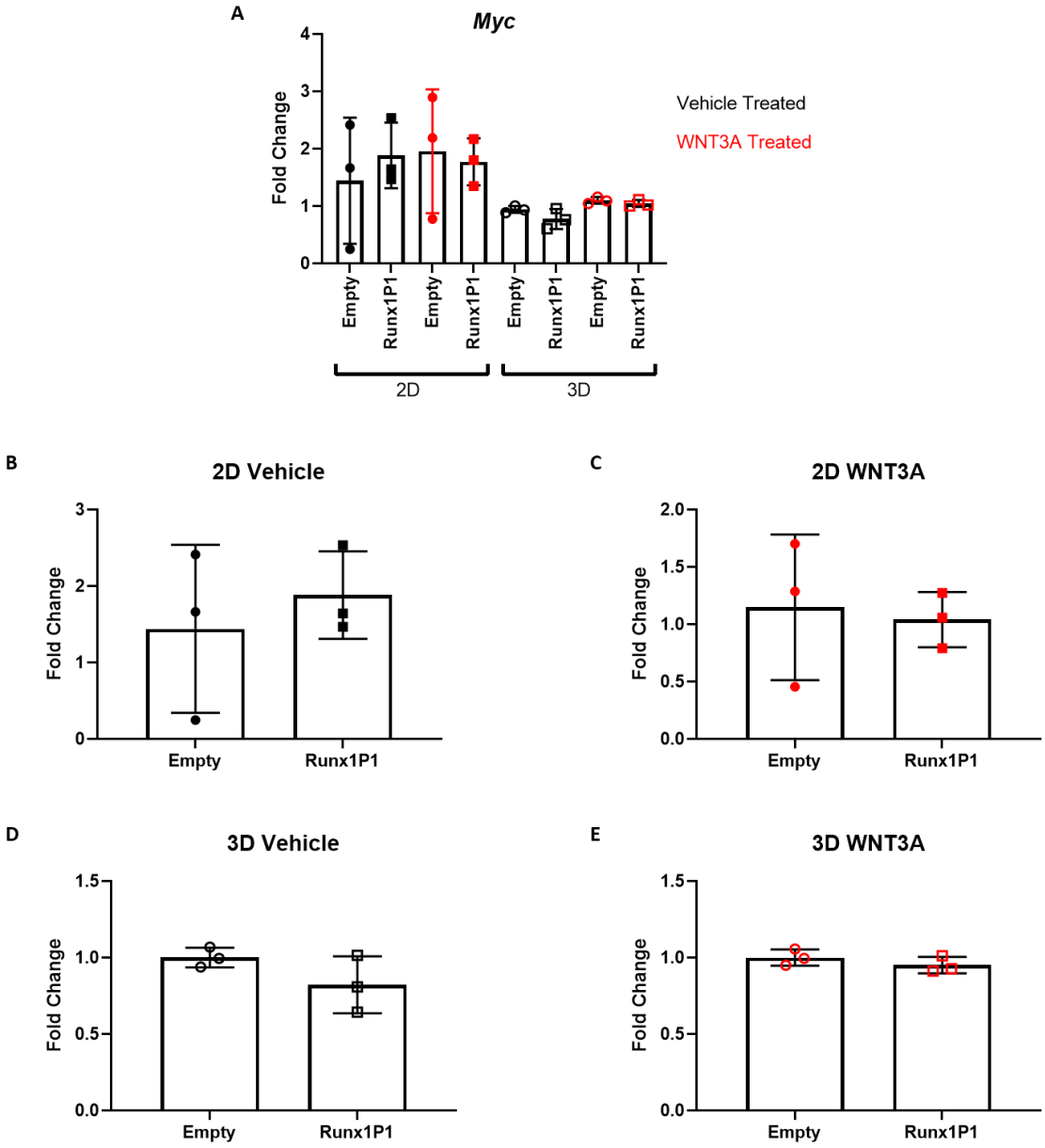


Appendix 5 qPCR analysis to assess the impact of RUNX1 overexpression on *Notum* levels.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Notum*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

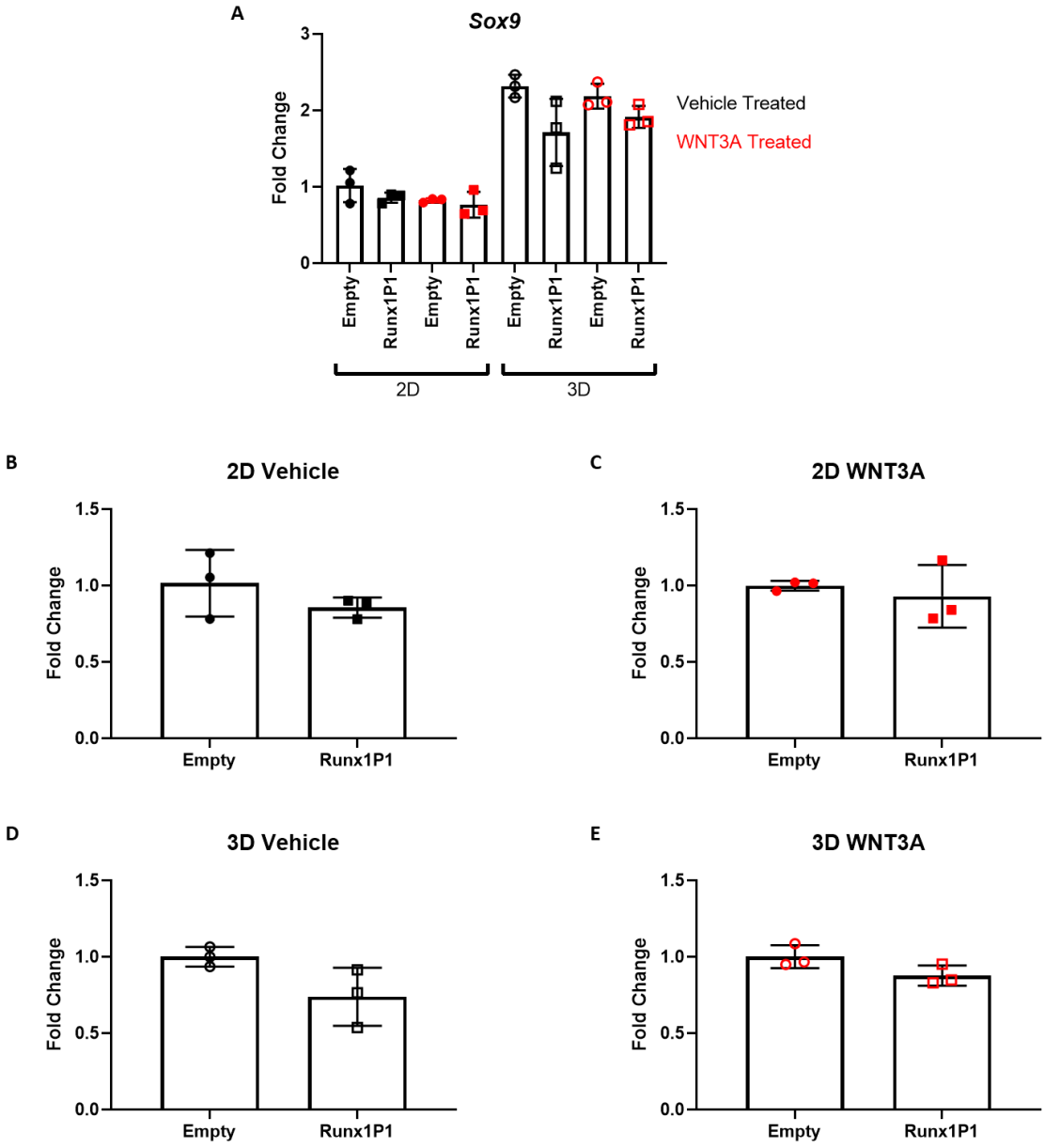
Unpaired t-tests with Welch's correction performed for statistical comparisons [** $p < 0.01$].



Appendix 6 qPCR analysis of *Myc* expression in Runx1P1 and Empty Vector control cells.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Myc*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons.

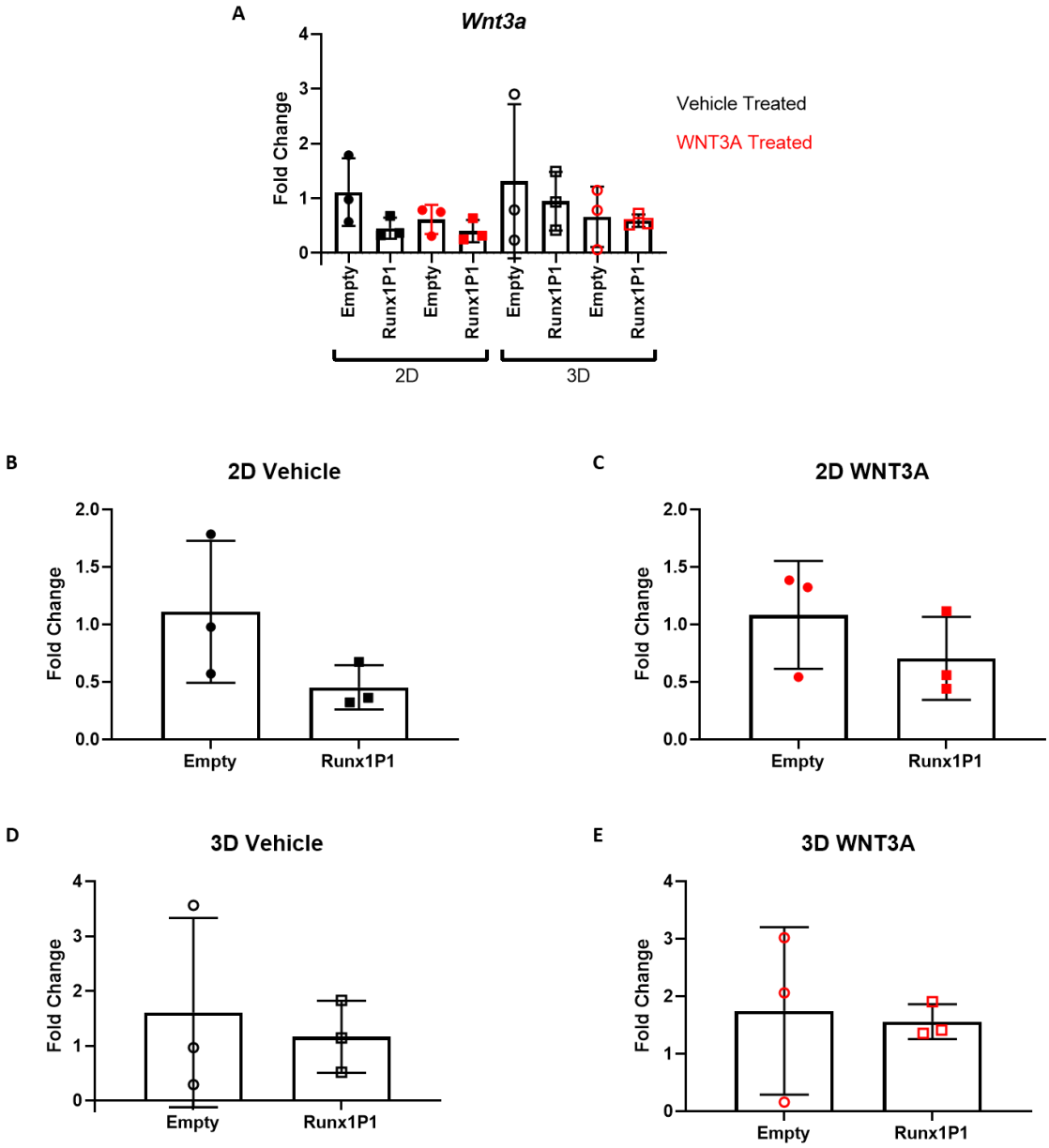


Appendix 7 Analysis of *Sox9* expression in Runx1P1 and Empty Vector control cells by qPCR.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Sox9*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons.

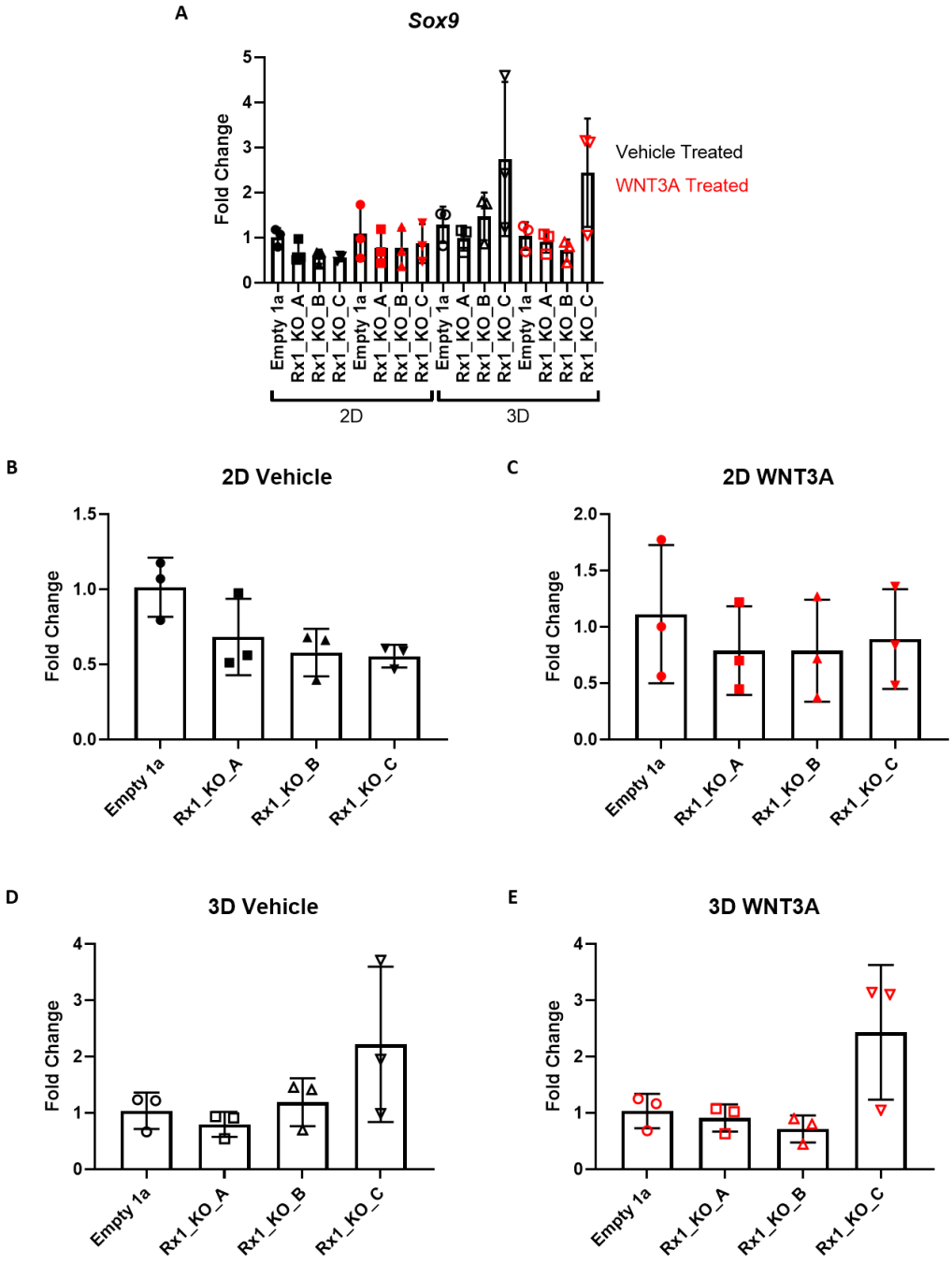


Appendix 8 Evaluating the effects of RUNX1 overexpression on *Wnt3a* levels in HC11 cells.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty Vector and Runx1P1 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Wnt3a*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

Unpaired t-tests with Welch's correction performed for statistical comparisons.

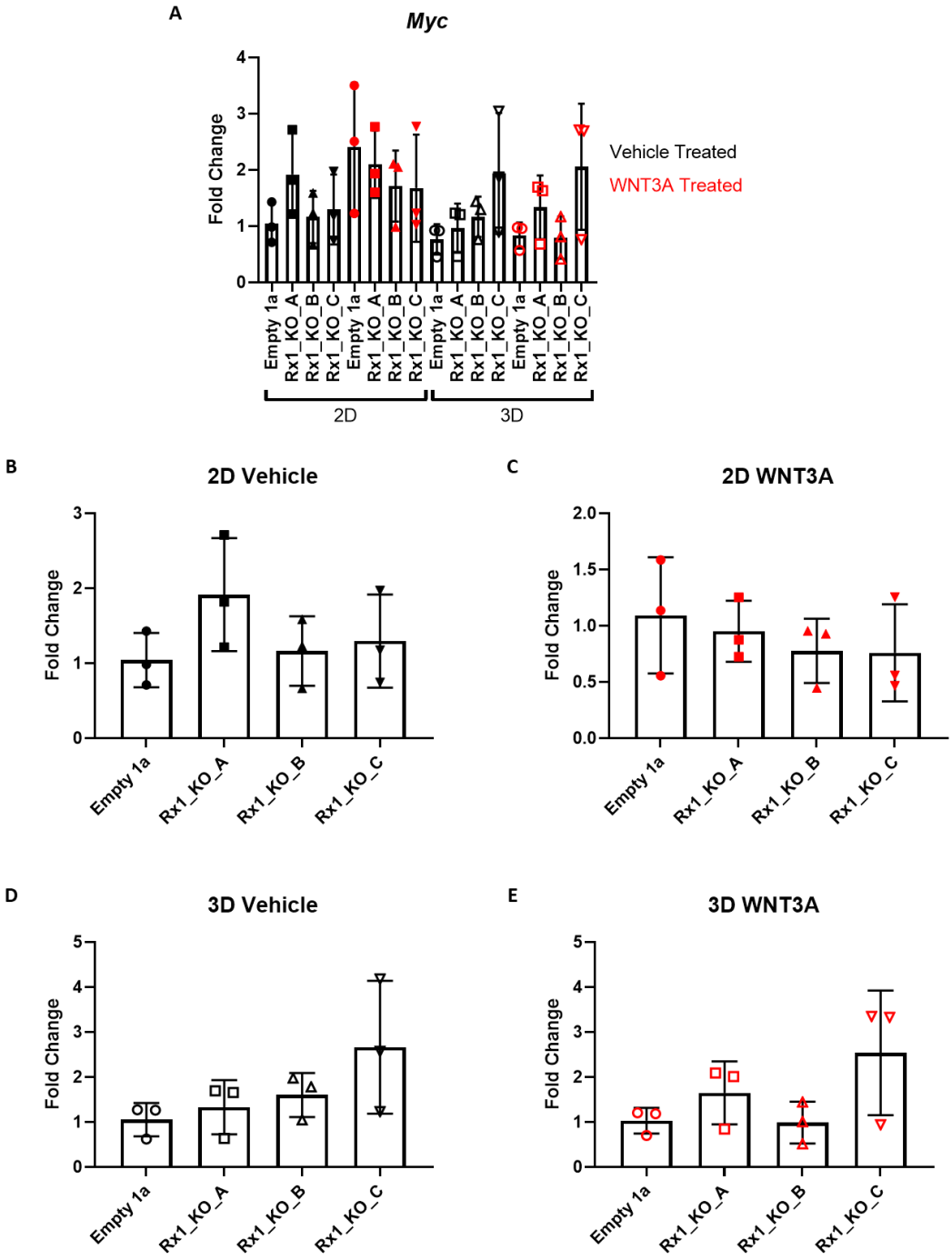


Appendix 9 RT-qPCR analysis to determine the effects of *Runx1* depletion on *Sox9* expression levels.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty vector and *Runx1*-deleted HC11 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Sox9*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons.



Appendix 10 *Myc* expression analysis in *Runx1*-deleted HC11 cell lines by qPCR.

Box and whisker plots summarising results from qPCR analysis of cell pellets, obtained from Empty vector and *Runx1*-deleted HC11 cells grown in either 2D or 3D (mammosphere) growth conditions for 7 days with additional WNT3A or vehicle (distilled water) treatment. *Gapdh* and *Actb* genes used as internal references to normalise the expression data for the test gene (*Myc*). Each data point represents the normalised relative expression (fold change) of the given test gene (calculated from an average of 3 technical repeats, and using a ΔΔCq method) obtained from an individual cell pellet (each of which was obtained from 1 of 3 individual experimental setups).

The results for all experimental conditions are plotted together in [A], while the results of each growth condition are individually plotted in [B] to [E].

One-way ANOVA, with Dunnett's multiple comparisons test, for statistical comparisons.

```
Rx2_KO_1

Query: None Query ID: lcl|Query_7599 Length: 955


>Mus musculus strain C57BL/6J chromosome 17, GRCm39
Sequence ID: NC_000083.7 Length: 95294699
Range 1: 45046176 to 45046731

Score:822 bits(445), Expect:0.0,
Identities:524/556(94%),  Gaps:30/556(5%), Strand: Plus/Minus

Query  19        GACTCTGTCCGGTCTCCAGTCGCCGGCCCCCAGCAGACTCCGCCCGGCACACGGCTTCCT  78
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046731  GACTCTGTCCGGTCTCCAGTCGCCGGCCCCCAGCAGACTCCGCCCGGCACACGGCTTCCT  45046672

Query  79        CGAACTTGATTTCTCACCTCCTCTGTCCCGTCAcctccatcctctttccgcccccgtctc  138
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046671  CGAACTTGATTTCTCACCTCCTCTGTCCCGTCACCTCCATCCTCTTTCCGCCCCCGTCTC  45046612

Query  139       gcctccacccccctcgatttcctcctcctcgccccccatttccaccctcctcccccctcccc  198
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046611  GCCTCCACCCCCTCGATTTCCTCCTCCTCGCCCCCCATTTCCACCCTCCTCCCCCTCCCC  45046552

Query  199       cGGCCACTTCGCTAACTTGTGGCTGTTGTGATGCGTATTCCTGTAGATCCGAGCACCAGC  258
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046551  CGGCCACTTCGCTAACTTGTGGCTGTTGTGATGCGTATTCCTGTAGATCCGAGCACCAGC  45046492

Query  259       CGGCGCTTCAGcccccccTCcagcagcctgcagcccggcaagatgagcgacgtgagcccg  318
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046491  CGGCGCTTCAGCCCCCCCTCCAGCAGCCTGCAGCCCGGCAAGATGAGCGACGTGAGCCCG  45046432

Query  319       g----------------------------cagcagcagcagcaacagcagcagcaacaa  349
                  |                            ||||||||||||||||||||||||||||||
Sbjct  45046431  GTGGTGGCTGCGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAACAA  45046372

Query  350       cagcaacagcaacaacagcagcagcagcagcagcagcaggaggcggccgcagcagcagcg  409
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046371  CAGCAACAGCAACAACAGCAGCAGCAGCAGCAGCAGCAGGAGGCGGCCGCAGCAGCAGCG  45046312

Query  410       gcggcagcggcggcggcagcagcggcggcggccgcagTGCCCCGATTGAGGCCGCCGCAC  469
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046311  GCGGCAGCGGCGGCGGCAGCAGCGGCGGCGGCCGCAGTGCCCCGATTGAGGCCGCCGCAC  45046252

Query  470       GACAACCGCACCATGGTGGAGGTCATCGCGGACCACCCGGCCGAACTGGTCGGCACCGAC  529
                  |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Sbjct  45046251  GACAACCGCACCATGGTGGAGATCATCGCGGACCACCCGGCCGAACTGGTCCGCACCGAC  45046192

Query  530       AGTCC-AACTTCCTGT  544
                  ||||| ||||||||||
Sbjct  45046191  AGTCCCAACTTCCTGT  45046176
```

Appendix 11 *Runx2* gene successfully targeted by CRISPR-Cas9-mediated deletion in HC11 cell clone 1.

DNA extracted from *Runx2* CRISPR clone 1 (Rx2_KO_1) was sequenced to determine whether there was successful targeting of the *Runx2* gene in these cells. The Basic Local Alignment Search Tool (BLAST) was used to align the obtained sequence with a reference sequence for the mouse *Runx2* gene. Highlighted in yellow is the largest area that was targeted in this CRISPR-Cas9-mediated gene deletion. Total number of gaps found to amount to 30.

```
Rx2_KO_2

Query: None Query ID: lcl|Query_204587 Length: 1063


>Mus musculus strain C57BL/6J chromosome 17, GRCm39
Sequence ID: NC_000083.7 Length: 95294699
Range 1: 45046176 to 45046731

Score:771 bits(417), Expect:0.0,
Identities:523/566(92%),  Gaps:39/566(6%), Strand: Plus/Minus

Query  20        GACTCTGTCCGGTCTCCAGTCGCCGGCCCCCAGCAGACTCCGCCCGGCACACGGCTTCCT  79
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046731  GACTCTGTCCGGTCTCCAGTCGCCGGCCCCCAGCAGACTCCGCCCGGCACACGGCTTCCT  45046672

Query  80        CGAACTTGATTTCTCACCTCCTCTGTCCCGTCAcctccatcctctttccgcccccgtctc  139
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046671  CGAACTTGATTTCTCACCTCCTCTGTCCCGTCACCTCCATCCTCTTTCCGCCCCCGTCTC  45046612

Query  140       gcctccacccccctcgatttcctcctcctcgccccccatttccaccctcctcccccctcccc  199
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046611  GCCTCCACCCCCTCGATTTCCTCCTCCTCGCCCCCCATTTCCACCCTCCTCCCCCTCCCC  45046552

Query  200       cGGCCACTTCGCTAACTTGTGGCTGTTGTGATGCGTATTCCTGTAGATCCGAGCACCAGC  259
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046551  CGGCCACTTCGCTAACTTGTGGCTGTTGTGATGCGTATTCCTGTAGATCCGAGCACCAGC  45046492

Query  260       CGGCGCTTCAGcccccccTCcagcagcctgcagcccggcaagatgagcgacgtgagcccg  319
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046491  CGGCGCTTCAGCCCCCCCTCCAGCAGCCTGCAGCCCGGCAAGATGAGCGACGTGAGCCCG  45046432

Query  320       g----------------------------cagcagcagcagcaacagcagcagcaacaa  350
                  |                            ||||||||||||||||||||||||||||||
Sbjct  45046431  GTGGTGGCTGCGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAACAA  45046372

Query  351       cagcaacagcaacaacagcagcagcagcagcagcagcaggaggcggccgcagcagcagcg  410
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45046371  CAGCAACAGCAACAACAGCAGCAGCAGCAGCAGCAGGAGGCGGCCGCAGCAGCAGCG  45046312

Query  411       gcggcagcgggggggcggcagcagcggcgggcggccgcagTGCCCCGATTGAGGCCGCCGC  470
                  |||||||||| || ||||||||||||| ||||||||||||||||||||||||||||||
Sbjct  45046311  GCGGCAGCGGCGG-CGGCAGCAGCGGCGG-CGGCCGCAGTGCCCCGATTGAGGCCGCCGC  45046254

Query  471       CACGACAACCCGCCCCATGGGGGGAGATCATCGCGGACCACCCCGGGCCGAACTGGGTCC  530
                   |||||||| || ||||||  |||||||||||||||||||| || ||||||||| |||
Sbjct  45046253  -ACGACAACC-GCACCATGGT-GGAGATCATCGCGGACCACCC-GG-CCGAACTGG-TCC  45046200

Query  531       CGCACCGGACAGTCCCAACTTCCGGT  556
                   ||||| ||||||||||||||| ||
Sbjct  45046199  -GCACC-GACAGTCCCAACTTCCTGT  45046176
```

Appendix 12 A second HC11 clone with successful CRISPR-Cas9-mediated targeting of *Runx2* gene expression.

DNA extracted from *Runx2* CRISPR clone 2 (Rx2_KO_2) was sequenced to determine whether there was successful targeting of the *Runx2* gene in these cells. The Basic Local Alignment Search Tool (BLAST) was used to align the obtained sequence with a reference sequence for the mouse *Runx2* gene. Highlighted in yellow is the largest area that was targeted in this CRISPR-Cas9-mediated gene deletion. Total number of gaps found to amount to 39.