The Effect of Sox9 on Cell Growth and Stemness in HER2-Positive Breast Cancer

Riana Zuccarini

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the Masters of Science degree in Cellular and Molecular Medicine

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa, Ontario, Canada

Abstract

Breast cancer is one of the most diagnosed and fatal malignancies, and its incidence increases each year. The HER2-positive subtype is associated with aggressive disease and poor outcome, with a high risk of acquired resistance to first-line treatment. A key contributor to these qualities is epithelial-to-mesenchymal transition (EMT). Sox9 has been linked to tumorigenesis and EMT in many cancers, but its role in HER2-positive breast cancer has yet to be fully characterized. Previous studies have identified a gene network downstream of HER2 that promotes a stem-like phenotype *in vitro* and reduces overall survival *in vivo* due to Sox10 upregulation through a PDK1-AKT-Sox9 axis, and we have shown that Sox10 deletion reduces stemness and inhibits hyperplasia. We hypothesized that Sox9 drives a distinct genetic program required for invasion and the maintenance of stem cells. To determine the effect of Sox9 on cell phenotype we have generated Sox9-deficient cells with a CRISPR/Cas9 knockout and Sox9-overexpressing cells by lentiviral transduction. *In vitro* characterization of these cells demonstrated significant decreases in proliferation, sphere-forming, and invasion in the Sox9-deficient cells with stronger effects observed in complete knockout cells compared to those with residual Sox9 expression. RNA-sequencing revealed an upregulation of epithelial-associated genes in Sox9-deficient cells. Lastly, we have shown that Sox9-deficiency inhibits expansion *in vivo* using orthotopic and tail-vein injection models. Together, our findings suggest that Sox9-knockout inhibits stemness in HER2-positive tumour-derived cells. Further insight into the role of Sox9 may reveal novel therapeutic targets for breast cancer treatment.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACVS	Animal Care and Veterinary Service
ADAM10	A disintegrin and metalloproteinase domain-containing protein 10
ADC	Antibody-drug conjugate
AKT	Protein kinase B
ANOVA	Analysis of variance
BSA	Bovine serum albumin
cDNA	Compliment DNA
CDS	Coding DNA sequence
CK8	Cytokeratin 8
CNS	Central nervous system
CSC	Cancer stem cell
DAPI	4',6-diamidino-2-phenylindole
DEG	Differentially expressed gene
DIM	SoxE dimerization domain
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dpc	Days *post coitum*
EGFR	Epidermal growth factor receptor
EMT	Epithelial-to-mesenchymal transition
ESR1	Estrogen receptor
FACS	Fluorescence-activated cell sorting

FBS	Fetal bovine serum
FISH	Fluorescent *in situ* hybridization
GFP	Green fluorescent protein
GSEA	Gene set enrinchment analysis
HER1	Human epidermal growth factor receptor 1
HER2	Human epidermal growth factor receptor 2
HER3	Human epidermal growth factor receptor 3
HER4	Human epidermal growth factor receptor 4
HMG	High mobility group
IHC	Immunohistochemistry
IRES	Internal ribosome entry site
ISH	*In situ* hybridization
KRT8	Cytokeratin 8, gene name
MAPK	Mitogen-activated protein kinase 1
MEGS	Mammary epithelial growth supplement
MET	Mesenchymal-to-epithelial transition
MMTV	Mouse mammary tumour virus
mRNA	Messenger RNA
NDL	Neu deletion
NES	Normalized enrichment scores
NES	Nuclear export signal
NIC	Ndl2-5-IRES-Cre
NTC	Non-targeting control

OHRI	Ottawa Hospital Research Institute
ORA	Overrepresentation analysis
OVOL2	Ovo like zinc finger 2
P95HER2	95kDa HER2 fragment
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGFB	Platelet derived growth factor subunit B
PDGFRA	Platelet-derived growth factor receptor alpha
PDGFRL	Platelet-derived growth factor receptor-like
PDK1	Phosphoinositide-dependent kinase-1
PEI	Polyethylenimine
PFA	Paraformaldehyde
PGR	Progesterone receptor
PI3K	Phosphoinositide-3 kinase
PIP2	Phosphatidylinositol-4, 5-bisphosphate
PIP3	Phosphatidylinositol-3, 4, 5-triphosphate
poly-HEMA	Poly(2-hydroxyethyl methacrylate)
PTEN	Phosphatase and tensin homolog
PyVMT	Polyomavirus middle T
qPCR	Quantitative polymerase chain reaction
RFP	Red fluorescent protein
RNA	Ribonucleic acid
ROI	Region of interest

SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
sgRNA	Single guide RNA
SLK	Ste-20 like kinase
SOX	Sry-HMG box
SRY	Sex-determining region Y
SUMO	Small ubiquitin-like modifier
TAC	C-terminal transactivation domain
TAM	Middle transactivation domain
TBST	Tris-buffered saline with TWEEN
T-DM1	Ado-trastuzumab emtansine
T-DXd	Fam-trastuzumab deruxtecan-nxki
TEB	Terminal end bud
TGF-α	Transforming growth factor alpha
TGF-β	Transforming growth factor beta
TIC	Tumour initiating cell
TNBC	Triple negative breast cancer
TPBC	Triple positive breast cancer
Wnt	Wingless-related integration site

Significant Contributions

Murine cell lines were originally established by Dr. Khalid Al-Zahrani (NDL3903) and Brennan Garland (NDL668). Dr. John Abou-Hamad designed the sgSox9^{K1} and sgSox9^{K2} CRISPR guides, and Brennan generated the knockout cells shown in Figure 3.1.

Brennan performed the flow cytometry of murine lungs shown in Figure 3.13 and sorted the Sox9-deficient cell lines. All other FACS was performed by OHRI StemCore Laboratories, who also performed the Sanger sequencing of plasmids prior to transduction and of the sequences shown in Figure 3.5.

Bulk RNA sequencing shown in Figure 3.10 and Figure 3.11 was performed by Genome Quebec. Samuel Delisle analyzed the resulting data and assisted with its interpretation.

Brennan and Samuel performed all murine injections shown in Figure 3.12 and Figure 3.13.

Paraffin embedding of all tissues and mammary gland sectioning was performed by the University of Ottawa Louise Pelletier Histology Core Facility.

Acknowledgments

First, I'd like to thank my supervisor, Dr. Luc Sabourin, for his guidance and mentorship throughout the project. Thank you for always having more than one idea ready to go, for having more than one way of testing them, and for pushing me to think the same. I admire your dedication to listening to the data, whether it fits with the original hypothesis or not, and your openness to new ideas and paths forward.

Thank you to all my friends and family who patiently handled my lack of commitment to plans because I was never sure if I'd have to feed my cells that day, and who endured my compulsive need to share anything and everything I learned in the form of fun facts. Thank you all for listening to my long ramblings of what was or wasn't working and thank you for pretending to understand what I was talking about. Although some of you pushed the limits of "no stupid questions" I really do appreciate you all for challenging my summarizing skills and making me think about my research from different perspectives.

Thank you to everyone in the South Wing, especially Henna, Marlena, Lauren, and our adoptee Emma G., for making the lab feel like a community I'm proud to be a part of. I'm grateful for every shared reagent, piece of advice, and conversation held by yelling over the TC fans. Thank you for keeping me motivated during the seemingly endless "just two more weeks" phase and for making sure my figures didn't hurt to look at. Some of you are sitting behind me as I write this part. I'm so glad I had so many people I knew I could always rely on and who continue to make the lab a place I want to be.

Thank you to the Sab Lads: John, Brennan, Sam, and Hong, for taking the important things seriously and nothing else. You kept the lab loud and alive even on late nights, weekends, and the rare early morning. Sincerely, your instruction, experience, and willingness to double check my

math have been just as valuable as your friendship. The research was cool and all but it wouldn't have been nearly as fun without you guys laughing with me from across the bench.

Finally, and arguably most importantly, thank you to everyone with a license and a car who's ever made the journey between the lab and Barrhaven with me. This literally would not have been possible without you.

1. Introduction

1.1. Mammary Gland Development

Mammary glands are among few tissues whose development occurs mostly post-natally. The final ductal structure branches radially from the nipple through the mammary fat pad, which is vascular and collagen-rich, and surrounds the structure in fibroblasts[1]. Each branch is comprised of an inner luminal layer made of polarized cells and an outer basal layer in direct contact with the basement membrane (Figure 1.1)[1]. Structures at the end of each branch, known as lobules, contain specialized luminal epithelial cells that secrete milk upon hormone stimulation during pregnancy[1]. During embryonic development, ectodermal cells form the mammary milk line and organize along it into placodes from which the ductal structure eventually sprouts[2]. Surrounding cells differentiate into fibroblasts and preadipocytes and stimulate further differentiation of mammary epithelial cells[3,4]. The last step of embryonic mammary development is proliferation through the fat pad precursor mesenchyme, which creates the foundation for the ductal structure[5]. The more complex and mature ductal structure forms at puberty, with most proliferative activity occurring in the terminal end buds (TEBs). TEBs consist of a single layer of undifferentiated cap cells, which give rise to the myoepithelium, surrounding the central body cells, which give rise to the luminal epithelium and eventually undergo apoptosis to form the lumen[6,7]. *In vivo* tracking of these cells has demonstrated that the branching process is driven by a hierarchy of progenitor cells[8]. Throughout adulthood, alveolar buds are formed from estrogen receptor (ESR1)/ progesterone receptor (PGR)-negative luminal progenitors and terminally differentiate into secretory cells during pregnancy to facilitate lactation[5,9]. Further branching of the ductal tree also occurs during this time. Involution occurs after weaning and is characterized by the death and clearing of alveolar and secretory cells that return the ductal tree to a pre-pregnancy-like state[10]. The ability of the

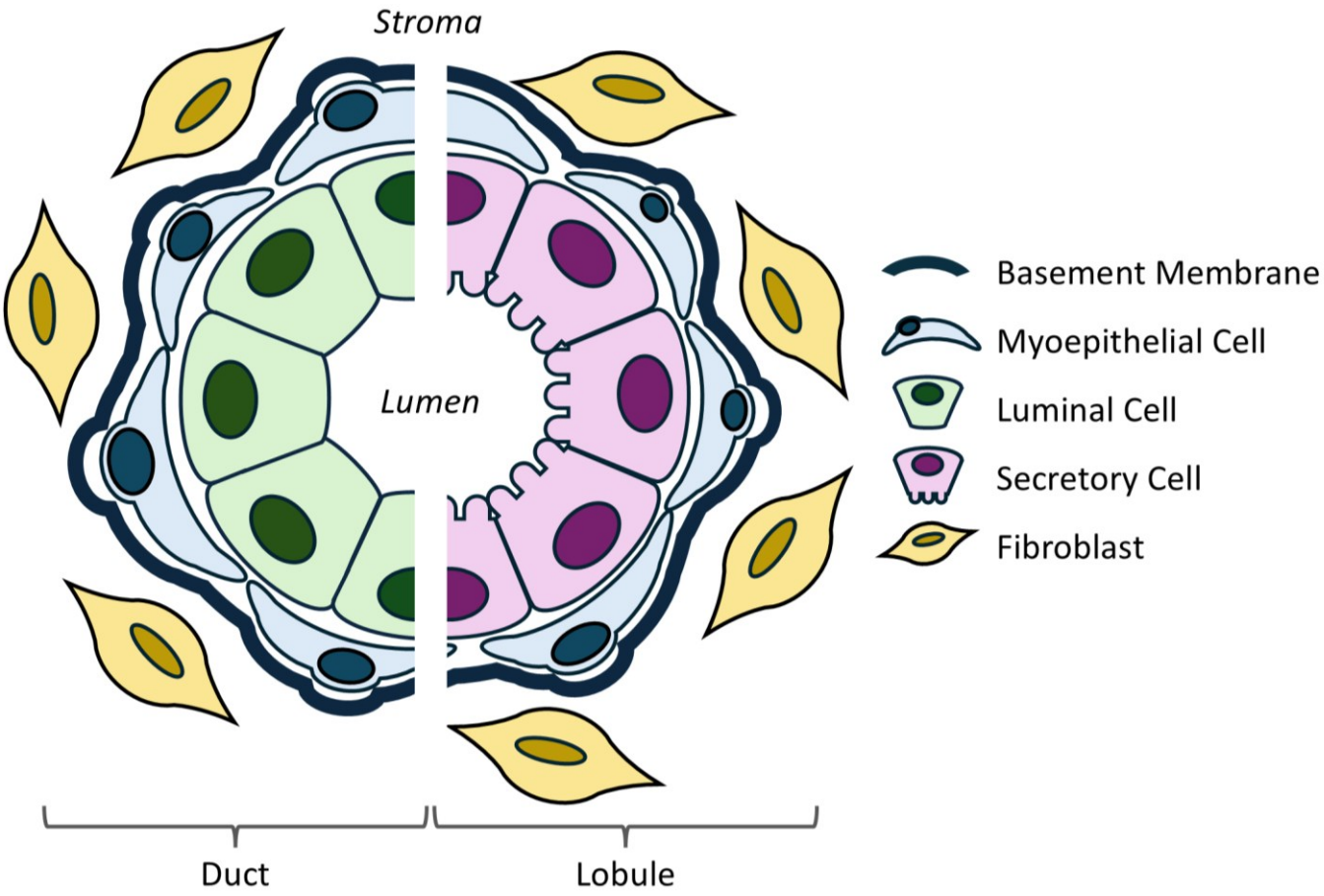


Figure 1.1. Schematic representation of a mature mammary duct cross section. Inner luminal (duct) or secretory (lobule) cells and outer myoepithelial cells are encased in a basement membrane surrounded by fibroblasts from the mammary fat pad stroma. Figure adapted from Pellacani et al[1].

mammary gland to regenerate each pregnancy cycle suggests the presence of a maintained pool of mammary stem cells in adult tissue[11].

1.2. Molecular Subtypes of Breast Cancer

Breast cancer remains one of the most prevalent and lethal malignancies, accounting for more than 2,000,000 new cases and 600,000 deaths globally each year[12]. In 2024, breast cancer accounted for an estimated 25% of new cancer cases and 13% of cancer deaths in female Canadians[13]. It is a highly complex and heterogenous disease. While each case arises from a different mutational path, subtypes sharing molecular characteristics have been established to direct prognosis and treatment decisions.

The molecular classification system of breast cancer is based on tumour gene expression profiles and the presence or overexpression of specific surface receptors on the tumour cells. Most commonly, breast cancer is divided into three molecular subtypes: hormone receptor positive breast cancer, human growth factor receptor 2 (HER2)-positive breast cancer, and triple negative breast cancer (TNBC)[14]. Recent studies point to the emergence of two additional subsets: triple positive breast cancer (TPBC), and HER2-low breast cancer, which share predominating molecular characteristics with other subtypes but still possess a potentially targetable level of HER2[15–17].

Most cases are hormone receptor positive and therefore categorized as one of the luminal subtypes. Together luminal A and B cases account for about 60% of diagnoses and are characterized by the overexpression of ESR1 and PGR receptors[18]. The luminal B subtype is more proliferative than luminal A, noted by the increased expression of Ki67, and tends to manifest with more intra-tumoural heterogenity[18]. These subtypes are associated with the least aggressive disease progression and most favourable prognosis. Patients with luminal breast cancer typically present

without secondary tumours, greatly improving their outcome. In those with metastatic disease, there is a preference for bone metastases, which is a prognostic marker of a more favourable outcome than other sites of metastasis[19]. Treatment typically consists of chemotherapy and endocrine therapy to better target ESR1/PGR mediated signalling, which has been found to improve the five-year recurrence by about 50% compared to chemotherapy alone[20–22].

HER2-positive breast cancer accounts for 20-30% of diagnoses and is associated with a less favourable prognosis and more aggressive disease progression than hormone receptor positive breast cancers[19,23–25]. This subtype is characterized by the overexpression of HER2, typically measured by immunohistochemistry (IHC) and *in situ* hybridization (ISH), where a score of 3+ by IHC or a positive ISH result is considered HER2-positive[23,24]. About 10% of patients present with metastases[19]. Visceral metastases tend to be the first site of secondary tumours, which are associated with a better outcome than brain metastases but a worse outcome than bone metastases[19]. First line treatment consists of chemotherapy in conjunction with a HER2-targeting monoclonal antibody, but up to 70% of patients will develop resistance[26–28]. HER2-positive breast cancer is also associated with a higher rate of recurrence than average[23].

TNBC is the most rare subtype, accounting for approximately 10-15% of diagnoses[29]. It is also considered the most aggressive subtype and is associated with the least favourable prognosis. TNBC is characterized by the lack of expression of ESR1, PGR, and HER2, and therefore possesses no unique receptors for targeted therapy[14,29]. Treatment options are extremely limited, and chemotherapy is the standard of care[14]. TNBC arises in the luminal mammary epithelium but adopts a basal gene signature and expresses basal markers such as cytokeratin 5/6, due to a later phenotype transition[18,30]. This subtype is highly metastatic and has an increased propensity to

metastasize to the brain compared to the other subtypes, contributing to its less favorable outcome[19].

Recent advances have identified HER2 expressing tumours that, while previously considered ineligible, are sensitive to novel HER2 targeted therapeutics, especially to antibody-drug conjugates (ADCs)[15–17]. This has led to the emergence of two classifications: TPBC and HER2-low breast cancer. TPBC is a subset of luminal A/B breast cancer and is defined by the expression of ESR1, PGR, and HER2[15]. It is more aggressive than the luminal A/B subtypes but is more targetable than luminal or HER2-postive tumours due to the expression of additional receptors[15,16]. Some studies, however, report an increased risk of developing resistance to hormone therapy, potentially due to crosstalk of HER2 and ESR1 pathways[31,32]. The HER2-low subtype is a subset of TNBC and is only defined by the presence of HER2, but at a relatively lower expression than what is seen in the HER2-postive subtype[15]. This is generally defined as an IHC score of 1+ or 2+, and a negative ISH result[33]. Since there is less expression of HER2, HER2-low tumors are less responsive to treatment than HER2-positive tumours, thus, the subtype is associated with a less favourable prognosis than HER2-positive breast cancer[15]. Taken together, HER2-positive, TPBC, and HER2-low cases make up about 50% of all tumours[34].

1.3. HER2-Positive Breast Cancer

1.3.1. The EGFR Family

The human epidermal growth factor receptor (EGFR) family comprises four receptor-tyrosine kinases: HER1 (EGFR, ErbB1), HER2 (ErbB2, Neu), HER3 (ErbB3) and HER4 (ErbB4)[35]. It has become one of the most studied receptor families since its discovery. Although each member is unique, they share a general structure (Figure 1.2). EGFR family proteins possess an N-terminal extracellular domain comprised of two leucine-rich ligand binding domains and cysteine-rich

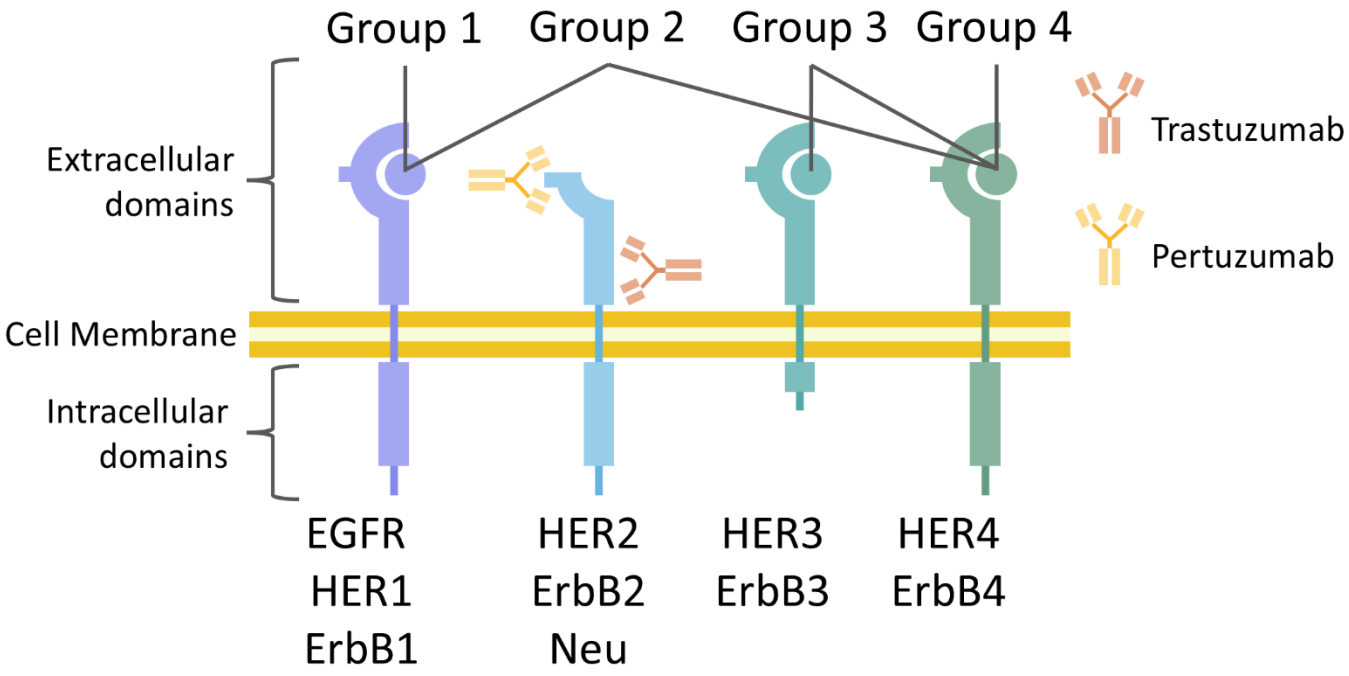


Figure 1.2. Schematic representation of EGFR family structure and ligands. HER1 and HER4 contain all functional domains. HER2 lacks a known ligand and is always in the open conformation, and HER3 lacks a functional intracellular protein kinase domain. Ligands are grouped according to their compatible receptor: group 1 binds exclusively HER1, group 2 binds HER1 and HER4, group 3 binds HER3 and HER4, and group 4 binds exclusively HER4. The two most common monoclonal antibody treatments for HER2, trastuzumab and pertuzumab, bind distinct epitopes at the extracellular domain near the cell membrane and dimerization domain, respectively. Figure adapted from Roskoski[40].

dimerization regions; a transmembrane domain; and a C-terminal intracellular domain that contains a kinase region and a carboxyterminal tail[36]. However, HER3 does not harbor a functional protein kinase domain[37]. These growth factor receptors are found in most non-hematopoietic tissues and are involved in signalling pathways that regulate cell cycle, differentiation, proliferation, apoptosis, and tissue development. Murine models with systemic EGFR family knockouts have been reported with embryonic or perinatal lethality due to developmental dysfunction in a variety of organs, especially in cardiac and brain tissue[38,39].

Ligands of the EGFR family are categorized according to their compatible receptor[40]. The first group consists of the HER1-exclusive ligands: epidermal growth factor, amphiregulin, epigen, and transforming growth factor alpha[41–44]. The second group are ligands that bind to HER1 and HER4: betacellulin, epiregulin, and heparin-binding epidermal growth-like factor[45–48]. The last two groups are comprised of neuregulins that bind to HER3 and HER4 (group 3) or exclusively HER4 (group 4)[40,49–53].

In addition to ligand binding, heterodimerization or homodimerization is required for activation[54,55]. Ligand binding induces a conformational change in HER1, HER3, and HER4 that allows for dimerization. Dimerization induces the autophosphorylation of multiple tyrosine residues in the c-terminal tail, triggering downstream signalling[56]. As HER2 has no known ligand, it is always in the open conformation and able to dimerize[54]. While HER3 is able to form a functional homodimer, it rarely occurs and only possesses a fraction of the kinase activity of other family members[57]. Generally, heterodimerization is preferred to activate the downstream signalling of both partners[54].

1.3.2. The HER2 Oncogene

HER2 is an orphan receptor and does not have any known ligands. It is the preferred heterodimerization partner of each of the other family members and preferentially dimerizes with HER3[54]. Due to its conformational capacity for dimerization, HER2 is able to homodimerize and activate its downstream signalling pathway independent of ligand binding, especially in conditions where it is overexpressed[54,58]. Heterodimerization or homodimerization with HER2 both result in the phosphorylation of tyrosine residues within the c-terminal tail, initiating a protein cascade that activates PI3K, MAPK, and src-signalling, promoting proliferative and pro-survival/ anti-apoptotic pathways. Under normal conditions, HER2 is involved in a variety of developmental processes, including normal mammary gland development and cardiac and skeletal muscle function. HER2 has an integral role in mammary ductal branching but the mechanism through which the signalling pathway is regulated has not been fully elucidated[59,60].

When overexpressed, HER2 activates pro-survival pathways contributing to the development of a number of cancers, especially breast, lung, and gastric cancers[61]. Fluorescent *in situ* hybridization (FISH) revealed that normal cells each have one-to-five gene copies of HER2 but HER2-positive breast cancer cells have at least 25 copies, and immunofluorescent staining and immunoblotting have shown up to a 100-fold increase in amplification[62,63]. Due to its preferential binding in the EGFR dimerization hierarchy, the high-level expression results in an increase in HER2-containing dimers, and ligand-independent hyperactivation of downstream HER2 signalling. This effect is further enhanced not only by HER2's ability to self-phosphorylate through homodimerization, but also its ability to avoid normal regulatory mechanisms. HER2 appears to inhibit dimer degradation, allowing its dimerization partners to avoid endocytic degradation, be recycled and quickly reactivated[64,65]. HER2 amplification and copy number in human breast

tumours remains constant throughout disease progression, and is associated with subtype-defining characteristics rather than providing staging information[66,67].

HER2 variants with constitutively active signalling or otherwise increased tumorigenicity have also been identified. The most characterised isoform is P95HER2, a fragment with an incomplete extracellular domain[68]. The fragment arises from two mechanisms: the proteolytic shedding of the extracellular domain catalyzed by ADAM10, and by alternative messenger RNA (mRNA) translation from internal initiation codons[69,70]. P95HER2 has been found in about 20% of all breast cancer cases and in 30% of HER2-positive breast cancer cases, and is associated with increased treatment resistance to HER2 targeted therapy as inhibitory molecules cannot bind to the external portion of the receptor[71,72]. Defects in the extracellular domain have been found to increase the transforming efficiency of HER2 by up to 100-fold[73]. For these reasons, P95HER2 and related variants are considered prognostic markers of worse outcome and may help shape treatment decisions.

Somatic mutations can contribute to HER2 driven tumorigenesis, although these are quite rare. Less than 5% of breast cancer cases are estimated to express mutant HER2, and the rate of HER2 mutations in other cancer types is not much higher[74]. HER2 mutations are most common in bladder cancer, where they are only found in an estimated 8% of cases[74]. The most common HER2 mutations across all cancer types are found in the protein kinase domain and result in constitutive activation[74]. The most prevalent mutation is a 12 base pair insertion in exon 20 resulting in a tandem repeat of tyrosine-valine-methionine-alanine (YVMA)[74]. $HER2^{YVMA}$ has been extensively studied in the context of lung cancer, where it has been found to increase the dimerization affinity of HER2 and the subsequent trans-phosphorylation of other EGFR family members, especially in dimers containing HER1 and HER3[75]. While the biological effects of HER2 mutants are clear,

there are conflicting reports in the literature as to whether HER2 mutations alone are a risk factor for cancer development, as reviewed by Galogre et. al[76]. Regardless, given the rarity of these occurrences, the main driving force behind HER2-mediated tumorigenesis is its overexpression.

The oncogenic potential of HER2 has been extensively characterized. The link between HER2 and cancer was first reported in a study where mouse fibroblast cells were transfected with DNA from rat neuroglioblastomas, resulting in their transformation with characteristics consistent with fibrosarcoma[77]. One of the transforming genes encodes a protein identified as Neu, the rat homolog of human HER2[78]. Further cloning of the wild-type form of Neu revealed that it was not sufficient to transform cells, instead the transforming ability was due to a point substitution mutation in the transmembrane domain, V664E, that increased tyrosine kinase activity[79,80]. This mutation has been named NeuT. Overexpression of the wildtype c-Neu in mouse mammary glands was only shown to have transformation activity after undergoing small deletions in the juxtamembrane domain, resulting in increased dimerization and tyrosine kinase activity[81]. This is in contrast to transformation by HER2. Overexpression of wildtype HER2 was shown to be sufficient to induce transformation *in vitro* in a number of cell lines, including mouse fibroblasts and human mammary epithelial cells. Transformation by HER2 had similar effects on proliferation, invasion, drug resistance, and apoptotic resistance[82–86].

1.3.3. HER2-Targeting Treatment

Treatment options for HER2-positive breast cancer have expanded over recent years but are still quite limited. First-line treatment consists of a chemotherapeutic in conjunction with a HER2 targeting monoclonal antibody. The first monoclonal antibody introduced for the treatment of HER2-positive breast cancer (trastuzumab) was initially approved in 1998 in combination with the chemotherapeutic paclitaxel but it has since been approved for use in other HER2-expressing solid

tumours, such as in gastric and colorectal cancer, and on its own as an adjuvant therapy in early HER2-positive breast cancer[28]. It was first isolated in mice and humanized to prevent an immune response[87,88]. Trastuzumab directly binds to the extracellular domain near the cell membrane, but the mechanisms through which it inhibits downstream HER2 signalling are not fully understood[89]. The use of trastuzumab in conjunction with chemotherapy greatly improves survival and recurrence outcomes over chemotherapy alone, however the rate of acquired resistance to trastuzumab is 70%[26–28]. As it specifically binds to the HER2 extracellular domain its strongest effects are seen in cases driven by HER2 homodimers[58]. It is not effective in cases where the P95HER2 fragment is already expressed, nor is it effective at preventing HER2-HER3 dimers[90,91]. These findings indicate that cells may overexpress HER3 as a compensatory mechanism, and that individuals with high levels of P95HER2 may be innately resistant.

Pertuzumab is a hamster derived-humanized monoclonal antibody that inhibits downstream signalling through steric dimerization inhibition. It has also been found to increase affinity for antibody-dependent cellular cytotoxicity, though this is not its main mechanism of action[92]. The epitope for pertuzumab does not overlap with the epitope for trastuzumab, and the combined use of both has synergistic effects[89,92]. Adding pertuzumab to the regimen was found to increase overall survival and decrease recurrence compared to trastuzumab, chemotherapy, and a placebo group with mild additional side effects[93]. Combining trastuzumab with pertuzumab blocks HER2-HER3 dimerization as a compensatory mechanism and reduces the rate of acquired resistance. Pertuzumab was approved in 2012 for use in combination with trastuzumab and the chemotherapeutic docetaxel to treat metastatic HER2-postive breast cancer, but the approval has since been expanded to include its use as an adjuvant and neoadjuvant treatment in early HER2-

positive breast cancer, still in combination with trastuzumab[93,94]. This combination is the most often prescribed treatment course for HER2-positive breast cancer.

The search for novel HER2-targeting therapeutic options is ongoing. Among the most promising options are ADCs. ADCs combine the targeting ability of monoclonal antibodies with the cytotoxic effects of chemotherapeutics delivered as one entity. For the treatment of HER2-positive breast cancer, two are currently used in the clinical setting. Ado-trastuzumab emtansine (T-DM1) was first approved in 2013, and fam-trastuzumab deruxtecan-nxki (T-DXd) was approved in 2019. T-DM1 combines trastuzumab with emtansine, a tubulin inhibitor, and was found to improve overall survival in patients with late-stage disease who have undergone prior treatment compared to patients treated with a "physician's choice" regimen[95]. T-DXd instead combines trastuzumab with a topoisomerase I inhibitor and has similarly been found to be effective in patients with pre-treated late-stage HER2-positive breast cancer[16]. Especially interesting is the apparent ability of these drugs to target cells expressing low amounts of HER2, with IHC scores of 1+ or 2+. A phase III clinical trial demonstrating T-DXd's efficacy in patients who express low-levels of HER2 have led to it being the first drug approved to target HER2-low breast cancer[33].

1.3.4. Transgenic Models of HER2-Positive Breast Cancer

Transgenic animal models are an invaluable tool for the study of disease progression. A number of models have been developed for the study of HER2-positive breast cancer. Most current transgenic models utilize the Neu oncogene driven by the mouse mammary tumour virus (MMTV) promoter[96]. The MMTV promoter fragment is primarily expressed in luminal mammary epithelium and myoepithelium but is also lowly expressed in the salivary glands and testes due to "leaky" expression[97].

A number of mutations in Neu have been identified for the study of breast cancer. The first was designated as Neu-T, which contains an activating point mutation in the transmembrane domain[79]. Expression of Neu-T was sufficient to transform cell lines *in vitro*, and MMTV-driven expression *in vivo* results in mammary tumour formation consistent with human breast cancer[79,98]. In Neu-T transgenic female mice, tumours involving the whole mammary gland epithelium would arise in approximately 90 days[98]. This model was representative of a somatic mutation, which is rare in human breast cancer, and this led to the investigation of the transforming potential of wild-type Neu. A separate study found that transgenic female mice expressing the wild-type c-Neu driven by MMTV would form tumours in 5-10 months, even though the level of expression was similar to that of Neu-T[97,99]. The tumours that formed in this model almost exclusively contained Neu mutants with in-frame 7-12 amino acid deletions in the juxtamembrane domain and an *in vitro* experiment demonstrated the increased transforming potential of these mutants compared to the wild-type [81]. These results indicate that tyrosine kinase activity is the rate determining factor in mammary tumorigenesis.

Neu deletion 2-5 (NDL2-5) is a mutated form of Neu containing a five amino acid deletion that results in constitutive activation through the promotion of stable dimers, due to a lack of intermolecular cysteine bridges[100]. This is similar to a splice variant of HER2 found in human breast cancers where a 16 amino acid deletion prevents cysteine bridges, resulting in the formation of a dimer stabilizing disulfide bond[81,100]. The MMTV-NDL2-5 model uses the MMTV promoter to drive expression of the NDL2-5 oncogene. Tumours in this model spontaneously arise after about 160 days and are pathologically consistent with human breast carcinoma[100]. Metastases to the lungs are evident in 65% of the mice[100]. This oncogene is also used in a model designed to facilitate knockout studies, NDL2-5-IRES-Cre (NIC). The MMTV-NIC model allows for the

conditional co-expression of the NDL2-5 oncogene and Cre-recombinase through an internal ribosome entry site (IRES)[101]. In this model, a conditional allele can be deleted upon expression of the NIC cassette[102].

Mouse background has been shown to have an effect on tumour progression in models of breast cancer. Typically, the two most common backgrounds are FVB or C57B1/6. FVB mice are desirable for their high fecundity while C57B1/6 mice are considered best suited for knockout studies due to their increased capacity for germline transmission facilitated by 129/SvJ embryonic stem cells[103]. C57B1/6 mice have been shown to be resistant to HER2-driven tumorigenesis in two transgenic models. In models of HER2-positive breast cancer driven by Neu or Polyomavirus middle T (PyVMT), a transgenic model that mimics downstream HER2 signalling, tumour initiation was observed to be nearly twice as long in the C57B1/6 mice compared to FVB mice despite consistent levels of transgene expression in both strains[103,104]. Multiple genes have been identified to have a role in the suppressive effect[104]. These studies demonstrate that maintaining a consistent background is integral for the interpretation of *in vivo* results.

1.4. Sox9

1.4.1. Sox Family Proteins

The sex-determining region Y high mobility group-box (SRY-HMG box; Sox) family consists of 20 transcription factors in mammals. They were first discovered to have a role in male embryonic sex determination and are named for their transcriptional regulation of Y-chromosome genes[105,106]. They have since become known to be critical in a variety of processes for cell differentiation, tissue development and homeostasis, and contribute to various pathologies. The Sox family members share a highly conserved single homology box domain that binds the minor groove of AT-rich DNA sequences and bends DNA to regulate transcription either directly or through chromatin

remodelling[107]. The Sox homology box consists of two α-helical domains and an acidic tail that negatively regulate DNA binding through intramolecular binding[107,108]. When in the open conformation, this structure allows Sox proteins to control the expression of target genes by selectively binding to the consensus motif: 5'-(A/T)(A/T)CAA(A/T)G-3'[109]. Sequences flanking this region are specific to different Sox proteins[110].

Sox family members are subgrouped by sequence homology and similarities in biological function[111]. Nine groups have been defined in mammals: A, B1, B2, and C through H. Sox proteins within a subgroup typically share more than 50% sequence homology[111]. There tends to be significant overlap in their functions and often evidence of compensatory mechanisms amongst them.

1.4.2. SoxE Subgroup

The SoxE subgroup consists of three proteins: Sox8, Sox9, and Sox10[111]. Initially they were discovered along with 4 other Sox proteins based on their sequence homology to SRY[111]. Structurally and functionally, the members are very similar. Each member has highly conserved dimerization, HMG, and two transactivation domains[112]. Within the HMG domain is a nuclear export signal (NES) and two nuclear localization signals that flank the NES at the N and C termini[113,114]. In addition to their structural similarities, SoxE members also share functional similarities such as in male sex determination, chondrogenesis, neural crest development, and glial cell differentiation pathways. Dimerization is a property unique to the SoxE group and is required for chondrogenesis but not sex determination[115]. In recent years the SoxE subgroup has been implicated in numerous cancer types by contributing to stem and progenitor cell regulation and tumour progression.

Sox8 is the least characterized member of the SoxE subgroup. Like most Sox proteins it has a role in embryonic development, and is expressed in mouse embryos starting at 9.5dpc[116]. In adult human tissues it is most highly expressed in the brain[117]. Mutations or partial deletion of Sox8 have been shown to contribute to intellectual disabilities and infertility[117,118]. Although it is associated with male infertility, global knockouts of Sox8 in mice are not associated with other severe dysfunction, indicating functional redundancy with other SoxE proteins[116,119]. Sox8 is overexpressed in many cancer types, especially glioblastostoma, medulloblastoma, and TNBC[120–122]. It has been implicated in a number of malignant pathways, particularly in the regulation of proliferation through Wnt/β-catenin signalling[123,124].

Sox9 is the most studied SoxE member. It was first investigated when a Sox9 mutation was observed to be a contributing factor to campomelic dysplasia, a syndrome defined by skeletal abnormalities, sex reversal, and neonatal lethality[125]. This highlights the main functions of Sox9 in development: sex determination and chondrocyte differentiation. Haploinsufficiency of Sox9 was also connected to campomelic dysplasia, indicating a dosage-dependent requirement for Sox9 during embryonic development[126]. Sox9 is first expressed in mouse embryos by 9dpc and initially is involved in early cartilage and skeletal formation[127,128]. Throughout development it becomes more generally expressed in other tissues and is involved in testis differentiation and nervous system development[129,130]. Global knockouts of Sox9 are embryonically lethal, and conditional knockouts exhibit stronger phenotypes when more than one SoxE member is affected rather than just Sox9 alone, suggesting compensatory activity by Sox8 and Sox10[131,132]. Sox9 has been implicated in a number of malignancies, especially by regulating proliferation, stemness, and polarity in breast cancer, non-small cell lung cancer, central nervous system (CNS) cancers**,** and

gastrointestinal cancers[133–137]. Its expression has also been found to contribute to drug resistance in luminal breast cancer[138].

The last member of the SoxE subgroup, Sox10, is more similar to Sox9 than to Sox8[111]. Sox10 is critical for neural crest and peripheral nervous system development. In murine embryos, Sox10 becomes activated around 9.5dpc, during peripheral nervous system development, and later becomes integral to Schwann and glial cell differentiation in the central nervous system[139,140]. Global knockout of Sox10 is embryonically lethal, and mutated forms have been linked to severe developmental disorders, most notably demyelinating syndromes[141–143]. Sox10 has been shown to be a marker of basal-like breast cancer or TNBC and contributes to a less favourable prognosis in breast cancer through regulation of proliferation and stemness[144–146].

1.4.3. Sox9 Structure and Regulation

Sox9 was first sequenced and identified during a screen for proteins related to SRY[111]. It was designated as part of the SoxE subgroup due to its sequence homology with two other novel genes, Sox8 and Sox10[111]. It was further isolated in mice and found to be located on chromosome 11 in a region syntenic to human chromosome 17q[125].

Murine and human Sox9 mRNA transcripts contain three exons separated by two introns (Figure 1.3)[147]. The coding sequence spans all three exons and in humans encodes a protein that is 509 amino acids long[147]. Like other SoxE proteins, Sox9 contains a dimerization (DIM) domain, the HMG, a central transactivation (K2/TAM) domain, and a PQS-rich c-terminal transactivation (TAC) domain[112]. Uniquely, it contains a PQA domain comprised exclusively of proline, glutamine and alanine[148]. The DIM is a region exclusive to SoxE subgroup members and facilitates homodimerization or heterodimerization necessary for the transcriptional regulation of target genes through interactions between the DIM of one monomer and the HMG of the other[149].

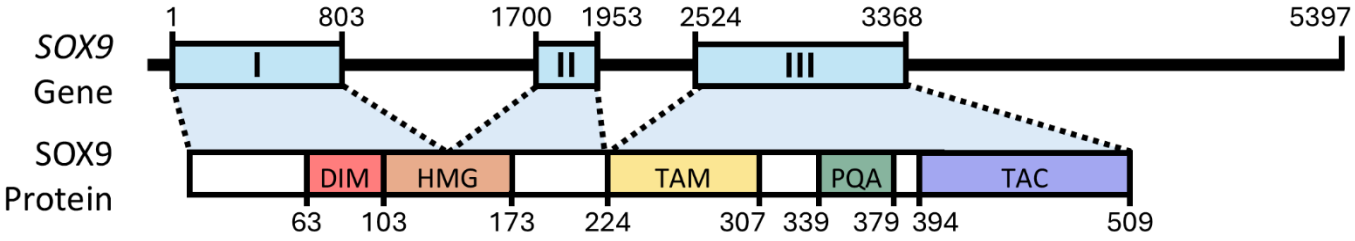


Figure 1.3. Schematic representation of human *SOX9* gene and SOX9 protein structure. The *SOX9* coding sequence spans all three exons and encodes a protein that contains dimerization (DIM), homology group (HMG), central transactivation (TAM), PQA, and c-terminal transactivation (TAC) domains. Roman numerals indicate exons, numbers above *SOX9* gene indicate base pair relative to the start codon, and numbers under SOX9 protein indicate amino acid position. Figure adapted from Angelozzi and Lefebvre[204].

Dimerization is only required in some processes, implying that it is a context dependent mechanism[115]. The TAM and TAC domains are both required for transactivation, but the TAC domain is more integral to the function of Sox9 and can directly interact with other proteins[150,151]. These two regions have been found to function independently or synergistically, depending on cellular context[112]. The PQA domain is able to enhance transactivation but the mechanism responsible is not fully known[148,152].

Transcriptional regulation of Sox9 is quite complex and has yet to be fully understood. Patterns of expression throughout early development indicate that the regulation of Sox9 begins at the transcriptional level. During embryonic development, Sox9 activity is first detected during chondrocyte differentiation in a number of progenitor cells. Case studies involving campomelic dysplasia patients with Sox9 mutations outside the coding DNA sequence (CDS) indicate that a 3MB region located up-and-downstream of the gene may be a regulatory region[153]. Various sequences within this region likely act as context-dependent enhancers for Sox9 transcription[153,154]. Some studies with transgenic mice have begun to elucidate this mechanism, but homology is not always conserved in humans[155–157]. Methylation also contributes to the transcriptional regulation of Sox9. The gene has been found to be hypomethylated and therefore over-activated in breast cancer, especially following chemotherapy[158].

The post-transcriptional regulation of Sox9 is better understood. Specific regulatory mechanisms are highly context dependent and involve multiple pathways that direct nuclear localization, binding affinity, and degradation of Sox9. The most common modification is phosphorylation, which increases the binding affinity of Sox9 for DNA and increases transcription of its downstream targets[159–161]. In some contexts, this step is required. Another mechanism is ubiquitination by a small ubiquitin-like modifier (SUMOylation). SUMOylation can be activating

or inhibitory depending on cellular context[162–164]. Further activating or inhibitory post-transcriptional regulation of Sox9 is mediated by microRNA and long noncoding RNA, as reviewed by Jana et. al[165].

1.4.4. The Role of Sox9 in Development and Homeostasis

During embryonic development Sox9 plays key roles in brain development, sex determination, and chondrogenesis. In the central nervous system, Sox9 expression begins at 9.5dpc and is required for the generation and maintenance of neural stem cells[130,166]. Later, Sox9 also regulates the expression of targets necessary for glial, astrocyte, and oligodendrocyte differentiation[166,167]. Following the initiation of the male sex determination process by SRY, Sox9 promotes the differentiation of Sertoli cells to drive testis development, with low-level expression evident as early as 10.5dpc[168]. Sox9 expression is initially dependent on SRY but continues after SRY expression decreases and is able to drive male gonadal differentiation without a Y-chromosome, indicating prolonged self-regulation[155,168,169]. Throughout this process, Sox9 activates a number of targets that are necessary for ovarian pathway repression and differentiation of Sertoli cells and Leydig cells[129,170].

At the onset of chondrogenesis, Sox9 is involved in the determination of chondrocytes from osteochondroprogenitor cells[163,171]. From 11.5dpc onward Sox9 is highly expressed in nearly all chondrocyte derived cells throughout the process of chondrogenesis and expression remains elevated in mature cartilage[128,172,173]. Sox9 is involved in a number of pathways throughout chondrocyte development, and many of its downstream targets are chondrocyte specific genes, but also other transcription factors, including Sox5 and Sox6[174]. Other organs in which Sox9 is expressed and promotes differentiation include the lungs, heart, gastrointestinal tract, liver, pancreas, and retina[175–180].

Sox9 remains expressed in a number of mature tissues to maintain homeostasis and normal function. In cartilage, Sox9 is necessary to facilitate repair after injury and prevent disorders like osteoarthritis by continuing to regulate stem cells[181,182]. In the liver, Sox9 is always expressed and plays a role in cell differentiation during tissue regeneration[183]. A similar mechanism has been observed in the pancreas and intestine where Sox9 expression is a requirement to maintain progenitor cells in adult tissue[184]. Sox9 is also involved in post-natal development. Conditional knockout of Sox9 was observed to result in impaired mammary gland development in mice[185]. Sox9 is expressed by luminal progenitors and is necessary for differentiation and proliferation of luminal cells[185]. The activation of Sox9's targets appear to be context dependent. However, pathways of stem cell maintenance and lineage commitment tend to involve targets from the TGF-β, Wnt/β-catenin and Notch pathways.

1.4.5. The Role of Sox9 in Cancer

Consistent with its role in proliferative and pro-survival pathways in numerous tissues, Sox9 has been implicated in the development and progression of many cancer types. Notably, Sox9 has been shown to be a biomarker for less favorable outcome in non-small cell lung cancer, hepatocellular carcinoma and CNS tumours [137,186–188]. It is also a regulator of cancer stem cells (CSCs) and promotes stemness in prostate cancer, renal cell carcinoma, colorectal cancer, liver cancer, and pancreatic cancer[134,188–190]. In many cancers, Sox9 is found to be upregulated with Sox10 or Sox2, indicating that they are promoting similar pathways, with evidence of cooperation in some contexts[191,192]. Regulation of stemness in cancer often involves targets or crosstalk with the TGF-β, Wnt/β-catenin and Notch pathways, the same pathways typically promoted by Sox9 for the maintenance of normal stem cell ratios.

In breast cancers, Sox9 has been shown to be co-expressed with Sox10. Together they have been associated with acquired treatment resistance to hormone therapy in luminal breast cancer[138]. In TNBC, Sox9 has been shown *in vitro* to promote both proliferation and stemness through growth and invasion assays[136]. Sox9 is also highly expressed in HER2-postive breast cancer, but its role has not been fully elucidated in the subtype.

1.4.6. Sox9 and Sox10 in Breast Cancers

Sox9 and Sox10 are active throughout mammary gland development. Sox9 is expressed in the ESR1/PGR-negative luminal progenitor population, and to a lesser degree in basal cells, where it appears to have a role in the dedifferentiation of unipotent luminal progenitor cells to bipotent progenitor cells and multipotent mammary stem cells[9,185,193]. Sox10 is present throughout the mammary gland and is strongly expressed in basal cells where it has been demonstrated to be required for myoepithelial progenitor function[194]. It has also been shown to contribute to stem cell maintenance through the dedifferentiation of basal progenitors to multipotent mammary stem cells[194]. Both transcription factors have been shown to be required for normal mammary gland development in mice[185,195].

Some studies have connected molecular signatures of breast cancer subtypes to normal mammary cells as their cell type of origin. TNBC appears most similar to early luminal progenitor cells, while HER2-positive breast cancer is most similar to ESR1/PGR-negative alveolar cells, and the luminal breast cancers are most similar to ESR1/PGR-positive luminal cells[196,197]. These genetic similarities suggest potential cell types of origin from differentiation states that are consistent with pathological features of each subtype, like stemness (Figure 1.4). However, due to the high degree of plasticity between cellular states in both normal mammary tissue and breast cancer this is likely an oversimplification of the actual oncogenic processes.

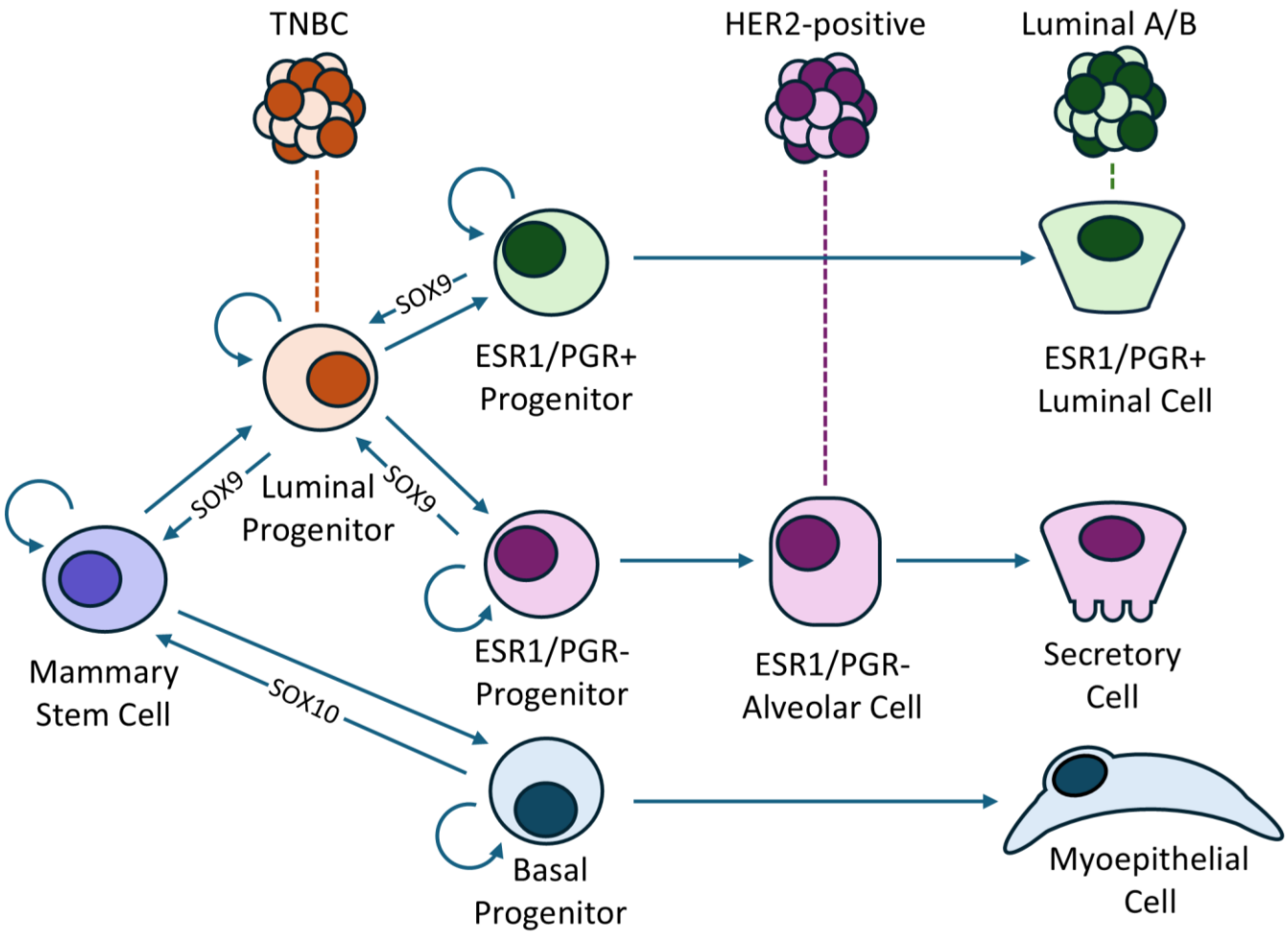


Figure 1.4. Schematic representation of potential relationships between normal mammary cell differentiation states and the origin of human breast tumours. Multipotent mammary stem cells give rise to unipotent progenitors that further differentiate into ductal, secretory, or myoepithelial cells. Various transcription factors facilitate the plasticity between cell states. Dotted lines connect the molecular subtypes to their closest normal cell based on genetic profiling. Figure adapted from Fu et. al[11].

1.5. Stemness and EMT

A key mechanism of acquired resistance in breast cancer is epithelial-to-mesenchymal transition (EMT). EMT generally describes the process through which epithelial-like tumour cells acquire mesenchymal-like characteristics. The resulting mesenchymal-like cells often display characteristics of CSCs or tumour initiating cells (TICs) due to their increased metastatic potential. In non-carcinoma malignancies, the process is known as dedifferentiation or phenotype switching. Epithelial-like cells are generally considered to be proliferative, treatment-sensitive, and non-invasive, with high expression of epithelial markers like E-cadherin and claudin[198,199]. Mesenchymal-like cells display opposing characteristics. They are known to be self-renewing, treatment resistant, immune evasive, highly migratory, and highly invasive[200,201]. During EMT, epithelial markers are down-regulated, and mesenchymal markers like N-cadherin and vimentin are upregulated[199]. A complementary process, mesenchymal-to-epithelial transition (MET), occurs at distant sites during metastasis when CSCs are establishing a secondary tumour[199]. Together, EMT/MET plasticity is an important mechanism contributing to tumour heterogeneity, treatment resistance, and metastasis. There are multiple pathways through which EMT can be mediated. Most result in the upregulation of EMT transcription factors Zeb, SnaiI, and TWIST. Zeb and SnaiI inhibit epithelial gene expression while TWIST promotes mesenchymal gene expression[200]. In HER2-positive breast cancers the transforming growth factor beta (TGF-β), Notch, and Wnt/β-catenin pathways play a major role in the EMT process.

1.6. Objectives and Hypothesis

Previous work in our lab has shown that Ste-20 like kinase (SLK), a serine-threonine kinase, is activated downstream of HER2. It was initially hypothesized that SLK deletion would inhibit tumour progression and increase overall survival. Surprisingly, our previous studies demonstrated

accelerated tumour onset and increased mammary stem and progenitor cell activity in MMTV-NDL2-5/NIC mice bearing a conditional SLK knockout[202]. These changes were attributed to the induction of Sox10 in mammary cells. SLK is activated by PDK1, but we have shown that it also regulates PDK1 through a negative feedback loop. In the absence of SLK, Sox10 was induced through activation of a PDK1-AKT-Sox9 axis[192]. *In vitro*, phosphorylation of Sox9 by AKT induced Sox10 expression. A summary of this pathway is shown in Figure 1.5.

One possibility is that SLK deletion led to an expansion of Sox10-expressing luminal progenitor cells. Recently, our lab has shown that luminal Sox10 deletion completely abrogates tumour initiation in SLK-expressing MMTV-NIC mice[203(submitted)]. The expression of Sox9 in the luminal epithelial tissue of mammary glands from these mice suggests that Sox9 is unable to compensate for the loss of Sox10. Furthermore, we have shown that Sox10 expressing NDL tumour-derived cell lines exhibit a more stem-like behaviour than Sox10-deficient cells[203(submitted)].

Our hypothesis is that Sox9 is driving a distinct genetic program required for invasion, stemness, and tumour progression following the Sox10-dependent induction of hyperplastic lesions. Our first objective was to determine the role of Sox9 on stemness *in vitro*. Using HER2-positive tumour-derived cell lines we have tested the effect of Sox9 on proliferation, self-renewal ability, migration, and invasion. Our second objective was to assess the effect of Sox9 on tumour progression and metastasis *in vivo.* Tumour growth was assessed by orthotopic injection of Sox9-expressing and deficient HER2-positive tumour cells. Metastasis was evaluated by tail-vein injection of the same cells.

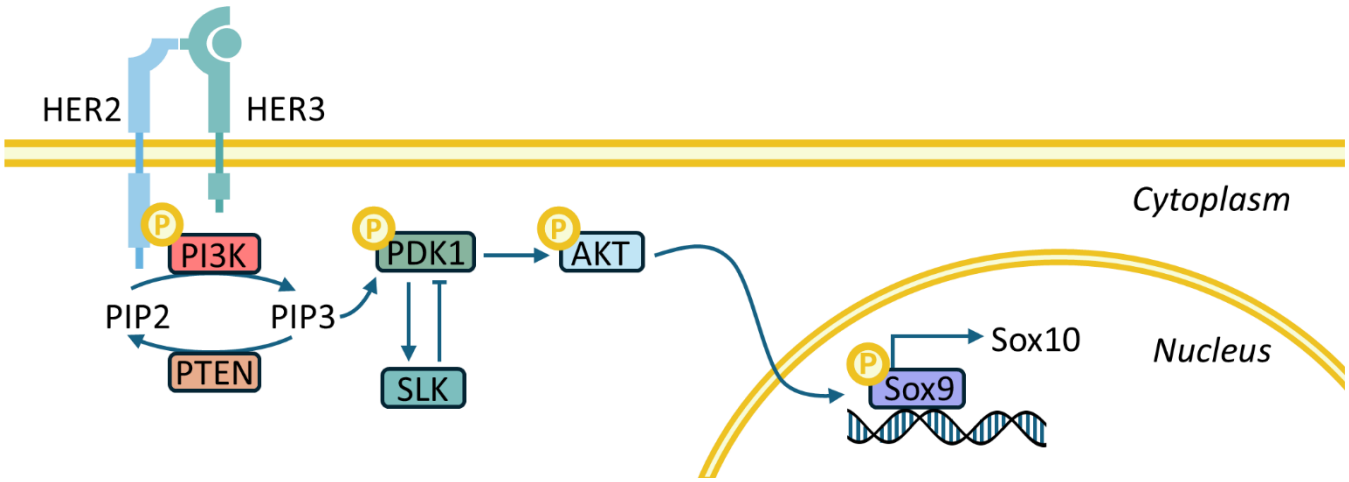


Figure 1.5. Summary of Sox9 activation downstream HER2 signalling. In HER2 mediated tumorigenesis, HER2 is activated through heterodimerization or homodimerization and phosphorylates PI3K. PI3K converts PIP2 into PIP3, which then phosphorylates PDK1. AKT is activated by PDK1. SLK is thought to negatively regulate PDK1 through a negative feedback loop. Loss of SLK increases AKT activity, Sox9 activation, and Sox10 upregulation. Figure adapted from Al-Zahrani[205].

2. Materials and Methods

2.1. Plasmids and Cloning

The following plasmids were purchased from Addgene: pMSCV-IRES-GFP II (pMIGII, 52107), pUMCV (8449), and pCMV-VSV-G (8454). The Sox9 mouse tagged ORF clone (MR237031) was purchased from Origene. The following plasmids were kindly gifted from Dr. Daniel Schramek (Lunenfeld-Tanenbaum Research Institute, Sinai Health): pLKO-H2B-mRFP-2A-puro and pLentiCas9-blast. Single guide RNA (sgRNA) targeting the exon 1 region of Sox9 were designed with Benchling (Benchling.com) and ChopChop (https://chopchop.cbu.uib.no). Non-targeting control sgRNA (sgNTC) were designed to not bind DNA. All oligonucleotide sequences are listed in Table 2.1.

To generate the murine and human Sox9 knockout guide plasmid the lentiviral vector pLKO-H2B-mRFP-2A-puro was digested with BsmBI-V2 (NEB: R0739S) and NEBuffer 3.1 (NEB: B7203S) at 55°C for 1 hour and the 9kb band was purified from a 0.8% agarose gel with a QIAQuick gel purification kit (QIAGEN) following the manufacturer's protocol. The purified vector was dephosphorylated with rSAP (NEB: M0371) by incubating at 37°C for 30 minutes and heat inactivating by incubating at 65°C for 5 minutes. The product was purified with a QIAQuick PCR purification kit (QIAGEN), following the manufacturer's protocol. Sense and antisense sgRNA were annealed using T4 ligase buffer (NEB: B0202A) and T4 polynucleotide kinase (NEB: M0201S) in a thermocycler with the following cycling protocol: 37°C for 30 minutes, 95°C for 5 minutes, ramp down to 20°C at a rate of 0.1°C/ second. The annealed sgRNA were ligated to the dephosphorylated vector using Quick Ligase (NEB: M2200S) and Quick Ligation Buffer (NEB: B2200S), following the manufacturer's protocol. To transform the ligated product, it was combined with 90µL NEB stable competent bug (NEB: C3040H), incubated on ice for 30 minutes,

and heat shocked at 42°C for 1 minute before incubating for 1 hour at 30°C, shaking at 250rpm. The broth was then spread on an agar plate with 100µg/mL ampicillin (Wisent Inc: 400-110-XG) and incubated for 24 hours at 30°C. Colonies were selected and grown in 3mL of LB broth (11% peptone [Gibco: 211677], 1% NaCl, 0.5% yeast extract [Thermo Fischer Scientific: BP1422-500]) for 24 hours at 30°C and 225rpm and DNA was extracted with a PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen) by following the manufacturer's protocol. Positive clones were identified by Sanger sequencing.

The Sox9 overexpression plasmid was previously constructed in our lab. Briefly, the retroviral vector, pMIGII, was digested with BamHI and XhoI. The Sox9 mouse tagged ORF was digested from a subcloning vector with BamHI and SaII and the Sox9 fragment was ligated to the pMIGII vector. The remaining steps were performed in the same manner as described above with the following exceptions: the plasmid was transformed with 90µL DH5α competent bacteria (produced in house), samples were incubated at 37°C instead of 30°C, and for 16 hours instead of 24.

2.2. Cell Culture

2.2.1. Cell Lines

All murine cell lines were isolated from endpoint MMTV-NDL mammary tumours and processed by our lab. Briefly, tumours were digested in collagenase and cells underwent lineage negative depletion by magnetic sorting to isolate luminal cells using the Mammary Stem Cell Enrichment Kit (Stemcell Technologies) according to the manufacturer's protocol without the addition of CD24 or CD49f antibodies. Cells were cultured in the media described in section 2.2.2. Media was changed every 48 hours and the cells were passaged at 80% confluency until they reached a doubling time of approximately 24 hours. The parental human cell line, BT474, was kindly gifted

from Dr. Christina Addison (Ottawa Hospital Research Institute [OHRI] and University of Ottawa).

2.2.2. Routine Cell Culture

Murine cell lines were maintained in culture media containing equal parts Dulbecco's modified eagle medium (DMEM, Life Technologies: 11995123) and F12 (Thermo Fisher Scientific: 11765070) supplemented with 10% fetal bovine serum (FBS, Corning: 35-077-CV), 1% penicillin/streptomycin (Gibco: 15140122), 1% L-Glutamine (Gibco: 25030081), and 1% mammary epithelial growth supplement (MEGS), a mixture of 0.003µg/mL human epithelial growth factor recombinant protein (Gibco: PHG0311), 0.01µg/mL human insulin-like growth factor 1 recombinant protein (Gibco: PHG0071), 10µg/mL bovine pituitary extract (Gibco: 13028014), and 0.5µg/mL hydrocortisone (Thermo Scientific Chemicals: A16292.03). Human cell lines were maintained in DMEM culture media supplemented with 10% FBS, 1% penicillin/streptomycin, and 1% L-Glutamine. All cell lines were cultured in a 37°C humidified incubator with 5% CO_2 and passaged at a dilution of 1:5 - 1:10 every 2-3 days. Mycoplasma testing was performed regularly by polymerase chain reaction (PCR) with the primers indicated in Table 2.1. Cells passaged more than 20 times were not included in experiments.

2.2.3. Transduction, Infection, and Selection

For virus production, HEK293T cells were plated at a density of 5,000,000 cells/ 10cm culture plate. The following day they were transduced with linear polyethylenimine (linear PEI, Sigma-Aldritch: 919012). 8µg packaging plasmid (pCMV-dR8.2 dvpr [lentivirus], or pUMCV [retrovirus]), 2µg of envelope plasmid (pCMV-VSV-G), and 10µg of viral construct were added to a 600µL solution of DMEM with 0.15M NaCl. The mixture was vortexed and incubated for 5 minutes at room temperature. It was then added dropwise to a 600µL solution of DMEM with

0.15M NaCl and 0.118µg/µL Linear PEI. The mixture was incubated at room temperature for 30 minutes and added to the cells. After 8-12 hours of incubation under normal conditions the media was replaced with 5mL fresh media. The cells were incubated under otherwise normal conditions for an additional 48 hours before the media was collected and filter sterilized through a 0.45µm pore filter.

Cells to be infected were seeded at a density of 100,000 cells / 60mm culture plate the day prior to infection. 1mL of filtered media with 8µg/mL polybrene (Sigma-Aldritch: 107689) was added and the cells were incubated under normal conditions for 48 hours. Cells were passaged once before and after drug selection or sorting prior to use.

Parental cells infected with Cas9 virus underwent blasticidin (Sigma-Aldritch: 203350) selection. Briefly, cells were plated in six well plates and cultured in their usual media with 5µg/mL blasticidin for 7 days. Cells infected with sgRNA-RFP virus or Sox9-GFP virus underwent selection by flow sorting. Briefly, cells were resuspended in fluorescence-activated cell sorting (FACS) buffer (phosphate buffered saline [PBS] with 1% FBS) and sorting was either performed with the Sony MA900 or by OHRI StemCore Laboratories with the MoFlo XDP.

To sequence the CRISPR-Cas9 induced mutation a 200bp region was amplified and cleaned with a QIAGEN PCR purification kit. The fragment was ligated into a pJET1.2/blunt plasmid using the CloneJET PCR cloning kit (Thermo Fisher Scientific: K1231) by following the manufacturer's protocol. It was transformed with DH5α competent bugs following the same method as previously stated in section 2.1. DNA extracted from selected colonies using a midiprep kit were sequenced by OHRI StemCore Laboratories with the primer supplied with the CloneJet kit. The sequences of the sequencing primer and all other mentioned oligonucleotides are listed in Table 2.1.

2.2.4. Clonal Expansion

For clonal selection, cells were plated at a density of 5 cells/mL in a 96 well plate and cultured under normal conditions. 4 days after cell seeding, wells were assessed by fluorescent microscopy to ensure they only contained one colony of RFP expressing cells. Only cells from these wells were counted and expanded. Cells were sequentially passaged into larger wells once they reached 80% confluency, until they were maintained in 10cm plates as described in section 2.2.2.

2.2.5. Immunocytochemistry

Sterilized glass microscope cover slips were coated with collagen I (Corning: 354236) diluted 1:100 in PBS and incubated at 37°C for 30 minutes before cells were seeded. Cells were cultured on the cover slips in a 6 well plate for at least 24 hours and washed three times with PBS before fixing with 1mL 4% paraformaldehyde (PFA). The cover slips were washed a further three times and concurrently underwent blocking and permeabilization in a buffer containing 0.2% Igepal (MP Biomedicals: 02198596-CF), 0.05% Tween 20 (Thermo Scientific Chemicals: J20605.AP), 1% bovine serum albumin (BSA, BioShop: ALB001.250) with 5% goat serum (Abcam: ab7481). The primary antibody was incubated overnight at 4°C in PBS with 5% goat serum, and the negative control was incubated with a mixture of only PBS and 5% goat serum. The coverslips were then washed three times with PBS and the appropriate fluorophore-conjugated secondary antibody was added and incubated in darkness at room temperature for 60 minutes in PBS with 5% goat serum. After three more PBS washes the nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI, Invitrogen: D1306) diluted 1:5000 in PBS for 10 minutes at room temperature and mounted on glass microscope slides with fluorescent mounting medium (Dako: S302380-2) and sealed with nail polish. All primary and secondary antibodies and their dilutions are listed in Table 2.2.

2.2.6. Proliferation Assay

Cells were seeded in duplicate at a density of 25,000 cells/well in a 12 well plate. Every 24 hours for 5 days cells were washed with 500µL PBS, recovered with 200µL trypsin, and diluted in 800µL media. Suspended cells were stained with trypan blue and counted in duplicate with a hemocytometer to determine the number of viable cells per well. Three independent experiments were conducted with biological replicates. Images were taken with an EVOS microscope.

2.2.7. Sphere-Forming Assay

Cells were trypsinized, washed, pelleted, and resuspended in mammosphere-forming media (DMEM/F12 supplemented with 2% B-27 [Gibco: 17504044] and 1% MEGS [Gibco: S0155]). A single cell suspension was generated by passing resuspended cells through a 25G syringe three-to-five times. Cells were then plated in triplicate in ultra-low attachment 24-well flat bottom plates (Corning: CLS3473) in 500µL of mammosphere-forming media. Cells were cultured under otherwise normal conditions for 7 days. Secondary spheres were seeded by disaggregating the primary spheres with 500µL trypsin and repeating the process of primary sphere seeding. On day seven following primary or secondary sphere seeding, phase contrast images of every sphere were taken at 4x magnification with an EVOS microscope and clusters of cells 40µm or larger were counted as a sphere. Sphere-forming efficiency was determined by the following calculation:

$$\frac{Number\ of\ spheres\ formed}{Number\ of\ cells\ seeded} \cdot 100\%$$

Three independent experiments were conducted with biological replicates.

2.2.8. Transwell Migration Assay

Prior to the experiment, cells were cultured overnight in serum-free media. They were counted, washed, and resuspended in 1% serum media in collagen coated 8.0µm pore chamber Transwell

inserts (Corning: 07-200-150). The chambers were placed in 1% serum media to measure haptotactic migration or 10% serum media to measure chemotactic migration. The chambers were cultured under otherwise normal conditions for 8 hours to allow cells to migrate. At the timepoint, all media was aspirated from the chambers and wells and the chambers were washed three times with PBS. Cells remaining in the chamber were removed by swabbing the inner surface with a sterile cotton swab. The chamber was placed in 400µL of 10% formalin for 10 minutes, followed by three PBS washes. To visualize the cells the chambers were incubated in darkness with 400µL of DAPI diluted 1:5000 in PBS for 10 minutes and washed three times with PBS. The membranes were cut from the chambers with a scalpel and mounted on glass slides with fluorescent mounting media and a cover slip. Images were taken with an EVOS microscope and migrated cells were quantified with ImageJ software from nine regions of interest (ROIs) per membrane and averaged. Three independent experiments were conducted with biological replicates.

2.2.9. Transwell Invasion

Transwell chambers were incubated at 37°C for 60 minutes with 100µL of Matrigel (Corning: CLS354234) diluted 1:40 in PBS in the chamber to coat the inner surface and a collagen I diluted 1:100 in PBS to coat the outer surface. Cells were prepared and seeded as described in section 2.2.8. All chambers were placed in 1% serum media and incubated under normal conditions for 24 hours to allow for invasion. Fixing, staining, mounting, and quantification was performed as described in section 2.2.8. Three independent experiments were conducted with biological replicates.

2.2.10. Droplet Invasion

V-bottom 96-well (Sarstedt) and 24-well flat-bottom (Corning) plates were coated with 60µL or 200µL 5mg/mL Poly(2-hydroxyethyl methacrylate) (poly-HEMA, Sigma-Aldritch: P3932),

respectively and allowed to evaporate overnight. Spheres were formed by seeding 5,000 cells per well in the poly-HEMA coated 96-well plate and culturing for 48 hours on a shaker at 120rpm. 10μL of media containing the sphere was gently combined with 20μL each of collagen and Matrigel. 40μL of this sphere-containing mixture was plated in the poly-HEMA coated 24-well flat-bottom plates. Five replicates were seeded per cell line. Every 24 hours images were taken on an EVOS microscope at 10x magnification until the area of one sphere extended beyond the field of view, at which point all the invasive arms on a focal plane were counted and measured in pixel units using ImageJ software. Three trials were conducted with independent biological replicates.

2.3. Protein Expression Analysis

2.3.1. Collection

Cells and tissues were collected and homogenized in RIPA lysis buffer (0.05% Sodium dodecyl sulfate [SDS], 1% Triton X-100, 1% NP-40, 50mM Tris Base [pH 7.5], 150mM NaCl, 2mM EDTA [pH 8.0], 12mM Na-Deoxycholate, 10mM NaF, 1mM DTT, 10mM β-glycerophosphate, 0.6mM NaVO3, 1mM PMSF, 10μg/mL leupeptin, 10μg/mL aprotinin, 10μg/mL pepstatin, and 100μM benzamide). All samples were either incubated on ice for at least 30 minutes or frozen immediately at -70°C and later thawed on ice for use. Thawed samples were centrifuged at 20,000 x g for 8 minutes at 4°C prior to use.

2.3.2. Protein Quantification and Western Blotting

Protein concentration was determined by spectrophotometry with protein assay dye reagent concentrate (Biorad: 5000006) and samples were diluted in PBS and 1x SDS buffer (50 mM Tris Base, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol) to a protein concentration of 28μg/μL. Samples were denatured at 100°C for 5 minutes prior to electrophoresis on

polyacrylamide (BioRad: 1610158) gels at 50mA in 1x running buffer (25mM Tris Base, 192mM glycine, 0.1% SDS). Proteins were transferred to a polyvinyldene difluoride membrane for 90 minutes with constant voltage at 100V in 1x transfer buffer (48mM Tris-HCl, 39mM glycine, 20% methanol). The membrane was incubated with a primary antibody diluted in Tris-buffered saline with Tween (TBST, 50mM Tris Base, 150mM NaCl, pH 7.4, 0.1% Tween 20) and 5% BSA on a rotator overnight at 4°C. The membrane was washed three times for 5 minutes with TBST prior to incubation for 60 minutes at room temperature on a nutator with the appropriate HRP-conjugated secondary antibody diluted in TBST with 5% BSA. The membranes were washed and target proteins were visualized with western ECL substrate (BioRad: 1705061) following the manufacturer's protocol. Membranes were then exposed to X-ray film or imaged with a ChemiDoc (BioRad). For subsequent probings, the membrane was incubated for 20 minutes with stripping buffer (200mM glycine, 0.005% SDS, 0.01% Tween 20, pH 2.0) three times at room temperature on a nutator. The membrane was washed with TBST three times for 5 minutes before the next primary antibody was added. A list of primary and secondary antibodies and their dilutions is found in Table 2.2.

2.4. RNA Expression Analysis

2.4.1. RNA Isolation

Cells from 10cm dishes at 80% confluency and tissues were collected in 1mL Trizol buffer (Invitrogen: 15596018). Tissue samples were homogenized in the buffer prior to freezing at -70°C. Samples were thawed on ice and insoluble material was pelleted from tissue samples by centrifugation at 12,000 x g for 10 minutes at 4°C. To all samples, 200µL of chloroform was added and samples were vortexed for 15 seconds then incubated at room temperature for 3 minutes. Samples were centrifuged at 12,000 x g for 15 minutes at 4°C. The aqueous phase was transferred

to a new tube and 500µL of isopropanol was added to precipitate nucleic acids. The mixture was incubated for 10 minutes at room temperature before nucleic acids were pelleted by centrifugation at 12,000 x g for 10 minutes at 4°C. The pellets were washed with 75% ethanol and centrifuged at 7500 x g for 5 minutes before drying for 10 minutes. The pellets were resuspended in 100µL of RNAse-free water and incubated at 60°C for 10 minutes. The RNeasy Mini Kit (QIAGEN) was used to treat samples with DNAse according the manufacturer's protocol. RNA samples were assessed for concentration and purity using A260/230 and A260/280 ratios with a Nanodrop One (Thermo Fisher Scientific).

2.4.2. cDNA Synthesis

First strand compliment DNA (cDNA) was synthesized by incubating a 13µL mixture of 500ng of RNA, 50ng of random primers, 250ng of Oligo(dT)12-18 Primer (Thermo Fisher Scientific: 18418012), and 10nmol of dNTP mix at 65°C for 5 minutes. Reverse transcription was performed with Superscript III Reverse Transcriptase (Invitrogen: 18080044) To each sample, 4µL of 5X First Strand Buffer, 1µL RNase OUT, 1µL 0.1M DTT, and 1µL of Superscript III Reverse Transcriptase were added and samples were incubated in a thermocycler at 25°C for 5 minutes, 55°C for 60 minutes, and 70°C for 15 minutes before cooling to 4°C. Samples were diluted 1:10 to a final volume of 200µL in nuclease-free water and stored at 4°C.

2.4.3. qPCR

For quantitative PCR (qPCR) analysis, 31.5µL of a mastermix for each gene being analyzed containing 5µL iTaq Universal SYBR Green Supermix (Biorad: 1725124), 0.6µL 10nM primers, and 3.4µL nuclease-free water per sample was added to 3.5µL of diluted cDNA to create a minimix. From each minimix, 10µL was plated in triplicate. Target transcript expression was quantified using a 7500 Fast Real Time PCR System (Applied Biosystems) thermocycler and the

following protocol: initial hold at 50°C for 20 seconds and initial denaturation at 95°C for 10 minutes followed by 40 cycles of denaturation at 95°C for 15 seconds and annealing and extension at 60°C for 1 minute. The melt curve was generated by holding at 95°C for 15 seconds, 60°C for 1 minute, 95°C for 30 seconds, and 60°C for 15 seconds. The derivative melt curve was used to verify that a single target was amplified in each reaction. Relative mRNA expression was determined using the $\Delta\Delta C_T$ method, and normalization to the expression of the housekeeping gene, ribosomal 18S. All qPCR primers are listed in Table 2.1.

2.5. Transcriptomic Profiling Bioinformatics

2.5.1. Sample Preparation and Bulk RNA Sequencing

RNA was extracted and purified as previously described in the RNA extraction and isolation section. Samples were sent for bulk RNA sequencing with polyA-enriched RNA (stranded) library preparation and 25,000,000 reads per sample by Genome Quebec.

2.5.2. RNA-Sequencing Processing and Quantification

Mouse RNA-sequencing reads were quality filtered using the rfastp R package. Filtered paired-end reads were aligned to the mm10 reference genome using subjunc (Rsubread v2.22.1) with a custom-built index based on primary chromosomes (chr1–19, X, Y, M)[206,207]. Gene-level quantification was performed using the summarizeOverlaps() function (GenomicAlignments v1.42.0) in union mode, with a BamFileList of sorted and indexed BAM files[208]. Exonic regions were defined using TxDb.Mmusculus.UCSC.mm10.knownGene, grouped by gene, and read counting was performed with ignore.strand = TRUE. The resulting SummarizedExperiment object was used for differential expression analysis.

2.5.3. Differential Gene Expression Analysis

Differentially expressed genes (DEGs) were identified with DESeq2 (v1.48.1)[209]. A DESeqDataSet was created from the raw count matrix with group information ($sgNTC^{K3}1$, $sgSox9^{K3}1$, $sgSox9^{K3}2$, $sgSox9^{K3}6$). Size-factor normalisation, dispersion estimation, and model fitting were performed using the DESeq() function. Differential expression for each knockout clone versus $sgNTC^{K3}1$ was extracted using the results() function with contrasts, and log2FC were shrunken with lfcShrink(method = "apeglm") to improve ranking and visualization[210]. Genes with adjusted $p < 0.05$ (Benjamini–Hochberg) and $|log2FC| > 1$ were considered significant.

2.5.4. Overrepresentation Analysis

Shared DEGs (adjusted $p < 0.05$, $|log2FC| > 1$) from all Sox9 knockout clones underwent overrepresentation analysis (ORA) using the enrichPathway() function from the Gene Ontology Biological Processes database[211]. The ten most significant (adjusted $p < 0.05$) pathways were selected for visualization.

2.5.5. Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed using the clusterProfiler package (v4.6.2) with mouse gene sets from the MSigDB C6: oncogenic signature collection obtained through msigdbr package (v25.1.0)[212–215]. Genes were ranked by the average Wald test statistic across each Sox9 knockout clone and enrichment was performed using the GSEA() function with “p cutoff = 0.05”. The top 10 positively and negatively enriched gene sets were visualized and ordered by normalized enrichment score (NES).

2.5.6. Heatmap of Selected Genes

Rlog-transformed gene expression values were used to generate a heatmap for a curated list of lineage-related genes using the pheatmap package (v1.0.13)[216]. Expression values were row-scaled (z-score) and hierarchical clustering was performed using Euclidean distance.

2.6. Animal Care

2.6.1. Husbandry

All mouse lines were maintained on an FVB/N background. Backcrossing was done with FVB/N mice (Jackson Laboratory) for at least two generations, with offspring DNA analyzed for FVB/N content by advanced speed congenic genotyping with microsatellite analysis. MMTV-NDL2-5 transgenic mice were kindly provided by Dr. William Muller (McGill University). Only female mice were used for cell line isolation and were not used as breeders. NCG mice were kindly generated by Dr. Rebecca Auer's laboratory (OHRI and University of Ottawa), and were immediately used in experiments without further husbandry. All husbandry and animal procedures were performed in accordance with the guidelines established by the Canadian Council on Animal Care and specific protocols were approved by the University of Ottawa Animal Care Committee.

2.6.2. Genotyping

Ear clips were taken from 3 week-old mice at the time of weaning. Samples were digested overnight and genomic DNA was isolated with the DNEasy Blood and Tissue kit (QIAGEN), following the manufacturer's protocol. A Taq mastermix (VWR: 5200300-1250) was used for genotyping by PCR. The cycling protocol was as follows: initial denaturation at 95°C for 5 minutes; 35 cycles of denaturation at 95°C for 30 seconds, annealing at 5°C lower than the primer melting temperature for 30 seconds, and elongation at 72°C for 45 seconds; and a final elongation

at 72°C for 5 minutes before being cooled to 4°C. Primer sequences used for genotyping are listed in Table 2.1.

2.6.3. Orthotopic Injections

Cells were cultured under normal maintenance conditions and confirmed to be negative for mycoplasma prior to injection. After trypsinization, cells to be injected were washed twice with PBS and serially diluted to the appropriate concentrations (10^4 or 10^5 cells/injection) in 100µL of a 1:1 mixture of PBS: Matrigel per injection. Right mammary glands 3 and 4 were injected with sgNTC cells while the corresponding left mammary glands were injected with sgSox9 cells. Each NCG mouse received four orthotopic injections with the same cell concentration, and two mice were injected per concentration. Mice were monitored once a week until tumours were detectable, at which point tumours were measured twice a week to assess their size. Tumour size was determined by taking two perpendicular diameter measurements and using the average as the variable *d* in the following calculation:

$$volume = \frac{4}{3} \cdot \pi \cdot \left(\frac{d}{2}\right)^3$$

The time when the first tumour reached a volume of $0.5cm^3$ was considered tumour onset and endpoint was defined by a cumulative tumour burden of $1.7cm^3$, as per University of Ottawa Animal Care and Veterinary Service (ACVS) Guidelines.

2.6.4. Tail Vein Injections

Cells were cultured under normal maintenance conditions and confirmed to be negative for mycoplasma prior to injection. After trypsinization, cells to be injected were washed twice with PBS and 10^6 cells were resuspended in 100µL of PBS per injection. Cells were injected into the tail vein of each of six NCG mice per cell line. Mice were sacrificed 12 weeks after the injection,

this endpoint was determined by a preliminary experiment using the same number of parental cells. During this time, mice were monitored for behavioural signs. At endpoint, the lungs were collected and one lobe was prepared for paraffin embedding and sectioning while the other was prepared for FACS.

For FACS, lungs were incubated in a 1:10 mixture of gentle collagenase/hyaluronidase (Stemcell Technologies**:** 07919): DMEM/F12 overnight at 37**°**C. Samples were then centrifuged at 350 x g for 5 minutes and resuspended in a 4:1 mixture of NH4Cl:1% BSA in PBS to lyse red blood cells during a five-minute incubation on ice. The samples were centrifuged at 300 x g for 5 minutes and resuspended in 1mg/mL DNAse and passed through a 40µm strainer. The samples centrifuged again and resuspended in FACS buffer. FACS for RFP was performed using the Aurora Spectral Flow Cytometer (Cytek) and SpectroFlo software.

2.7. Histology

2.7.1. Tissue Preparation

All murine tissues were fixed in 10% formalin for 48-72 hours then transferred to 70% ethanol and stored at 4°C. Samples were further processed and paraffin embedded by the University of Ottawa Louise Pelletier Histology Core Facility, and paraffin blocks were stored at room temperature. Mammary glands and mammary tumours were also sectioned to 4µm by the facility. Lungs were sectioned in house, with ten 5µm sections being taken per level with 100µm between levels.

2.7.2. Hematoxylin And Eosin Staining

Sections were deparaffinized with three 5-minute xylene washes followed by rehydration with five graded ethanol washes (100%, 100%, 95%, 95%, 80%) for 5 minutes each. Samples were

submerged in hematoxylin for 30-45 seconds and drained for an additional 15 seconds before three washes in deionized (DI) water. Samples were briefly submerged in acid alcohol, DI water, ammonium water, and DI water before counter-staining by submersion in alcoholic eosin (Leica: 3801615) for 2 minutes. Samples were dehydrated in five graded ethanol washes (80%, 95%, 95%, 100%, 100%) for 1 minute each, followed by three 5-minute washes in xylenes. Slides were sealed with coverslips using Cytoseal 60 (Thermo Fisher Scientific: 8310-4) and images were taken at 10x magnification with a Zeiss Axio Scan.Z1**.**

2.7.3. Immunohistochemistry

Sections were deparaffinized according to the process described in section 2.7.2. Sections were washed in PBS and underwent antigen retrieval with 10mM sodium citrate for 10 minutes in a pressure cooker. After cooling, they were quenched with 3% hydrogen peroxide. Permeabilization was concurrent with blocking, as described in section 2.2.5. Primary antibody was incubated overnight at 4°C in either the permeabilization buffer or PBS with 5% goat serum. The slides were washed in PBS before the addition of the appropriate secondary HRP-conjugated antibody diluted in PBS with 5% goat serum and incubation for 60 minutes. The secondary antibody was washed with PBS and DAB Substrate (Vector Laboratories: SK-4100) was applied following the manufacturer's protocol. Tissues were counterstained with hematoxylin, dehydrated, and sealed according to the process described in section 2.7.2. Images were taken at 10x magnification with a Zeiss Axio Scan.Z1.

2.8. Reproducibility And Statistics

For *in vitro* experiments, data shown is the mean of three independent biological replicates unless otherwise specified, and error bars represent the standard error of the mean (SEM). For fluorescent images, nine random fields of view were quantified per image and data is presented as the mean

of three independent biological replicates with error bars representing SEM unless otherwise specified. For orthotopic tumour injections, two mice per cell type were injected twice, and data shown is the mean tumour volume per mouse with error bars representing SEM. For tail-vein injections, at least three mice were included in the statistical analysis per cell type.

For statistical analysis of two comparisons, a student's t-test was performed. For statistical analysis of more than two comparisons a one-way or two-way ANOVA was performed with a Dunnett's post-hoc test. Significance is defined as *P<0.05, **P<0.01, and ***P<0.001.

Table 2.1. List of oligonucleotides. A list of oligonucleotide sequences and their application used throughout the thesis. All sequences are given 5' – 3'.

Name	Type/ Application	Sequence
Sox9 Forward	Primer, CRISPR mutation cloning	CCAGCAAGAACAAGCCACAC
Sox9 Reverse	Primer, CRISPR mutation cloning	GGGGTCGGTGAGTGCAGCCC
Sox9 Forward	Primer, genotyping	GGGGCTTGTCTCCTTCAGAG
Sox9 Reverse	Primer, genotyping	TGGTAATGAGTCATACACAGTAC
NDL Forward	Primer, genotyping	GTTTCCTGCAGCAGCCTACGC
NDL Reverse	Primer, genotyping	TTCCGGAACCCACATCAGGCC
Sox9K1 Sense	sgRNA	CACCGGATGACCGACGAGCAGGAGA
Sox9K1 Antisense	sgRNA	AAACTCTCCTGCTCGTCGGTCATCC
Sox9K2 Sense	sgRNA	CACCGGTGTCCGAGCCGGAGCCCGA
Sox9K2 Antisense	sgRNA	AAACTCGGGCTCCGGCTCGGACACC
Sox9K3 Sense	sgRNA	CACCGGTACCCGCATCTGCACAACG
Sox9K3 Antisense	SgRNA	AAACCGTTGTGCAGATGCGGGTACC
NTC Sense	sgRNA	CACCGATCGGTATGTTTAGGGTT
NTC Antisense	sgRNA	AAACAACCCTAAACATAACCGATC
MGSO	Primer, PCR mycoplasma testing	TGCACCATCTGTCACTCTGTTAACCTC
GPO-3	Primer, PCR mycoplasma testing	GGGAGCAAACAGGATTAGATACCCT
CloneJet Forward	Primer, PCR	CGACTCACTATAGGGAGAGCGGC
18S Forward	Primer, qPCR	AGTCCCTGCCCTTTGTACACA
18S Reverse	Primer, qPCR	GATCCGAGGCCTCACTAAAC
OVOL2 Forward	Primer, qPCR	TCTCCCGACACCCTAATGAGC

OVOL2 Reverse	Primer, qPCR	CTCTGGAGTTTCAGGCTCCG
KRT8 Forward	Primer, qPCR	GAAGCCAGGGCTTAGTGAGT
KRT8 Reverse	Primer, qPCR	GGACATCGAGATCACCACCT
PDGFRA Forward	Primer, qPCR	AGAGTTACACGTTTGAGCTG
PDGFRA Reverse	Primer, qPCR	GTCCCTCCACGGTACTCCT
PDGFB Forward	Primer, qPCR	CGCACAGAGGTGTTCCAGAT
PDGFB Reverse	Primer, qPCR	GTGGTCCTCCAAGGTCACTG
PDGFRL Forward	Primer, qPCR	AGGAGCAGGGAGAAAACAGG
PDGFRL Reverse	Primer, qPCR	TCCATTCCACTTTGCTCCCC

Table 2.2. List of antibodies. A list of antibodies and their dilutions used in each application throughout the thesis. WB: western blotting, IF: immunofluorescence, IHC: immunohistochemistry.

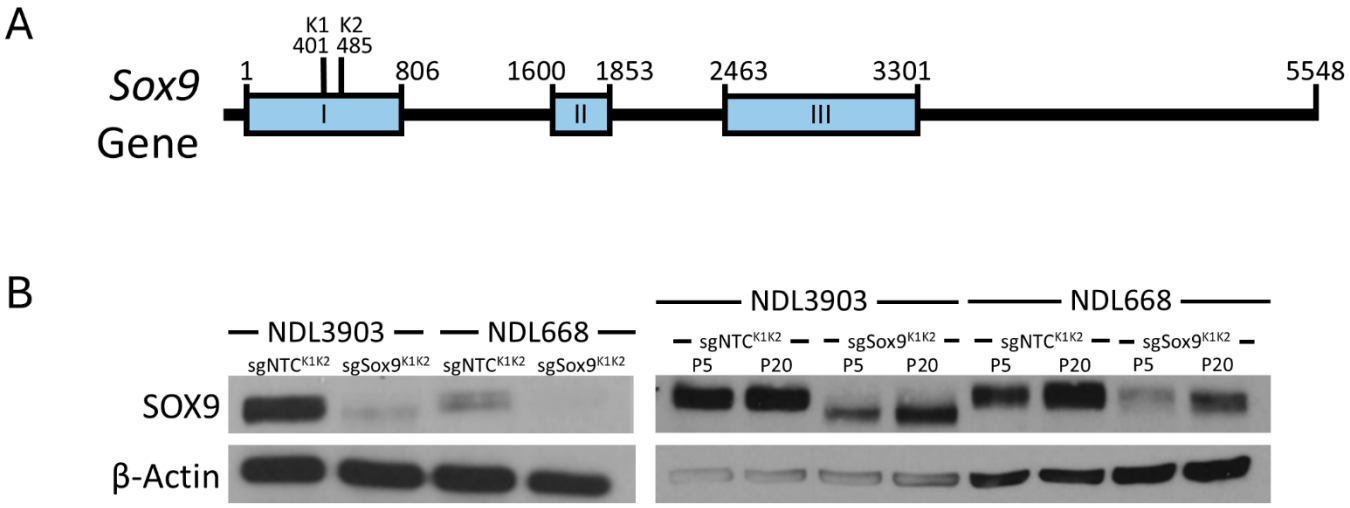
Antibody Name	Manufacturer	Catalog Number	Dilution		
			WB	IF	IHC
Sox9	CST	82630S	1:1000	1:100	1:100
Sox10	CST	89356S	1:1000	-	-
Neu	Sigma	OP15	1:1000	-	1:100
β-Actin	SigmaAldritch	A5316	1:5000	-	-
α-Tubulin	SigmaAldritch	T5168	1:5000	-	-
Cas9	CST	14697S	-	-	1:200
CK8	Abcam	ab53280	-	-	1:200
Goat anti-Mouse IgG HRP-conjugated	BioRad	1706516	1:5000	-	-
Goat anti-Rabbit IgG HRP-conjugated	BioRad	1706515	1:5000	-	-
Goat anti-Rabbit IgG (H+L) Alexa Fluor™ 488	Thermo Fisher Scientific	A-11008	-	1:1000	-
SignalStain® Boost IHC Detection Reagent (HRP, Mouse)	CST	8125S	-	-	Undiluted
SignalStain® Boost IHC Detection Reagent (HRP, Rabbit)	CST	8114S	-	-	Undiluted

3. Results

3.1. Effect of Sox9 in tumour cell pools

The Sox9 transcription factor has been shown to be critical for the maintenance of stem cell populations and tumour progression in a number of cancer types, including TNBC[134,136,188–190]. Therefore, to gain insight into the effect of Sox9 on cell stemness and tumour progression we selected two Neu+ murine cell lines previously established in our lab from MMTV-NDL2-5 endpoint tumours. Western blotting showed that the NDL3903 line expressed relatively high levels of endogenous Sox9, while the NDL668 displayed lower expression (Figure 3.1B). To first assess the effect of Sox9 inactivation, two sgRNA-IRES-RFP plasmids with sgRNA targeting exon 1 at 401bp/ 9aa (sgSox9^{K1}) and 485bp/37aa (sgSox9^{K2}) relative to the start codon, were transduced into cell lines stably expressing Cas9 and RFP-positive cells were selected by flow sorting (Figure 3.1.A). Control cells were generated in parallel by transduction with a non-targeting control sgRNA-IRES-RFP (sgNTC^{K1K2}). Expression of Sox9 was assessed in the bulk-transduced populations by immunoblotting. Western blot analyses demonstrated efficient knockdown in both cell lines (Figure 3.1.B). Surprisingly, routine Western blot analysis revealed that bulk populations regained Sox9 expression after prolonged passaging in culture, suggesting that Sox9-expressing subpopulations outgrew Sox9-deficient cells (Figure 3.1.B). Thus, low passage cultures (<P5 after sorting) were used in subsequent studies.

The role of Sox9 on tumour growth and stem cell function is well documented in a variety of malignancies[134,136,188,190]. To assess the effect of Sox9 inactivation in HER2-positive breast cancer, we began by investigating the proliferative ability and tumoursphere-forming ability of our cell lines *in vitro*. Proliferation in bulk populations was not found to be significantly different between the sgNTC^{K1K2} and sgSox9^{K1K2} cells of either line (Figure 3.1.C). However, significantly


A
K1 K2
401 485
1
806
1600
1853
2463
3301
5548
Sox9
Gene
I
II
III
B
NDL3903
NDL668
sgNTC^K1K2
sgSox9^K1K2
P5
P20
SOX9
β-Actin

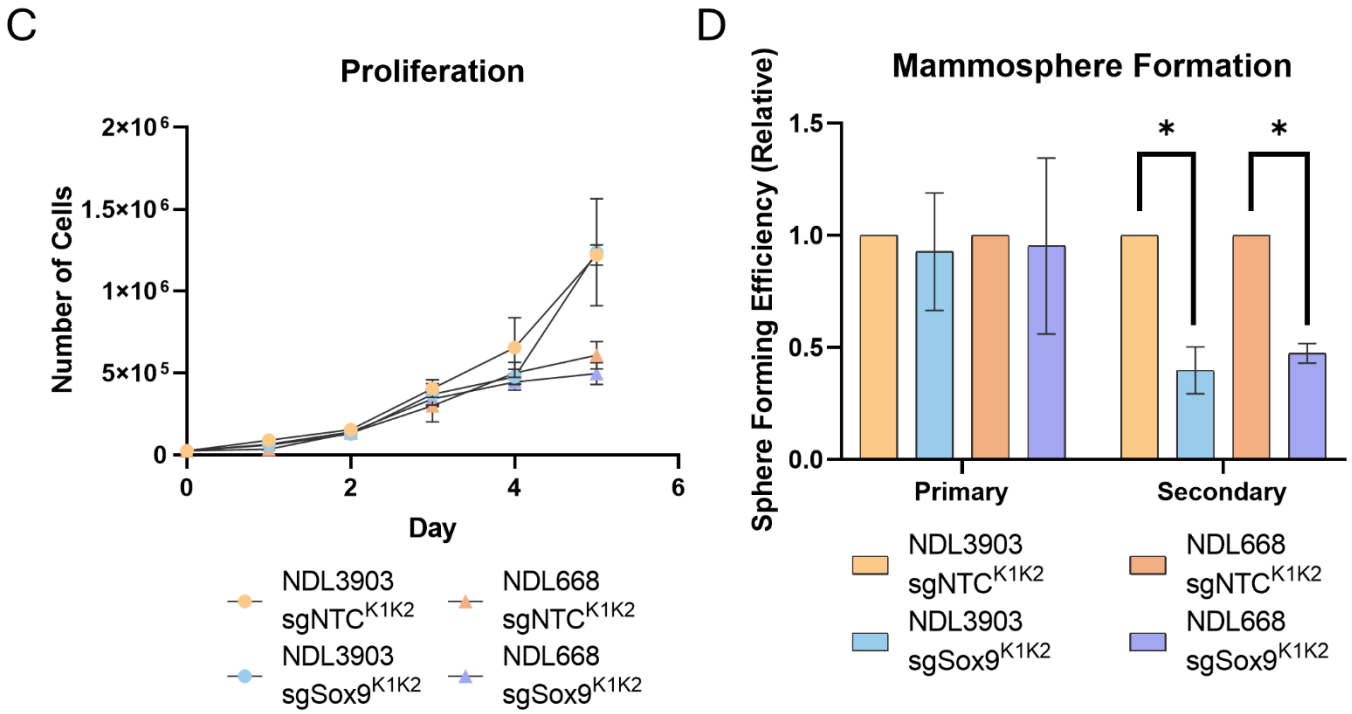

C
Proliferation
Number of Cells
2×10^6
1.5×10^6
1×10^6
5×10^5
0
0
2
4
6
Day
NDL3903 sgNTC^K1K2
NDL668 sgNTC^K1K2
NDL3903 sgSox9^K1K2
NDL668 sgSox9^K1K2
D
Mammosphere Formation
Sphere Forming Efficiency (Relative)
1.5
1.0
0.5
0.0
*
*
Primary
Secondary
NDL3903 sgNTC^K1K2
NDL668 sgNTC^K1K2
NDL3903 sgSox9^K1K2
NDL668 sgSox9^K1K2

Figure 3.1. Generation of initial knockout cell lines and the effect of Sox9 deletion on stemness. A) Schematic representation of the murine *Sox9* gene with two sgRNA targeting exon I, numbers above indicate bp after start codon and roman numerals indicate exon. **B)** Parental 3903 and 668 murine cell lines were treated with Cas9 lentivirus, underwent selection with blastocidin and were then treated with $sgNTC^{K1K2}$ or $sgSOX9^{K1K2}$ lentivirus to stably express Cas9 and inactivate Sox9. SOX9 protein expression was validated by immunoblotting after culturing for two, five, and 20 passages after flow sorting for RFP. **C)** Sox9 knockout does not affect proliferation in bulk cells. Cells were seeded in duplicate at a density of 25,000 cells/ well in a 12 well plate and counted every 24 hours for 5 days. Significance of three independent biological replicates was determined by one-way ANOVA where $P<0.05$. **D)** Sox9 knockout reduces self-renewal ability. Cells were seeded as a single suspension at an initial density of 2000 cells/ well in a 24 well ultra-low attachment plate. They were cultured in serum-free media supplemented with B27 and MEGS but otherwise normal conditions for 7 days. Secondary spheres were seeded by disaggregating the primary spheres and culturing for a further 7 days under the same conditions. Sphere-forming efficiency was calculated as a percentage of the number of spheres with a diameter greater than 40μm per initial number of seeded cells. Significance of three independent biological replicates was determined by two-way ANOVA where $*P<0.05$.

fewer secondary spheres were formed in the sgSox9^{K1K2} cells, suggesting that their self-renewal ability is reduced (Figure 3.1.D).

Sox9-overexpressing murine and human cells were also generated by retroviral transduction of a Sox9-IRES-GFPII virus previously constructed in our lab. Vector control cells were infected in parallel with the IRES-GFPII plasmid without the Sox9 insert. Murine Sox9-overexpression cells (3903OE) were generated from the same NDL3903 parental cells as the NDL3903 knockout lines and expression was validated by immunoblotting (Figure 3.2.A). Although significant overexpression was achieved, no significant differences were observed in proliferation or sphere-forming efficiency between the 3903OE cells and control (3903Ctrl) cells (Figure 3.2.B). These experiments were repeated with a human HER2-positive breast cancer cell line (BT474), also with no significant differences in proliferation or sphere-forming efficiency between Sox9 overexpressed (BT474OE) or control (BT474Ctrl) cells (Figure 3.2.C). These data suggest that Sox9 overexpression in Sox9-expressing cells does not impact cell growth or self-renewal ability, possibly due to Sox9 targets already bound by endogenous Sox9 with no effect by exogenous Sox9 expression.

To circumvent the potential expansion of Sox9-expressing subpopulations in knockout pools, a new sgRNA (sgSox9^{K3}, 803bp/ 143aa relative to the start codon) was designed, while the sgNTC sequence used in the generation of the previous knockout lines was kept (Figure 3.3.A). Our hypothesis was that this new sgSox9^{K3} would result in more efficient knockdown with a single guide. To mitigate the effects of prolonged cell culture, new pools were generated from early passage parental cell lines. The additional cell lines were generated as stated above and Sox9 expression was validated by immunoblotting (Figure 3.3.B). Significant knockdown in both cell lines was achieved with sgSox9^{K3}, but the effect of Sox9 deficiency on the bulk population was

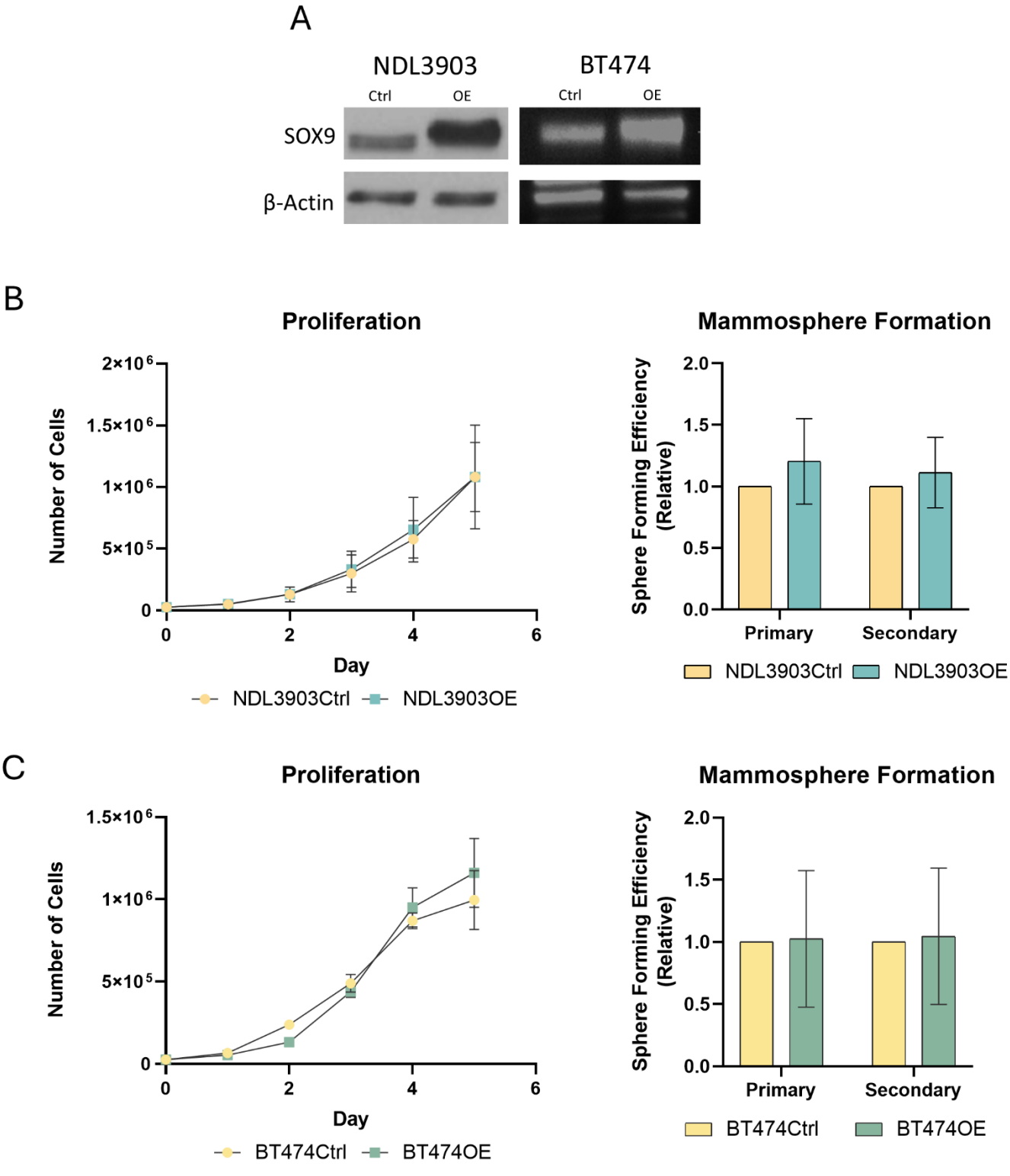
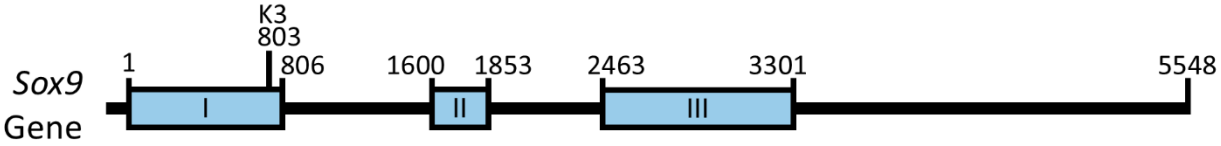
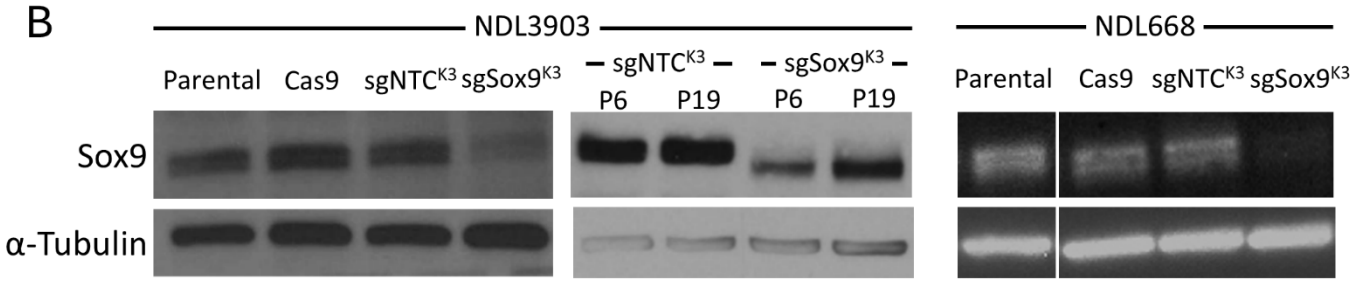

A
NDL3903
BT474
Ctrl
OE
Ctrl
OE
SOX9
β-Actin
B
Proliferation
Number of Cells
2×10^6
1.5×10^6
1×10^6
5×10^5
0
0
2
4
6
Day
NDL3903Ctrl
NDL3903OE
Mammosphere Formation
Sphere Forming Efficiency (Relative)
2.0
1.5
1.0
0.5
0.0
Primary
Secondary
NDL3903Ctrl
NDL3903OE
C
Proliferation
Number of Cells
1.5×10^6
1×10^6
5×10^5
0
0
2
4
6
Day
BT474Ctrl
BT474OE
Mammosphere Formation
Sphere Forming Efficiency (Relative)
2.0
1.5
1.0
0.5
0.0
Primary
Secondary
BT474Ctrl
BT474OE

Figure 3.2. Sox9 overexpression does not affect growth or stemness. A) Parental cells were infected with a Sox9 retrovirus and protein expression was confirmed by flow sorting for GFP and immunoblotting after culturing for two passages. **B)** Sox9 overexpression does not affect proliferation. Cells were seeded in duplicate at a density of 25,000 cells/ well in a 12 well plate and counted every 24 hours for 5 days. Significance of three independent biological replicates was determined by one-way ANOVA where $P<0.05$. **C)** Sox9 overexpression does not affect primary or secondary sphere-formation. Cells were seeded as single suspension at an initial density of 2000 cells/ well in a 24 well ultra-low attachment plate. They were cultured in serum-free media supplemented with B27 and MEGS but otherwise normal conditions for 7 days. Secondary spheres were seeded by disaggregating the primary spheres and culturing for a further 7 days under the same conditions. Sphere-forming efficiency was calculated as a percentage of the number of spheres with a diameter greater than 40μm per initial number of seeded cells. Significance of three independent biological replicates was determined by two-way ANOVA where $*P<0.05$.

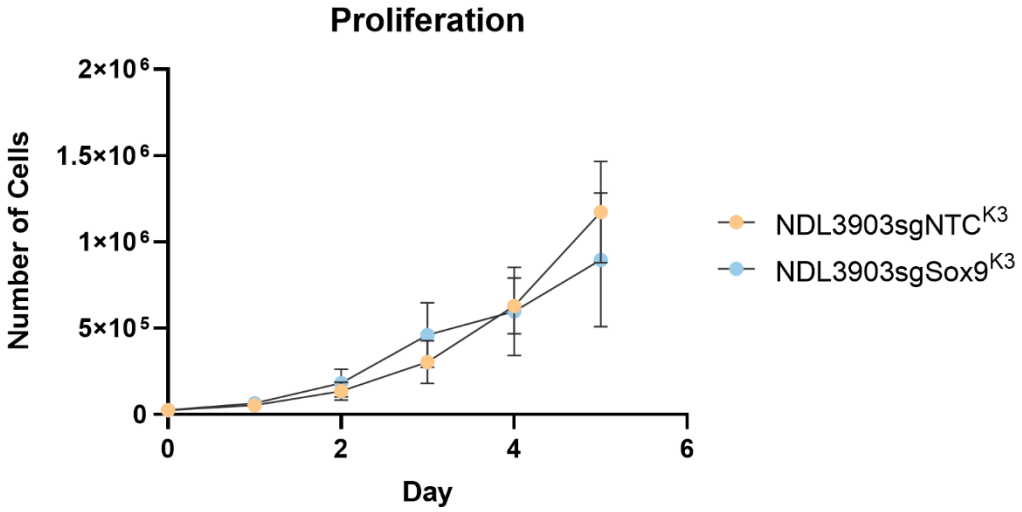
A


K3
803
1
806
1600
1853
2463
3301
5548
Sox9
Gene
I
II
III

B


NDL3903
NDL668
Parental
Cas9
sgNTCK3
sgSox9K3
sgNTCK3
sgSox9K3
P6
P19
P6
P19
Parental
Cas9
sgNTCK3
sgSox9K3
Sox9
α-Tubulin

C


Proliferation
Number of Cells
2×10⁶
1.5×10⁶
1×10⁶
5×10⁵
0
0
2
4
6
Day
NDL3903sgNTCK3
NDL3903sgSox9K3

D

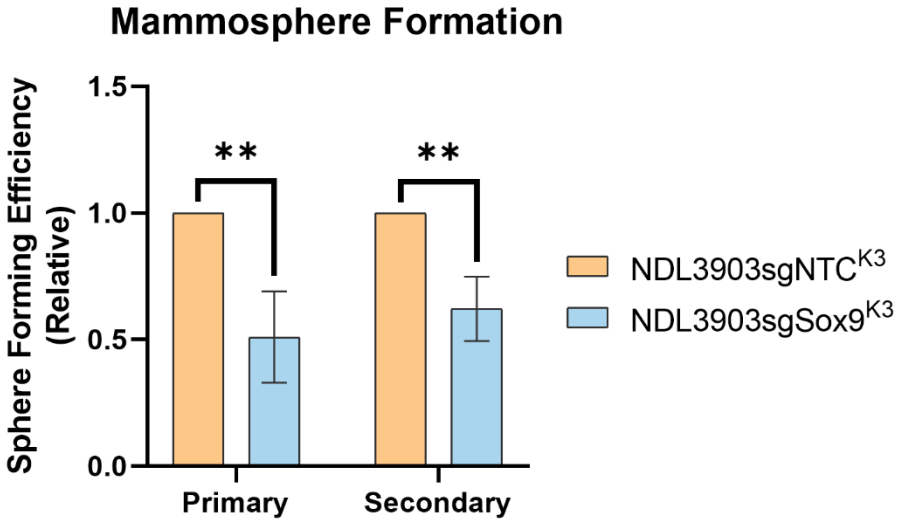

Mammosphere Formation
Sphere Forming Efficiency (Relative)
1.5
1.0
0.5
0.0
**
**
Primary
Secondary
NDL3903sgNTCK3
NDL3903sgSox9K3

Figure 3.3. Generation of early-passage bulk Sox9 knockout cell lines. A) Schematic representation of murine Sox9 gene with $sgSox9^{K3}$ targeting exon I, numbers above indicate bp after start codon and roman numerals indicate exon. **B)** Parental cell lines were treated with Cas9 lentivirus, underwent selection with blastocidin and were then treated with $sgNTC^{K3}$ or $sgSox9^{K3}$ lentivirus to stably express Cas9 and inactivate Sox9. Sox9 protein expression was validated by immunoblotting after culturing for two passages after flow sorting for RFP. **C)** Sox9 deletion does not have an effect on proliferation in bulk cell populations. Cells were seeded in duplicate at a density of 25,000 cells/ well in a 12 well plate and counted every 24 hours for 5 days. Significance of three independent biological replicates was determined by one-way ANOVA where $P<0.05$. **D)** Sox9 deletion reduces primary and secondary sphere-formation in bulk cell populations. Cells were seeded as a single suspension at an initial density of 2000 cells/ well in a 24 well ultra-low attachment plate. They were cultured in serum-free media supplemented with B27 and MEGS but otherwise normal conditions for 7 days. Secondary spheres were seeded by disaggregating the primary spheres and culturing for a further 7 days under the same conditions. Sphere-forming efficiency was calculated as a percentage of the number of spheres with a diameter greater than 40μm per initial number of seeded cells. Significance of three independent biological replicates was determined by two-way ANOVA where $**P<0.01$.

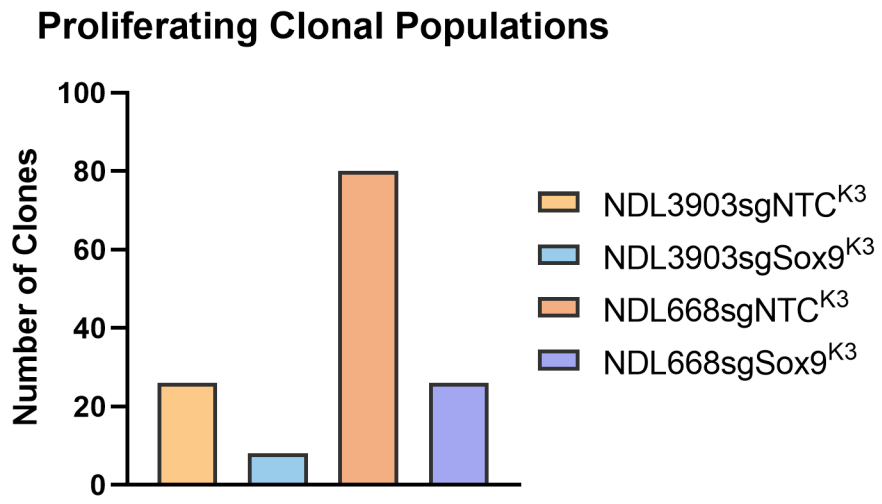
only assessed in NDL3903 cells. No significant differences between the proliferation of 3903sgNTC^{K3} and 3903sgSox9^{K3} cells were observed (Figure 3.3.C). However relative sphere-forming efficiency of both primary and secondary spheres was significantly reduced in the 3903sgSox9^{K3} cells, supporting our findings with the previous independent sgRNAs (Figure 3.3.D).

3.2. Sox9 knockout reduces stem-like phenotype in clonal cell populations

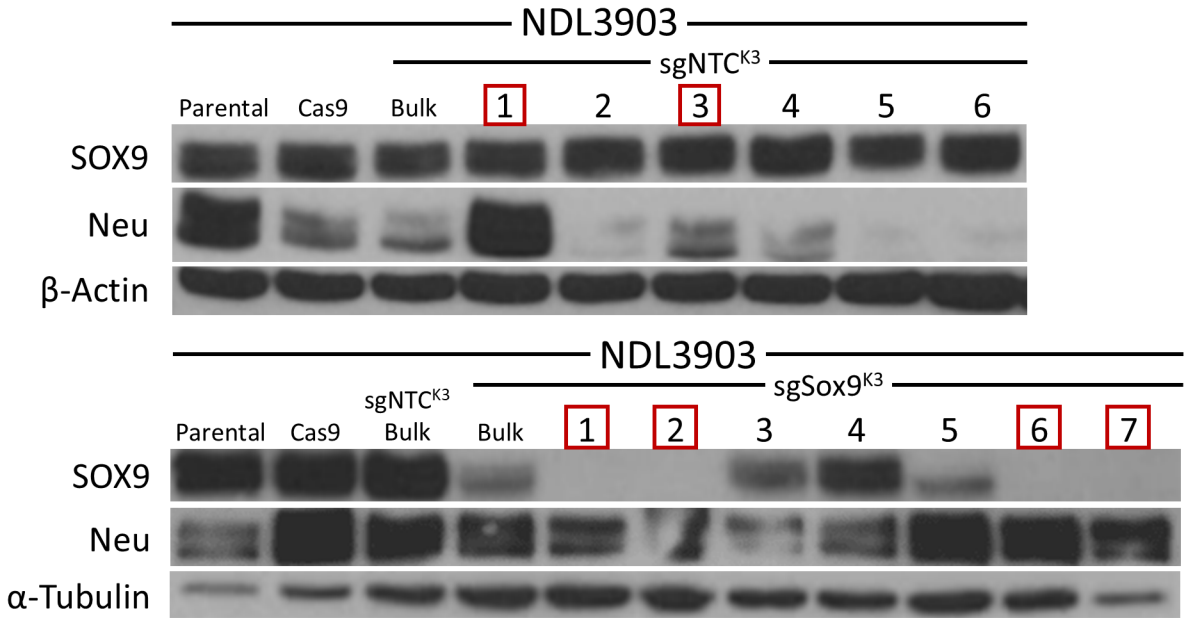
To enrich the sgSox9^{K3} for Sox9-deficient cells and limit the residual Sox9-expressing cells, clonal populations were expanded from single cells. Interestingly, more clones of sgNTC^{K3} cells were recovered than of sgSox9^{K3} cells. At the time the plates were assessed for wells containing single cells, 26 clones of 3903sgNTC^{K3} cells were present and RFP positive, while only 8 clones of 3903sgSox9^{K3} cells were RFP positive (Figure 3.4.A). There were 88 clones of RFP 668sgNTC^{K3} cells and 26 clones of 668sgSox9^{K3} cells (Figure 3.4.A). This supports our observations that Sox9-deletion significantly affects the self-renewal capacity and growth of cells.

Clonal populations were expanded and assessed for Sox9 expression by immunoblotting. In both cell lines, some of the clonal populations expressed Sox9 despite expression of the sgSox9^{K3}/RFP guides, suggesting inefficient inactivation at high frequency or preferential expansion of Sox9-expressing cells (Figure 3.4.B-C). The Sox9-deficient NDL3903 clones were sequenced, and all were found to contain a single base insertion resulting in the same frame shift (Figure 3.5.A). Knockout clones 1 (3903sgSox9^{K3}1), 6 (3903sgSox9^{K3}6), and 7 (3903sgSox9^{K3}7) contained an adenosine insertion while knockout clone 2 contained a thymidine insertion at the same locus. Clone 2 (3903sgSox9^{K3}2) also contained a number of substitutions downstream of the CRISPR/Cas9 cut site.

A


Proliferating Clonal Populations
Number of Clones
100
80
60
40
20
0
NDL3903sgNTC^{K3}
NDL3903sgSox9^{K3}
NDL668sgNTC^{K3}
NDL668sgSox9^{K3}

B


NDL3903
sgNTC^{K3}
Parental
Cas9
Bulk
1
2
3
4
5
6
SOX9
Neu
β-Actin
NDL3903
sgSox9^{K3}
sgNTC^{K3} Bulk
Bulk
1
2
3
4
5
6
7
SOX9
Neu
α-Tubulin

C

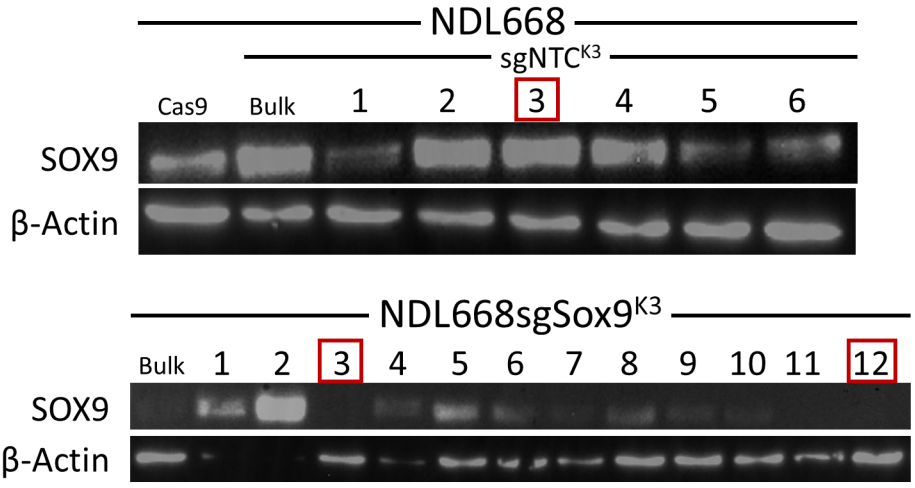
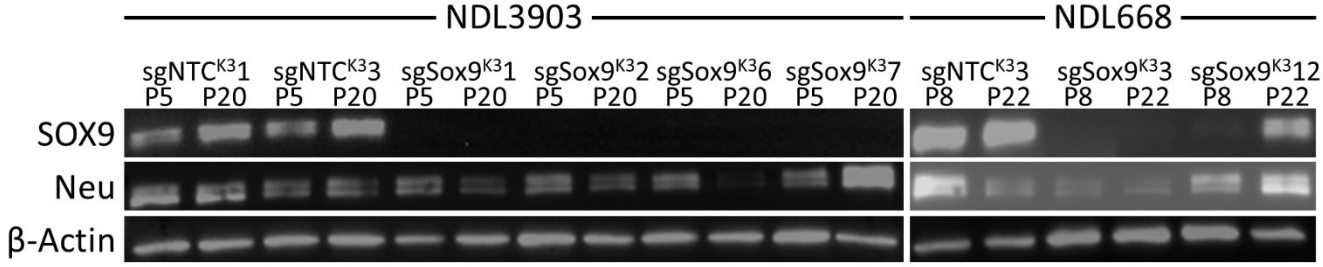

NDL668
sgNTC^{K3}
Cas9
Bulk
1
2
3
4
5
6
SOX9
β-Actin
NDL668sgSox9^{K3}
Bulk
1
2
3
4
5
6
7
8
9
10
11
12
SOX9
β-Actin

Figure 3.4. Clonal population selection. A) Sox9-expressing cells are more likely to form a clonal population. Cells were seeded at a density of 5 cells/mL in a 96 well plate and cultured under normal conditions. After 4 days, wells were assessed for RFP expression and colony number. Only cells from wells containing one colony with RFP expression were further passaged. **B-C)** Sox9 expression in NDL3903 and NDL668 clones was confirmed by immunoblotting. Squares indicate clones selected for characterization.

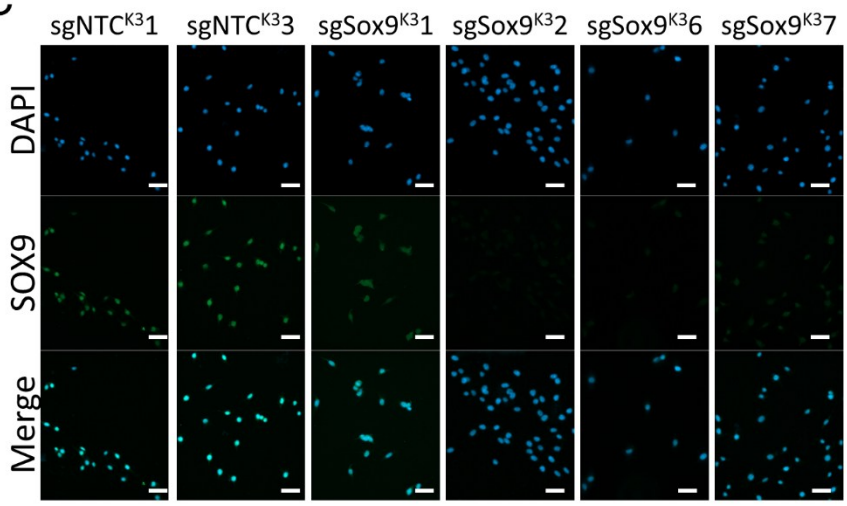
A

Reference	GTACCGCATCTGCACA ACGCGGAGCTCAGCAAGACT
3903sgNTC^{K3}1	GTACCGCATCTGCACA ACGCGGAGCTCAGCAAGACT
3903sgNTC^{K3}3	GTACCGCATCTGCACA ACGCGGAGCTCAGCAAGACT
3903sgSox9^{K3}1	GTACCGCATCTGCACA **A** ACGCGGAGCTCAGCAAGACT
3903sgSox9^{K3}2	GTACCGCATCTGCACA **T** ACGCGGAGCTCA**C**C**GG**GACT
3903sgSox9^{K3}6	GTACCGCATCTGCACA **A** ACGCGGAGCTCAGCAAGACT
3903sgSox9^{K3}7	GTACCGCATCTGCACA **A** ACGCGGAGCTCAGCAAGACT

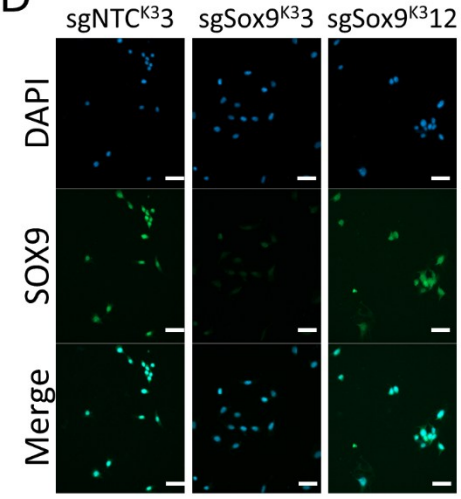
B



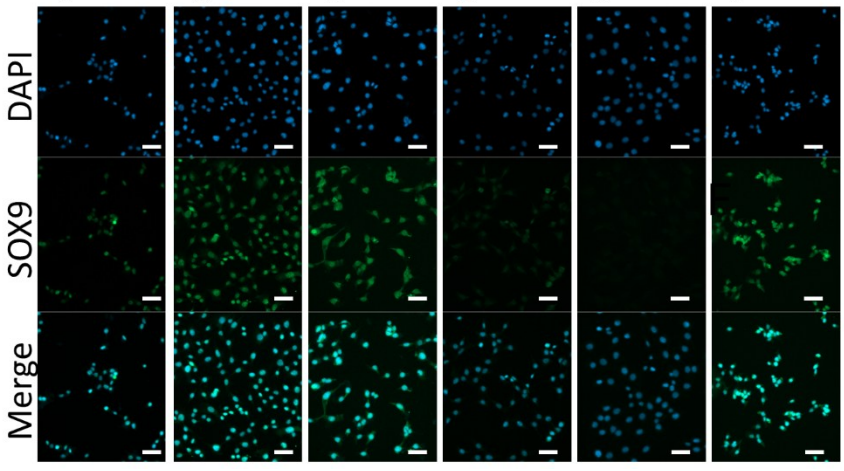
C



D



E



F

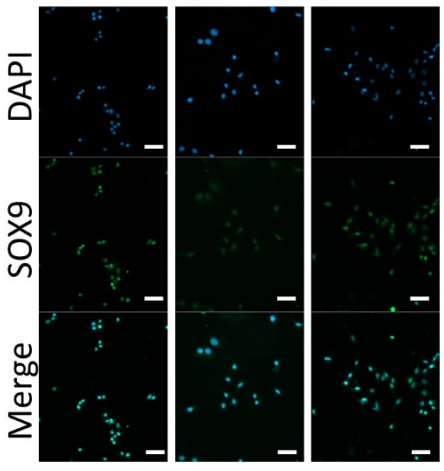


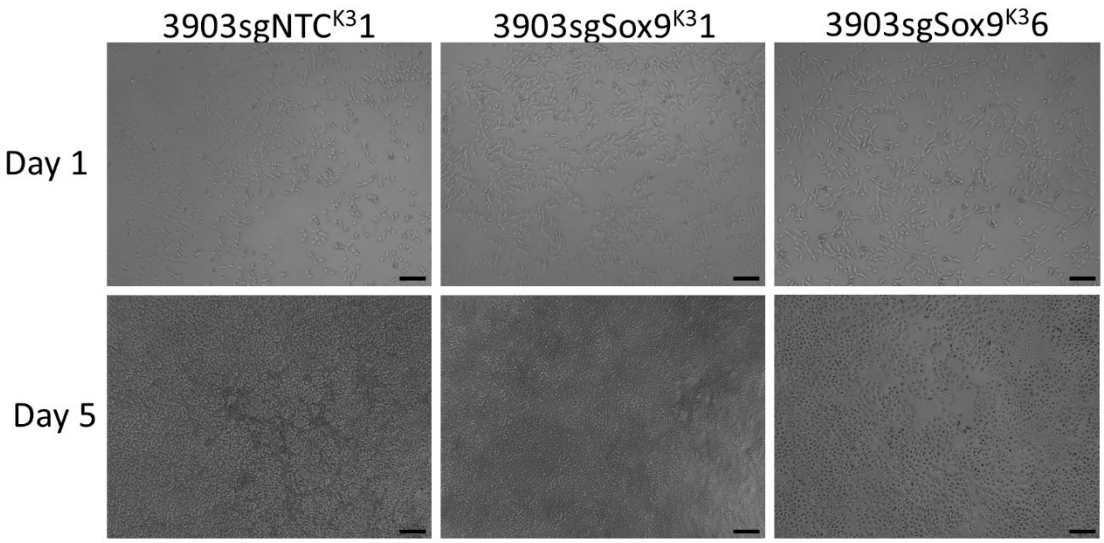
Figure 3.5. Sox9 in clonal cell populations. A) 200bp regions flanking the CRISPR/Cas9 cut site were amplified and sequenced. No mutation was detected in the control clones but frame shifts were observed in the knockout clones. **B)** Protein expression of Sox9 and Neu were analyzed by immunoblotting after culturing for 5 and 20 passages after flow sorting for RFP. **C-F)** The Zeiss M2 was used to obtain 20x magnification images of NDL3903 (C) or NDL668 (D) cells five passages after sorting or NDL3903 (E) or NDL668 (F) cells 20 passages after sorting. Scale bars represent 100μm.

Two Sox9-negative NDL3903 clones, 3903sgSox9^{K3}1 and 3903sgSox9^{K3}6, were randomly selected for *in vitro* characterization. Control clone 1 (3903sgNTC^{K3}1) was also randomly selected for further studies. Further analyses of Sox9-deficient clones showed that all 3903sgSox9^{K3}1 cells expressed low residual levels of Sox9 as assessed by immunofluorescence staining (Figure 3.5.C,E). While the expression of Sox9 in the sgNTC^{K3} clones is distinctly nuclear, Sox9 was observed to be localized throughout the 3903sgSox9^{K3}1 cells. The other NDL3903 knockout clones did not express detectable levels of Sox9 at low passage. 3903sgSox9^{K3}7 showed Sox9 expression at high passages, with a similar nuclear and cytosolic localization as 3903sgSox^{K3}1 (Figure 3.5.E). To eliminate the possibility of confounding results due to residual Sox9 expression in 3903sgSox9^{K3}1 cells, some assays were repeated with the additional two knockout clones and a second sgNTC^{K3} clone (3903sgNTC^{K3}3) at a low passage.

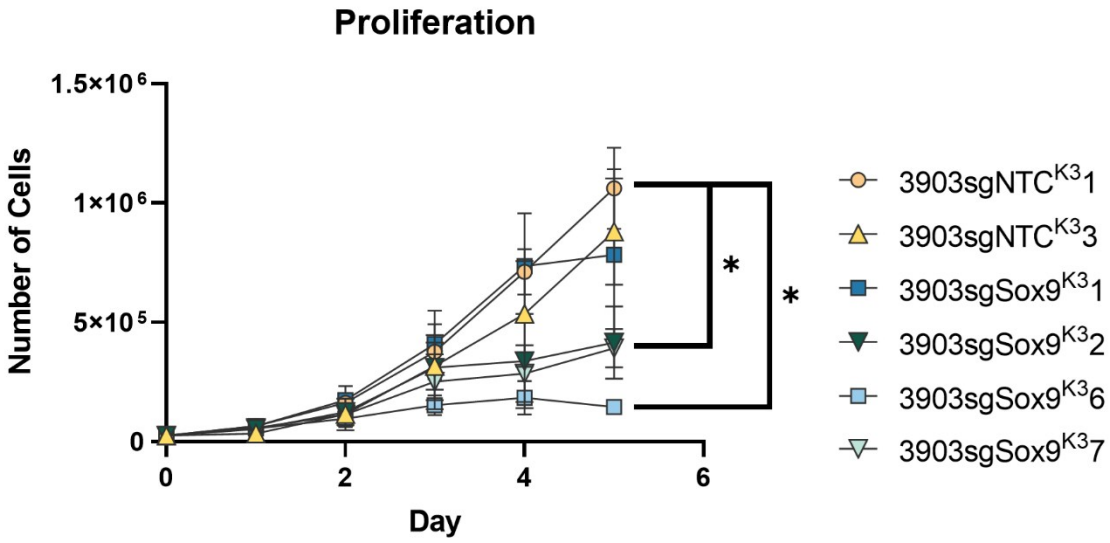
668sgNTC^{K3}2 cells were selected as the control clonal population. Although many clonal populations of the RFP-positive 668sgSox9^{K3} cells were expanded, only three knockout clones were found to be Sox9-deficient by immunoblotting and two (668sgSox9^{K3}3 and 668sgSox9^{K3}12) were randomly selected for characterization (Figure 3.4.C). However, as in the 3903sgSox9^{K3}1 cells, 668sgSox9^{K3}12 was also found to express Sox9 by immunofluorescence throughout the cell, unlike the nuclear localization in the 668sgNTC^{K3}2 cells (Figure 3.5.D,F). At high passage, Sox9 was readily detectable by immunoblotting in these cells (Figure 3.5.B).

In the clonal populations, Sox9 deletion was found to have a significant effect on proliferation (Figure 3.6). In a five-day growth assay, significantly reduced growth rates were observed between the 3903sgSox9^{K3}2, 3903sgSox9^{K3}6, and 3903sgSox9^{K3}7 cells and their respective controls (Figure 3.6.B). No changes were observed between the 3903sgNTC^{K3} clones .and 3903sgSox9^{K3}1 cells, possibly due to residual Sox9 expression. Interestingly, growth plateaus

A


3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6
Day 1
Day 5

B


Proliferation
Number of Cells
1.5×10^6
1×10^6
5×10^5
0
0
2
4
6
Day
*
*
3903sgNTC^{K3}1
3903sgNTC^{K3}3
3903sgSox9^{K3}1
3903sgSox9^{K3}2
3903sgSox9^{K3}6
3903sgSox9^{K3}7

C

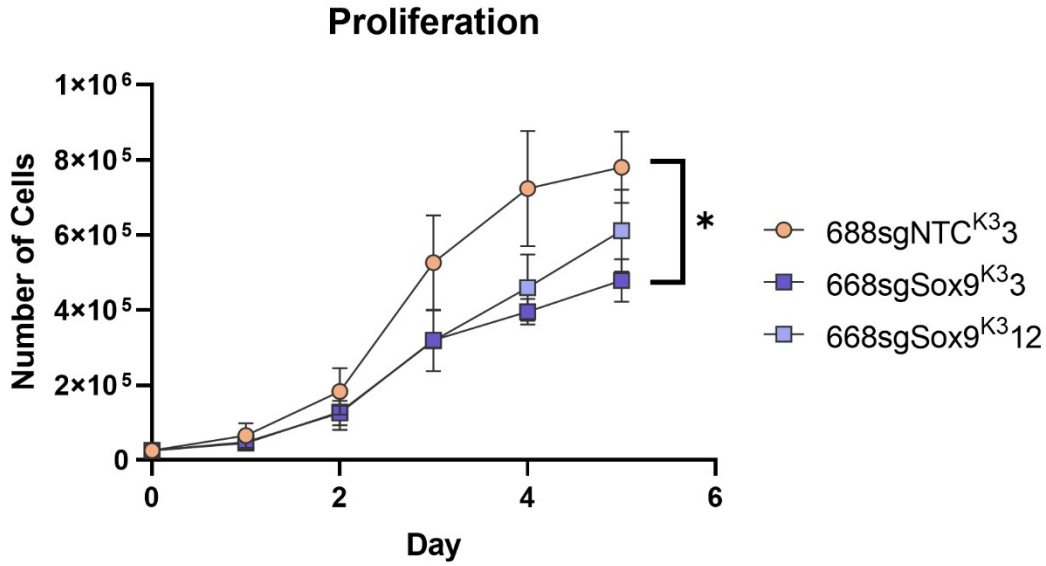

Proliferation
Number of Cells
1×10^6
8×10^5
6×10^5
4×10^5
2×10^5
0
0
2
4
6
Day
*
688sgNTC^{K3}3
668sgSox9^{K3}3
668sgSox9^{K3}12

Figure 3.6. Sox9 deletion reduces proliferation. A) Representative images of cell cultures 1 and 5 days after seeding. Phase contrast images were taken with an EVOS microscope. Scale bars represent 100µm. **B-C)** Cells were seeded in duplicate at a density of 25,000 cells/ well in a 12 well plate and counted every 24 hours for 5 days. Significance of three independent biological replicates was determined by one-way ANOVA where *P<0.05.

after 3 days in the complete knockout clones and they never reach confluency, whereas the 3903sgSox9^{K3}1 cells grow to confluency, and the 3903sgNTC^{K3} clones form foci after becoming confluent. These results suggest that the presence of Sox9 may overcome contact inhibition. Similar results were observed in the NDL668 clones. Proliferation was significantly different between the 668sgSox9^{K3}3 and 668sgNTC^{K3}2 cells, but there no significance was observed in the 668sgSox9^{K3}12 clone (Figure 3.6.C). As with the 3903sgSox9^{K3}1 cells, residual Sox9 expression in the 668sgSox9^{K3}12 cells could account for this.

We then re-assessed the self-renewal ability of the clonal populations *in vitro* (Figure 3.7). Compared to the 3903sgNTC^{K3} clones, relative primary sphere-forming efficiency was found to be reduced by about 50% in each 3903sgSox9^{K3} clone, except 3903sgSox9^{K3}1, where the efficiency is unchanged. Again, this is likely due to residual Sox9 expression in this clone. The difference is significant in the 3903sgSox9^{K3}2 and 3903sgSox9^{K3}7 clones (Figure 3.7.E). Similarly, compared to the 3903sgNTC^{K3} clones, the relative secondary sphere-forming efficiency was significantly reduced (~50%) in the three knockout clones with no Sox9 expression. The relative sphere-forming efficiency of the 3903sgSox9^{K3}1 cells is not significantly different from any of the clones, but is about 75% of the 3903sgNTC^{K3}1 efficiency, suggesting that even reduced Sox9 levels may affect self-renewal ability (Figure 3.7.D). The relative primary and secondary sphere-forming efficiencies of the 668sgSox9^{K3} clones are not significantly different from the efficiencies of the control cells. A trend similar to the NDL3903 cells is apparent in the secondary spheres, where the relative secondary sphere-forming efficiencies of the 668sgSox9^{K3}3 and 668sgSox9^{K3}12 cells is about 50% and 75% of the 668sgNTC^{K3}2, respectively (Figure 3.7.F).While not significantly reduced with Sox9 depletion, migration follows a similar downward trend. In a haptotactic Transwell assay, migration trends negatively following Sox9 knockout

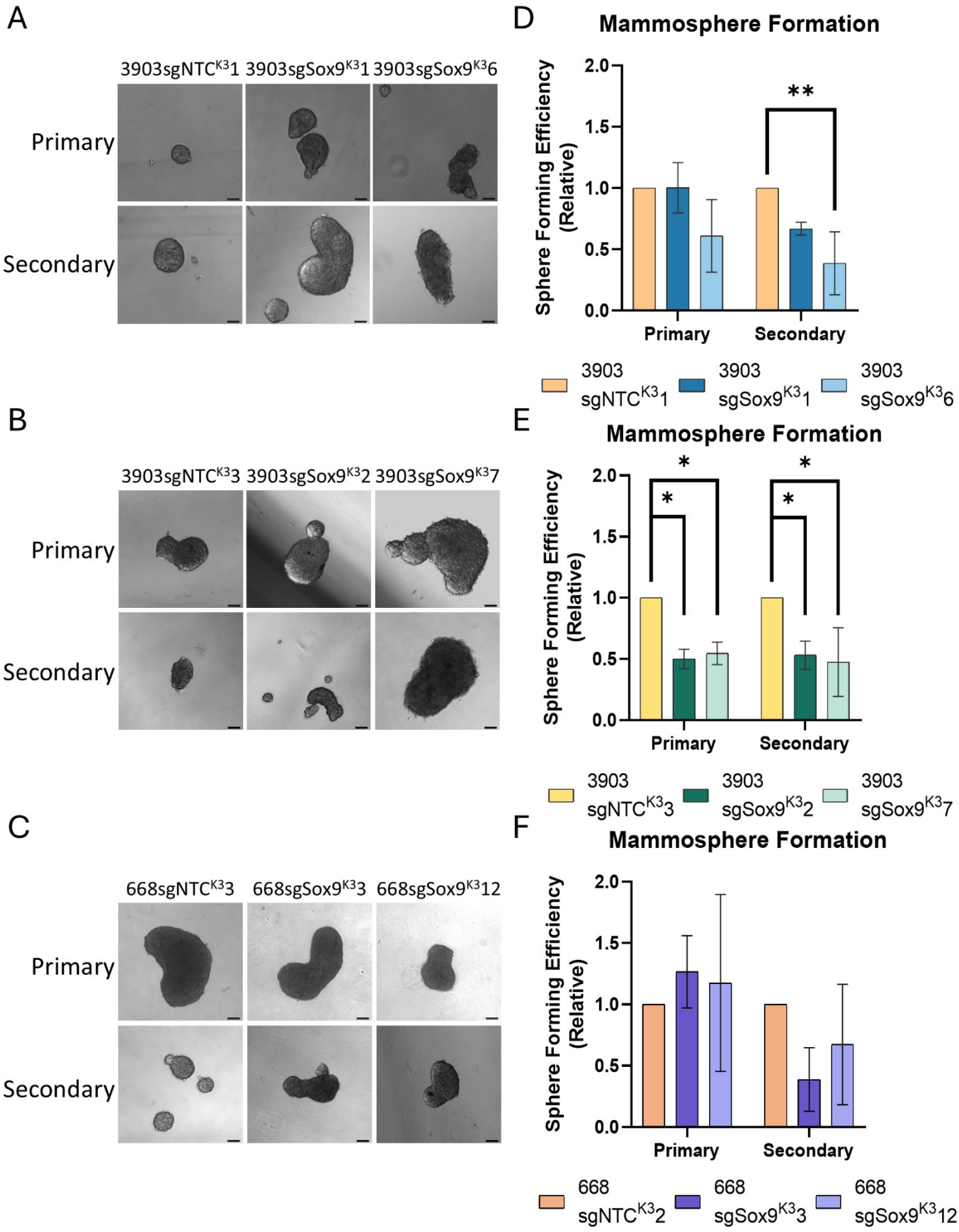

A
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6
Primary
Secondary
B
3903sgNTC^{K3}3
3903sgSox9^{K3}2
3903sgSox9^{K3}7
Primary
Secondary
C
668sgNTC^{K3}3
668sgSox9^{K3}3
668sgSox9^{K3}12
Primary
Secondary
D
Mammosphere Formation
Sphere Forming Efficiency (Relative)
2.0
1.5
1.0
0.5
0.0
**
Primary
Secondary
3903 sgNTC^{K3}1
3903 sgSox9^{K3}1
3903 sgSox9^{K3}6
E
Mammosphere Formation
Sphere Forming Efficiency (Relative)
*
*
*
*
Primary
Secondary
3903 sgNTC^{K3}3
3903 sgSox9^{K3}2
3903 sgSox9^{K3}7
F
Mammosphere Formation
Sphere Forming Efficiency (Relative)
Primary
Secondary
668 sgNTC^{K3}2
668 sgSox9^{K3}3
668 sgSox9^{K3}12

Figure 3.7. Sox9 deletion reduces self renewal ability. A-C) Representative images of primary and secondary spheres from each cell line. Phase contrast images were taken at 10x magnification with an EVOS microscope. Scale bars represent 40µm. **D-F)** Cells were seeded as a single cell suspension at an initial density of 3000 cells/ well in a 24 well ultra-low attachment plate. They were cultured in serum-free media supplemented with B27 and MEGS but otherwise normal conditions for 7 days. Secondary spheres were seeded at a density of 2000 cells/ well by disaggregating the primary spheres and culturing for a further 7 days under the same conditions. Sphere-forming efficiency was calculated as a percentage of the number of spheres with a diameter greater than 40µm per initial number of seeded cells. Significance of three independent biological replicates was determined by two-way ANOVA where *$P<0.05$.

(Figure 3.8.A-B). Relative to the 3903sgNTC^{K3}1 cells, the migration in the 3903sgSox^{K3}6 cells was reduced by approximately 50%. As in the sphere-forming assay, the effect of Sox9 knockout is stronger in the 3903sgSox^{K3}6 cells than in the 3903sgSox9^{K3}1 cells. In a chemotactic assay, the same trend was apparent in the 3903sgSox9^{K3}6 cells, where chemotactic migration was reduced by about 50% relative to the 3903sgNTC^{K3}1 cells (Figure 3.8.C-D).

The effect of Sox9 deletion on invasiveness was assessed with two independent assays. In a Matrigel-Transwell assay, invasion was significantly reduced in both knockout lines (Figure 3.9.A-B). In contrast to the haptotactic and chemotactic migration assays, invasion was significantly reduced in both 3903sgSox9^{K3}1 and 3903sgSox9^{K3}6 cells relative to the 3903sgNTC^{K3}1 cells, suggesting that Sox9 may control genetic programs required for invasion.

The invasiveness of the cells was also measured by their ability to send invasive processes through Matrigel in a droplet assay. Although not significant, a downward trend was observed in the number of invasive protrusions produced per sphere, with half as many forming in 3903sgSox9^{K3}6 relative to the control (Figure 3.9.D). However, a significant reduction in the average protrusion length is apparent between those formed by the 3903sgSox9^{K3}6 spheres and the 3903sgNTC^{K3}1 spheres (Figure 3.9.E). The average length of protrusions formed by the 3903sgSox9^{K3}1 spheres was not significantly different from the control cells. The protrusion lengths were also binned to assess size distribution (Figure 3.9.F). Although a trend towards shorter arms was observed in 3903sgSox9^{K3}6 cells, differences in the distribution of arm length were only significant for protrusions measuring 350-450px in length. The 3903sgSox9^{K3}6 spheres did not form any protrusions larger than 650px. Together, this supports our observations from the Matrigel-Transwell assay and suggests that Sox9-deficient cells may be less invasive.

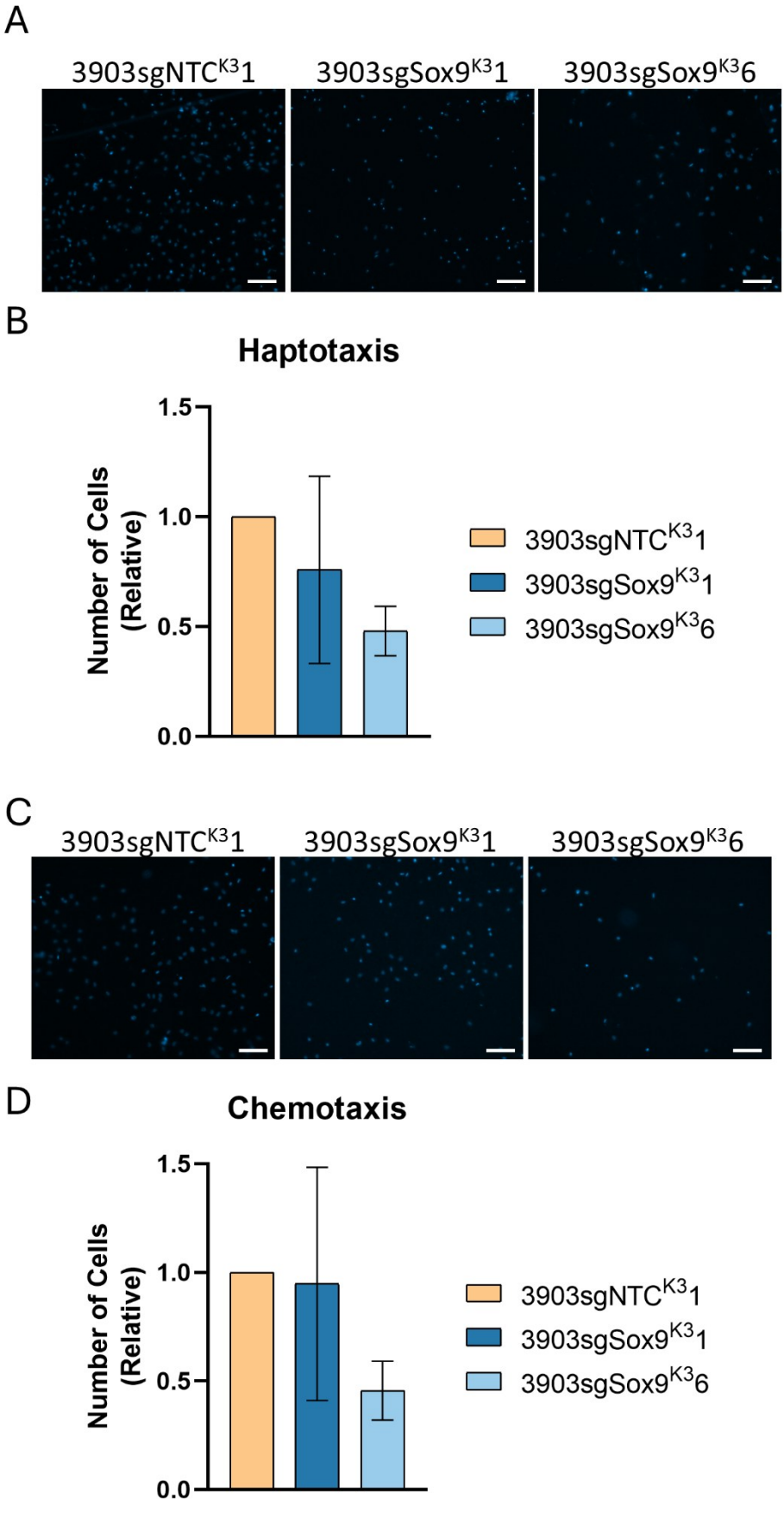

A
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6
B
Haptotaxis
Number of Cells (Relative)
1.5
1.0
0.5
0.0
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6
C
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6
D
Chemotaxis
Number of Cells (Relative)
1.5
1.0
0.5
0.0
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6

Figure 3.8. The effect of Sox9 on migration. A) Representative fluorescent images of cells fixed on Transwell chamber (8µm pore) membranes after haptotactic migration. 50,000 cells were seeded per collagen-coated chamber with 1% serum media in both chambers. Cells were incubated under normal conditions for 8 hours. Cells were fixed with 4% PFA and stained with DAPI to visualize the nuclei. Scale bars represent 100µm **B)** Sox9 deletion results in a negative trend in haptotactic migration in clonal populations. ImageJ software was used to count the nuclei of 9 ROI from each of three membranes per cell line. Cell counts were normalized to the average of the sgNTC.1 replicates and significance of three independent biological replicates was determined by one-way ANOVA where $P<0.05$. **C)** Representative images of cells fixed on Transwell chamber (8µm pore) membranes after chemotactic migration. 30,000 cells were seeded per chamber with 10% serum media in the bottom chamber and all other conditions as described in the haptotactic assay. **D)** Sox9 deletion results in a negative trend in chemotactic migration in clonal populations. Cells were counted and significance was calculated as described in (B).

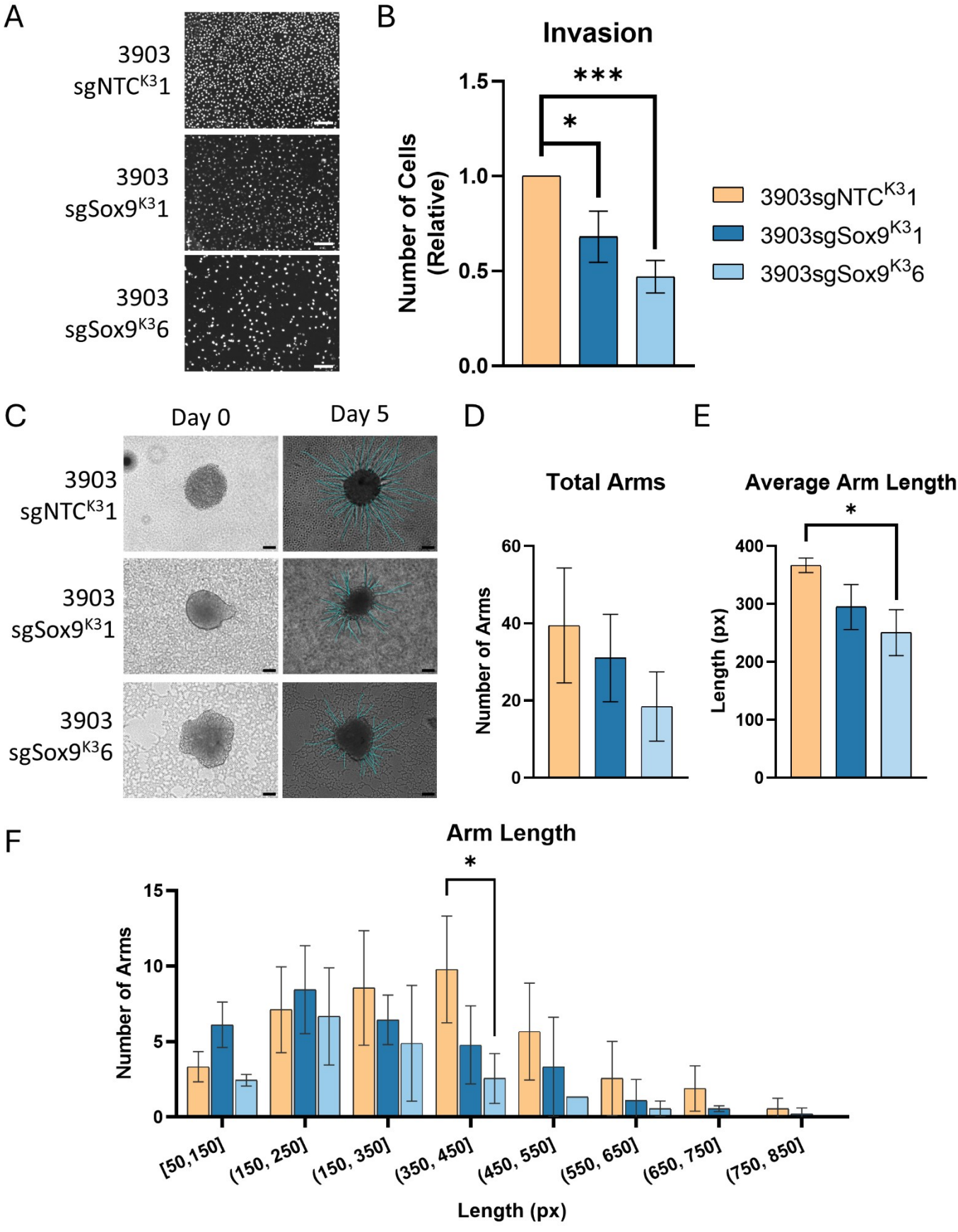

A
3903 sgNTC^{K3}1
3903 sgSox9^{K3}1
3903 sgSox9^{K3}6
B
Invasion
***
*
Number of Cells (Relative)
1.5
1.0
0.5
0.0
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}6
C
Day 0
Day 5
3903 sgNTC^{K3}1
3903 sgSox9^{K3}1
3903 sgSox9^{K3}6
D
Total Arms
Number of Arms
60
40
20
0
E
Average Arm Length
*
Length (px)
400
300
200
100
0
F
Arm Length
*
Number of Arms
15
10
5
0
[50,150]
(150, 250]
(150, 350]
(350, 450]
(450, 550]
(550, 650]
(650, 750]
(750, 850]
Length (px)

Figure 3.9. Sox9 deletion inhibits invasive ability. A) Representative fluorescent images of cells fixed on Matrigel coated Transwell chamber (8µm pore) membranes after invasion. 50,000 cells were seeded per Matrigel and collagen-coated chamber with 1% serum media in both inserts. Cells were incubated under normal conditions for 24 hours. Cells were fixed with 4% PFA and stained with DAPI to visualize the nuclei. Scale bars represent 50µm. **B)** Sox9 deletion reduces invasion in clonal populations. ImageJ software was used to count the nuclei of 9 ROI from each of three membranes per cell line. Cell counts were normalized to the average of the sgNTC.1 replicates and significance of three independent biological replicates was determined by one-way ANOVA where *P<0.05. **C)** Representative images of spheres suspended in Matrigel droplets. Cells were initially seeded at a density of 5,000 cells per well in a 5mg/mL poly-HEMA coated 96-well plate and cultured on a shaker at 120 RPM and otherwise normal conditions for 48 hours. Spheres were then suspended in a droplet of 1:1 Matrigel: collagen, covered with 1% media and cultured under normal conditions until an invasive protrusion extended beyond the field of view (5 days). **D-F)** The invasive protrusions formed by 5 spheres per cell line were counted and measured in pixels with ImageJ software. Significance of three independent biological replicates was determined by one-way ANOVA where *P<0.05 and ***P<0.005.

3.3. Effect of Sox9 knockout on gene expression

To identify potential Sox9-dependent genetic programs, we performed bulk RNA-sequencing on 3903sgNTC^{K3}1, 3903sgSox9^{K3}1, 3903sgSox9^{K3}2, and 3903sgSox9^{K3}6 cells. We reasoned that the use of 3 independent clones would reduce false positives due to clonal variability. Analyses of all replicates showed that there were only 343 positively regulated and 165 negatively regulated genes in common between all three knockout clones, suggesting that a significant number of changes may be due to clonal variability (Figure 3.10.A). Generally, epithelial markers were upregulated and stemness markers were downregulated in the knockout cells, however, the specific gene networks that were most affected varied in each clone. GSEA of each knockout clone relative to the control clone revealed activated gene signatures associated with differentiated, less metastatic cancers and apoptosis in 3903sgSox9^{K3}2 and 3903sgSox9^{K3}6 cells (Figure 3.10.C-D)[217–224]. Similar pathways were activated in the 3903sgSox9^{K3}1 cells, but with an upregulation of pro-survival pathways that are not as apparent in the other two clones (Figure 3.10.B)[225–227]. A similar trend is seen in the downregulated gene signatures. Suppressed genes in the complete knockout clones are related to undifferentiated, metastatic cancers, growth, and motility, while in 3903sgSox9^{K3}1 cells a contradictory suppression of pathways related to growth inhibition was observed (Figure 3.10.B-D)[217,219,220,224,228–235]. It is possible that Sox9 dosage affects the activation of some transcriptional networks. The low expression level of Sox9 in the 3903sgSox9^{K3}1 cells may favour binding to high affinity sites within distinct sets of genes such as growth suppression. We further investigated the differentially expressed genes by conducting OVA on the grouped knockout clones (Figure 3.11.B). We found an overrepresentation of genes related to migration, invasion, and cell signalling, suggesting that these systems were disproportionately affected by the deletion of Sox9. Sox9 is elevated in TNBC, therefore, we also investigated known markers of

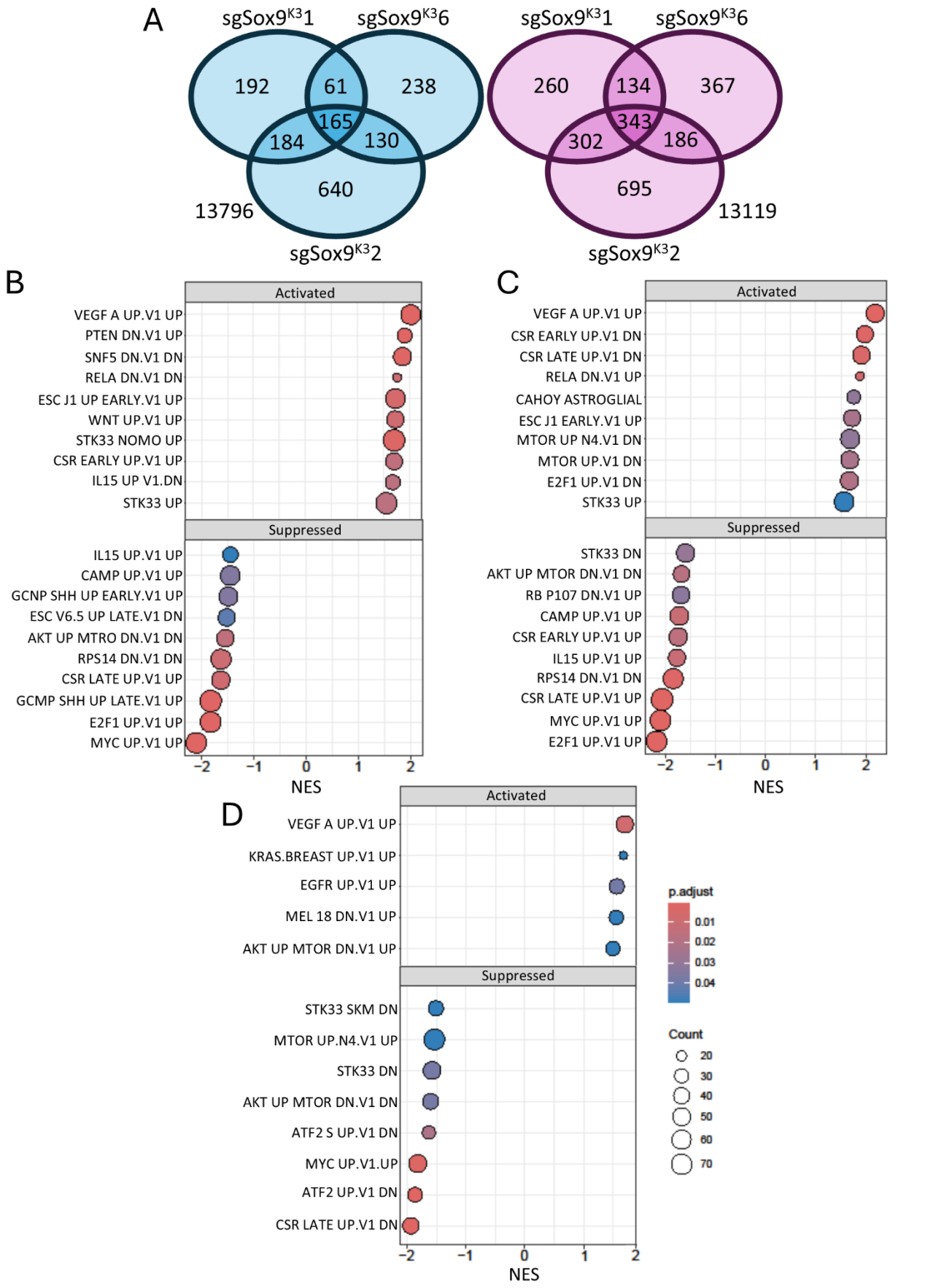

A
sgSox9K3 1
sgSox9K3 6
192
61
238
165
184
130
640
13796
sgSox9K3 2
sgSox9K3 1
sgSox9K3 6
260
134
367
343
302
186
695
13119
sgSox9K3 2
B
Activated
VEGF A UP.V1 UP
PTEN DN.V1 UP
SNF5 DN.V1 DN
RELA DN.V1 DN
ESC J1 UP EARLY.V1 UP
WNT UP.V1 UP
STK33 NOMO UP
CSR EARLY UP.V1 UP
IL15 UP V1.DN
STK33 UP
Suppressed
IL15 UP.V1 UP
CAMP UP.V1 UP
GCNP SHH UP EARLY.V1 UP
ESC V6.5 UP LATE.V1 DN
AKT UP MTRO DN.V1 DN
RPS14 DN.V1 DN
CSR LATE UP.V1 UP
GCMP SHH UP LATE.V1 UP
E2F1 UP.V1 UP
MYC UP.V1 UP
−2 −1 0 1 2
NES
C
Activated
VEGF A UP.V1 UP
CSR EARLY UP.V1 DN
CSR LATE UP.V1 DN
RELA DN.V1 UP
CAHOY ASTROGLIAL
ESC J1 EARLY.V1 UP
MTOR UP N4.V1 DN
MTOR UP.V1 DN
E2F1 UP.V1 DN
STK33 UP
Suppressed
STK33 DN
AKT UP MTOR DN.V1 DN
RB P107 DN.V1 UP
CAMP UP.V1 UP
CSR EARLY UP.V1 UP
IL15 UP.V1 UP
RPS14 DN.V1 DN
CSR LATE UP.V1 UP
MYC UP.V1 UP
E2F1 UP.V1 UP
−2 −1 0 1 2
NES
D
Activated
VEGF A UP.V1 UP
KRAS.BREAST UP.V1 UP
EGFR UP.V1 UP
MEL 18 DN.V1 UP
AKT UP MTOR DN.V1 UP
Suppressed
STK33 SKM DN
MTOR UP.N4.V1 UP
STK33 DN
AKT UP MTOR DN.V1 DN
ATF2 S UP.V1 DN
MYC UP.V1.UP
ATF2 UP.V1 DN
CSR LATE UP.V1 DN
−2 −1 0 1 2
NES
p.adjust
0.01
0.02
0.03
0.04
Count
20
30
40
50
60
70

Figure 3.10. Summary of differentially regulated genes in Sox9 knockout clones. A) Venn diagram representing significantly downregulated (blue) or upregulated (pink) differentially expressed genes (adjusted $p<0.05$ $|log2FC|>1$) in common identified in the independent knockout clones relative to the control. **B-D)** Gene Set Enrichment Analysis of genes ranked by the average Wald statistic of 3903sgSox9^{K3}1 (B), 3903sgSox9^{K3}2 (C), or 3903sgSox9^{K3}6 (D). Gene sets were derived from the MSigDB C6: Oncogenic Signatures collection for *Mus musculus*. The dot plot displays the top positively (activated) and negatively (suppressed) enriched gene sets normalized enrichment scores (NES) with adjust $p<0.05$. Dot size reflects the number of contributing genes and colour indicates adjusted p-value. Figures by Samuel Delisle.

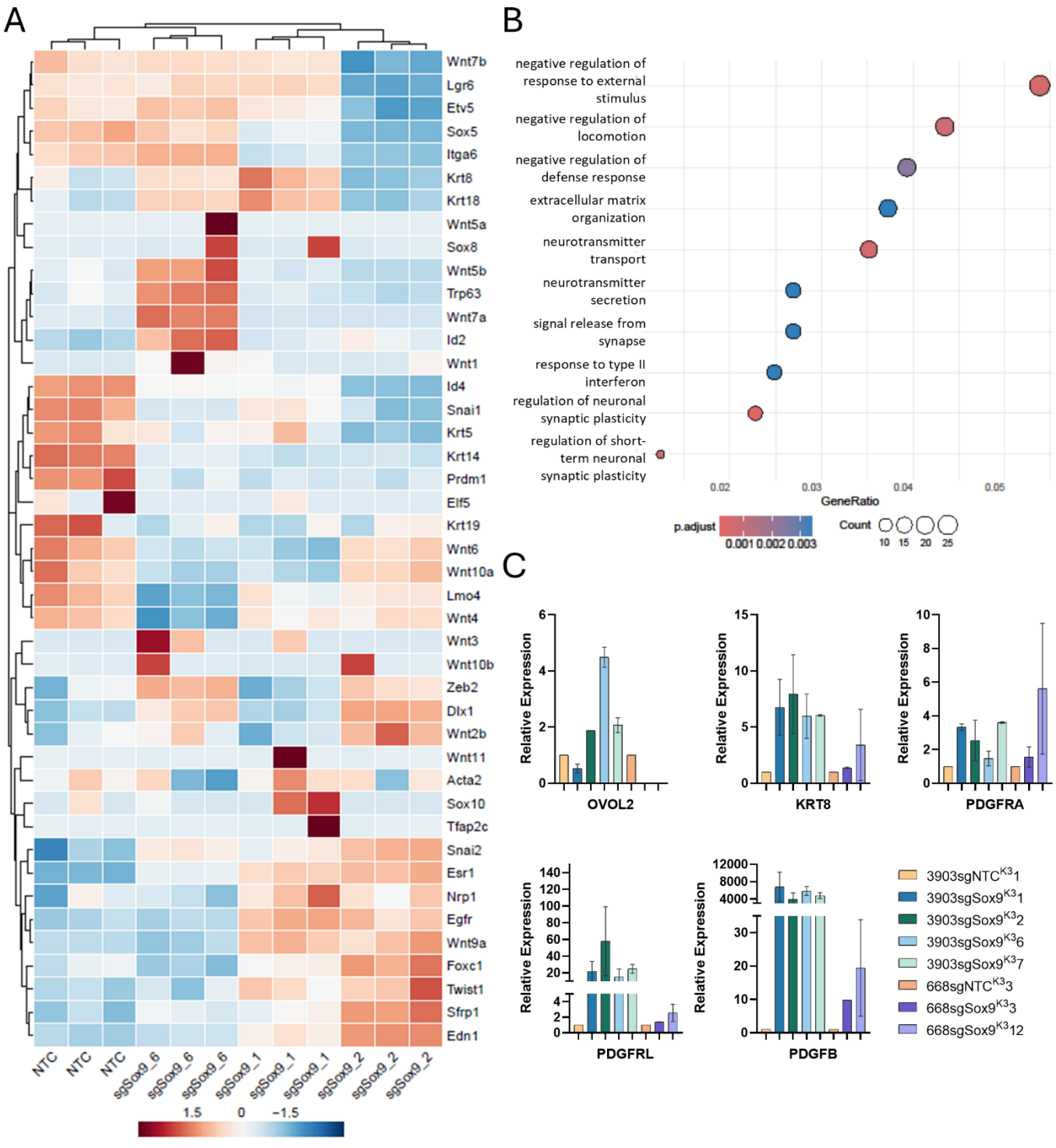

A
Wnt7b
Lgr6
Etv5
Sox5
Itga6
Krt8
Krt18
Wnt5a
Sox8
Wnt5b
Trp63
Wnt7a
Id2
Wnt1
Id4
Snai1
Krt5
Krt14
Prdm1
Elf5
Krt19
Wnt6
Wnt10a
Lmo4
Wnt4
Wnt3
Wnt10b
Zeb2
Dlx1
Wnt2b
Wnt11
Acta2
Sox10
Tfap2c
Snai2
Esr1
Nrp1
Egfr
Wnt9a
Foxc1
Twist1
Sfrp1
Edn1
NTC
NTC
NTC
sgSox9_6
sgSox9_6
sgSox9_6
sgSox9_1
sgSox9_1
sgSox9_1
sgSox9_2
sgSox9_2
sgSox9_2
1.5
0
−1.5
B
negative regulation of response to external stimulus
negative regulation of locomotion
negative regulation of defense response
extracellular matrix organization
neurotransmitter transport
neurotransmitter secretion
signal release from synapse
response to type II interferon
regulation of neuronal synaptic plasticity
regulation of short-term neuronal synaptic plasticity
0.02
0.03
0.04
0.05
GeneRatio
p.adjust
0.001 0.002 0.003
Count
10 15 20 25
C
Relative Expression
OVOL2
KRT8
PDGFRA
PDGFRL
PDGFB
3903sgNTC^{K3}1
3903sgSox9^{K3}1
3903sgSox9^{K3}2
3903sgSox9^{K3}6
3903sgSox9^{K3}7
668sgNTC^{K3}3
668sgSox9^{K3}3
668sgSox9^{K3}12

Figure 3.11. Sox9 deletion affects differentiation, not luminal to basal reprogramming. A) Triplicate Rlog-transformed gene expression values were used to generate a heatmap for a curated list of luminal and basal markers in 3903sgSox9^{K3}1, 3903sgSox9^{K3}2, and 3903sgSox9^{K3}6 cells. B) Overrepresentation analysis of shared differentially expressed genes (adjusted $p<0.05$, $|\log_2 FC|>1$) using the gene ontology biological processes database. The top 10 enriched pathways were based on adjusted p-value and are displayed ordered by gene ratio. Dot size represents the number of enriched genes per pathway, and dot colour indicates adjusted p-value. **C)** Quantitative PCR (qPCR) validation of selected epithelial markers in all clonal cells. Figures A and B by Samuel Delisle.

luminal and basal epithelial cells but found no clear trends (Figure 3.11.A)[136]. This finding suggests that while Sox9-deficient cells may be more differentiated, Sox9 is unlikely to have a significant effect on luminal to basal reprogramming. We validated our findings and assessed epithelial-like gene expression in additional clones by qPCR (Figure 3.11.C). While the fold change was variable between the clones, an upregulation was observed in all genes with the exception of OVOL2, which was not expressed in the 668sgSox9^{K3} clones. Corroborating the results of the bulk RNA sequencing, epithelial-associated genes were upregulated in the Sox9-deficient clones relative to the control clones. These results, together with our *in vitro* findings, suggest that Sox9-deficient cells are less stemlike than Sox9-expressing cells. To further test the effect of Sox9-inactivation we conducted *in vivo* experiments to functionally assess the stemness of the cell lines.

3.4. Sox9 knockout inhibits cell growth *in vivo*

Our *in vitro* data suggest that Sox9 deletion results in defects in proliferation, self renewal, and invasion. To test this *in vivo* we used functional xenograft assays in immunocompromised NCG mice. For these studies, as it is a complete Sox9 knockout, only 3903sgSox9^{K3}6 cells were tested alongside the 3903sgNTC^{K3}1 control cells. Orthotopic injections were performed as a model of tumour initiation and progression. Two different cell densities of Sox9-expressing and Sox9-deficient clonal populations were injected into the mammary fat pads of mice. Supporting our mammosphere data, no tumours were detected in mammary glands injected with Sox9-deficient 3903sgSox9^{K3}6 cells, suggesting a significant depletion of tumour-initiating capacity (Figure 3.12.A). The time to tumour onset in glands injected with 3903sgNTC^{K3}1 was found to correlate with the number of cells injected. Most tumours ulcerated and resulted in humane endpoint being reached before the cumulative tumour burden dictated by ACVS (1.7cm^3). Unexpectedly, no tumours strongly expressed SOX9 or SOX10 at endpoint (Figure 3.12.B), suggesting that the

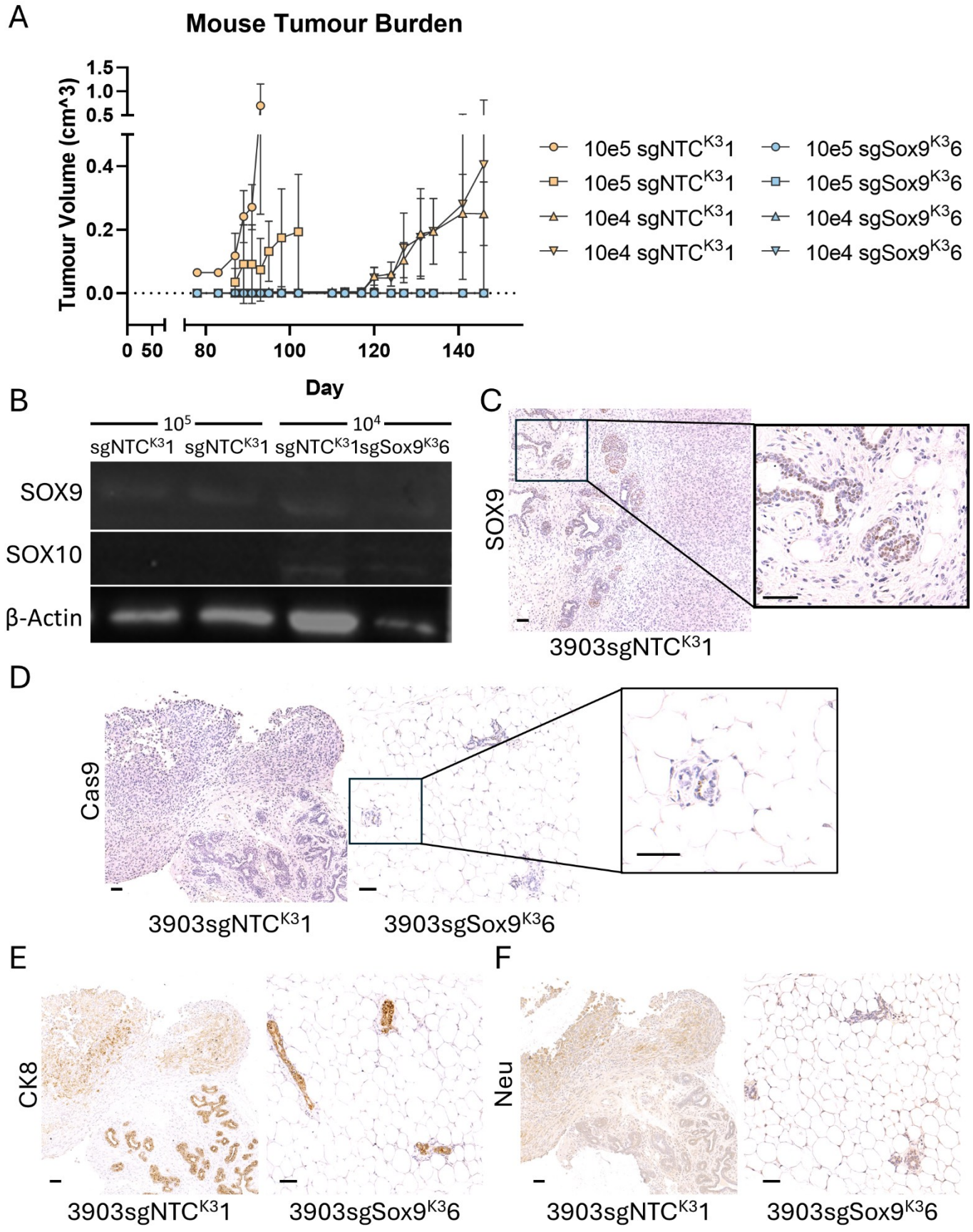

A
Mouse Tumour Burden
Tumour Volume (cm^3)
1.5
1.0
0.5
0.4
0.2
0.0
0 50 80 100 120 140
Day
10e5 sgNTC^K3 1
10e5 sgNTC^K3 1
10e4 sgNTC^K3 1
10e4 sgNTC^K3 1
10e5 sgSox9^K3 6
10e5 sgSox9^K3 6
10e4 sgSox9^K3 6
10e4 sgSox9^K3 6
B
10^5
10^4
sgNTC^K3 1 sgNTC^K3 1 sgNTC^K3 1sgSox9^K3 6
SOX9
SOX10
β-Actin
C
SOX9
3903sgNTC^K3 1
D
Cas9
3903sgNTC^K3 1
3903sgSox9^K3 6
E
CK8
3903sgNTC^K3 1
3903sgSox9^K3 6
F
Neu
3903sgNTC^K3 1
3903sgSox9^K3 6

Figure 3.12. Sox9 deficient cells do not form tumours when injected orthotopically. A) Each NCG mouse received four injections of the same concentration, with two injections of 3903sgNTC^{K3}1 cells and two injections of 3903sgSox9^{K3}6 cells. After onset, tumours were measured twice weekly until humane endpoint. Point shape indicates individual mouse. **B)** Protein expression in tumours and mammary glands were assessed by immunoblotting. **C-F)** Immunohistochemical staining of SOX9 (C, brown), Cas9 (D, brown), CK8 (E, brown), and Neu (F, brown) counterstained with hematoxylin (purple). Images were taken with a Zeiss Axio Scan.Z1 at 100x magnification. Scale bars represent 50µm.

3903sgNTC^{K3}1 cells may have become Sox9-deficient during expansion. The low level of SOX9 detected is likely from the normal mammary gland tissue, as visualized by IHC (Figure 3.12.C). Cas9 was only evident in small clusters of Cas9-positive cells in the mammary glands injected with 3903sgSox9^{K3}6 cells (Figure 3.12.D). Cytokeratin 8 (CK8), a luminal marker, was evident throughout the tumours but was localized to the luminal layer of ductal structures in areas of normal tissue (Figure 3.12.E). A similar CK8 expression pattern was apparent in the mammary glands injected with 3903sgSox9^{K3}6 cells. Neu was expressed throughout luminal structures but was more strongly expressed in the tumour tissue (Figure 3.12.F). Together, with a reduced self renewal capacity *in vitro*, these data suggest that Sox9-deficient cells are unable to expand *in vivo* and have reduced tumour initiating potential.

To test the ability of Sox9-deficient cells to colonize distant organs the same cell lines were injected into the tail-veins of NCG mice to assess extravasation and secondary tumour establishment in lung tissue. Within 2 weeks of injections, three of the six mice injected with 3903sgNTC^{K3}1 cells expired due to causes unrelated to tumour burden and were excluded. At endpoint, lung tissues were harvested and single cell preparations were subjected to RFP flow sorting to quantitate the percentage of injected cells. Flow cytometry showed that the lung colonization ability of Sox9-deficient cells was about 50% of that of controls (Figure 3.13.A-B). Macrometastatic lesions identifiable by hematoxylin and eosin staining were only apparent in the lungs of mice injected with 3903sgNTC^{K3}1 cells, and Cas9 expression was detected throughout these structures (Figure 3.13.C). Lungs of mice injected with 3903sgSox9^{K3}6 did not show any obvious metastases identifiable by hematoxylin and eosin staining but did contain small clusters of Cas9-expressing cells scattered throughout the tissue (Figure 3.13.D). Together our data suggest

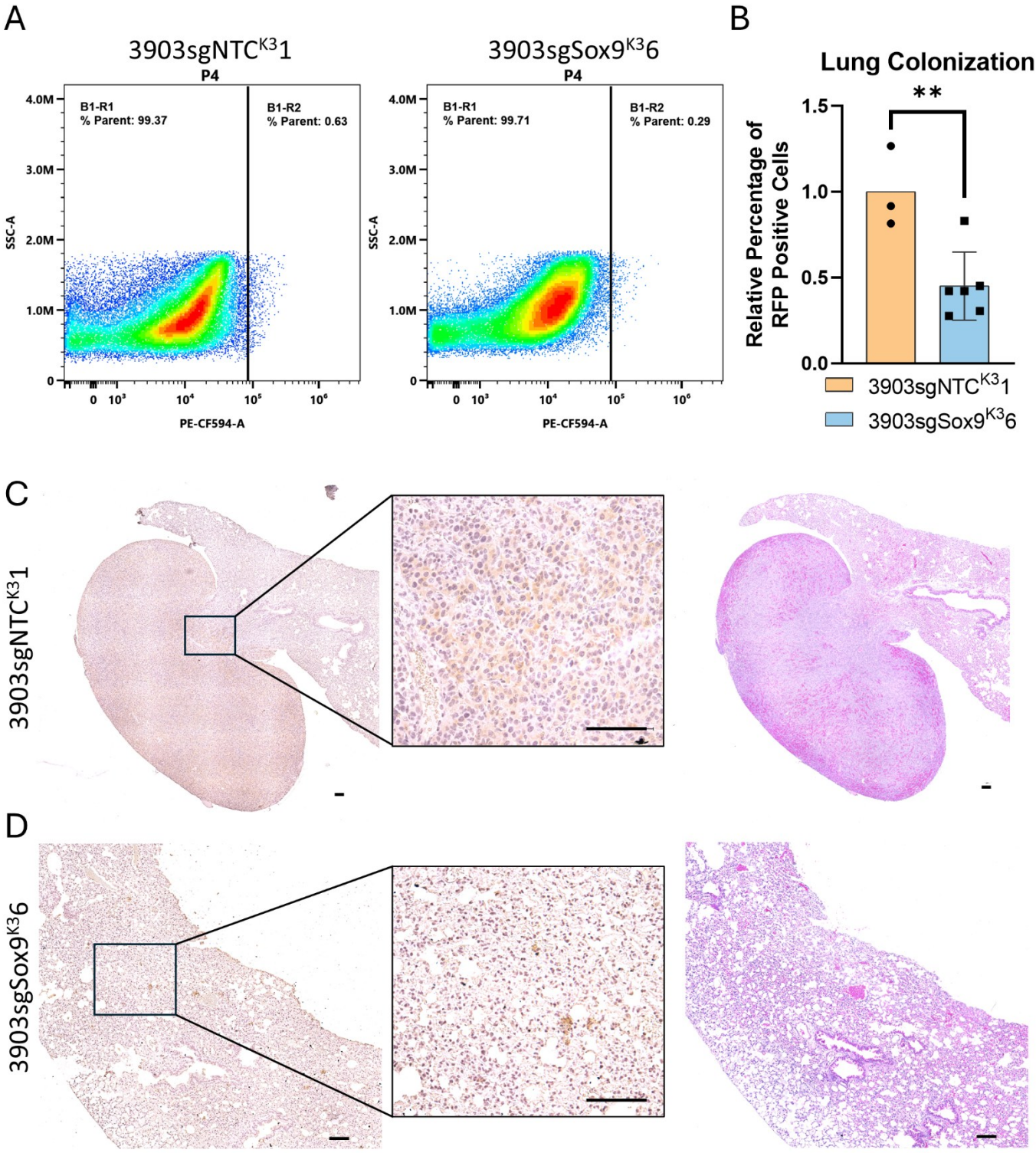

A
3903sgNTC^{K3}1
P4
B1-R1
% Parent: 99.37
B1-R2
% Parent: 0.63
SSC-A
PE-CF594-A
3903sgSox9^{K3}6
P4
B1-R1
% Parent: 99.71
B1-R2
% Parent: 0.29
SSC-A
PE-CF594-A
B
Lung Colonization
**
Relative Percentage of RFP Positive Cells
3903sgNTC^{K3}1
3903sgSox9^{K3}6
C
3903sgNTC^{K3}1
D
3903sgSox9^{K3}6

Figure 3.13. SOX9 knockout inhibits lung colonization. A) Representative plots of FACS for RFP. Gating strategy is based on unstained lungs and cultured cells used in the injections. **B)** Fewer RFP positive cells were found in the lungs of mice injected with 3903sgSox9^{K3}6 cells. Each point represents a mouse injected with 1.0x10^{6} cells. Summary data is normalized to the unstained lungs and the average of 3903sgNTC^{K3}1 samples. Significance was determined with an unpaired t-test where **P<0.01. **C-D)** Representative images of immunohistochemical staining for Cas9 (brown) counterstained with hematoxylin (purple), and hematoxylin counterstained with eosin (pink) in lungs from mice injected with 3903sgNTC^{K3}1 (C) or 3903sgSox^{K3}6 (D) cells. Images were taken with a Zeiss Axio Scan.Z1 at 100x magnification. Scale bars represent 100μm.

that Sox9 knockout impairs tumour initiating activity and the establishment of distant niches following extravasation in lung tissues.

4. General Discussion

4.1. Overview

Metastatic breast cancer is a leading cause of cancer death globally, and its incidence is increasing each year[12]. The disease is particularly heterogenous and has been divided into molecular subtypes to help direct prognosis and treatment decisions. The HER2-positive subtype is driven by the overexpression of HER2 and is associated with worse outcome than hormone receptor-positive cancers[14]. Elevated HER2 expression results in an increase of HER2 containing dimers and promotion of pro-survival pathways downstream of HER2 signalling, contributing to a more aggressive disease progression[19,23]. This subtype is also associated with a 70% risk of developing resistance to first-line therapy, typically chemotherapy in conjunction with a HER2-targeting monoclonal antibody[26–28]. A key mechanism of metastasis and acquired treatment resistance in carcinoma is the process of EMT. Through EMT, epithelial-like cells acquire stem-like characteristics with increased self-renewal capacity, migration, invasion, and immune- and therapeutic resistance[200,201]. Further understanding of the gene and protein systems that are driving this process in HER2-positive breast cancer may reveal novel therapeutic options for HER2-targeted therapies. Sox9 is integral to terminal differentiation in embryonic development and has been implicated as a regulator of stemness in a variety of cell types in both normal and pathological processes[166,167,236]. Notably, it has been implicated as a driver of tumorigenesis, EMT, and metastasis in many malignancies, but its role in HER2-positive breast cancer has not been fully investigated[134,136,188–190]. The involvement of Sox9 in these processes would identify it and its downstream networks as potential candidates for the development of novel therapeutic options. Here, we aimed to characterize the effect of Sox9-deficiency on the growth and stemness of HER2-positive tumour-derived cell lines.

4.2. The role of Sox9 on tumour cell phenotype *in vitro*

4.2.1. The effect of Sox9 deletion on Neu+ cell populations

We began our investigation by identifying two Neu+ cell lines with high (NDL3903) and low (NDL668) endogenous Sox9 expression to assess the role of Sox9 on cellular phenotypes, and first used CRISPR/Cas9 to inactivate Sox9 in each. Although significant knockdown of Sox9 was achieved, residual Sox9 expression was observed. Serial passaging of those cells resulted in the return of Sox9 expression to near-control levels, suggesting that Sox9-expressing subpopulations have a growth advantage (Figure 3.1). The re-expressed Sox9 band appeared smaller than the wildtype when assessed by immunoblotting, suggesting that the protein was truncated in some way. It is possible that a protein is being translated from an internal start codon downstream of the CRISPR/Cas9 induced frame-shift mutation. This residual expression likely contributes to the lack of significant phenotypic changes observed in the bulk cells that are apparent in experiments using clonally derived cells. Sox9 has been demonstrated to affect proliferation and self-renewal ability in some cancer types, therefore, we assessed the effect of its deletion on these properties with growth and sphere-forming assays, respectively[136,189]. Although proliferation in the Sox9-deficient bulk cell populations was unchanged, the increase in Sox9 expression over time suggests that Sox9-expressing subpopulations are outgrowing the culture. Similar findings have been reported *in vivo*, where a mosaic Sox9-knockout model returned to a normal phenotype in just 8 weeks[185]. Primary sphere-formation is indicative of substrate independent growth, a progenitor cell characteristic, whereas secondary sphere-formation is indicative of self-renewal ability, a classic stem cell characteristic[237]. In both bulk Sox9-knockout populations, only secondary sphere-forming efficiency was reduced, suggesting that while the bulk $sgNTC^{K1K2}$ control cells were more stem-like than their $sgSox9^{K1K2}$ counterparts, they were not more progenitor-like.

In an effort to obtain a more efficient knockout, we designed new sgRNA targeting the exon 1 region of Sox9 (sgSox9^{K3}, Figure 3.3). The NDL3903 knockout line was generated from the same early-passage parental cells as the overexpressed line and a comparably early passage of NDL668 cells was used to generate control and knockout cells from an innately-low Sox9 expressing line. We then reassessed cell growth and sphere-forming efficiency with the 3903sgSox9^{K3} bulk populations and observed similar effects as in the s3903sgSox9^{K1K2} cells, although primary sphere-forming efficiency was significantly reduced in the knockout cells. Again, although significant Sox9 knockdown was achieved, residual Sox9 expression was still detectable and increased with cell passage, likely due to outgrowth of Sox9-expressing subpopulations. Therefore, clonal populations were expanded from Sox9-deficient cells.

We found that more clones of RFP-positive sgNTC^{K3} cells could be recovered than clones of RFP-positive sgSox9^{K3} cells in both NDL lines, supporting our previous finding of reduced self-renewal ability in the Sox9-deficient cells (Figure 3.4). Sox9 knockdown has been shown to induce apoptosis in human TNBC and colorectal cancer, and its re-expression has been shown to rescue microRNA induced apoptosis in lung cancer[136,238,239]. It is possible that Sox9 inactivation in sgSox9^{K3} cells could induce cell death. This would also suggest that Sox9 deficient cells recovered after expansion may have been able to overcome a Sox9-driven requirement for growth and viability. RNA-sequencing analysis showed that there is significant variability in gene expression between the clones (Figure 3.10 and Figure 3.11). This could reflect adaptive differences in the activation of compensatory mechanisms in the absence of Sox9. The activation of gene pathways related to survival and cell growth suggests that the cells may be promoting pathways to compensate for Sox9's dysfunction. Nonetheless, Sox9-deficient clones could be expanded and exhibited reduced proliferation, stemness and invasion compared to Sox9-expressing cells.

Many unsuccessful attempts were made to generate CRISPR/Cas9 mediated Sox9-knockout cells from human cell lines. At best, Sox9 appeared to be minimally reduced and the expression returned to control-like levels after one or two passages (Appendix Figure 1). Any attempts to expand clonal populations only resulted in Sox9-expressing colonies. These findings strongly suggest that the requirement for Sox9 is more critical in human cells than in murine cells, or that human cells are less likely to activate pathways to compensate for the loss of Sox9. This would be beneficial in the therapeutic targeting of Sox9, where compensatory mechanisms would be less likely to be activated.

Unexpectedly, in all clones with residual Sox9 expression, Sox9 localization was apparent throughout the cell in contrast to the distinctly nuclear localization observed in the sgNTC^{K3} clones. Alterations to the Sox9 protein due to the CRISPR/Cas9 mutation may have affected its regulation of nucleo-cytoplasmic shuttling, and this dysregulation may have contributed to the modest phenotypic difference observed in clones that express Sox9 at a low passage, referred to as incomplete knockout clones[113]. It is possible that the cell is unable to recognize the truncated protein as a fragment and the apparent increase of Sox9 at high passage in the knockout clones is due to the lack of degradation.

In the complete knockout clones, 3903sgSox9^{K3}2, 3903sgSox9^{K3}6, 3903sgSox9^{K3}7, and 668sgSox9^{K3}3, we observed growth arrest at high density, also known as contact inhibition (Figure 3.6)[240]. As the sgNTC^{K3} cells continued to grow they formed a relatively consistent lawn of cells and began to form foci after reaching 100% confluency. In contrast, the complete Sox9 knockout cells reached a growth plateau before reaching more than 90% confluency. The incomplete knockout cells did not form foci but did grow to a higher density than the complete knockout cells. These growth patterns suggest that cells with Sox9 expression are not only more proliferative but

also able to overcome contact inhibition of growth. Cells that do not respond to contact inhibition tend to be more mesenchymal, and have a greater ability for substrate independent growth, as there is crosstalk between the regulatory pathways that promote proliferation and inhibit anoikis in cells undergoing EMT[241].

Generally, relative primary and secondary sphere-forming efficiency were reduced in the 3903sgSox9^{K3} clones, relative to the 3903sgNTC^{K3} clones (Figure 3.7). Consistent with the observations from the proliferation assays, the complete knockout cells displayed reduced substrate independent growth as evidenced by their lower primary sphere-forming efficiency[237]. Results in the NDL668 cells were also consistent with the proliferation assay, as these cells were not observed to have their growth plateau following Sox9-deletion they were also less likely to have their growth inhibited by non-adherent conditions[241]. As in the bulk cell experiments, secondary sphere-forming efficiency was significantly reduced in all complete 3903sgSox9^{K3} clones and trended towards reduced efficiency in the 668sgSox9^{K3} clones. These results suggest that Sox9 may promote both a progenitor-like and stem-like phenotype with a stronger, dose-dependent effect on overall stemness[237]. These results are corroborated by our findings of differentially expressed genes, where the downregulation of metastatic signatures and upregulation of differentiated signatures in the Sox9-deficient clones suggests that they are less-stemlike without Sox9 (Figure 3.10 and Figure 3.11).

The effect of Sox9 deletion on proliferation and sphere-forming efficiency was stronger in the NDL3903 cells than in the NDL668 cells, which may indicate a difference in their dependence on Sox9 mediated pathways. The NDL3903 cells have a higher endogenous expression of Sox9 and therefore may be more dependent on Sox9 than the NDL668 cells.

Migration and invasion are essential steps for metastasis. To form a tumour at a secondary site, cells must be capable of migrating away from the primary tumour and invading into surrounding tissue before entering the vascular system. The migratory and invasive abilities of cells are dependent on a number of cellular signalling pathways related to adhesion, polarization at the leading edge, and contraction[242]. In TNBC, the reduced focal adhesions in stem-like cells compared to epithelial-like cells allowed for the more mesenchymal cells to form more protrusions which, along with microtubule dysregulation, contributed to increased migratory ability[243]. Migration mechanisms can be divided into broad categories, two of which are haptotaxis and chemotaxis. Haptotaxis is a measure of migration across a substrate while chemotaxis is a measure of migration due to a chemical gradient, and each are activated by different signalling pathways[244]. Invasion is the ability of cells to move through a tissue or substrate, and involves activation of migratory mechanisms in addition to modification of cell-matrix interactions[245].

Migratory ability in the clonal populations was not significantly affected by Sox9 deletion in either the haptotactic or chemotactic assays, although a similar trend to those apparent in the other characterization assays was observed in the haptotactic assay (Figure 3.8). Sox9-deletion significantly reduced the invasive ability of the cells in both the Matrigel-Transwell and Matrigel-droplet invasion assays (Figure 3.9). Consistent with our findings, Sox9 inhibition has been shown to reduce invasion and haptotaxis in a number of cancers, including TNBC, colorectal, and gastric cancer by reducing stemness[246–248]. Stemness has also been shown to be correlated with chemotaxis in breast cancers[249,250]. Again, these results are supported by our findings from the RNA sequencing. Matrix modification and regulation of locomotion are disproportionately altered in the Sox9-deficient clones relative to the $sgNTC^{K3}$ control cells, suggesting that Sox9 has an

effect on the expression of genes required for invasion (Figure 3.11). Together, these results suggest that Sox9 may promote invasion, and therefore overall stemness *in vitro*.

Throughout nearly all characterization assays, a trend was apparent where a more significant phenotypic change was observed between the complete knockout and control clones than between the clones with residual Sox9 expression and control clones. This suggests that Sox9 is affecting those phenotypes in a dose-dependent manner. However, a lack of dose dependent effect on primary sphere-forming efficiency was observed, suggesting that low levels of Sox9 expression may promote a progenitor-like phenotype, allowing cells to survive in low-attachment conditions and overcome restrictions of density-inhibited growth[241]. Since this trend was apparent in the haptotactic assay but not the chemotactic assay, the migration deficiency observed in the 3903sgSox9^{K3}1 cells may be overcome by the presence of a chemoattractant, which activates pathways of directed migration that differ from those activated during haptotaxis[244].

4.2.2. The effect of Sox9 overexpression on bulk cell populations

To test the effect of Sox9 overexpression on those cellular properties we stably infected one of the mouse cell lines, NDL3903, and a HER2-positive human cell line, BT474, with a Sox9-GFP retrovirus. Neither growth nor sphere-forming ability was found to be significantly affected by Sox9 overexpression (Figure 3.2). Sox9 is known to require heterodimerization and the recruitment of co-factors to transcriptionally regulate gene targets[115,150]. It is possible that the overexpression of Sox9 alone is not sufficient to further promote stemness pathways in HER2-positive tumour-derived cell lines. It is likely that the relevant cofactors would also need to be overexpressed to maintain the required stoichiometry and observe a phenotypic change. Literature suggests that Sox9 is subject to tissue-dependent alternative splicing in a number of species, including mice and humans[251]. It is possible that our testes-derived Sox9 cDNA may not be fully

functional in mammary cells. As visualized by immunoblotting, the exogenous Sox9 expressed in both lines appears to be a larger protein than the endogenous Sox9 expressed in the control cells (Figure 3.2.A). The lack of alternative splicing in the testes' isoform may render it biologically irrelevant in the context of HER2-positive breast cancer.

4.3. The requirement of Sox9 *in vivo*

4.3.1. Sox9 deficient cells do not expand *in vivo*

Orthotopic injections are used to model tumour initiation and progression by implanting cells in their tissue of origin. This model provides a more pathologically relevant environment for tumour growth compared to an ectopic injection model[252]. As such, we injected 3903sgNTC^{K3}1 or 3903sgSox9^{K3}6 cells directly into the mammary glands of mice to test the effects of Sox9 deletion on tumour onset and growth. Supporting our mammosphere observations, tumours only formed in mammary glands injected with control cells. This is in agreement with the RNA-sequencing data that suggests that Sox9 deletion results in a more differentiated phenotype, which would reduce the TIC potential of Sox9-deficient cells (Figure 3.10 and Figure 3.11). Although a substantial reduction in the tumour initiating capacity of Sox9-deficient cells was predicted, and clonal expansion should have removed the possibility of subpopulations, limiting dilution assays would need to be performed to conclusively determine the presence of a TIC subset[253]. Unexpectedly, the endpoint tumours did not strongly express Sox9 or Sox10, although we did detect a low level of Sox9 from the normal tissue (Figure 3.12).

Immunohistochemistry revealed the presence of CK8 and Neu-positive cells throughout the tumours formed in control cell-injected mammary glands, suggesting that the tumours were luminal cells and driven by Neu (Figure 3.12.D-F). Cas9 expression was distinctly nuclear in the 3903sgSox9^{K3}6 cells but not evident in the 3903sgNTC^{K3}1-injected tumours, and faintly evident

throughout the 3903sgNTC^{K3}1 cells in the lungs (Figure 3.12 and Figure 3.13). This difference may be due to our sgNTC sequence, which was designed not to bind DNA and results in Cas9 not actively being recruited to the nucleus in the control cells[254]. The difference between the Cas9 staining in the orthotopic tumours and the tail-vein tumours may be due to antibody availability. The orthotopic tumours were larger and more dense, and a higher concentration of antibody may be necessary to visualize Cas9. Recent work in our lab suggests that Sox10 is required for tumour initiation *in vivo*, and while its expression may be lost *in vitro* it returns after *in vivo* implantation[203(submitted)]. It is possible that the relatively high expression of Sox9 in these cells allows for Sox9 to compensate for the requirement of Sox10. For instance, at high levels, Sox9 may turn on Sox10 targets even if it has low affinity binding for those SoxE sites. It is also possible that Sox9 is not required for proliferation *in vivo* after tumour onset due to interactions with the tumour microenvironment that are not represented *in vitro.* Further experiments would be required to determine the temporal expression of each protein and the mechanism through which the cells downregulate SOX9.

4.3.2. Sox9 deficient cells do not form metastatic lesions

A commonly used murine model to study metastasis in breast cancer is by cell implantation via lateral tail vein injection[255]. For metastasis to occur, tumour cells must migrate away from the primary tumour and invade into surrounding tissues before intravasating, surviving in circulation, extravasating through the capillary bed at a secondary site and establishing a niche in the new tissue. Tail vein injected-cells will first pass through the capillary network in the lungs, likely allowing successful extravasation and growth[256]. Lung tumours formed in this model are the result of cells that have successfully extravasated and formed a niche, and its key advantages are consistency and timing. As cells are injected into the tail vein, lung tumours can form much quicker

due to bypassing the requirements of primary tumorigenesis and intravasation and there is no possibility of mice reaching humane endpoint due to primary tumour burden without metastasis. Although tail vein injection does not recapitulate the metastatic process, it is a good measure of the invasive potential and TIC content of cell lines[255].

Here, we used tail vein injections as a model of lung colonization to gain insight into the role of Sox9 in metastasis. Flow cytometry analysis showed that lungs from mice injected with 3903sgSox9^{K3}6 cells contained half as many tumour cells as those injected with 3903sgNTC^{K3}1 cells, and this finding was corroborated with hematoxylin and eosin staining (Figure 3.13). Staining of lung sections with Cas9 revealed the presence of few knockout cells in the lungs. These results suggest that while Sox9 may not be required for extravasation, it is necessary for the expansion of tumour cells at secondary sites. Further work with a more robust and complete model of metastasis is required to fully elucidate the role of Sox9 in the metastatic cascade in HER2-positive breast cancer.

4.4. Future Directions

The role of Sox9 in cellular processes is complex and is likely to involve the control of multiple pathways regulating stemness and invasion. Here, we have begun to investigate the role of Sox9 in HER2-positive breast cancer and have uncovered many paths forward. To strengthen the findings of this study, *in vitro* experiments should be repeated with additional HER2-positive cell lines to ensure that the data is not specific to murine NDL cells.

To better understand the temporal expression of Sox9 in tumour onset and progression *in vivo,* tumours from orthotopic injections should be collected at various timepoints from pre-onset to endpoint for histological analysis of protein expression and RNA expression analysis. In depth analysis of Sox9 and related target expression over the course of tumour progression may reveal

insights into the underlying mechanisms of Sox9's role in tumorigenesis. A particularly interesting experiment would be to assess the compensatory ability of Sox9 and Sox10 in tumour initiation. Expressing Sox10 in Sox9-deficient cells, and expressing Sox9 in Sox10-deficient cells prior to injection may determine if the initiation deficiency caused by the lack of one transcription factor could be rescued by the expression of the other.

Further insights into the effect of Sox9 on tumour initiation, progression, and metastasis could be obtained using transgenic models. Sox9 is essential for embryonic survival and remains necessary to maintain homeostasis in a number of cell types[131]. It is also critical for mammary gland development[185]. Therefore, a knockout model must be both conditional and inducible. The PyVMT-IRES-Cre model co-expresses the PyVMT oncogene and Cre recombinase through an IRES, under the control of a doxycycline-inducible Tet-On system[257]. Doxycycline can be used to induce conditional oncogenic expression and Sox9 excision simultaneously with a conditional allele[257]. Differences in tumour initiation, progression, and metastasis between the Sox9-expressing and Sox9-knockout mice can be compared and the dose dependent effect of Sox9 can be assessed in mice heterozygous for the Sox9-flox allele[126].

Another question outside the scope of this study is how Sox9 affects the process of EMT. Due to its transitional nature, EMT is a difficult process to study, and there are no robust functional assays. A potential experimental model of EMT *in vitro* would be to treat cells in culture with an EMT effector and monitor the protein and RNA expression of downstream targets in Sox9-expressing and Sox9-deficient cells.

4.5. Conclusion

EMT is a key mechanism contributing to the high rates of acquired chemo-resistance in HER2-positive breast cancer. Sox9 has been linked to cell stemness through a number of pathways in

malignancies where it has been shown to promote proliferation, migration, and invasion, but its role has not been fully characterized in HER2-positive breast cancer. Here, we have shown in HER2/Neu-positive tumour cell lines that Sox9 deficiency significantly inhibits proliferation, self-renewal, and invasion in a dose-dependent manner but that the overexpression of Sox9 alone is insufficient to induce stem-like phenotypic changes in murine and human cell lines. In support of a role for Sox9 in cancer stem cell expansion, we have observed that Sox9 deletion suppresses EMT-related pathways and inhibits the ability of cells to expand *in vivo*. As Sox9 is also frequently expressed in TNBC, further work into investigating the role of Sox9 in EMT and acquired resistance may reveal novel candidates for the development of therapeutics targeting several breast cancer subtypes, including HER2-positive breast cancers.

5. References

1. Pellacani, D., Tan, S., Lefort, S. & Eaves, C. J. Transcriptional regulation of normal human mammary cell heterogeneity and its perturbation in breast cancer. *EMBO J* **38**, e100330 (2019).
2. Propper, A. Y. Wandering epithelial cells in the rabbit embryo milk line: A preliminary scanning electron microscope study. *Developmental Biology* **67**, 225–231 (1978).
3. Kimata, K., Sakakura, T., Inaguma, Y., Kato, M. & Nishizuka, Y. Participation of two different mesenchymes in the developing mouse mammary gland: synthesis of basement membrane components by fat pad precursor cells. *Development* **89**, 243–257 (1985).
4. Sakakura, T., Nishizuka, Y. & Dawe, C. J. Mesenchyme-Dependent Morphogenesis and Epithelium-Specific Cytodifferentiation in Mouse Mammary Gland. *Science* **194**, 1439–1441 (1976).
5. Hovey, R. C., Trott, J. F. & Vonderhaar, B. K. Establishing a Framework for the Functional Mammary Gland: From Endocrinology to Morphology. *J Mammary Gland Biol Neoplasia* **7**, 17–38 (2002).
6. Williams, J. M. & Daniel, C. W. Mammary ductal elongation: Differentiation of myoepithelium and basal lamina during branching morphogenesis. *Developmental Biology* **97**, 274–290 (1983).
7. Humphreys, R. C. Programmed Cell Death in the Terminal Endbud. *J Mammary Gland Biol Neoplasia* **4**, 213–220 (1999).
8. Sreekumar, A. *et al.* WNT-Mediated Regulation of FOXO1 Constitutes a Critical Axis Maintaining Pubertal Mammary Stem Cell Homeostasis. *Developmental Cell* **43**, 436-448.e6 (2017).
9. Wang, C., Christin, J. R., Oktay, M. H. & Guo, W. Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages. *Cell Reports* **18**, 2825–2835 (2017).
10. Hennigar, S. R., Seo, Y. A., Sharma, S., Soybel, D. I. & Kelleher, S. L. ZnT2 is a critical mediator of lysosomal-mediated cell death during early mammary gland involution. *Sci Rep* **5**, 8033 (2015).
11. Fu, N. Y., Nolan, E., Lindeman, G. J. & Visvader, J. E. Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiological Reviews* **100**, 489–523 (2020).
12. Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229–263 (2024).
13. Brenner, D. R. *et al.* Projected estimates of cancer in Canada in 2024. *CMAJ* **196**, E615–E623 (2024).
14. Goldhirsch, A. *et al.* Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* **22**, 1736–1747 (2011).

15. Schettini, F. *et al.* Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. *NPJ Breast Cancer* **7**, 1 (2021).

16. Modi, S. *et al.* Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. *N Engl J Med* **387**, 9–20 (2022).

17. Tarantino, P., Curigliano, G. & Tolaney, S. M. Navigating the HER2-Low Paradigm in Breast Oncology: New Standards, Future Horizons. *Cancer Discovery* **12**, 2026–2030 (2022).

18. Zhang, X. Molecular Classification of Breast Cancer: Relevance and Challenges. *Archives of Pathology & Laboratory Medicine* **147**, 46–51 (2022).

19. Kast, K. *et al.* Impact of breast cancer subtypes and patterns of metastasis on outcome. *Breast Cancer Research and Treatment 2015 150:3* **150**, 621–629 (2015).

20. Waks, A. G. & Winer, E. P. Breast Cancer Treatment: A Review. *JAMA* **321**, 288–300 (2019).

21. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *The Lancet* **365**, 1687–1717 (2005).

22. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *The Lancet* **378**, 771–784 (2011).

23. Slamon, D. J. *et al.* Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene. *Science* **235**, 177–182 (1987).

24. Schalper, K. A., Kumar, S., Hui, P., Rimm, D. L. & Gershkovich, P. A Retrospective Population-Based Comparison of HER2 Immunohistochemistry and Fluorescence In Situ Hybridization in Breast Carcinomas: Impact of 2007 American Society of Clinical Oncology/ College of American Pathologists Criteria. *Archives of Pathology & Laboratory Medicine* **138**, 213–219 (2013).

25. Gómez, H. L. *et al.* Prognostic effect of hormone receptor status in early HER2 positive breast cancer patients. *Hematology/Oncology and Stem Cell Therapy* **3**, 109 (2010).

26. Gajria, D. & Chandarlapaty, S. HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Rev Anticancer Ther* **11**, 263–275 (2011).

27. Swain, S. M., Shastry, M. & Hamilton, E. Targeting HER2-positive breast cancer: advances and future directions. *Nat Rev Drug Discov* **22**, 101–126 (2023).

28. Slamon, D. J. *et al.* Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *https://doi.org/10.1056/NEJM200103153441101* **344**, 783–792 (2001).

29. Won, K.-A. & Spruck, C. Triple-negative breast cancer therapy: Current and future perspectives (Review). *Int J Oncol* **57**, 1245–1261 (2020).

30. Molyneux, G. *et al.* BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. *Cell Stem Cell* **7**, 403–417 (2010).

31. Miller, T. W. *et al.* Hyperactivation of phosphatidylinositol-3 kinase promotes escape from hormone dependence in estrogen receptor-positive human breast cancer. *J Clin Invest* **120**, 2406–2413 (2010).

32. Massarweh, S. *et al.* Tamoxifen Resistance in Breast Tumors Is Driven by Growth Factor Receptor Signaling with Repression of Classic Estrogen Receptor Genomic Function. *Cancer Research* **68**, 826–833 (2008).

33. Bardia, A. *et al.* Trastuzumab Deruxtecan after Endocrine Therapy in Metastatic Breast Cancer. *New England Journal of Medicine* **391**, 2110–2122 (2024).

34. Tarantino, P. *et al.* HER2-Low Breast Cancer: Pathological and Clinical Landscape. *JCO* **38**, 1951–1962 (2020).

35. Cohen, S., Ushiro, H., Stoscheck, C. & Chinkers, M. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. *J Biol Chem* **257**, 1523–1531 (1982).

36. Ullrich, A. *et al.* Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309**, 418–425 (1984).

37. Sierke, S. L., Cheng, K., Kim, H. H. & Koland, J. G. Biochemical characterization of the protein tyrosine kinase homology domain of the ErbB3 (HER3) receptor protein. *Biochem J* **322**, 757–763 (1997).

38. Erickson, S. L. *et al.* ErbB3 is required for normal cerebellar and cardiac development: a comparison with ErbB2- and heregulin-deficient mice. *Development* **124**, 4999–5011 (1997).

39. Gassmann, M. *et al.* Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature* **378**, 390–394 (1995).

40. Roskoski, R. Small molecule inhibitors targeting the EGFR/ErbB family of protein-tyrosine kinases in human cancers. *Pharmacological Research* **139**, 395–411 (2019).

41. Cohen, S. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *J Biol Chem* **237**, 1555–1562 (1962).

42. Shoyab, M., McDonald, V. L., Bradley, J. G. & Todaro, G. J. Amphiregulin: a bifunctional growth-modulating glycoprotein produced by the phorbol 12-myristate 13-acetate-treated human breast adenocarcinoma cell line MCF-7. *Proceedings of the National Academy of Sciences* **85**, 6528–6532 (1988).

43. Strachan, L. *et al.* Cloning and biological activity of epigen, a novel member of the epidermal growth factor superfamily. *J Biol Chem* **276**, 18265–18271 (2001).

44. de Larco, J. E. & Todaro, G. J. Growth factors from murine sarcoma virus-transformed cells. *Proceedings of the National Academy of Sciences* **75**, 4001–4005 (1978).

45. Shing, Y. *et al.* Betacellulin: a mitogen from pancreatic beta cell tumors. *Science* **259**, 1604–1607 (1993).

46. Sasada, R. *et al.* Cloning and Expression of cDNA Encoding Human Betacellulin, a New Member of the EGF Family. *Biochemical and Biophysical Research Communications* **190**, 1173–1179 (1993).

47. Higashiyama, S., Abraham, J. A., Miller, J., Fiddes, J. C. & Klagsbrun, M. A heparin-binding growth factor secreted by macrophage-like cells that is related to EGF. *Science* **251**, 936–939 (1991).

48. Toyoda, H. *et al.* Epiregulin. A novel epidermal growth factor with mitogenic activity for rat primary hepatocytes. *J Biol Chem* **270**, 7495–7500 (1995).

49. Wen, D. *et al.* Neu differentiation factor: A transmembrane glycoprotein containing an EGF domain and an immunoglobulin homology unit. *Cell* **69**, 559–572 (1992).

50. Higashiyama, S. *et al.* A Novel Brain-Derived Member of the Epidermal Growth Factor Family That Interacts with ErbB3 and ErbB41. *The Journal of Biochemistry* **122**, 675–680 (1997).

51. Busfield, S. J. *et al.* Characterization of a Neuregulin-Related Gene, Don-1, That Is Highly Expressed in Restricted Regions of the Cerebellum and Hippocampus. *Molecular and Cellular Biology* **17**, 4007–4014 (1997).

52. Zhang, D. *et al.* Neuregulin-3 (NRG3): A novel neural tissue-enriched protein that binds and activates ErbB4. *Proceedings of the National Academy of Sciences* **94**, 9562–9567 (1997).

53. Harari, D. *et al.* Neuregulin-4: a novel growth factor that acts through the ErbB-4 receptor tyrosine kinase. *Oncogene* **18**, 2681–2689 (1999).

54. Tzahar, E. *et al.* A Hierarchical Network of Interreceptor Interactions Determines Signal Transduction by Neu Differentiation Factor/Neuregulin and Epidermal Growth Factor. *Molecular and Cellular Biology* **16**, 5276–5287 (1996).

55. Freed, D. M. *et al.* EGFR ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell* **171**, 683-695.e18 (2017).

56. Dankort, D. L., Wang, Z., Blackmore, V., Moran, M. F. & Muller, W. J. Distinct tyrosine autophosphorylation sites negatively and positively modulate neu-mediated transformation. *Mol Cell Biol* **17**, 5410–5425 (1997).

57. Shi, F., Telesco, S. E., Liu, Y., Radhakrishnan, R. & Lemmon, M. A. ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proceedings of the National Academy of Sciences* **107**, 7692–7697 (2010).

58. Ghosh, R. *et al.* Trastuzumab Has Preferential Activity against Breast Cancers Driven by HER2 Homodimers. *Cancer Research* **71**, 1871–1882 (2011).

59. Jackson-Fisher, A. J. *et al.* ErbB3 is required for ductal morphogenesis in the mouse mammary gland. *Breast Cancer Res* **10**, R96 (2008).

60. Andrechek, E. R., White, D. & Muller, W. J. Targeted disruption of ErbB2/Neu in the mammary epithelium results in impaired ductal outgrowth. *Oncogene* **24**, 932–937 (2005).

61. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).

62. Venter, D., Kumar, S., Tuzi, N. & Gullick, W. OVEREXPRESSION OF THE c-erbB-2 ONCOPROTEIN IN HUMAN BREAST CARCINOMAS: IMMUNOHISTOLOGICAL ASSESSMENT CORRELATES WITH GENE AMPLIFICATION. *The Lancet* **330**, 69–72 (1987).

63. Kallioniemi, O. P. *et al.* ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* **89**, 5321–5325 (1992).

64. Harari, D. & Yarden, Y. Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. *Oncogene 2000 19:53* **19**, 6102–6114 (2000).

65. Haslekås, C. *et al.* The Inhibitory Effect of ErbB2 on Epidermal Growth Factor-induced Formation of Clathrin-coated Pits Correlates with Retention of Epidermal Growth Factor Receptor-ErbB2 Oligomeric Complexes at the Plasma Membrane. *MBoC* **16**, 5832–5842 (2005).

66. Niehans, G. A., Singleton, T. P., Dykoski, D. & Kiang, D. T. Stability of HER-2/neu Expression Over Time and at Multiple Metastatic Sites. *JNCI: Journal of the National Cancer Institute* **85**, 1230–1235 (1993).

67. Carlsson, J. *et al.* HER2 expression in breast cancer primary tumours and corresponding metastases. Original data and literature review. *Br J Cancer* **90**, 2344–2348 (2004).

68. Zabrecky, J. R., Lam, T., McKenzie, S. J. & Carney, W. The extracellular domain of p185/neu is released from the surface of human breast carcinoma cells, SK-BR-3. *J Biol Chem* **266**, 1716–1720 (1991).

69. Liu, P. C. C. *et al.* Identification of ADAM10 as a major source of HER2 ectodomain sheddase activity in HER2 overexpressing breast cancer cells. *Cancer Biology & Therapy* https://doi.org/10.4161/cbt.5.6.2708 (2006) doi:10.4161/cbt.5.6.2708.

70. Anido, J. *et al.* Biosynthesis of tumorigenic HER2 C-terminal fragments by alternative initiation of translation. *EMBO J* **25**, 3234–3244 (2006).

71. Maria, A. M., El-Shebiney, M., El-Saka, A. M. & Zamzam, Y. Expression of truncated HER2 and its prognostic value in HER2-positive breast cancer patients. *J Egypt Natl Canc Inst* **30**, 49–55 (2018).

72. Christianson, T. A. *et al.* NH2-terminally Truncated HER-2/neu Protein: Relationship with Shedding of the Extracellular Domain and with Prognostic Factors in Breast Cancer1. *Cancer Research* **58**, 5123–5129 (1998).

73. Segatto, O., King, C. R., Pierce, J. H., Di Fiore, P. P. & Aaronson, S. A. Different structural alterations upregulate in vitro tyrosine kinase activity and transforming potency of the erbB-2 gene. *Mol Cell Biol* **8**, 5570–5574 (1988).

74. Robichaux, J. P. *et al.* Pan-Cancer Landscape and Analysis of ERBB2 Mutations Identifies Poziotinib as a Clinically Active Inhibitor and Enhancer of T-DM1 Activity. *Cancer Cell* **36**, 444-457.e7 (2019).

75. Wang, S. E. *et al.* HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. *Cancer Cell* **10**, 25–38 (2006).

76. Galogre, M., Rodin, D., Pyatnitskiy, M., Mackelprang, M. & Koman, I. A review of *HER2* overexpression and somatic mutations in cancers. *Critical Reviews in Oncology/Hematology* **186**, 103997 (2023).

77. Shih, C., Padhy, L. C., Murray, M. & Weinberg, R. A. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature* **290**, 261–264 (1981).

78. Schechter, A. L. *et al.* The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature* **312**, 513–516 (1984).

79. Bargmann, C. I., Hung, M. C. & Weinberg, R. A. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. *Cell* **45**, 649–657 (1986).

80. Weiner, D. B., Liu, J., Cohen, J. A., Williams, W. V. & Greene, M. I. A point mutation in the neu oncogene mimics ligand induction of receptor aggregation. *Nature* **339**, 230–231 (1989).

81. Siegel, P. M., Dankort ,David L., Hardy ,William R. & and Muller, W. J. Novel Activating Mutations in the neu Proto-oncogene Involved in Induction of Mammary Tumors. *Molecular and Cellular Biology* **14**, 7068–7077 (1994).

82. Ignatoski, K. M. W., Grewal, N. K., Markwart, S., Livant, D. L. & Ethier, S. P. p38MAPK Induces Cell Surface α4 Integrin Downregulation to Facilitate erbB-2-Mediated Invasion. *Neoplasia* **5**, 128–134 (2003).

83. Muthuswamy, S. K., Gilman ,Michael & and Brugge, J. S. Controlled Dimerization of ErbB Receptors Provides Evidence for Differential Signaling by Homo- and Heterodimers. *Molecular and Cellular Biology* **19**, 6845–6857 (1999).

84. Benz, C. C. *et al.* Estrogen-dependent, tamoxifen-resistant tumorigenic growth of MCF-7 cells transfected with HER2/neu. *Breast Cancer Res Tr* **24**, 85–95 (1992).

85. Chazin, V. R., Kaleko, M., Miller, A. D. & Slamon, D. J. Transformation mediated by the human HER-2 gene independent of the epidermal growth factor receptor. *Oncogene* **7**, 1859–1866 (1992).

86. Hudziak, R. M., Schlessinger, J. & Ullrich, A. Increased expression of the putative growth factor receptor p185HER2 causes transformation and tumorigenesis of NIH 3T3 cells. *Proc Natl Acad Sci U S A* **84**, 7159–7163 (1987).

87. Hudziak, R. M. *et al.* p185HER2 monoclonal antibody has antiproliferative effects in vitro and sensitizes human breast tumor cells to tumor necrosis factor. *Mol Cell Biol* **9**, 1165–1172 (1989).

88. Carter, P. *et al.* Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proceedings of the National Academy of Sciences* **89**, 4285–4289 (1992).

89. Hao, Y., Yu, X., Bai, Y., McBride, H. J. & Huang, X. Cryo-EM Structure of HER2-trastuzumab-pertuzumab complex. *PLOS ONE* **14**, e0216095 (2019).

90. Wehrman, T. S. *et al.* A system for quantifying dynamic protein interactions defines a role for Herceptin in modulating ErbB2 interactions. *Proceedings of the National Academy of Sciences* **103**, 19063–19068 (2006).

91. Scaltriti, M. *et al.* Expression of p95HER2, a truncated form of the HER2 receptor, and response to anti-HER2 therapies in breast cancer. *J Natl Cancer Inst* **99**, 628–638 (2007).

92. Scheuer, W. *et al.* Strongly Enhanced Antitumor Activity of Trastuzumab and Pertuzumab Combination Treatment on HER2-Positive Human Xenograft Tumor Models. *Cancer Research* **69**, 9330–9336 (2009).

93. Minckwitz, G. von *et al.* Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *New England Journal of Medicine* **377**, 122–131 (2017).

94. Swain, S. M. *et al.* Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer. *New England Journal of Medicine* **372**, 724–734 (2015).

95. Krop, I. E. *et al.* Trastuzumab emtansine versus treatment of physician's choice in patients with previously treated HER2-positive metastatic breast cancer (TH3RESA): final overall survival results from a randomised open-label phase 3 trial. *Lancet Oncol* **18**, 743–754 (2017).

96. Ursini-Siegel, J., Schade, B., Cardiff, R. D. & Muller, W. J. Insights from transgenic mouse models of ERBB2-induced breast cancer. *Nat Rev Cancer* **7**, 389–397 (2007).

97. Bouchard, L., Lamarre, L., Tremblay, P. J. & Jolicoeur, P. Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-*neu* oncogene. *Cell* **57**, 931–936 (1989).

98. Muller, W. J., Sinn, E., Pattengale, P. K., Wallace, R. & Leder, P. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. *Cell* **54**, 105–115 (1988).

99. Guy, C. T. *et al.* Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proceedings of the National Academy of Sciences* **89**, 10578–10582 (1992).

100. Siegel, P. M., Ryan, E. D., Cardiff, R. D. & Muller, W. J. Elevated expression of activated forms of Neu/ErbB-2 and ErbB-3 are involved in the induction of mammary tumors in transgenic mice: implications for human breast cancer. *EMBO J* **18**, 2149–2164 (1999).

101. Ursini-Siegel, J. *et al.* ShcA signalling is essential for tumour progression in mouse models of human breast cancer. *EMBO J* **27**, 910–920 (2008).

102. Orban, P. C., Chui, D. & Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *Proc Natl Acad Sci U S A* **89**, 6861–6865 (1992).

103. Davie, S. A. *et al.* Effects of FVB/NJ and C57Bl/6J strain backgrounds on mammary tumor phenotype in inducible nitric oxide synthase deficient mice. *Transgenic Res* **16**, 193–201 (2007).

104. Rowse, G. J., Ritland, S. R. & Gendler, S. J. Genetic Modulation of neu Proto-oncogene-induced Mammary Tumorigenesis1. *Cancer Research* **58**, 2675–2679 (1998).

105. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117–121 (1991).

106. Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240–244 (1990).

107. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C. & Norman, D. G. Solution structure of a DNA-binding domain from HMG1. *Nucleic Acids Research* **21**, 3427–3436 (1993).

108. Stott, K., Watson, M., Howe, F. S., Grossmann, J. G. & Thomas, J. O. Tail-Mediated Collapse of HMGB1 Is Dynamic and Occurs *via* Differential Binding of the Acidic Tail to the A and B Domains. *Journal of Molecular Biology* **403**, 706–722 (2010).

109. Harley, V. R., Lovell-Badge, R. & Goodfellow, P. N. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res* **22**, 1500–1501 (1994).

110. Mertin, S., McDowall, S. G. & Harley, V. R. The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Res* **27**, 1359–1364 (1999).

111. Wright, E. M., Snopek, B. & Koopman, P. Seven new members of the Sox gene family expressed during mouse development. *Nucleic Acids Res* **21**, 744 (1993).

112. Haseeb, A. & Lefebvre, V. The SOXE transcription factors—SOX8, SOX9 and SOX10—share a bi-partite transactivation mechanism. *Nucleic Acids Research* **47**, 6917–6931 (2019).

113. Gasca, S. *et al.* A nuclear export signal within the high mobility group domain regulates the nucleocytoplasmic translocation of SOX9 during sexual determination. *Proceedings of the National Academy of Sciences* **99**, 11199–11204 (2002).

114. Südbeck, P. & Scherer, G. Two Independent Nuclear Localization Signals Are Present in the DNA-binding High-mobility Group Domains of SRY and SOX9. *Journal of Biological Chemistry* **272**, 27848–27852 (1997).

115. Bernard, P. *et al.* Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. *Human Molecular Genetics* **12**, 1755–1765 (2003).

116. Sock, E., Schmidt ,Katy, Hermanns-Borgmeyer ,Irm, Bösl ,Michael R. & and Wegner, M. Idiopathic Weight Reduction in Mice Deficient in the High-Mobility-Group Transcription Factor Sox8. *Molecular and Cellular Biology* **21**, 6951–6959 (2001).

117. Pfeifer, D., Poulat, F., Holinski-Feder, E., Kooy, F. & Scherer, G. The *SOX8* Gene Is Located within 700 kb of the Tip of Chromosome 16p and Is Deleted in a Patient with ATR-16 Syndrome. *Genomics* **63**, 108–116 (2000).

118. Portnoi, M.-F. *et al.* Mutations involving the SRY-related gene SOX8 are associated with a spectrum of human reproductive anomalies. *Hum Mol Genet* **27**, 1228–1240 (2018).

119. Kennedy, C. L., Koopman, P., Mishina, Y. & O'bryan, M. K. Sox8 and Sertoli-cell Function. *Annals of the New York Academy of Sciences* **1120**, 104–113 (2007).

120. Cheng, Y.-C., Lee, C.-J., Badge, R. M., Orme, A. T. & Scotting, P. J. *Sox8* gene expression identifies immature glial cells in developing cerebellum and cerebellar tumours. *Molecular Brain Research* **92**, 193–200 (2001).

121. Dong, P., Yu, B., Pan, L., Tian, X. & Liu, F. Identification of Key Genes and Pathways in Triple-Negative Breast Cancer by Integrated Bioinformatics Analysis. *BioMed Research International* **2018**, 2760918 (2018).

122. Schlierf, B. *et al.* Expression of SoxE and SoxD genes in human gliomas. *Neuropathology and Applied Neurobiology* **33**, 621–630 (2007).

123. Li, C. *et al.* SOX8 promotes tumor growth and metastasis through FZD6-dependent Wnt/β-catenin signaling in colorectal carcinoma. *Heliyon* **9**, e22586 (2023).

124. Tian, W. *et al.* The oncogenic role of SOX8 in endometrial carcinoma. *Cancer Biol Ther* **21**, 1136–1144 (2020).

125. Wagner, T. *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *SOX9*. *Cell* **79**, 1111–1120 (1994).

126. Bi, W. *et al.* Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proceedings of the National Academy of Sciences* **98**, 6698–6703 (2001).

127. Wright, E. *et al.* The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nat Genet* **9**, 15–20 (1995).

128. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & De Crombrugghe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes & development* **16**, 2813–2828 (2002).

129. Chaboissier, M.-C. *et al.* Functional analysis of Sox8 and Sox9 during sex determination in the mouse. *Development* **131**, 1891–1901 (2004).

130. Scott, C. E. *et al.* SOX9 induces and maintains neural stem cells. *Nat Neurosci* **13**, 1181–1189 (2010).

131. Barrionuevo, F. *et al.* Homozygous Inactivation of Sox9 Causes Complete XY Sex Reversal in Mice1. **74**, 195–201 (2006).

132. Barrionuevo, F. *et al.* Testis cord differentiation after the sex determination stage is independent of *Sox9* but fails in the combined absence of *Sox9* and *Sox8*. *Developmental Biology* **327**, 301–312 (2009).

133. Darido, C. *et al.* Defective Claudin-7 Regulation by Tcf-4 and Sox-9 Disrupts the Polarity and Increases the Tumorigenicity of Colorectal Cancer Cells. *Cancer Research* **68**, 4258–4268 (2008).

134. Hu, B., Wang, J. & Jin, X. MicroRNA-138 suppresses cell proliferation and invasion of renal cell carcinoma by directly targeting SOX9. *Oncology Letters* **14**, 7583–7588 (2017).

135. Liu, S., Dong, H., Dai, H., Liu, D. & Wang, Z. MicroRNA-216b regulated proliferation and invasion of non-small cell lung cancer by targeting SOX9. *Oncology Letters* **15**, 10077–10083 (2018).

136. Ma, Y. *et al.* SOX9 is essential for triple-negative breast cancer cell survival and metastasis. *Molecular Cancer Research* **18**, 1825–1838 (2020).

137. Wang, L. *et al.* Oncogenic role of SOX9 expression in human malignant glioma. *Med Oncol* **29**, 3484–3490 (2012).

138. Jeselsohn, R. *et al.* Embryonic transcription factor SOX9 drives breast cancer endocrine resistance. *Proc Natl Acad Sci U S A* **114**, E4482–E4491 (2017).

139. Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I. & Wegner, M. Sox10, a Novel Transcriptional Modulator in Glial Cells. *J. Neurosci.* **18**, 237–250 (1998).

140. Pusch, C. *et al.* The SOX10/Sox10 gene from human and mouse: sequence, expression, and transactivation by the encoded HMG domain transcription factor. *Hum Genet* **103**, 115–123 (1998).

141. Kapur, R. P. Early death of neural crest cells is responsible for total enteric aganglionosis in Sox10(Dom)/Sox10(Dom) mouse embryos. *Pediatr Dev Pathol* **2**, 559–569 (1999).

142. Osaka, H. *et al.* Disrupted SOX10 regulation of GJC2 transcription causes Pelizaeus-Merzbacher-like disease. *Annals of Neurology* **68**, 250–254 (2010).

143. Izumi, Y. *et al.* Hypogonadotropic hypogonadism in a female patient previously diagnosed as having waardenburg syndrome due to a sox10 mutation. *Endocrine* **49**, 553–556 (2015).

144. Jin, L. *et al.* Clinicopathological significance of Sox10 expression in triple-negative breast carcinoma. *Transl Cancer Res* **9**, 5603–5613 (2020).

145. Weisman, P., Yu, Q. & Xu, J. Clinicopathological characteristics of the SOX10+ subset of HER2+ breast cancer. *Annals of Diagnostic Pathology* **63**, 152087 (2023).

146. Yang, K., Yun, F., Shi, L., Liu, X. & Jia, Y. F. SOX10 promotes the malignant biological behavior of basal-like breast cancer cells by regulating EMT process. *Heliyon* **9**, e23162 (2023).

147. Kwok, C. *et al.* Mutations in SOX9, the Gene Responsible for Campomelic Dysplasia and Autosomal Sex Reversal. *Am J Hum Genet* **57**, 1028–1036 (1995).

148. McDowall, S. *et al.* Functional and Structural Studies of Wild Type SOX9 and Mutations Causing Campomelic Dysplasia. *Journal of Biological Chemistry* **274**, 24023–24030 (1999).

149. Huang, Y. H., Jankowski, A., Cheah, K. S. E., Prabhakar, S. & Jauch, R. SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. *Scientific Reports 2015 5:1* **5**, 1–12 (2015).

150. Südbeck, P., Schmitz, M. L., Baeuerle, P. A. & Scherer, G. Sex reversal by loss of the C–terminal transactivation domain of human SOX9. *Nat Genet* **13**, 230–232 (1996).

151. Akiyama, H. *et al.* Interactions between Sox9 and β-catenin control chondrocyte differentiation. *Genes Dev* **18**, 1072–1087 (2004).

152. Deng, W. *et al.* SOX9 inhibits β-TrCP-mediated protein degradation to promote nuclear GLI1 expression and cancer stem cell properties. *J Cell Sci* **128**, 1123–1138 (2015).

153. Gordon, C. T. *et al.* Long-range regulation at the SOX9 locus in development and disease. *Journal of Medical Genetics* **46**, 649–656 (2009).

154. Mochizuki, Y. *et al.* Combinatorial CRISPR/Cas9 Approach to Elucidate a Far-Upstream Enhancer Complex for Tissue-Specific Sox9 Expression. *Developmental Cell* **46**, 794-806.e6 (2018).

155. Sekido, R. & Lovell-Badge, R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930–934 (2008).

156. Yao, B. *et al.* The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. *Nucleic Acids Research* **43**, 5394–5408 (2015).

157. Mead, T. J. *et al.* A far-upstream (−70 kb) enhancer mediates Sox9 auto-regulation in somatic tissues during development and adult regeneration. *Nucleic Acids Research* **41**, 4459–4469 (2013).

158. Luo, Y. *et al.* Regional methylome profiling reveals dynamic epigenetic heterogeneity and convergent hypomethylation of stem cell quiescence-associated genes in breast cancer following neoadjuvant chemotherapy. *Cell Biosci* **9**, 16 (2019).

159. Huang, W., Zhou ,Xin, Lefebvre ,Véronique & and de Crombrugghe, B. Phosphorylation of SOX9 by Cyclic AMP-Dependent Protein Kinase A Enhances SOX9's Ability To Transactivate aCol2a1 Chondrocyte-Specific Enhancer. *Molecular and Cellular Biology* **20**, 4149–4158 (2000).

160. Liu, J. A. J. *et al.* Phosphorylation of Sox9 is required for neural crest delamination and is regulated downstream of BMP and canonical Wnt signaling. *Proceedings of the National Academy of Sciences* **110**, 2882–2887 (2013).

161. Coricor, G. & Serra, R. TGF-β regulates phosphorylation and stabilization of Sox9 protein in chondrocytes through p38 and Smad dependent mechanisms. *Sci Rep* **6**, 38616 (2016).

162. Oh, H. J., Kido, T. & Lau, Y.-F. C. PIAS1 interacts with and represses SOX9 transactivation activity. *Molecular Reproduction and Development* **74**, 1446–1455 (2007).

163. Zuo, C. *et al.* SHP2 regulates skeletal cell fate by modifying SOX9 expression and transcriptional activity. *Bone Res* **6**, 12 (2018).

164. Hattori, T. *et al.* Interactions between PIAS Proteins and SOX9 Result in an Increase in the Cellular Concentrations of SOX9. *Journal of Biological Chemistry* **281**, 14417–14428 (2006).

165. Jana, S. *et al.* SOX9: The master regulator of cell fate in breast cancer. *Biochemical Pharmacology* **174**, 113789 (2020).

166. Stolt, C. C. *et al.* The Sox9 transcription factor determines glial fate choice in the developing spinal cord. *Genes Dev.* **17**, 1677–1689 (2003).

167. Molofsky, A. V. *et al.* Expression profiling of Aldh1l1-precursors in the developing spinal cord reveals glial lineage-specific genes and direct Sox9-Nfe2l1 interactions. *Glia* **61**, 1518–1532 (2013).

168. Sekido, R., Bar, I., Narváez, V., Penny, G. & Lovell-Badge, R. SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. *Developmental Biology* **274**, 271–279 (2004).

169. Vidal, V. P. I., Chaboissier, M.-C., de Rooij, D. G. & Schedl, A. Sox9 induces testis development in XX transgenic mice. *Nat Genet* **28**, 216–217 (2001).

170. Ottolenghi, C. *et al.* Loss of Wnt4 and Foxl2 leads to female-to-male sex reversal extending to germ cells. *Human Molecular Genetics* **16**, 2795–2804 (2007).

171. Akiyama‡, H. *et al.* Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. *Proceedings of the National Academy of Sciences* **102**, 14665–14670 (2005).

172. Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & de Crombrugghe, B. Sox9 is required for cartilage formation. *Nat Genet* **22**, 85–89 (1999).

173. Henry, S. P., Liang, S., Akdemir, K. C. & de Crombrugghe, B. The postnatal role of Sox9 in cartilage. *Journal of Bone and Mineral Research* **27**, 2511–2525 (2012).

174. Ohba, S., He, X., Hojo, H. & McMahon, A. P. Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell Reports* **12**, 229–243 (2015).

175. Chang, D. R. *et al.* Lung epithelial branching program antagonizes alveolar differentiation. *Proceedings of the National Academy of Sciences* **110**, 18042–18051 (2013).

176. Garside, V. C. *et al.* SOX9 modulates the expression of key transcription factors required for heart valve development. *Development* **142**, 4340–4350 (2015).

177. Blache, P. *et al.* SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. *Journal of Cell Biology* **166**, 37–47 (2004).

178. Poncy, A. *et al.* Transcription factors SOX4 and SOX9 cooperatively control development of bile ducts. *Developmental Biology* **404**, 136–148 (2015).

179. Seymour, P. A. *et al.* SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proceedings of the National Academy of Sciences* **104**, 1865–1870 (2007).

180. Muto, A., Iida, A., Satoh, S. & Watanabe, S. The group E Sox genes *Sox8* and *Sox9* are regulated by Notch signaling and are required for Müller glial cell development in mouse retina. *Experimental Eye Research* **89**, 549–558 (2009).

181. Kou, I. & Ikegawa, S. SOX9-dependent and -independent Transcriptional Regulation of Human Cartilage Link Protein*. *Journal of Biological Chemistry* **279**, 50942–50948 (2004).

182. Carmon, I. *et al.* HU308 Mitigates Osteoarthritis by Stimulating Sox9-Related Networks of Carbohydrate Metabolism. *J Bone Miner Res* **38**, 154–170 (2023).

183. Han, X. *et al.* Lineage Tracing Reveals the Bipotency of SOX9+ Hepatocytes during Liver Regeneration. *Stem Cell Reports* **12**, 624–638 (2019).

184. Furuyama, K. *et al.* Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nat Genet* **43**, 34–41 (2011).

185. Malhotra, G. K. *et al.* The role of Sox9 in mouse mammary gland development and maintenance of mammary stem and luminal progenitor cells. *BMC Developmental Biology* **14**, 1–11 (2014).

186. Chen, W. *et al.* Expression and correlation of MALAT1 and SOX9 in non-small cell lung cancer. *The Clinical Respiratory Journal* **12**, 2284–2291 (2018).

187. Swartling, F. J., Ferletta, M., Kastemar, M., Weiss, W. A. & Westermark, B. Cyclic GMP-dependent protein kinase II inhibits cell proliferation, Sox9 expression and Akt phosphorylation in human glioma cell lines. *Oncogene* **28**, 3121–3131 (2009).

188. Liu, C. *et al.* Sox9 regulates self-renewal and tumorigenicity by promoting symmetrical cell division of cancer stem cells in hepatocellular carcinoma. *Hepatology* **64**, 117 (2016).

189. Chen, X. *et al.* Defining a Population of Stem-like Human Prostate Cancer Cells That Can Generate and Propagate Castration-Resistant Prostate Cancer. *Clin Cancer Res* **22**, 4505–4516 (2016).

190. Eberl, M. *et al.* Hedgehog-EGFR cooperation response genes determine the oncogenic phenotype of basal cell carcinoma and tumour-initiating pancreatic cancer cells†. *EMBO Molecular Medicine* **4**, 218–233 (2012).

191. Chen, Q. *et al.* SOX9 modulates the transformation of gastric stem cells through biased symmetric cell division. *Gastroenterology* **164**, 1119-1136.e12 (2023).

192. Al-Zahrani, K. N. *et al.* AKT-mediated phosphorylation of Sox9 induces Sox10 transcription in a murine model of HER2-positive breast cancer. *Breast Cancer Research* **23**, 1–13 (2021).

193. Guo, W. *et al.* Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* **148**, 1015–1028 (2012).

194. Dravis, C. *et al.* Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells. *Cell Reports* **12**, 2035–2048 (2015).

195. Mertelmeyer, S. *et al.* The transcription factor Sox10 is an essential determinant of branching morphogenesis and involution in the mouse mammary gland. *Sci Rep* **10**, 17807 (2020).

196. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907–913 (2009).

197. Shehata, M. *et al.* Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Research* **14**, R134 (2012).

198. Santamaria, P. G., Moreno-Bueno, G., Portillo, F. & Cano, A. EMT: Present and future in clinical oncology. *Mol Oncol* **11**, 718–738 (2017).

199. Akhmetkaliyev, A., Alibrahim, N., Shafiee, D. & Tulchinsky, E. EMT/MET plasticity in cancer and Go-or-Grow decisions in quiescence: the two sides of the same coin? *Molecular Cancer 2023 22:1* **22**, 1–16 (2023).

200. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* **16**, 488–494 (2014).

201. Liu, Y. *et al.* A SOX9-B7x axis safeguards dedifferentiated tumor cells from immune surveillance to drive breast cancer progression. *Developmental Cell* **58**, 2700-2717.e12 (2023).

202. Al-Zahrani, K. N. *et al.* Loss of the Ste20-like kinase induces a basal/stem-like phenotype in HER2-positive breast cancers. *Oncogene 2020 39:23* **39**, 4592–4602 (2020).

203. Garland, B. *et al.* Loss of Sox10 in MMTV-NIC mice prevents tumor initiation and induces luminal-to-basal reprograming in Neu+ tumor cells. *Submitted*.

204. Angelozzi, M. & Lefebvre, V. SOXopathies: Growing Family of Developmental Disorders Due to SOX Mutations. *Trends in Genetics* **35**, 658–671 (2019).

205. Al-zahrani, K. The Role of the Ste20-like Kinase in Embryonic Development and Neu-induced Mammary Tumorigenesis. (2018).

206. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* **41**, e108 (2013).

207. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

208. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**, e1003118 (2013).

209. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

210. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

211. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. BioSyst.* **12**, 477–479 (2016).

212. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

213. Dolgalev, I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. (2025).

214. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

215. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *cels* **1**, 417–425 (2015).

216. Kolde, R. pheatmap: Pretty Heatmaps. (2019).

217. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* **2**, E7 (2004).

218. Isakoff, M. S. *et al.* Inactivation of the Snf5 tumor suppressor stimulates cell cycle progression and cooperates with p53 loss in oncogenic transformation. *Proc Natl Acad Sci U S A* **102**, 17745–17750 (2005).

219. Hailesellasse Sene, K. *et al.* Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* **8**, 85 (2007).

220. Scholl, C. *et al.* Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* **137**, 821–834 (2009).

221. KRAS.BREAST_UP.V1_UP. https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/KRAS.BREAST_UP.V1_UP.html.

222. Creighton, C. J. *et al.* Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors. *Cancer Res* **66**, 3903–3911 (2006).

223. Wiederschain, D. *et al.* Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. *Mol Cell Biol* **27**, 4968–4979 (2007).

224. Majumder, P. K. *et al.* mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nat Med* **10**, 594–601 (2004).

225. Vivanco, I. *et al.* Identification of the JNK signaling pathway as a functional target of the tumor suppressor PTEN. *Cancer Cell* **11**, 555–569 (2007).

226. Elkon, R. *et al.* Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis. *Genome Biol* **6**, R43 (2005).

227. Ziegler, S. *et al.* Novel target genes of the Wnt pathway and statistical insights into Wnt target promoter regulation. *FEBS J* **272**, 1600–1615 (2005).

228. Marzec, M. *et al.* Differential effects of interleukin-2 and interleukin-15 versus interleukin-21 on CD4+ cutaneous T-cell lymphoma cells. *Cancer Res* **68**, 1083–1091 (2008).

229. van Staveren, W. C. G. *et al.* Gene expression in human thyrocytes and autonomous adenomas reveals suppression of negative feedbacks in tumorigenesis. *Proc Natl Acad Sci U S A* **103**, 413–418 (2006).

230. Zhao, Q. *et al.* Identification of genes expressed with temporal-spatial restriction to developing cerebellar neuron precursors by a functional genomic approach. *Proc Natl Acad Sci U S A* **99**, 5704–5709 (2002).

231. Ma, Y., Croxton, R., Moorer, R. L. & Cress, W. D. Identification of novel E2F1-regulated genes by microarray. *Arch Biochem Biophys* **399**, 212–224 (2002).

232. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).

233. Lara, M. F. *et al.* Gene profiling approaches help to define the specific functions of retinoblastoma family in epidermis. *Mol Carcinog* **47**, 209–221 (2008).

234. Wei, G. *et al.* Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* **10**, 331–342 (2006).

235. Bailey, J., Tyson-Capper, A. J., Gilmore, K., Robson, S. C. & Europe-Finner, G. N. Identification of human myometrial target genes of the cAMP pathway: the role of cAMP-response element binding (CREB) and modulator (CREMalpha and CREMtau2alpha) proteins. *J Mol Endocrinol* **34**, 1–17 (2005).

236. Johnson, K. *et al.* A Stem Cell–Based Approach to Cartilage Repair. *Science* **336**, 717–721 (2012).

237. Shaw, F. L. *et al.* A detailed mammosphere assay protocol for the quantification of breast stem cell activity. *Journal of Mammary Gland Biology and Neoplasia* **17**, 111–117 (2012).

238. Lizárraga-Verdugo, E. *et al.* Cell Survival Is Regulated via SOX9/BCL2L1 Axis in HCT-116 Colorectal Cancer Cell Line. *J Oncol* **2020**, 5701527 (2020).

239. Liu, S., Li, S., Yu, X., Wang, Q. & Sun, H. microRNA-133b represses the progression of lung cancer through inhibiting SOX9/β-catenin signaling pathway. *Int J Clin Exp Pathol* **13**, 2270–2279 (2020).

240. Levine, E. M., Becker, Y., Boone, C. W. & Eagle, H. Contact inhibition, macromolecular synthesis, and polyribosomes in cultured human diploid fibroblasts*. *Proceedings of the National Academy of Sciences* **53**, 350–356 (1965).

241. Guberman, E., Sherief, H. & Regan, E. R. Boolean model of anchorage dependence and contact inhibition points to coordinated inhibition but semi-independent induction of proliferation and migration. *Comput Struct Biotechnol J* **18**, 2145–2165 (2020).

242. Soon, L. L. A discourse on cancer cell chemotaxis: Where to from here? *IUBMB Life* **59**, 60–67 (2007).

243. Heussner, R. K., Zhang, H., Qian, G., Baker, M. J. & Provenzano, P. P. Differential contractility regulates cancer stem cell migration. *Biophysical Journal* **122**, 1198–1210 (2023).

244. SenGupta, S., Parent, C. A. & Bear, J. E. The principles of directed cell migration. *Nat Rev Mol Cell Biol* **22**, 529–547 (2021).

245. Cheung, K. J. & Ewald, A. J. Illuminating breast cancer invasion: diverse roles for cell–cell interactions. *Current Opinion in Cell Biology* **30**, 99–111 (2014).

246. Xue, M. *et al.* MicroRNA-613 induces the sensitivity of gastric cancer cells to cisplatin through targeting SOX9 expression. *Am J Transl Res* **11**, 885–894 (2019).

247. Lamichhane, S., Mo, J.-S., Sharma, G., Joung, S.-M. & Chae, S.-C. MIR133A regulates cell proliferation, migration, and apoptosis by targeting SOX9 in human colorectal cancer cells. *Am J Cancer Res* **12**, 3223–3241 (2022).

248. Wang, Y.-F. *et al.* Downregulation of SOX9 suppresses breast cancer cell proliferation and migration by regulating apoptosis and cell cycle arrest. *Oncol Lett* **22**, 517 (2021).

249. Pang, M.-F. *et al.* TGF-β1-induced EMT promotes targeted migration of breast cancer cells through the lymphatic system by the activation of CCR7/CCL21-mediated chemotaxis. *Oncogene* **35**, 748–760 (2016).

250. Gao, J. B., Zhu, M. N. & Zhu, X. L. miRNA-215-5p suppresses the aggressiveness of breast cancer cells by targeting Sox9. *FEBS Open Bio* **9**, 1957–1967 (2019).

251. Agrawal, R., Wessely, O., Anand, A., Singh, L. & Aggarwal, R. K. Male-specific expression of Sox9 during gonad development of crocodile and mouse is mediated by alternative splicing of its proline-glutamine-alanine rich domain. *The FEBS Journal* **276**, 4184–4196 (2009).

252. Zhang, Y. *et al.* Establishment of a murine breast tumor model by subcutaneous or orthotopic implantation. *Oncol Lett* **15**, 6233–6240 (2018).

253. Kurani, H. *et al.* DOT1L Is a Novel Cancer Stem Cell Target for Triple-Negative Breast Cancer. *Clin Cancer Res* **28**, 1948–1965 (2022).

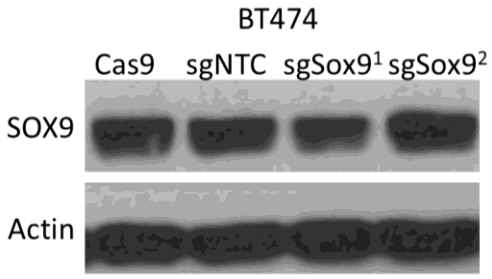
254. Zhang, W. *et al.* Hitchhiking of Cas9 with nucleus-localized proteins impairs its controllability and leads to efficient genome editing of NLS-free Cas9. *Molecular Therapy* **32**, 920–934 (2024).

255. Rashid, O. M. *et al.* Is tail vein injection a relevant breast cancer lung metastasis model? *J Thorac Dis* **5**, 385–392 (2013).

256. Khanna, C. & Hunter, K. Modeling metastasis in vivo. *Carcinogenesis* **26**, 513–523 (2005).

257. Rao, T. *et al.* Inducible and coupled expression of the polyomavirus middle T antigen and Cre recombinase in transgenic mice: an in vivo model for synthetic viability in mammary tumour progression. *Breast Cancer Research : BCR* **16**, R11 (2014).

6. Appendix



Appendix Figure 1: Sox9 knockout was not achieved in human cell lines. Parental BT474 cells were treated with Cas9 lentivirus, underwent selection blastocidin and were then treated with sgNTC, $sgSox9^1$, or $sgSox9^2$ lentivirus to stably express Cas9 and inactivate Sox9. SOX9 protein expression was visualized by immunoblotting after culturing for two passages after sorting.