PhD degree in Molecular Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Settore disciplinare: MED/04

Identification of Epigenetic Inhibitors of Physiological Cellular Plasticity as Novel Tumor Suppressors

Xieraili Aobuli

European Institute of Oncology (IEO), Milan

Matricola n. R10757

Supervisor: Prof. Pier Giuseppe Pelicci, MD, PhD

European Institute of Oncology (IEO), Milan

Added Supervisor: Dr. Maria Cristina Moroni, PhD

European Institute of Oncology (IEO), Milan

Academic year 2017-2018

Acknowledgements

I take this special opportunity to express my sincere and deepest gratitude to my PhD supervisor, Pier Giuseppe Pelicci, for giving me the opportunity to work on this project, for his support, guidance and continuous encouragement.

Next, I would like to thank the members of the group directly involved in this project: Maria Cristina Moroni and Errico D'Elia. Maria Cristina Moroni has been a supervisor who has guided day-by-day during the years. I am indebted to her for her endless support, kindness and patience, for always being available to help, and for her perseverance especially during the hard times. I am grateful to Errico D'Elia for teaching me practical skills and for his contribution in demanding experiments. He was a wonderful human being, companying to me throughout the hard times both in the lab and outside for daily life.

A special word of thanks goes to Thalia Vlachou, with whom I had a wonderful and fruitful collaboration. She has been always inspiring and giving positive energy to stay motivated throughout the years.

I would like to thank Laura Riva, a former member of our group, who did the bioinformatics analysis on screening to identify the genes, Lucilla Luzi for her work on the RNA-seq analysis, and Roman Hillje who did the single-cell RNA-seq analysis.

I am grateful to former scientific editor of our group Paola Dalton and present Stefania Averiamo for their help on scientific writing. Finally, thank all other our lab members, especially Ivan Gaetano Dellino, Mario Faretta, Paul Edward Massa, Barbara Gallo, Maria Elena Boggio, Katsiaryna Belenkaya, Rani Pallavi, Rossana Piccioni. Thanks to Simona Ronzoni in the FACS sorting facility, and all the other members of facilities in the institute.

Last but not least, I offer my best regards to my loving parents, my mum Sayipjamalhan, dad Obul Yusun, and my sister and brothers for their consistent support. Lastly, and most of all, I owe my deepest gratitude to my wife, Dilhumar, for her affection, encouragement, and extremely supportive of me throughout this entire process and has made countless sacrifices to help me got this point. My thesis acknowledgment would be incomplete without thanking to my son, Zulqar, who brings up my life to another level. In the same time, my heartfelt apologies to him for not being with him for two years due to our study, and I owe my parent's indispensable help for that period.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACK	Ammonium chloride potassium
CSC	Cancer stem cell
DEG	Differentially expressed gene
DIA	Digital image analysis
DMEM	Dulbecco's modified Eagle's medium
EIC	Intestinal epithelial cells
Elf-5	Elongation factor 1α
ELDA	Extreme limiting dilution analysis
EMT	Epithelial-mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
ESC	Embryonic stem cell
ETRR	Early transcriptional reprogramming response
EV	Empty vector
FBS, NA	Foetal bovine serum, North-American
FBS, SA	Foetal bovine serum, South-American
FDR	False discovery rate
FVB	Friend leukaemia virus B strain
GBM	Gliomablastoma multifoma
GFP	Green fluorescent protein
GO	Gene Ontology
GSC	Gelioma stem cell
GSEA	Gene set enrichment analysis
HME	Human mammary epithelial
HSC	Haematopoietic stem cell
iPSC	Induced pluripotent stem cell
K14	Cytokeratin 14
K18	Cytokeratin 18
K5	Cytokeratin 5
K8	Cytokeratin 8
KEGG	Kyoto Encyclopaedia of Genes and Genome
LUM	Luminal cell
MaSC	Mammary stem cell
MEC	Mammary epithelial cell
mEpi1	Mouse epigenetic library 1
mEpi2	Mouse epigenetic library 2
MRU	Mammary repopulating unit
MBEM	Mammary epithelial basal medium
miRNA	Micro RNA
MOI	Multiplicity of infection
MSigDB	Molecular signature database
mTert	Mouse telomerase reverse transcriptase

nCSC	Non-Cancer stem cell
NES	Normalized enrichment score
NI	Not infected
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
NMuMG	Normal murine mammary gland epithelial
NSG	NOD SCID IL2RG null
OSKM	Oct4, Sox2, Klf4, Myc
PuroR	Puromycin resistance marker
ProcR	Protein C receptor
PRC1	Polycomb repressive complex1
RFP	Red fluorescent protein
RRE	Rev response element
SC	Stem Cell
Sca-1	Stem cell antigen-1
shRNA	Short hairpin RNA
TEB	Terminal end bud
TF	Transcription factor
TSS	Transcription start site
WT	Wild type

1. Abstract

Cellular plasticity, the inter-conversion of cells between differentiated cells and stem cells (SCs), can lead to tissue regeneration and restoration of homeostasis after injury. Conversely, inappropriate induction of cellular plasticity could be involved in tumor initiation and progression, through de-novo generation of cancer stem cells (CSCs) by de-differentiation of normal or non-tumorigenic bulk tumor cells. This intrinsically dangerous potential must be tightly controlled by genetic and epigenetic mechanisms, to prevent unscheduled de-differentiation. However, their systematic identification by large-scale screenings is just beginning to be exploited.

In order to identify physiological inhibitors of cell plasticity, that could play a function as tumor suppressors, we performed short-hairpin RNA (shRNA) screens. In the screens, we used pooled lentiviral shRNA libraries targeting 234 epigenetic regulators to identify shRNAs endowing mouse mammary progenitors with the SC-specific ability to regenerate mammary gland tissue upon *in vivo* orthotopic transplantation. Sequencing of genomic DNA extracted from the regenerated mammary glands led to the identification of 38 hits/genes. We individually validated 7 hits by *in vivo* regeneration assays, showing that their down-regulation leads to conversion of mammary progenitors into stem cells able to regenerate mammary glands. We performed also *in vitro* mammosphere and phenotypic cell conversion assays to examine the hits' function in self-renewal and cell plasticity, respectively. Next, among the validated hits, we showed that the inhibition of Cbx5 and Kmt2d induced efficient reprogramming of mammary progenitors. Therefore, we focused to investigate their mechanistic function in reprogramming at the transcriptome level, by RNA sequencing. We identified a specific enrichment for pro-inflammatory signalling pathways as an early transcriptional reprogramming response (ETRR), followed by up-regulation of Myc target genes.

Finally, in order to set the basis for further characterization of reprogramming mechanisms induced by the validated shRNAs, we established an organoid assay using MycER, our positive control for mammary progenitors reprogramming, showing that MycER over-expression bestows mouse luminal cells with enhanced self-renewal ability and differentiation capacity. Moreover, we performed a single cell transcriptome analysis in mouse primary mammospheres that revealed considerable heterogeneity as 20 clusters identified and led to the identification of Cd36 (glycoprotein, collagen type I receptor) as a putative mammary SC-specific surface marker. We expect that further analyses of these data, together with single cell transcriptomic analysis of mammospheres interfered for the validated hits, will shed light on the mechanisms involved in physiological cell plasticity.

2. Introduction

2.1 Cellular Plasticity in Physiology and Cancer

During organism development and homeostasis, acquiring and maintaining a cellular identity over time is critical for the proper functioning of organs and tissues. In particular, during cell differentiation within a tissue, stem cells (SCs) residing at the apex of the hierarchy lose their plasticity and narrow down their identity into differentiated (specialized) cell types to function. For many years this differentiated state was believed to be irreversible in mammals. In recent years, seminal studies showed that cellular identity is not set in stone, but it can be reprogrammed either in experimental settings or physiological circumstances (Do and Scholer 2004, Takahashi and Yamanaka 2006, Tata, Mou et al. 2013).

Cellular plasticity refers to the ability of cells to adopt an alternative cellular identity in response to intrinsic or extrinsic stimuli. During its life span, an organism is exposed to constant environmental damages, thus requiring mechanisms to respond to injury and tissue repair in order to restore homeostasis. In adult tissues, homeostasis is maintained by undifferentiated and self-renewing resident SCs, that are able to generate all cell types of the tissue (Fig. 2-1). After tissue injury, damaged or lost cells can be replenished by mobilization of quiescent SCs and increased proliferation of surviving SCs. Alternatively, accumulating studies showed that tissue cells exhibit remarkable cellular plasticity after severe tissue injury, and that already specialized cells can revert/dedifferentiate into SCs that generate all cell types. Lost cells can also be replenished via transdifferentiation into another differentiated cell lineage. All these processes are often referred to as cellular plasticity or reprogramming (Tetteh, Farin et al. 2015, Merrell and Stanger 2016).

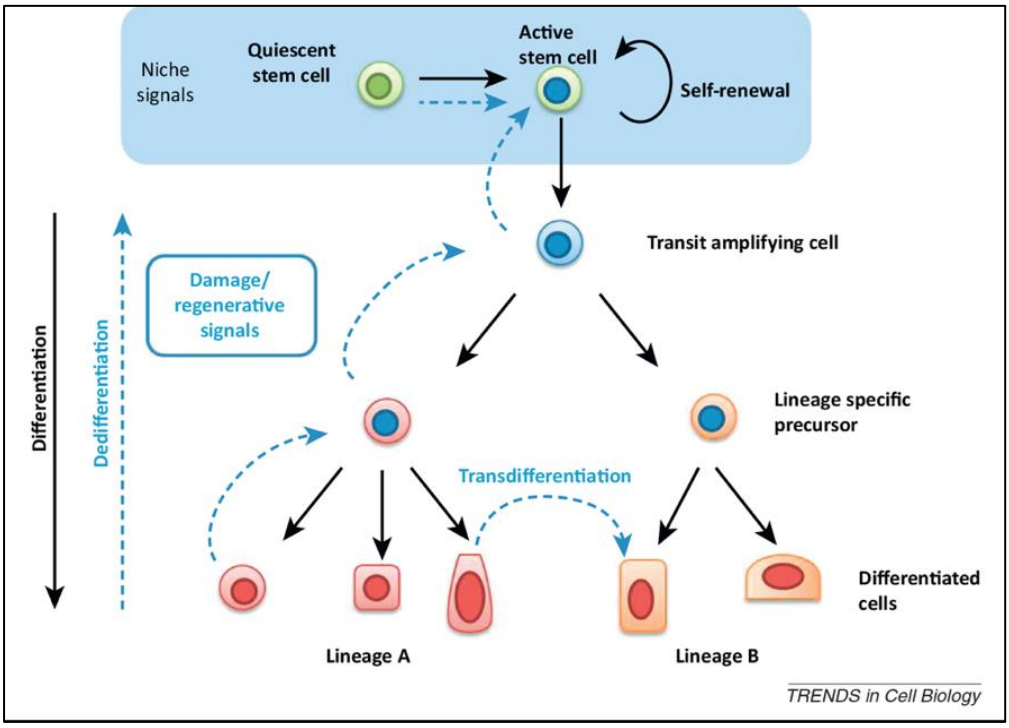


Figure 2-1. **Stem cell hierarchy in epithelial tissue homeostasis and regeneration**
(Adapted from Paul et al. 2014) © 2014 Elsevier Inc. All rights reserved

In tumors, the "classical" cancer stem cell (CSC) model suggests that, at least during the early phase of carcinogenesis, tumor tissue preserves a hierarchical organization similar to the normal tissue. The CSCs residing at the apex in the tumor cell hierarchy and as the only one which able to give rise both CSC and non-CSCs. However, a growing body of evidence in the last years suggests that considerable plasticity exists between the non-CSC and CSC compartments, and, specifically, that tumorigenic CSCs can be *de novo* generated from the bulk tumor cells, most of which are non- tumorigenic (Marjanovic, Weinberg et al. 2013).

2.1.1 Cellular plasticity in physiology

Cellular plasticity has become increasingly evident as a principal mode of injury response in many organs in mammals, in particularly in epithelial tissues where the prominent examples are well documented and a focused effort has been made to identify plasticity phenomenon (Lin, Srikanth et al. 2018). Epithelia are cellular sheets often residing at the interface between the external environment and internal body organs, including skin, gut, airway tracts, kidney, liver, mammary glands, and prostate.

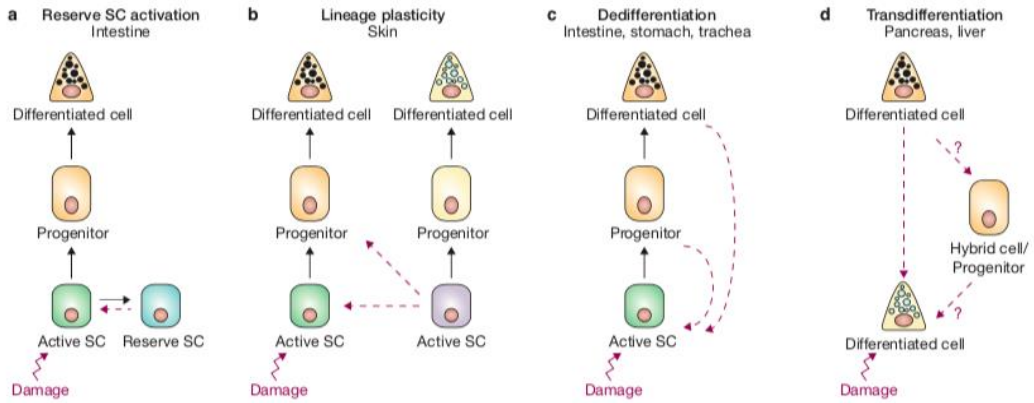


Figure 2-2. **Cell plasticity in tissue homeostasis and repair**
(Adapted from Julia et al. 2017)

Recent studies focusing on mammalian plasticity have well characterized the process thanks to use of the elegant lineage tracing studies in mouse epithelial tissues. These studies revealed that genetic ablation of SCs (diphtheria toxin receptor (DTR) gene knock-in) or physiological damage to tissues can lead lineage committed cells to revert into SCs or functional facultative SCs to restore homeostasis (Blanpain and Fuchs 2014). Cellular plasticity, in particular epithelial tissues, is present in different forms including interconversion between different stem cell pools, activation of reserve SCs, and dedifferentiation, transdifferentiation or phenotypic transition of differentiated cells within a tissue (Fig. 2-2). Conceivably, all these forms of plasticity are strictly controlled to maintain physiological organ homeostasis and ensure proper healing (Varga and Greten 2017).

Reserve stem cell activation plasticity: The intestinal epithelium is the fastest self-renewing tissue in mammals, and thus it requires continuous regeneration mediated by adult SCs that reside in two specific, functionally different intestinal stem cell compartments. The First is composed of actively cycling Lgr5+ SCs present at the crypt bottom, which are responsible for the daily generation of all differentiated cells (Barker, Ridgway et al. 2009). The second intestinal stem cell pool is instead constituted of quiescent SCs, located at the +4 position of the crypt, and was originally identified by Bmi1, mTert and HopX expression. This pool is required after severe tissue damage, and compensates for cell loss by generating new Lgr5+ SCs (Tian, Biehs et al. 2011)(Fig. 2-2a).

Lineage plasticity of functionally distinct stem cell populations: In the epidermis, bulge SCs are restricted to hair follicles, but extensive damage induces them to acquire lineage plasticity and transiently contribute to the epidermal compartment irrespectively of their origin during the acute phase of wound healing (Ito, Liu et al. 2005). This is a robust fail-safe mechanism to maintain the regenerative ability in case of stem cell loss when tissue is injured (Fig. 2-2b).

Transdifferentiation: Liver and pancreas regenerative ability is very well-established, still a debate is ongoing, regarding the existence of stem/progenitor cells in these organs, and their function in homeostasis and regeneration. During liver regeneration, hepatocytes seem to restore homeostasis by self-duplication as facultative SCs. Despite their proliferative potential, hepatocytes initiate an *in vivo* differentiation program following injury that results in their transdifferentiation into biliary epithelial cells (Fan, Malato et al. 2012, Yanger, Zong et al. 2013). In the pancreas, still no definitive evidence supports the existence of facultative SCs, however, recent work identified bi-phenotypical cells expressing both acinar and ductal markers upon severe inflammation and damage. Moreover, pancreatic acinar cells can be reprogrammed into proliferative duct-like cells expressing markers for acinar, ductal, and progenitor cells (Fig. 2-2d) (Kopp, Grompe et al. 2016). In pancreas, in another study using lineage-tracing to label the glucagon-producing α-cells before β-cell ablation showed that tracked large fractions of regenerated β-cells were derived from α-cells, revealing a previously disregarded degree of pancreatic cell plasticity through transdifferentiation (Thorel, Damond et al. 2011).

De-differentiation: It is a well-documented form of plasticity in various organs including lung, gastrointestinal tract, and stomach, in response to tissue injury (Fig. 2-2c). In the airway epithelium, depletion of basal SCs in the mouse trachea induces the dedifferentiation of secretory Clara cells into multipotent SCs that are morphologically and functionally indistinguishable from basal SCs (Tata, Mou et al. 2013).

In the intestine, a lineage tracing study showed that Delta-like 1 (*Dll1*) expressing

progenitor cells (*Dll1-GFP-IRES-CreERT2* alleles), during homeostasis, generate short lived clones including Paneth cells, Tuft cells, and enteroendocrine cells. Following ablation of SCs by irradiation, $Dll1^+$ cells yield fully labeled crypt units, indicating that committed secretory progenitor cells can de-differentiate into SCs (van Es, Sato et al. 2012). In the stomach, using a lineage tracing approach with the *Troy-Egfp-ires-CreERT2* allele, single marked differentiated Troy+ chief cells are shown to generate complete labeled gastric units at very low frequency (over a period of months), with their contribution to the corpus gland increasing after damage. Moreover, Troy+ cells can generate long-lived gastric organoids, indicating that these cells function as reserve SCs actively responsive to injury (Stange, Koo et al. 2013).

More recently, it was shown that by lineage tracing that surgical depletion of K15-GFP+ labeled corneal epithelial basal SCs (limbal SCs) located at the corneal regeneration site, induces corneal-committed cells to dedifferentiated into bona fide limbal SCs, thus restoring normal tissue dynamics and marker expression (Nasser, Amitai-Lange et al. 2018). Clearly, these examples of different forms of plasticity in epithelial tissue injury responses suggest that dedifferentiation is widely utilized among many epithelial tissues. However, it is not the only mechanism for tissue regeneration and repair.

Epithelial-Mesenchymal-Transition (EMT): is a process through which epithelial cells become mesenchymal cells via the loss of cell polarity and intercellular adhesion to acquire migratory properties and is also associated with the acquisition of a stem-like phenotype (Shibue and Weinberg 2017). EMT plays fundamental functions in many physiological processes such as embryonic development, tissue regeneration and is reactivated in pathological conditions such as wound healing, organ fibrosis and cancer (Thiery 2003, Hay 2005, Savagner 2015). EMT is a distinct form of cellular plasticity from the aforementioned forms as explained in more detail in the section below.

2.1.2 Cellular plasticity in cancer

Cancer is no longer considered a bulk of un-differentiated cells but is instead a complex

heterogeneous “tissue” manifested in numerous distinct subtypes, each with its own distinct histopathological and biological features. Cancers often overcome treatments and become resistant to therapy, and patients are left with few or no effective treatments. A new and further level of complexity was added by the recognition of cell state plasticity, in particular, non-CSCs can give rise to CSC-like populations with various degrees of efficiency.

The previous paragraphs highlight that the cell plasticity is an essential phenomenon encompassing diverse processes that mediate tissue regeneration and repair. However, its’ aberrantly activation can contribute to malignant transformation and endow tumors with the properties required to compensate for reconstruct the heterogeneity and to adapt to stress (Ye and Weinberg 2015, Merrell and Stanger 2016). Evidently, cell plasticity can contribute to cancer at different levels from tumor initiation, maintenance and heterogeneity to metastasis.

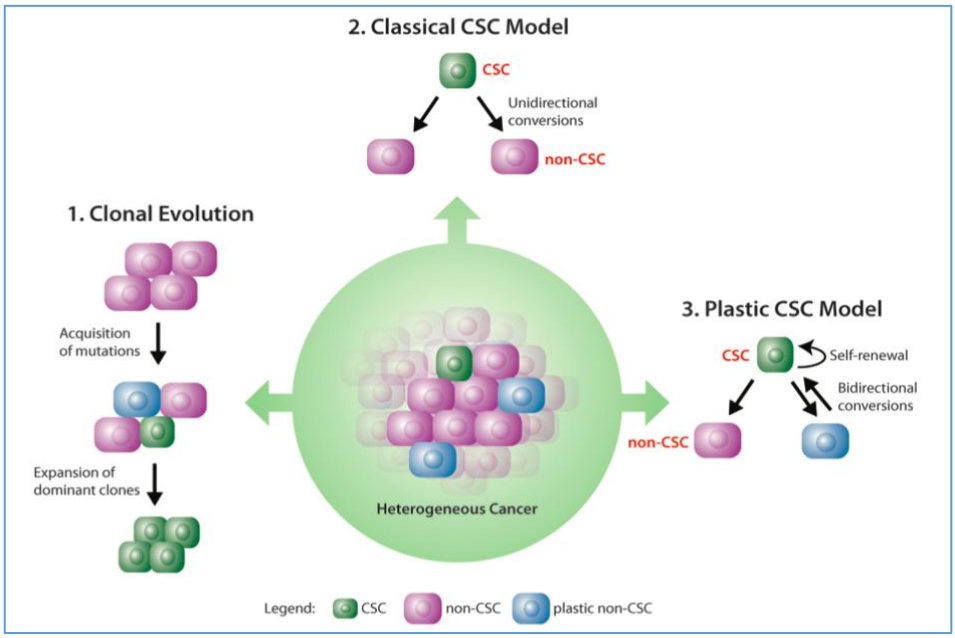


Figure 2-3. **Models of tumor heterogeneity**
(Adapted from Marjanovic et al. 2013)

Cell plasticity in Tumor Initiation: Two models of tumor initiation and progression to advanced disease have been proposed: Clonal Evolution and Cancer Stem Cell (CSC) models (Fig. 2-3). The clonal evolution theory suggests that successive mutations accumulating in a given cell enable it to generate clonal outgrowth that thrive in response

to microenvironmental selection pressures, dictating the phenotype of the tumor. The cancer stem cell model, instead suggests that cancer cells with similar genetic backgrounds can be hierarchically organized based on their tumorigenic potential as a complex "tissue". This model implies that tumors contains differentiated cells that are fueled by CSCs which resembling to their normal tissue counterpart, are placed at the apex of the tumor hierarchy and are thought to be possess the ability to self-renewal and initiate tumors. Furthermore, CSCs have been implicated in the eventual metastasis into distant organs and subsequent therapy resistance. Considerable evidence has been accumulated in recent years to suggest an evolving third model, the plastic cancer stem cell theory (non-canonical CSC model), in which bidirectional conversions exists between non-CSCs and CSCs. This model implies that tumorigenic CSCs can be *de novo* generated from non-CSCs populations throughout tumorigenesis (Marjanovic, Weinberg et al. 2013).

CSCs may represent transformed counterparts of normal adult SCs that are relatively long-lived and self-renewing. Both these properties would allow them to accumulate and propagate mutations over time to originate a tumor. For instance, in intestinal cancer, depletion of adenomatous polyposis *(Apc) in Lgr5*+ long-lived SCs induces their transformation within a day. Interestingly, these transformed SCs remain localized at the crypt bottoms, fueling a growing microadenoma, which can further develop into macroscopic adenomas within 3-5 weeks (Barker, Ridgway et al. 2009). Also in gastric cancer, it has been shown that Mist1+ SCs bearing the combination of *Kras* and *Apc* mutations, were shown to serve as a cell-of-origin for intestinal-type cancer (Hayakawa, Ariyama et al. 2015).

Another important mechanism of CSC generation could involve the ability of epithelial lineage committed or differentiated progeny, to dedifferentiate into SCs. In particular, the loss of specific tumor suppressors or activation of certain oncogenes can lead to the dedifferentiation of terminally differentiated cells and thus favor tumorigenesis by

increasing the number of tumor-initiating cells. For instance, in the mammary gland the deletion of the tumor suppressor *Brca1* in mammary epithelial luminal progenitors leads to the formation of tumors that resemble the majority of sporadic basal-like breast tumors, replete with a heterogeneous expression of both basal and luminal markers (Molyneux, Geyer et al. 2010).

In the intestine, elevated NF-κB signaling in intestinal epithelial cells (IEC) enhances Wnt activation and induces dedifferentiation of non-SCs (enterocytes) that can then acquire tumor-initiating capacity (Schwitalla, Fingerle et al. 2013).

Cell Plasticity in Cancer Maintenance and Heterogeneity: As cited earlier, the CSC model explains how tumor heterogeneity arises and contributes to tumor progression and maintenance. Alternative to this is the plastic CSC model, in which stem properties can be acquired by non-CSCs, rather than a cell-intrinsic property to distinct cell type. Among the plasticity forms, cancer cell dedifferentiation, and transdifferentiation significantly contribute to tumor maintenance and heterogeneity by generating *de-novo* CSCs.

De-differentiation: In invasive colon cancer, depletion Lgr5+ cells (CSCs) via diphtheria toxin led to reduced tumor growth, when drug administration was halted tumors regrow rapidly with Lgr5+ cells immediately reappearing, indicating that non-CSCs generate CSCs responsible for tumor growth (De Sousa, Wang et al. 2013). Moreover, fibroblasts and tumor-associated macrophages (TAMs) produce factors that induce stemness-associated pathways, such as Wnt, Notch, NF-kB or YAP/TAZ-dependent signaling. These factors can induce a stem-like state in differentiated tumor cells (Grivennikov, Greten et al. 2010, Zanconato, Battilana et al. 2016).

Transdifferentiation: apart from the dedifferentiation form of plasticity, transdifferentiation has also been documented to enhance tumor growth and metastasis. For example, in glioblastomas (GBMs), lethal brain tumors characterized by extensive abnormal vasculature, lineage tracing demonstrated that glioma SCs (GSCs) give rise to vascular pericytes and that the elimination of GSC-derived pericytes inhibits tumor

growth (Cheng, Bao et al. 2010). A further study in GBM showed that GSCs cells give rise to endothelial lineages, by *in vivo* lineage tracing analysis (Wang, Chadalavada et al. 2010).

EMT: in cancer, EMT is mainly associated with circulating tumor cells, metastasis, and cancer SCs. The first suggestion of both EMT and stemness at the invasive front of tumors was made in 2005 to explains how CSCs disseminate from the primary tumors and regenerate fully heterogeneous tumors at the secondary sites (Brabletz, Jung et al. 2005). Later studies showed, in head and neck cancer, hypoxia or over-expression of hypoxia-inducible factor-1 alpha (HIF-1α) promotes EMT resulting in the formation of a large number of lung metastases in mice (Yang, Wu et al. 2008). In mammary gland, EMT induction *in vitro/in vivo* in immortalized human mammary epithelial cells (HMLEs), nontumorigenic and tumorigenic via the ectopic expression of defined TFs (i.e. Twist and Snail), endows the cells with mammosphere or tumorsphere forming ability, respectively. Moreover, Twist/Snail-expression in transformed HMLE cells by introduction of an activated form of *HER2/neu* (HMLEN) leads to an increased number of tumor initiating cells which is reminiscent of both human breast CSCs and normal mammary epithelial SCs (Mani, Guo et al. 2008, Chaffer, Brueckmann et al. 2011).

The correlation between EMT and CSCs indicates that EMT induction is necessary not only for the dissemination of tumor cells to distant tissues but also for cell conversion into the CSC state. Therefore, it is not surprising that EMT activation induces autocrine signaling loops, TGFβ–SMAD and Wnt/ß-catenin pathways, that are known to contribute to the stemness of non-neoplastic and neoplastic cells (Scheel, Eaton et al. 2011). In addition, EMT activation confers to tumor cells a resistance to numerous of therapeutic drugs, which is another important hallmark of CSCs.

Overall, Cell plasticity contributes to cancer in a myriad of different ways. Thus, uncovering the mechanisms underlying plasticity has important implication for targeted cancer therapy and will provide us with new insights into clinically relevant questions.

2.1.3 Cellular plasticity and epigenetics

Epigenetic is characterized by inheritable, non-genetic histone and DNA alterations, which (explicitly) explain why, despite sharing the same genome, different cell types within the same organism respond differently to environmental, developmental, or metabolic signals. In a simplistic explanation, epigenetics informs a gene where and when to be expressed by epigenetic modifiers to read, write or erase DNA and histone modifications. When promoters or transcription start sites (TSS) are methylated, activating transcription factors are unable to bind these regulatory regions, or repressive complexes are recruited to these regions to repress of the gene expression. The fundamental unit of chromatin is the nucleosome which is composed of histone proteins and wrapped by 146bp of DNA. The N-terminal tails of histones are relatively accessible to enzymatic modifications that causes alterations in the distance between nucleosomes. The alterations impact on chromatin compaction and resulting in the recruitment of histone-modifying complexes that activate or repress gene expression. Three major categorizes of epigenetic modifications that control gene expression: (1) DNA methylation, (2) histone marks, and (3) non-coding RNAs (Paksa and Rajagopal 2017).

As described in previous paragraphs, cell plasticity both in normal and cancer tissue is a responsive mechanism of cells to injury, environmental and genetic perturbation. Moreover, the maintenance of a stable cellular fate is crucial for normal tissue function, and it is subjected to tight control by genetic and epigenetic regulators (Tetteh, Farin et al. 2015). Clearly, there must be a rewiring of the epigenetic landscape in which cell fate can be redirected into another distinct fate in response to intrinsic and extrinsic stimuli (Paksa and Rajagopal 2017).

The epigenetic basis of cell plasticity was well established and investigated in a non-physiological model: the direct reprogramming of differentiated cells into induced pluripotent SCs (iPSCs), induced by ectopic co-expression of defined factors Oct4, Sox2, Klf4 and Myc, generates a chromatin landscape that is highly similar to that of embryonic

SCs (ESCs) (Takahashi and Yamanaka 2006, Okita, Ichisaka et al. 2007, Rais, Zviran et al. 2013). However, iPSC reprogramming is a relatively inefficient process. Importantly, the reprogramming efficiencies are dependent on the somatic cell type, and might be inversely correlated with their degree of differentiation (Maherali and Hochedlinger 2008). This suggests that there are important epigenetic barriers that are imposed during differentiation to define the cell fate, and that they must be compromised in order to successfully reprogram a cell to reacquire a pluripotent state.

2.2 Mammary Gland Biogenesis and Mammary Epithelial Cell Hierarchy

The mammary gland is a unique organ in mammals that distinguishes them from all other animal and is among the few ones in which most of the developments occur during the postnatal life. The mammary gland is a system of hollow interconnecting ducts, to provide nourishment and passive immunity to offspring until they become able to feed themselves. The mammary gland originates from an invasive branching cord of epithelium. There are different developmental stages in both prenatal (embryonic) and postnatal glands during the lifespan of mammals. In particular, after birth it undergoes numerous cyclic morphogenetic changes from massive expansion to involution in response to local and systemic signals, especially in females. Therefore, mammary gland ducts must contain cells with significant proliferative, invasive and multi-lineage generative properties allowing these numerous morphogenetic cycles. Indeed, a fully functioning mammary gland can be generated from a single mammary stem cell (MaSC) upon transplantation (Kordon and Smith 1998, Stingl, Eirew et al. 2006). Thus, the mammary gland is a highly useful model system to study stemness, differentiation and cell plasticity.

In our study, we chose the mouse mammary gland as a model system, due to its structural and functional similarity to the human mammary gland, and widely used as a model for human breast cancer development and progression (Cardiff and Wellings 1999). Therefore, I will mainly focus on mouse mammary gland biogenesis in this chapter of the introduction.

2.2.1 Mouse mammary gland development

The mouse mammary gland originates simply from the epithelial rudiment. During the embryonic development, the overlying ectoderm is induced and differentiated from mammary mesenchyme to form mammary buds which form a placode in the ventral skin

around E12.5 (Inman, Robertson et al. 2015, McNally and Stein 2017). After birth, the mouse mammary gland undergoes multistage development during puberty, pregnancy, lactation and involution.

After birth, the mammary ductal epithelium remains quiescent until puberty (Fig. 2-4 A). At the onset of puberty, approximately 4 weeks of age in the mouse, under the control of hormones and other stimuli, the mammary epithelium gives rise to accelerated ductal extension and branching morphogenesis and large club-shaped terminal end buds (TEB) appear at the tip of growing ducts (Fig. 2-4B). TEBs are specialized structures containing histologically distinct and highly proliferative cell types: body cells and cap cells. Body cells give rise to mammary epithelial cells, Cap cells instead are the precursors of myoepithelial cells. TEBs drive ductal tree extension at the invading front of the branch and lumen formation by highly regulated processes of cell proliferation and death in the TEBs (Humphreys, Krajewska et al. 1996).

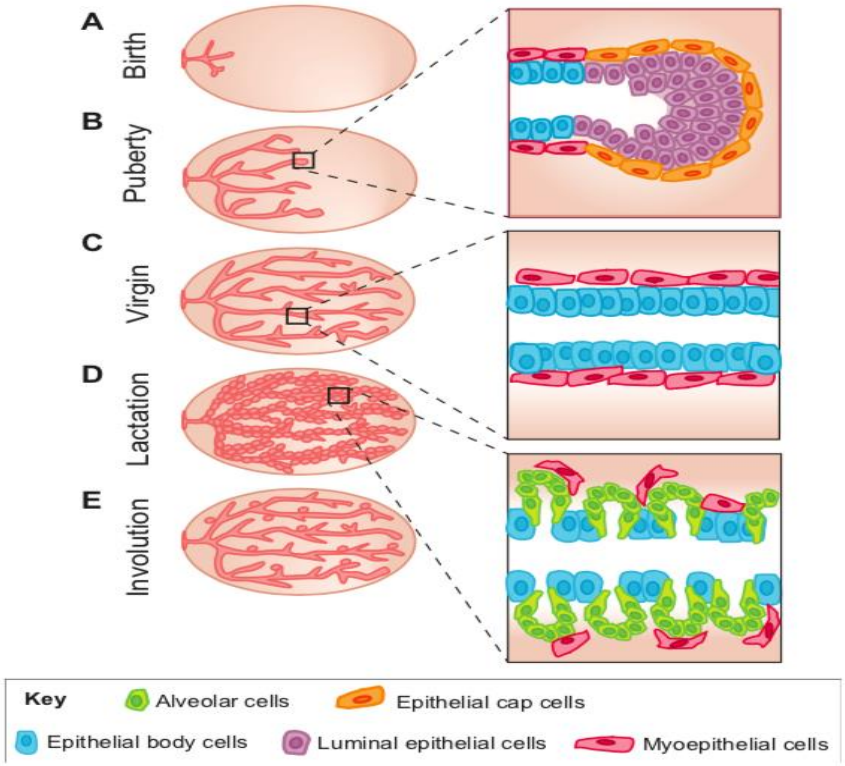


Figure 2-4. **The multistage of mouse mammary gland development after birth**
(Adapted from Inman J.L et al. 2015) © 2015 The Company of Biologist Ltd.

The final ductal tree at the adult virgin stage, which ultimately serves as a channel for milk, is a bilayer structure that is composed of an apically oriented single layer of luminal

epithelial cells that are surrounded on the basal side by contractile myoepithelial cells (Fig. 2-4C). During pregnancy, circulating hormones induce additional branching and extensive alveolar epithelium proliferation, gradually resulting in differentiated secretory alveoli that are able to produce milk (Fig. 2-4D).

During the lactation stage, milk proteins synthesized by the luminal epithelial cells are secreted into the lumen of the secretory alveoli, then the milk moves through ductal trees into the nipple by contraction of the surrounding myoepithelial cells induced by a suckling infant (Fig. 2-4D). During the process of weaning an involution occurs, due to the loss of stimuli for milk production, which leads to a complete remodeling of the alveolar epithelial compartment: apoptosis takes place, leading to regression of the mammary glands closely resembles the virgin state (Fig. 2-4E). At each pregnancy, the epithelial component and the tissue structure go through this process of cyclic remodeling (Inman, Robertson et al. 2015).

2.2.2 Mouse mammary gland epithelial hierarchy

The mammary gland epithelium is composed of two main cellular lineages: the basally oriented contractile myoepithelial/basal cells which are in contact with the basement membrane, and the apically oriented luminal cells that line the ducts (Fig. 2-4C) (Visvader 2009). As described in the preceding paragraph, mammary epithelium undergoes highly dynamic changes during the morphogenetic cycle, showing dramatic regenerative potential and the ability to undergo many cycles of growth and involution. These properties strongly suggest the existence of MaSCs. Similar to the hierarchical model in the hematopoietic system, in the mammary gland tissue, the MaSCs are thought to be at the apex of the hierarchy, responsible to maintain tissue homeostasis by giving rise to progenitor cells and differentiated cells of both lineages (Visvader 2009, Visvader and Stingl 2014, Koren and Bentires-Alj 2015).

However, recent studies led to two different models about the nature of MaSCs in the

postnatal mammary gland. The first one proposes, according to the classical transplantation assays, the existence of bipotent MaSCs (generate both basal and luminal lineages) that have an important function in coordination of considerable morphogenetic changes in the adult glands. The second model postulates, based on lineage tracing studies, that only unipotent MaSCs (generate either basal or luminal lineages) contribute to postnatal mammary gland morphogenesis. These opposing views were reconciled in recent studies and the current knowledge of mammary gland biology suggests the coexistence of bipotent and unipotent adult MaSCs in the adult mammary gland (Visvader and Clevers 2016, Celia-Terrassa 2018).

In brief, in 1998 the first evidence that the mammary gland could originate from a single stem cell was obtained by Kordon and Smith using serial transplantation of retrovirally labeled tissue fragments into cleared mammary fat pads (Kordon and Smith 1998). Later, in 2006, two pioneering works showed the prospective isolation of mouse MaSC-enriched subpopulations within basal cells. They characterized the basal cells as enriched for $CD29^{hi}$ $CD49f^{hi}CD24^{med/+}$ or $EpCAM^{+}$ $Sca\text{-}1^{neg}$, a combination of cell surface markers, and demonstrated their ability to reconstitute an entire functional mammary gland upon transplantation (Shackleton, Vaillant et al. 2006, Sleeman, Kendrick et al. 2006, Stingl, Eirew et al. 2006). These studies suggested that bipotent MaSCs reside at the top of the epithelial hierarchy in the adult mammary gland.

In 2011, Van Keymeulen et al. using an elegant inducible Cre *in vivo* lineage tracing approach to track mammary stem/progenitor cells during embryogenesis, adulthood and pregnancy, demonstrated that the presence of both basal (Keratin14 positive cells, $K14^{+}$) and luminal (Keratin18 positive cells, $K18^{+}$) unipotent MaSC, revealing an unexpected degree of complexity within the hierarchy (Van Keymeulen, Rocha et al. 2011). This study originated and supported the second model mentioned earlier.

In 2012, another layer of complexity was observed in a lineage tracing study, which showed that cell fate switches could be observed throughout the mammary gland

developmental stages. In this study, the authors used an *Axin2*^{CreERT2} allele was used to label and track the Wnt/ß-catenin responsive cells. *Axin2* was characterized as a functional mammary stem cell marker as a direct target gene of Wnt/ß-catenin pathways by showing the *Axin2+* cells were able to regenerate mammary gland in transplantation assays. Depending upon the developmental stages, *Axin2+* cells contribute to basal and luminal epithelial lineages. In the pre-pubertal mammary gland, *Axin2+* cells are restricted to the basal lineage as unipotent SCs, while during pregnancy these cells contributed to both basal/myoepithelial and luminal cells constituting the alveolar structure. This latter evidence is somehow coherent with the classical transplantation assay (van Amerongen, Bowman et al. 2012). These findings suggest the existence of both bipotent and unipotent MaSCs. In particular, basal/myoepithelail-restricted MaSCs in the virgin gland are long-lived SCs that can be switched to a bipotent SC state during multiple pregnancy cycles. However, the results from the lineage tracing studies raise a question on the existence of bipotent MaSCs along with unipotent MaSCs.

Subsequently, in 2014, Rios et al. provide further evidence for the alternative model of coexistence of bipotent MaSCs and long-lived unipotent MaSCs (Rios, Fu et al. 2014), using a novel 3D confocal imaging strategy combined with a stochastic multicolor Cre reporter for clonal cell mapping studies. One year later, in 2015, Wang et al. identified the protein C receptor (Procr) as a novel Wnt target in the mammary gland. Procr+ cells showed high regenerative ability upon transplantation and differentiated into both basal and luminal lineages as visualized by lineage tracing. Thus, the Procr^{+} cells are characterized as a unique population of a multipotent/bipotent mouse MaSCs (Wang, Cai et al. 2015).

Therefore, these combined approaches helped to resolve the controversy between the classical transplantation assay and lineage tracing assay, supporting the model (Fig. 2-5) whereby both bipotent MaSCs and unipotent MaSCs coexist in adult glands, and are responsible for driving morphogenesis during puberty, whereas bipotent MaSCs

orchestrate the remodeling of the mammary gland during the involution and contribute to ductal homeostasis of the mouse adult glands (Fig. 2-4).

Noteworthy, it is crucial to acknowledge the *in vivo* reconstitution (transplantation) and *in vitro* colony formation assays as functional assay to measure the potential of cells particular activity, particularly under non-native conditions. Whereas, lineage-tracing experiments measure the lineage fate of cells only in unperturbed native structures thus the multi-lineage potential of cells may not be evident (Wahl and Spike 2017, Dravis, Chung et al. 2018).

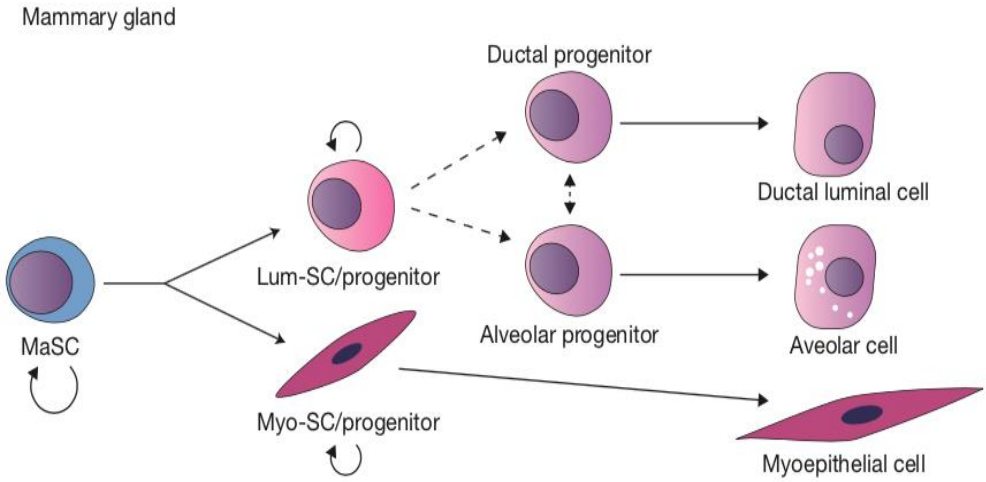


Figure 2-5. **Schematic model of mammary gland epithelial hierarchy**
Adapted from Visvader J.E and H. Clevers © 2016 Macmillan Publisher Limited. All rights reserved.

A recent study, in 2017, conducted by N. Fu et al. uncovered three distinct, largely quiescent MaSC subsets in the adult mammary gland, $Lgr5^{+}$ $Tspan8^{hi}$, $Lgr5^{-}$ $Tspan8^{hi}$, $Lgr5^{+}$ $Tspan8^{-}$, revealing an unexpectedly high degree of complexity within the adult MaSC compartment, using transplantation assays and measuring cell cycle status by FACS analysis for pyroninY (RNA content) and 7-AAD (DNA content). These cells are poised for activation in response to physiological stimuli such as steroid hormones during pregnancy. This suggests that these dormant MaSCs hold an essential function as reserve cells, and can be activated by hormones to be functioning as active MaSCs during the lifespan (Fu, Rios et al. 2017).

Historically, the presence of quiescent/slow-cycling MaSCs *in vivo* was documented as

long-lived label retaining MaSCs (Shackleton, Vaillant et al. 2006). Later, a pioneer work from our lab, Cicalese and Bonizzi et al. in 2009, developed a strategy to isolate and characterize MaSCs by exploiting their slow-cycling properties in *in vitro* mammosphere cultures. Using the PKH26 lipophilic fluorescent dye to label freshly isolated mammary epithelial cell's cell membrane, fluorescence intensity gets diluted at each successive cell division. As a result, highly proliferating cells lose the fluorescent intensity to become negative for the dye, while slowly-cycling cells retain it. Concordant with previous observations, the slow-cycling cells which retain high PKH26 fluorescent intensity was shown to be a MaSC-enriched population (SC frequency=1:3,4) by limiting dilution transplantation, whereas $PKH^{Neg/low}$ cells fail to regenerate mammary gland upon transplantation (Cicalese, Bonizzi et al. 2009). In 2013, Dos Santos et al. employed an inducible expression system in which the H2b histone, linked to GFP, which was regulated by a TRE and a tTA under control of the endogenous keratin K5 promoter (K5tTA-H2b-GFP) to mark slowly dividing cells in mouse mammary gland. This study led to the identification of a small subset (about 0.2%) of slow-cycling cells as $H2b\text{-}GFP^{hi}$ cells within MaSC population with high repopulating ability (dos Santos, Rebbeck et al. 2013).

These intensive studies on the identity, heterogeneity, location and differentiation capacity of MaSC had a tremendous impact on the understanding of the mammary epithelial cell hierarchy. However, as we already reviewed the cellular hierarchy within epithelial tissues (in the first chapter) that cellular differentiation from stem cells to differentiated cells is not unidirectional, instead committed progenitors or "terminally differentiated" cells can change their fate under certain conditions (i.e. tissue damage or stem cell ablations). This is also true in the mammary epithelial cell hierarchy wherein the cell fate plasticity becomes more evident in its contribution towards tumor malignancy.

2.2.3 Mammary epithelial cell plasticity

To date, mammary epithelial cell plasticity under physiological conditions has not yet been directly addressed. However, EMT has been extensively addressed in recent studies as a distinct form of epithelial cell plasticity, correlated with stemness both in normal and cancer cells (Shibue and Weinberg 2017). In 2008, Mani et al. showed the first evidence of a direct link between EMT and the gain of stem cell properties, both in normal and transformed cell lines. They found that the Basal/MaSC enriched subpopulation of non-tumorigenic immortalized human mammary epithelial cells (HMLEs) expresses several EMT markers and high levels of the EMT transcription factors (EMT-TFs): *Snai2/Slug* and *Twist.* Ectopic expression of *slug* or *twist* in HMLEs or in their transformed derivative HMLEN (expressing a activated form of the oncogene *HER2/neu*) endowed them with normal or cancer stem cell properties, respectively, as increased ability to form mammospheres in suspension culture, and increased frequency of CSCs (Mani, Guo et al. 2008).

In 2012, the same group, Guo et al. showed that transient co-expression of two key MaSC transcription factors, Sox9 and Slug (Snail2), within mouse primary luminal differentiated cells, could convert them into MaSCs with long-term reconstituting capacity (Guo, Keckesova et al. 2012). In the same year, Chakrabarti et al. showed that the loss of Elf5 (tumor suppressor gene) induces EMT through transcriptional up-regulation of EMT-TFs, Twist1/2, Zeb1/2 and Snail2. The authors dissected a regulatory axis linking the Elf5 and Snail2 to EMT and metastasis, and showed that the loss of Elf5 increases Snail2-dependent MaSCs' number and promotes breast cancer metastasis. These findings suggested that Elf5 plays a function in preventing dedifferentiation of progenitor cells to SCs (Chakrabarti, Hwang et al. 2012).

A year later, in 2013, Chaffer et al. described a key EMT-TF, ZEB1, in mammary epithelial cell plasticity. They showed the generation of new cancer SCs (CSCs) from

non-cancer SCs (nCSCs) in breast cancer cell lines, both *in vitro* and *in vivo* (Chaffer, Marjanovic et al. 2013).

In 2011, Chaffer et al. previously observed that in adhesion culture of human mammary epithelial (HME) cells, a small proportion of cells grew in suspension termed as HME-floating population of cells (HME-flopcs). Within this floating cell fraction there are two populations: $CD44^{high}$ cells (putative stem-like cells) and $CD44^{low}$ cells (a more differentiated population). They showed that the HME-flopcs-$CD44^{low}$ cells can spontaneously de-differentiate into stem-like cells ($CD44^{hi}$) with both reconstitution ability in humanized mouse mammary fat pad and differentiation ability (Chaffer, Brueckmann et al. 2011). In the same year, an involvement for the deregulation of the Hippo signaling pathway in cell fate plasticity was uncovered by Cordenonsi et al. They demonstrated that elevated expression level of hippo signaling pathway transcription factor TAZ, in non-CSCs, endowed them with CSCs properties, such as self-renewal and tumor initiation ability (Cordenonsi, Zanconato et al. 2011).

In 2015, two groups independently showed that the oncogenic mutant PIK3ca ($PIK3ca^{H1047R}$) expression in normal mammary epithelial leads to distinct cell state changes: basal and luminal lineage restricted cells can de-differentiate into multipotent stem-like states at the early stage of tumor initiation, priming the stage for upcoming intratumor heterogeneity. This was revealed using lineage tracing in adults (Lgr5-$CreER^{T2}$/Tomato-reporter for basal lineage restricted cells and K8-$CreER^{T2}$/Tomato-reporter mice for luminal lineage restricted cells) with or without $PIK3ca^{H1047R}$ and limiting dilution transplantation approaches (Koren, Reavie et al. 2015, Van Keymeulen, Lee et al. 2015). In 2016, Panciera et al. showed that transient over-expression of YAP/TAZ in luminal differentiated cells ($EpCAM^{high}CD49f^{low}CD61^{-}$) directly reprograms them into *de novo* MaSC-like cells with reconstitution ability *in vivo* and self-renewal and full differentiation ability, as assessed *in vitro* by organoid assay (Panciera,

Azzolin et al. 2016).

A year later, in 2017, Britschgi et al. used high content confocal imaging coupled with shRNA screens to discover that large tumor suppressor kinases 1 and 2 (LATS1 and LATS2), part of the Hippo pathways, play essential functions in human mammary epithelial cell fate determination, via direct interaction with ERα. In particular, shRNA-mediated knockdown of LATSs in human and mouse primary mammary epithelial cells, resulted in an increased number of bipotent and luminal progenitors, that are thought to be the cells-of-origin of most human breast cancer (Britschgi, Duss et al. 2017).

More recently, in 2018, Dravis et al. studied the epigenetic and transcriptomic profiles of mouse mammary epithelial cells enriched for E18 fetal MaSCs (fMaSCs-EpCAM^{high}CD49f^{high}) and adult Basal cells (Ba-EpCAM$^{med\text{-}low}$CD49f^{high}), luminal progenitors (LPs-EpCAM^{high}CD49f$^{med\text{-}low}$CD61^{high}), mature luminal cells (MLs-EpCAM^{high}CD49f$^{med\text{-}low}$CD61^{low}) to better understand the mechanisms underlying the ability of differentiated cells to gain of stem-like properties. They identified SOX10 as a regulator of mammary stem and progenitor identity in normal mammary glands. In cancer, SOX10^{high} expressing cells showed dedifferentiation and an EMT-like phenotype, suggesting that SOX10 regulates cell state plasticity (reprogramming) in normal and transformed mammary epithelial cells (Dravis, Chung et al. 2018).

2.3 Epigenetics of Mammary Epithelial Cell Hierarchy

During the lineage commitment processes, epigenetic programs are known to play a fundamental work in lineage specification through remodeling of the chromatin architecture (Margueron and Reinberg 2011). Multiple epigenetic mechanisms are known to contribute to specific gene expression or silencing during differentiation, including DNA methylation, post-translational modification of histone tails, chromatin remodeling and nucleosome positioning.

The epigenetic mechanisms underlying the self-renewal and differentiation of SCs and progenitor cells are still largely unknown, except for the embryonic SCs (ESCs), induced pluripotent SCs (iPSCs) and the hematopoietic system (Liang and Zhang 2013, Sharma and Gurudutta 2016). The mammary gland is a particularly useful model to study these complex epigenetic processes, due to the fact that most of its development occurs postnatal, making it easily accessible for study. Therefore, a number of epigenetic modifiers controlling the balance between the MaSC self -renewal and their differentiation have been identified. Polycomb proteins, histone methylation readers and DNA methyltransferases were shown to play important functions in maintaining MaSC and committed progenitor cells, and Lysine demethylases were shown to be involved in luminal lineage commitment (Holliday, Baker et al. 2018).

2.3.1 DNA methylation

DNA methyltransferases: DNA methylation was the first example of an epigenetic mark found to govern the mammary epithelial cell fate. In 2011, Lee et al using bisulphite sequencing of *Elf5* promoters in different lineages and showed a lineage specific hypermethylation in the basal/stem (88%), as compared to luminal progenitor cells (29%) and mature luminal cells (22%). This was the first example of an epigenetic mark that governs mammary cell fate, indicating that DNA methylation is important for the expression of lineage specific transcription factors in the mammary gland (Lee,

Hinshelwood et al. 2011).

In 2015, Pathania et al. investigated the function of *Dnmt1* in the mouse mammary gland. Dnmt1is a de-novo methyltransferase, principally responsible for the restoration of the original methylation pattern present prior to DNA replication. FACS analysis of mammary epithelial cells from Dnmt1 knock-out mice showed a reduced number of MaSC-enriched basal (Lin-Cd49f^{high}CD24^{+}) and luminal (Lin-Cd49f^{low}CD24^{+}) cells. Dnmt1$^{-/-}$ MaSCs showed reduced frequency of re-populating units and mammosphere forming efficiencies, suggesting that Dnmt1 is required for the maintaining of MaSCs (Pathania, Ramachandran et al. 2015).

2.3.2 Histone marks and their readers

Polycomb group of proteins: In 2008, Pietersen et al. studied a member of the Polycomb Repressive Complex 1 (PRC1), Bmi1, in the regulation of mouse mammary SCs self-renewal and the differentiation of committed progenitor cells. Bmi1 knock-out led to a 14-fold reduction in re-populating ability of MaSCs upon serial transplantation, implying it has function in MaSC self-renewal. Loss of Bmi1 causes precocious alveolar cell differentiation, whereas its over-expression prevents it (Pietersen, Evers et al. 2008). In 2013, Pal et al. using conditional targeting of Ezh2 in the mouse mammary gland, showed a profound impairment in ductal morphogenesis that lead to reduced alveolar unit density (Pal, Bouras et al. 2013). In the same year, another study conducted by Mickalak et al. concomitantly showed that expression of a doxycycline-inducible shRNA targeting *Ezh2 in vivo,* delays ductal elongation to penetrate into the surrounding adipose stroma in the fat pad as a consequence of a reduced luminal progenitor pool. Mammary epithelial cells from these mice showed both lower re-populating ability upon transplantation and colony forming efficiency *in vitro,* suggesting a critical function of Ezh2 in progenitor proliferation and differentiation (Michalak, Nacerddine et al. 2013).

Histone methylation reader: In 2013, Gu et al. showed that specific deletion of Pygo2

in basal cells, a histone methylation reader that is also a co-activator of the Wnt/ß catenin pathway, results in a two-fold reduction of the basal/MaSC population and a decreased re-populating ability. Mechanistically, Pygo2 deletion allowed Notch-mediated luminal differentiation. This suggested that Pygo2 has an important function in preventing the differentiation of MaSCs towards the luminal lineage, in order to maintain MaSCs' basal fate via the coordination of the Wnt and Notch pathways (Gu, Watanabe et al. 2013).

Lysine demethylases: In 2014, Zuo et al. generated knockout mice for the *Jarid1b gene*, encoding for a lysine demethylase that removes tri- and di-methylation of H3K4. Jarid1b$^{-/-}$ mouse showed a delayed in mammary ductal growth, during pubertal development. The authors mechanistically showed that the Jarid1b is critical for the expression of key regulators, of pubertal mammary gland development, namely *Foxa1, Wnt7b, Fgr2 and Prlr* (Zou, Cao et al. 2014). In 2016, Yoo et al. investigated the function of the H3K27me3 demethylase Kdm6A in the mammary gland. Deletion of Kdm6a in mammary epithelial cells lead to defects in mammary gland development during pregnancy, in particular, to excessive expansion of the basal compartment and lack of luminal lineages. These findings highlight the importance of Kdm6a in establishing a balanced expansion of the basal and luminal lineages (Yoo, Oh et al. 2016).

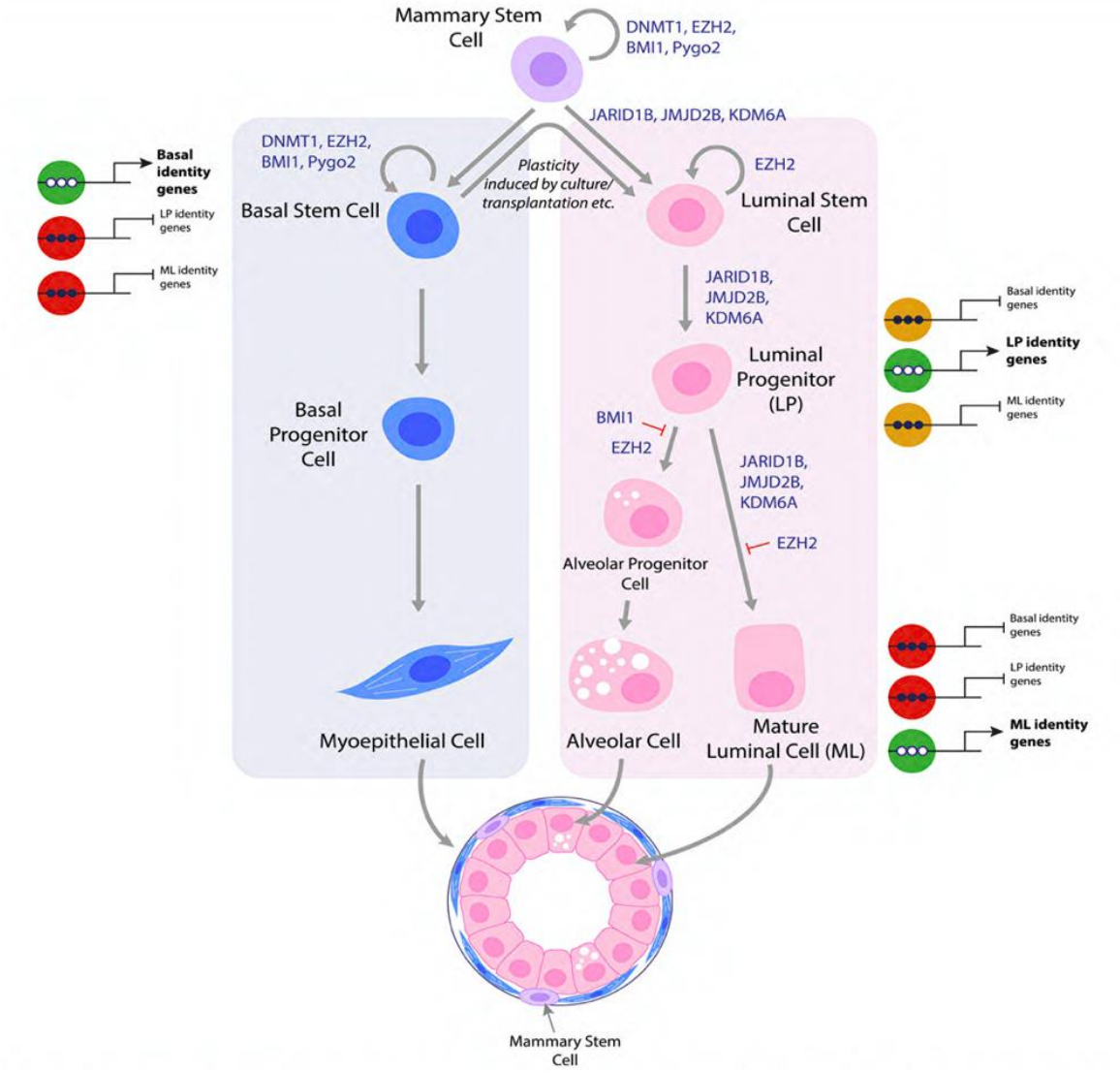


Figure 2-6. **Schematic model for the epigenetic regulation of mammary gland epithelial hierarchy**
Adapted from Holly Holliday et al. 2018 © The Author(s).

In summary, epigenetic modifiers have been shown to exert important functions in governing mammary gland development and hierarchy (Fig. 2-6). Focusing on the mechanism of the epigenetic process underlying normal mammary gland development, leads to more profound understanding of breast cancer wherein perturbation of this epigenetic processes could lead to cell state plasticity. Therefore, more global strategies such as functional screens in mammary epithelial cells using epigenetic shRNA or CRISPR-cas9 libraries to identify and elucidate the genes and pathways that control cell plasticity could be critical to design truly effective cancer therapies.

2.4 Background, Rationale and Aim of the Study

Maintenance of a stable cellular differentiation state is crucial for normal tissue function and conceivably it is subjected to tight control by genetic and epigenetic regulators (Tetteh, Farin et al. 2015).This hypothesis was mainly investigated in a non-physiological model, namely the reprogramming of differentiated cells into Induced Pluripotent SCs (iPSCs) by co-expression of Oct4, Sox2, Klf4 and Myc (Rais, Zviran et al. 2013). Both candidate approach and shRNA screens led to the identification of genes involved in iPSCs reprogramming (Qin, Diaz et al. 2014, Cheloufi, Elling et al. 2015). In more physiological model systems, some genes regulating liver regeneration and pancreatic cell trans-differentiation were recently published (Wuestefeld, Pesic et al. 2013, Sancho, Gruber et al. 2014). However, systematic identification of physiological inhibitors of the de-differentiation process through genetic screenings is just beginning to be exploited (Sun, Chuang et al. 2016).

In order to identify the physiological inhibitors of cellular plasticity in the mouse mammary gland model system, *in vivo* screens of pooled shRNA libraries were performed in our group. The screenings aimed at the induction of a functional SC phenotype in mammary progenitor cells, as assessed by the ability of a single stem cell to regenerate the mammary gland *in vivo*. Several hits were identified in the screenings as putative inhibitors of reprogramming progenitors into SCs.

The general aim of our study was to validate individual shRNAs identified in the screens, both by *in vivo* and *in vitro* assays, and select some of the most relevant validated hits to further investigate their function in reprogramming and the mechanisms involved.

Importantly, accumulating evidence strongly suggests a function of cellular plasticity in the generation of highly malignant CSCs from non-stem, more differentiated, tumor cells. Therefore, a deeper knowledge of the mechanisms controlling cellular plasticity in physiological conditions and their de-regulation in cancer, could lead to the design of

therapies specifically aimed at its inhibition, to achieve more effective target therapy.

2.4.1 Cbx5 and Kmt2d are the selected hits in this study

Cbx5: the gene encodes the HP1α protein and is a member of the heterochromatin protein family, that includes HP1α, HP1β, and HP1γ. They share a highly conserved "chromo-shadow" domain that binds specifically to di-and tri-methylated histone H3 lysine 9 (H3K9me2/3), leading to heterochromatin compaction and transcriptional repression, and a "chromo" shadow domain that is involved in protein-protein interaction (Eissenberg and Elgin 2014). HP1α has various functions such as heterochromatin formation, mitotic progression, gene-regulation, and forms complex networks of gene, RNA, and protein interactions (Vad-Nielsen and Nielsen 2015). HP1α is implicated in cancer, including lung, colon and breast carcinoma (De Lange, Burtscher et al. 2001, De Koning, Savignoni et al. 2009, Yu, Chiou et al. 2012).

Kmt2d is a histone lysine(K) *N*-methyltransferase 2d, formerly named Mll2. Kmt2d is the catalytic component of the mammalian COMPASS (COMplex of Proteins ASsociated with Set1) complex, which enhances transcription through lysine K-specific mono-and di-methylation (H3K4me1/2) at enhancer regions in DNA (Herz, Mohan et al. 2012). Kmt2d plays important functions in regulating gene transcription and is frequently mutated in a variety of cancers, including lymphoma (Morin, Mendez-Lago et al. 2011), medulloblastoma (Jones, Jager et al. 2012), and gastric cancer (Zang, Cutcutache et al. 2012). Kmt2d is also mutated in other human diseases, such as Kabuki syndrome (Paulussen, Stegmann et al. 2011). The identified Kmt2d mutations belong to the loss-of-function type, suggesting that it is a tumor suppressor in various tissues (Zaidi, Choi et al. 2013, Rao and Dou 2015).

3. Materials and Methods

3.1 Animal Experiment

3.1.1 Mice

7 or 8 week-old FVB wild-type female mice for mammary epithelial cell preparation and 3 week-old mice for transplantation were purchased from either Envigo or Charles River.

3.1.2 Mammary cleared fat pad transplantation assay

For mammary gland regeneration assay, the sample material to be injected (dissociated mammospheres or infected PKHneg cell subset) was collected; cells were counted and re-suspended in PBS at the proper cell density in 20µl of final volume per injection. 3 week-old female FVB mice were anesthetized with 2.5% Avertin in PBS (100% avertin: 10g of 2,2,2- tribromoethanol, Sigma, in 10 ml of 2-methyl-2-butanol, Sigma) and the fat-pad of their inguinal mammary gland was depleted of the endogenous mammary epithelium. At 3 weeks of age the mammary epithelial tree has not undergone the puberty-driven development that results in the penetration of the fat pad, and can be easily removed by surgical cut of the area spanning from the nipple to the lymph node, leaving the fat-pad clear for the injection of exogenous cells (Fig. 3-1).

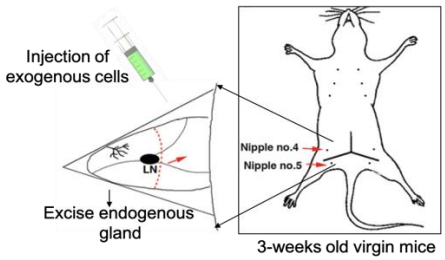


Figure 3-1. **Cleared mammary fat pad transplantation assay**
Modified from Maria del Mar Vivanco (ed.), *Mammary Stem Cells: Methods and Protocols*, Methods in Molecular Biology, vol. 1293, © Springer Science+Business Media New York 2015.

After 12 weeks, the mice were sacrificed with CO_2, and the transplanted mammary glands were taken and outgrowths were quantified as a percent of fat pad filling after carmine alum whole mount staining (range, 25% to 80%).

3.1.3 Carmine alum whole mount staining

The transplanted mammary glands were stretched out onto slides and fixed for at least 2hrs or overnight in 4% formaldehyde. Slides were washed twice in distilled water for 10 minutes and then stained overnight or more at room temperature with Carmine Alum solution (0.2% carmine, 0.5% aluminium potassium sulphate in water, Sigma). De-staining was performed in 70% ethanol for 30 minutes, followed by two 30 minutes' washes in 95% and 100% ethanol. Finally, samples were soaked in a 1:2 solutions of benzylalcohol/benzylbenzoate (Sigma) until the fat pad color clarified.

3.1.4 Evaluation of positive transplants and statistical analysis

Using stereo microscope, fat pads were scored as positive if the ductal branching had originated from a central region of the cleared fad pad, with radial directionality of the ductal branching. Estimation of stem cell frequency was performed using the Extreme Limiting Dilution Analysis (ELDA) web tool (http://bioinf.wehi.edu.au/software/elda/) (Hu and Smyth 2009). ELDA computes a 95% confidence interval for the active cell frequency in each population group and it implements a likelihood ratio test for the acceptance of the single-hit hypothesis (p-value).

3.2 Primary Mammary Epithelial Cells and Cell lines

3.2.1 Isolation of mouse primary mammary epithelial cells

The inguinal and axillar normal mammary glands were collected from 7 to 8 week-old virgin WT FVB mice. Mammary tissues were mechanically dissected into small pieces approximately 1-3mm size with scissors, and enzymatically digested with the following digestion mixture: Dulbecco's modified Eagle's medium (DMEM, Bio Whittaker), 2mM glutamine, 100U/ml penicillin and 100µg/ml streptomycin, supplemented with 200 U/ml collagenase (Sigma) and 100 U/ml hyaluronidase (Sigma) on the rotating wheel for 2-3

hours at 37°C in a humid atmosphere containing 5% CO2. When the digestion was complete, the cell suspensions were centrifuged at 600 rpm for 5 minutes and then re-suspended in PBS and filtered through 100, 70 and 40, 20µm cell strainers to eliminate cell aggregates. Red blood cells were lysed with ACK lysis buffer (Lonza) for 2 minutes in ice. Cell suspension depleted of red blood cells was washed with cold PBS and cultured in suspension in SC medium, as described below, to obtain mammospheres, stained with the PKH26 fluorescent dye or used for growth assays.

3.2.2 Primary mammosphere culture

Primary mammary cells were stained with PKH26 and plated onto ultralow attachment 6-well plates (Falcon) at a density of 400,000 viable cell/ml (to obtain primary mammospheres) in a serum-free mammary epithelial basal medium (MEBM, Bio Whittaker), supplemented with 2 mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin, 5 µg/ml insulin, 0.5 µg/ml hydrocortisone, 2% B27 (Invitrogen), 20 ng/ml EGF and βFGF (Peprotech), and 4 µg/ml heparin (Sigma) and cultured at 37°C in 5% CO2. In these conditions mammary epithelial cells grow as clonal colonies called mammospheres that reach their maximum size in 5-6 days. After 7 days of culture, primary mammospheres (obtained from freshly isolated mammary cells) were dissociated mechanically using a Gilson® pipette and re-plated at a density of 20,000 cells/ml in 6-well low-adhesion plates, to obtain secondary mammospheres. The same procedure was repeated at each serial re-plating passage for *in vitro* stemness assay.

3.2.3 PKH26 label retaining assay

WT FVB primary mammary cells were re-suspended at the concentration of $1x10^7$ cells/ml and stained for 5 min at room temperature by adding an equal volume of a PKH26 mix (1:2500 PKH-26 dye in PBS) (Sigma, PKH26-GL), moderate shaking in light protected for 5min. Then, the labelling blocked with 5% BSA, washed twice with PBS, and plated in SC medium to obtain primary mammospheres.

3.2.4 FACS sorting of PKH26 labelled cells

PKH26-labeled mammospheres were collected after 7 days and mechanically dissociated to obtain single cell suspension. After a filtering step with a 70µm cell strainer cells were subjected to FACS sorting. The gate for PKH26 negative population was selected according to the basal fluorescence of unstained cells. The obtained PKHneg cells were subjected to be infected by shRNA virus followed by transplantation after a short time (3-4 days) culture as mammospheres or in adhesion.

Cell sorting was performed either on BD MoFlo cell sorter equipped with a 488 nm laser and with a band pass 575/26 nm optical filter for PKH26 fluorescence detection, or on BD FACSAria^{TM} Fusion (Biosciences).

3.2.5 FACS sorting of mammary epithelial cell sub-populations

Single cells freshly isolated from the mammary tissue were mixed with digested organoids. Organoids were derived from the collection of aggregates that did not pass through each of the cell strainers used. This material was further digested with trypsin/EDTA (Lonza), dispase (5 U/ml, Stem Cell Technologies) and DNase (1 mg/ml, Stem Cell Technologies). Inactivation of the enzymes was performed with cold PBS supplemented with 2% FBS. The cell suspension of single cells and digested "organoids" was blocked in BSA 10% and count then re-suspend in PBS+1%BSA. Staining the cells 1h on ice in dark with mammary epithelial cell surface markers as listed below:

- Lineage cocktail (Lin-): anti-CD45 (eBioscience, clone RA3-6B2); ant-Ter119 (eBioscience, clone Ter119); anti-CD31 (eBioscience, clone 390); all PE-Cy7 conjugated (1:300)
- Anti-CD24 (eBioscience, clone M1/69) PE conjugated (1:200)
- Anti-CD49f (eBioscience, clone GoH3) APC conjugated (1:100)

After 1h staining, cells were washed once with 1xPBS, and re-suspended to be sorted by BD FACSAriaTM Fusion (Biosciences).

3.2.6 Cell lines and culture

3.2.6.1 NMuMG cell line culture

The NMuMG mouse mammary gland epithelial cell line was obtained from IEO Tissue Culture unit at passage number 21. Cells were cultured with DMEM supplemented with 10% FBS (South America origin), 2mM L-Glutamine and 10ug/ml Insulin. Splitting ratio 1:3 to 1:5 twice a week.

3.2.6.1.1 Test shRNA Down-Regulation Efficiencies

NMuMG cells were cultured by plating at 200,000 cells/well concentration into 6-well tissue culture plates, and infected with concentrated and titrated shRNA lentivirus at M.O.I. 2 with supplied polyberene. Next day stop infection after 16-17h by changing fresh medium. At 72h post-infection, start puromycin selection with 3ug/ml concentration to purify infected cells for the following RT-qPCR analysis. Infected and selected cells were collected with Zymo-lysis buffer and proceed the following steps to purify total RNA or stored at -80°C if not immediately proceed to RNA purification.

3.2.6.2 CommaDß mouse mammary epithelial cell line

CommaDß cells were kindly provided by Dr. Paola Bonetti from IIT, cultured in DMEM/F12 medium supplemented with 2% FCS, 2mM L-glutamine, 10 µg/ml bovine insulin, and 5ng/ml murine EGF. Splitting ratio 1:12 twice a week.

3.2.6.2.1 FACS Analysis of Sca-1high Stem/Progenitor cell Sub-population

CommaDß cells were plated at 200,000 cells/well into 6-well plates and infected with concentrated and titrated shRNA lentivirus at M.O.I. 2 with supplied polyberene. Stop infection after 16-17h by changing fresh medium. At 72h post-infection, start puromycin selection with 2,5ug/ml concentration to purify infected cells for the following FACS analysis. Infected and selected cells at 6-7 days after post-infection, plated 200,000 cells/well into 6-well plates as Day0. Day2 cells were detached with trypsin incubating 5min at 37°C when cells reached 70-80% confluences (don't allow the cells to reach over-

confluences). Count the cells and re-suspend 200,000 cells in 200ul PBS as single cell suspension, stained for Sca-1 and Cd24 cell surface markers as listed below:

- Anti Ly-6a/e(Sca-1) Ab (eBioscience, clone D7) PE-Cy7 conjugated (1:200)
- Anti-CD24 Ab (eBioscience, clone M1/69) APC conjugated (1:400)

FACS analysis were performed on MACSQuant analyser 10 (Miltenyi Biotec)

3.2.6.3 MCF10A

MCF-10A cell line was obtained from IEO tissue culture unit, DMEM/F12 supplemented with 5% Horse Serum, 100 mg/ml EGF, 1mg/ml Hydrocortisone, 1mg/ml Cholera toxin AND 10mg/ml Insulin. Split ratio 1:4.

3.2.6.3.1 MCF10A mammosphere assay

MCF10A cells were detached from culture plates by incubation with Trypsin for 20min incubation at 37°C, collected and counted. Re-suspend the cells at 1500cells/ml concentration with stem medium and plate into 24-well plate 1ml/well for suspension culture. Every 5-7 days, collect and count the mammospheres, wash once with 1xPBS, incubate with Trypsin at 37°C for 5min followed by mechanical disassociation with P200 pipette to obtain single cell suspension. Count the cell number and re-suspend again at 1500cells/ml concentration with stem medium, plating into 24-well plate with 1ml/well. Same procedure repeated for subsequent passaging.

All mammalian cell cultures were maintained at 37°C, 5% CO^2 and 20% O^2 and handled according to the principles of aseptic technique.

3.3 shRNA Libraries and Screening

3.3.1 shRNA Libraries

Custom-made mouse epigenetic libraries were purchased from Cellecta Inc. The shRNAs were cloned into the pRSI17-U6-(sh)-UbiC-GFP-2A-Puro lentiviral vector containing the puromycin-resistance and the GFP fluorescent marker. shRNAs were under the control

of a constitutive U6 promoter and univocally associated to a bar-code cassette (BC) of 18 degenerated, non-overlapping nucleotides. The libraries contained 1180 (mEpi1) and 1160 shRNAs (mEpi2) targeting 234 epigenetic genes, each with 10 different shRNAs, and 30 scrambles for each sub-libraries.

3.3.1.1 PRSI17 Lentiviral Vector

The pRSI17-U6-sh-UbiC-TagGFP2-2A-Puro shRNA Cloning and Expression Vector is a human immunodeficiency virus (HIV) lentiviral vector with a constitutive U6 promoter to express shRNA constructs. The linearized vector purchased from Cellecta Inc, and is used for cloning shRNA oligos.

3.3.2 ShRNA cloning into PRSI17 lentiviral vector

Individual shRNAs identified from the screens were cloned into same vector used in the screen, the pRSI17-U6-(sh)-UbiC-TagGFP-2A-Puro vector (Cellecta Inc.), following the manufacturer's protocol. The pRSI17 empty vector or Luciferase sh cloned into same vector were used as neutral controls.

Table 3- 1. The list of shRNAs and sequences used for both Mouse and Human cells

	Name	Sense and antisense sequences
Mouse	Kmt2d sh1	CCAGATCTTCCTATTCTGTGT ACACAGAATGGGAAGGTCTGG
	Kmt2d sh2	GTTCATCGAGTTGCGATATAA TTATGTCGCAACTCGATGAAC
	Kmt2d sh3	GCGCTTTGAGTTGTCATTTGA TCAGATGGCAACTCAAAGCGC
	Cbx5 sh11	CGTCAGGCATTATGTCATTTA TGAATGACGTAATGCCTGACG
	Cbx5 sh 15	CCTCCCAAAGCATTTCTGTAT CCTCCCAAAGCATTTCTGTAT
	LUC	CTTCGAAATGTTCGTTTGGTT AACCGAACGGACATTTCGAAG
Human	KMT2D sh1	CTGCAGGAAACTTGGGAATGG TCATTCCCAGGTTTCCTGCAG
	KMT2D sh2	GCAGGAAACTTGGGAATGATT AGTCATTCCCAGGTTTCCTGC
	CBX5 Sh1	GCCGATGACATCAAATTTAAA GCCGATGACATCAAATCTAAA
	CBX5 sh2	GAGAGAGCAGAGTAATGATAT GAGAGAGCAGAGCAATGATAT
	TET2 sh1	GGGTAAGCCAAGAAAGAAA CCCATTCGGTTCTTTCTTT
	TET2 sh2	AAACAAAGAGCAAGAGATT TTTGTTTCTCGTTCTCTAA

3.3.2.1 Plasmid isolation and sequencing

Plasmid DNA was purified from large-scale bacterial cultures with the use of the

NucleoBond® Xtra Maxi kit (MACHEREY-NAGEL), according to the manufacturer's instructions. Cells were resuspended, lysed in alkaline conditions and subsequently neutralised with the addition of appropriate buffers (Birnboim and Doly, 1979 (Birnboim and Doly 1979)). Lysates were centrifuged at ≥ 5,000 x g for 10 minutes in order to precipitate cell debris, protein- SDS complexes, chromosomal DNA and high molecular weight RNA. Supernatants, containing the plasmid DNA, were then transferred to a NucleoBond® Xtra Column, included in the kit, and cleared by the specially designed column filter. Next, plasmid DNA was bound to the NucleoBond® Xtra Silica Resin and, after a washing step, eluted in a high-salt concentration buffer. Finally, the eluted DNA was precipitated with the addition of isopropanol and the resulting pellet re-dissolved in a low volume of dH_2O. The purified plasmids were quantified on the NanoDrop 1000 spectrophotometer (Thermo Scientific) and sequencing of inserts were performed, before any use in further applications.

3.3.3 Lentivirus preparation

For lentiviral production HEK293-T cells were kept in DMEM supplemented with 10% FBS (Life Technologies), Glutamine and Antibiotics (HEK medium). Lentiviral particles were prepared by transiently transfecting HEK293T with the calcium-phosphate procedure with a mixture of lentiviral vectors: 7µg of dR8.2, 7µg of pENV (VSV- G), and 13µg of the lentiviral vector per plate. 62.5µl of 2M $CaCl_2$ were added to the DNA mix and brought to a total volume of 500 µl with water. The mix was added drop-wise to 500µl of 2X HBS (HEPES buffered saline: 250mM HEPES pH 7.0, 250mM NaCl and 150mM Na_2HPO_4) by bubbling. After 15 minutes of incubation, the precipitate was distributed on 70% confluent exponentially growing cells. The medium was replaced 16 hours later with DMEM medium. Viral supernatant was collected 24 and 48 hours after and filtered through a 0.45µm syringe-filter. The viral supernatant was concentrated by ultra-centrifugation for 2h at 22,000 rpm at 4°C and the viral pellet obtained was re-suspended in PBS at 1000X concentration and frozen at -80°C.

3.3.3.1 Lentivirus titration

The functional titre was determined for each batch of viral stock independently, by infection of 293T cells with serial dilutions of the frozen stock corresponding to 0.0001-1 µl of virus (Barde, Salmon et al. 2010). More precisely, 250,000 cells per well were plated in a 12-well plate in 0.5 ml of medium and transduced in duplicate with 1, 0.1, 0.01, 0.001 or 0.0001 µl of concentrated virus. The cells were collected at 72h post-infection, washed and fixed in a 2% formaldehyde solution (diluted in PBS from 37% stock, VWR). The percentage of the GFP+ infected cells was estimated by flow cytometry analysis and the titre was, then, calculated with the aid of the equation below, considering only the dilutions of viral stock which yielded 1-20% GFP+ cells:

$$Titre\ (\frac{TU}{\text{ml}}) = \frac{\text{Number of target cells (count at day 1)} x [\frac{\%\ positive\ cells}{100}]}{volume\ of\ vector\ (ml)}$$

The titre was finally reported as transducing units per ml (TU/ml) and typically ranged between 10^8-10^{10} TU/ml.

3.3.4 Lentiviral infection

The FACS sorted PKH26$^{Neg/low}$ cells or luminal cells (Lin$^-$ Cd24^{med} Cd49f^{hi}) were infected with the concentrated virus in suspension or adhesion, in their own medium added with polybrene, prior to transplantation, mammosphere or organoid assay. Infection efficiency was assessed by GFP expression followed by manual counting of GFP positive cells.

3.4 *In vivo* shRNA Screening

The purified PKH26$^{Neg/low}$ mammary progenitors were infected with pooled shRNA lentiviral libraries that targets 234 epigenetic regulators and for technical reasons were split in two sub-libraries mEpi1 and mEpi2 targeting 118 genes and 116 genes, respectively, each with 10 different sh constructs, transplanted into 3-week old mouse

mammary fat pads, pre-cleared of endogenous tissue. After 12-weeks, the mice were sacrificed, the genomic DNA was prepared from the regenerated mammary glands, and subjected to PCR amplification and Next Generation Sequencing (NGS) for barcode (BC) identification and quantification. In order to control for the representation of the 1200 shRNA constructs of each library in the target cells, a fraction of the transduced cells was collected at the same time of the injection, frozen, and subjected to High Throughput Sequencing (HTS) in parallel with the regenerated mammary glands.

3.5 Validation of shRNAs identified in the screen

3.5.1 *In vivo* validation

For the individual *in vivo* validation, PKH26$^{Neg/low}$ cells were infected with individual sh constructs, and transplanted into 3-week old mouse mammary fat pads, pre-cleared of endogenous tissue. Each gene was silenced by two different shRNAs, to exclude off-target effects. After 12-weeks the regenerated mammary epithelial tissue was harvested and stained by carmine alum staining for evaluation.

3.5.2 *In vitro* validation

For the *in vitro* validation, PKH26neg cells were infected with individual sh constructs, and subjected to mammosphere self-renewal assay and Matrigel organoid assay.

3.5.2.1 Mammosphere self-renewal assay and growth curve

In order to test the self-renewal and growth ability of mammary cells, primary mammospheres were dissociated mechanically and re-plated (at 20,000 cells/ml) to obtain secondary mammospheres in 6-well low-adhesion plates coated with poly-HEMA (Sigma). After 7 days, the newly formed mammospheres were counted, collected and manually dissociated by pipetting. At each passage, the number of retrieved mammospheres reflects the number of mammosphere initiating cells present in the

original culture and the number of cells counted after dissociation allows for the evaluation of the number of cells per sphere that was formed. At each passage spheres were enumerated using digital image analysis (Image J; object threshold 100 microns). Cumulative sphere and cell numbers curves were calculated based on the ratio between plated spheres and obtained spheres and cells, respectively. The number of plated spheres was derived from the total number of cells divided by the size of the mammospheres (nr. of cells/number of spheres) over the passages, under the assumption that the average mammosphere size in a culture does not change(Cicalese, Bonizzi et al. 2009).

The cumulative curves were plotted in a semi-logarithmic scale and they approximated an exponential curve, as expected for a cell population that grows or dies with a constant rate during the time. Growth rates (GRs) were evaluated as the slope of the trend-line of the exponential curves. The exponential regression of the data resulted in the value of the coefficients of determination (R2), which approximate 1 in each of the measured curve, thus indicating the goodness of the fitting model.

3.5.2.2 Matrigel culture of mammary organoids

After infection in 2D cultures, mammary luminal cells (Lin^{-} $Cd24^{med}$ $Cd49f^{hi}$) were detached with trypsin and seeded at a density of 10,000 cells/well in 6-well ultralow attachment plates (Corning) in mammary colony medium (DMEM/F12 containing glutamine, antibiotics, 5% Matrigel, 5% FBS, 10 ng/ml murine EGF, 20 ng/ml murine bFGF, and 4 µg/ml heparin). Primary colonies were counted 14 days after seeding. To test the self-renewal ability of MycER induced MaSCs, primary colonies were recovered from the MG-colony medium by1 hour incubation with an excess volume of ice cold HBSS on ice in order to solubilize Matrigel. Colonies were then rinsed 3 times in cold HBSS by centrifugation at 1000 rpm for 5 min and incubated with accutase for 5 min, followed by mechanical disassociation to obtain a single cell suspension. Cells were counted and re-seeded at 10,000 cells/well in 6-well ultralow attachment plates in MG colony medium for further passaging (Panciera, Azzolin et al. 2016).

3.6 Expression Analysis

3.6.1 Quantitative PCR (RT-qPCR)

3.6.1.1 RNA extraction and measurment

For quantitative PCR experiment, total cellular RNA was extracted using Quick- RNA MiniPrep kit ZymoResearch and RNA concentration measured by Nano-drop.

3.6.1.2 Reverse transcription PCR and qPCR

Reverse transcription was done using EasyScript Plus Reverse Transcriptase and cDNA Synthesis kit. Quantitative RT-PCR analyses were done in triplicate on the Applied Biosystems 7500 Fast Real-Time PCR System with the fast-SYBR Green PCR kit as instructed by the manufacturer (Applied Biosystems) or Bio-Rad. The transcription level of RPLP0 was used as housekeeper. The genes whose expression was under analysis in this study and the relative primers used for their amplification are listed in Table:

3.6.2 Bulk RNA-seq library preparation

3.6.2.1 Total RNA Picopure extraction and RNA quality check

A total amount of 50,000-70,000 cells were collected mammospheres for each experimental time points in biological replicates and total RNA from mammosphere cells was extracted using Arcturus Picopure RNA Isolation Kit from life sciences, following manufacturer's instruction. The quality of isolated RNA was evaluated by running Bioanalyzer RNA Pico chip.

3.6.2.2 cDNA library preparation

cDNA libraries of template RNA molecules suitable for subsequent sequencing were prepared from 0.2 µg high quality input RNA using the Illumina® TruSeq® RNA Sample Preparation Kit v2 and following the manufacturer instructions. Sequencing was performed on NovaSeq 6000 illumina sequencer, read length was 100 base pairs, paired end, sequencing depth was 35 million reads per sample.

3.6.3 10x Chromium Single cell RNA sequencing

Cells were collected from mouse primary mammosphere, PKH26 labeled or not stained (Bulk) mammosphere after cultured 7 days in suspension, disassociated into single cell suspension, then subjected to FACS sorting. Freshly sorted cells as $PKH26^{positive}$, $PKH26^{medium}$ and $PKH26^{negative}$ cells, and, Bulk cells were diluted to the final concentration in 1XPBS+0.02% BSA. Sorted cells were loaded on 10X Chromium System, single cell capture and complementary DNA (cDNA) library preparation according to the Single Cell 3′V2 Protocol recommended by the manufacturer. Briefly, the cells are first washed and prepared in an ideal concentration. The machine partitions thousands of cells into nanoliter-scale Gel Bead-In-EMulsions (GEMs), where all generated cDNA share a common 10x Genomics Barcode but uses a pool of ~750,000 bar-codes to separately index each cell's transcriptome. The silane magnetic beads and Solid Phase Reversible Immobilization beads were used to clean up the GEM reaction mixture and the barcoded cDNA was then amplified in a PCR step. Optimal cDNA amplicon size was achieved using Covaris machine prior to library construction. The P7 and R2 primers were added during the GEM incubation and the P5, and R1 during library construction via end repair, A-tailing, adapter ligation and PCR. The final libraries contain the P5 and P7 primers used in Illumina bridge amplification. Sequencing was carried out on a NovaSeq 6000 Illumina sequencer using asymmetric sequencing to achieve R1 26bp and R2 91bp with reading depth 50,000 reads/cell.

3.6.4 Western blot analysis

500,000 to 1,000,000 adherent cells were collected after washed in cold PBS and lysed in 300 to 500 µl of RIPA buffer (Tris-HCl 50mM; NaCl 150mM; 1% NP-40; EDTA 1mM; 0.5% Sodium Deoxylcholate; 0.1% SDS) supplemented with protease inhibitors (Roche). Incubate the lysis 30min on ICE, then centrifuge for 30min at 13,000 rpm at 4°C, collect supernatant. Proteins were quantified with the use of the DCTM Protein Assay (Bio-Rad) in a 96-well format and the absorbance was measured at 590 nm with the GloMax® 96

Microplate Luminometer (Promega). SDS-PAGE was performed using the NuPage® Novex® Gel System apparatus (Invitrogen) at a constant current of 120 V for approximately 2 hours. Samples were loaded on precast gels Nupage Novex 4-12% Bis-Tris (Invitrogen) and the 1X NuPAGE® MOPS SDS was used as running buffer (Invitrogen). Following SDS-PAGE electrophoresis, proteins were transferred to nitrocellulose membranes (Protan; Schleicher & Schuell) by electroblotting for 1.5 hours at 100 V and then were stained with Ponceau S to verify the efficiency of the transfer. Membranes were blocked for 1 hour in blocking solution: 10% low fat milk in TBS-T (Tris Buffered Saline, 0.1% Tween 20) for all the antibodies used were in 2-5% BSA (Bovine serum albumin). The membranes were washed three times in TBS-T (10 minutes each) and incubated with a secondary antibody linked to horseradish peroxidase for 1 hour at room temperature. After three washes in TBS-T, the proteins were visualized using enhanced ClarityTM Western ECL Blotting Substrate (Bio-Rad) and the ChemiDocTM MP System (Bio-Rad).

3.6.5 Immunofluorescence

Cells grown on cover slips were fixed with 4% formaldehyde for 5 minutes at room temperature, washed three times in PBS. Cells were then permeabilized for 10 minutes with 0.1% Triton-X100 in PBS at room temperature, washed three times in PBS and blocked with donkey serum (blocking solution) for 45 minutes. Staining with primary antibodies was performed in a humid chamber for 1 hour at room temperature and followed by three washes in PBS. Coverslips were then stained with secondary antibodies for 30 minutes at room temperature, washed three times in PBS, counter stained with DAPI and mounted with mowiol. Samples were analysed under an UpRight BX61 (Olympus) fluorescence microscope with a 60X/1.35 oil objective (Olympus). Acquired images were analysed through MetaMorph® Microscopy Automation & Image Analysis Software.

3.7 Bioinformatics and Statistical Analysis

3.7.1 Bulk RNA-seq data analysis

Quality control checks on raw sequences was done using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), then each sequenced samples were aligned on the mouse reference genome (Mus Musculus UCSC, mm10) using TopHat software (Trapnell, Pachter et al. 2009)), which uses the bowtie2 aligner (Langmead and Salzberg 2012). Further quality controls derived from alignment information, such GeneBodyCoverage, were derived using the RSeQC package (Wang, Wang et al. 2012). Mapped sequences were processed with HTSeq software (Anders, Pyl et al. 2015) with parameters: (-m) intersection strict, (-a) skip quality reads less than 1, to count the reads per gene, as a first measure of the expression level. Differential gene expression analyzes including size-factor normalization, shrinkage estimation for the distribution's variance and negative binomial distribution were performed using the R package edgeR (Robinson and Oshlack 2010). The read count was normalized via Loess normalization by the DESeq2 algorithm. DESeq is an R package to analyse count data from high-throughput sequencing assays such as RNA-seq to test for differential expression. In details, DESeq2 calculates the total counts in each condition, and then it performs test statistics under the assumption of a negative binomial distribution from which it calculates the p value. The statistics is then adjusted for Benjamini-hochberg correction to calculate the False Discovery Rate (FDR or *p* value). The final step of the DESeq algorithm calculates the log2 fold change between the two samples under comparison. Additionally, a filter on the average low represented genes in terms of read count (cut-off =10) is applied. The differentially expressed genes (DEGs) were obtained, as such, from pairwise confrontation of the selected sample and its relative control.

For the Gene Set Enrichment Analysis (GSEA, http://www.broadinstitute.org/gsea/index.jsp) we applied a filter on the q value setting

the threshold as lower or equal to 0.01. The analysis was performed on GSEA v2.2.0 platform running the all the DEGs of the Kmt2d, Cbx5, Luc list. As gene sets to overlap and calculate the normalized enrichment score (NES) we used the DEGs of selected experimental groups (Luc, Kmt2d sh1 and Kmt2d sh2 or Cbx5 sh11 and Cbx5 sh15, respectively). P-values of GSEA were calculated by performing 1.000 random permutations of gene labels to create ES null distribution.

Pathway and Gene Ontology analyses were performed by overlap of our gene set with the Molecular signature database (MSigDB v5.0) on the GSEA website (www.broadinstitute.org/gsea). In the Gene Set Enrichment analyses, a "query" dataset, is challenged with "gene-sets". The "query" dataset is composed by the entire list of the expressed genes obtained from RNA-seq experiments from the two conditions to be compared. The list is pre-ranked from the most up-regulated to the most down-regulated ones. A gene-set is a curated list of genes, known to be involved in specific biological processes. A very large ensemble of these gene-sets (more than 17,000) derived from the most used functional annotation source (i.e. GO, KEGG, BIOCARTA, REACTOME), but also extracted from specific reference papers, is available at the MSigDB database. If a specific gene-set under analysis is not enriched for up-regulated or down-regulated genes within the query dataset, genes are distributed randomly ranked in the query dataset. Conversely, the enrichment of a gene-set in one of the two tails of the ranked gene list (left tail represents up-regulated genes, whereas, right tail contains the down-regulated genes) suggests the presence of a group of regulated genes within the query dataset, which is expressed/regulated coherently with the transcriptional signature associated with a certain biological process, hence possibly involved in it. The GSEA algorithm computes the overlap between the query dataset and gene-set, calculates a score at each iteration, builds a cumulative score curves, a normalized enrichment score (NES), which expresses the overlap within the two signatures, and a corresponding FDR (or q-value) as a measure of the statistical significance.

3.7.2 ScRNA-seq data analysis

Alignment of the R2 reads was performed with STAR 2.5.3a to annotate the reads to the Mm10 reference genome (NCBI reference sequence database annotation) and assigning genes (annotation: GENCODE ver M16) to reads using feature Counts 1.6.1(remove reads that are assigned to more than 1 gene). The generation of transcript counts table by using umi-tools 0.5.3. Down-stream analysis performed with R package Seurat 2.3 (R version 3.4.1).

Differential gene expression analysis of single cell data was performed using the scde (v2.0.1) and DESeq2 (v1.12.4) packages, with data quality and default filtering performed using the Scatter package (v1.1.8). Normalized counts from DESeq2 were used for single cell clustering, PCA and tSNE analysis.

For unsupervised clustering of scRNA-seq data, counts of uniquely mapped reads in every protein-coding gene were calculated using SeqMonk (www.bioinformatics.bbsrc.ac.uk/projects/seqmonk) and exported for downstream analysis. Cells were filtered based on a minimum number of 2000 expressed genes per cell. Clusters and marker genes were obtained using the SC3 package (Kiselev, Kirschner et al. 2017).

4. RESULTS

4.1 Identification of Putative Inhibitors of Mammary Progenitor Reprogramming through *in vivo* shRNA Screening Targeting Epigenetic Modifiers

To identify genes that physiologically control cell plasticity, an *in vivo* assay has been set up in our group, using normal mammary progenitors and pooled shRNA libraries. Mouse mammary progenitors were isolated from normal mammary epithelial cells by a well established PKH26 label retaining assay. Briefly, the PKH26 lipophilic fluorescent dye is used to label freshly isolated mammary epithelial cell, that are then grown as mammospheres in suspension culture. The PKH26 fluorescence intensity gets diluted at each successive cell division. As a result, highly proliferating cells lose fluoresce, and become negative for the dye as $PKH^{Neg/low}$, while slowly-cycling cells retain it as $PKH26^{High}$. $PKH^{Neg/low}$ cells were characterized as mammary progenitors which are not able to regenerate the mammary gland tissue upon transplantation into cleared mammary fat pads (de-epithelized fat pad) of recipient pre-pubertal syngeneic mice (Cicalese, Bonizzi et al. 2009).

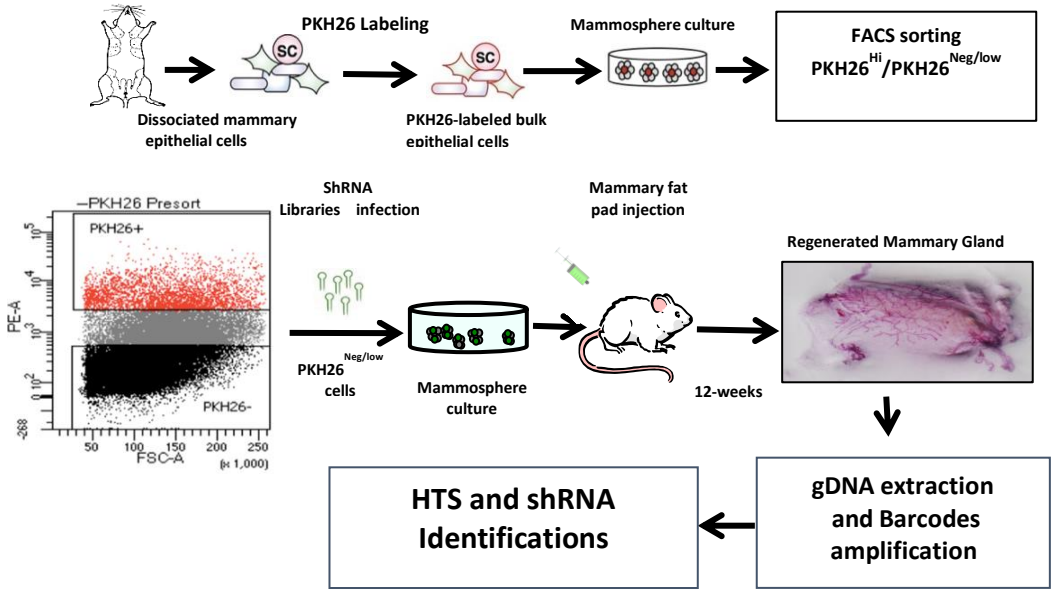


Figure 4-1. **Experimental scheme of the pooled *in vivo* screenings of mouse lentiviral shRNA libraries**

Pooled *in vivo* shRNA screens were performed in our Group by MC. Moroni and E. D'Elia. Using PKH26$^{Neg/low}$ mammary progenitor cells transduced with each custom-designed shRNA libraries (library design: L. Lanfrancone and S. Minucci) and cleared mammary fat pad transplantation. After 12 weeks, mice were sacrificed, the genomic DNA prepared from the regenerated mammary glands, and subjected to PCR amplification and Next Generation Sequencing (NGS) for barcode (BC) identification and quantification. A scheme of screening approach is shown in Figure 4-1 (see also section 3.4).

The libraries targeted 234 epigenetic regulators, each gene targeted by 10 different shRNAs, that were cloned in the pRSI17-U6-(sh)-UbiC-GFP-2A-PuroR lentiviral vector (Cellecta, Inc). In order to enhance the feasibility of the *in vivo* screens, the complexity of the library was reduced by dividing it into two smaller libraries, mEpi1 and mEpi2, targeting 118 and 116 genes, respectively (Fig. 4-2). As a positive control, each library included shRNAs against p53, since p53 down-regulation has been previously shown to confer *in vivo* regenerative ability to mammary progenitors, though at low frequency (MC. Moroni unpublished). Moreover, deletion of p53 or components of its pathway was shown to strongly enhance the generation of iPSCs from human or mouse fibroblasts (Hong, Takahashi et al. 2009, Utikal, Polo et al. 2009).

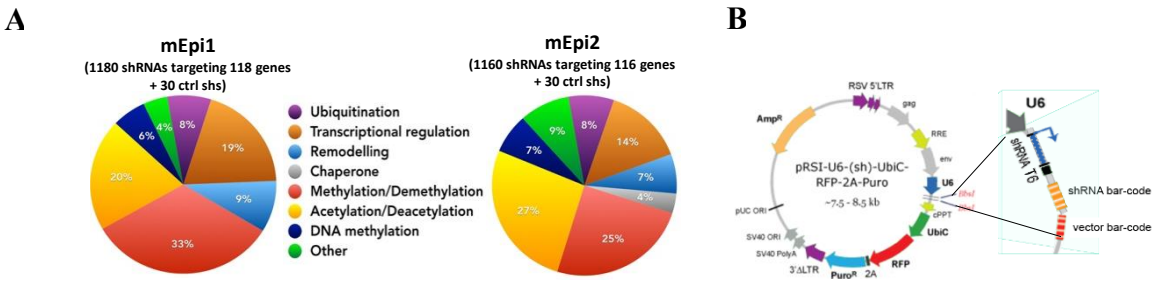


Figure 4-2. **Mouse lentiviral shRNA library composition and map of lentiviral vector pRSI17**

A. Categories of epigenetic regulators included in the two shRNA epigenetic libraries mEpi1 and mEpi2; B. Map of the pRSI17 lentiviral vector in which the barcoded shRNA libraries were cloned, and zoom of the U6 promoter and the double barcodes: i. Sh barcode, ii. Vector barcode to track individual clones;

4.1.1 Hit identification from the screens

The shRNA constructs present in regenerated mammary glands were identified by

bioinformatics analyses of NGS data.

Each shRNA is univocally associated to a single 18nt barcode (BC), to allow its identification and quantification by PCR amplification, independently from the hairpin structure (Fig. 4-1B). As mentioned earlier, in the mEpi1 and mEpi2 libraries (Fig. 4-2), each gene was targeted by 10 different shRNAs. In the screens, as expected, multiple shRNAs targeting the same gene were found in different reconstituted mammary glands, and in some cases even in the same gland. Figure 4-3 shows two examples of the hit/gene identification, the data from the regenerated glands as pooled (A) or individually (B) sequenced.

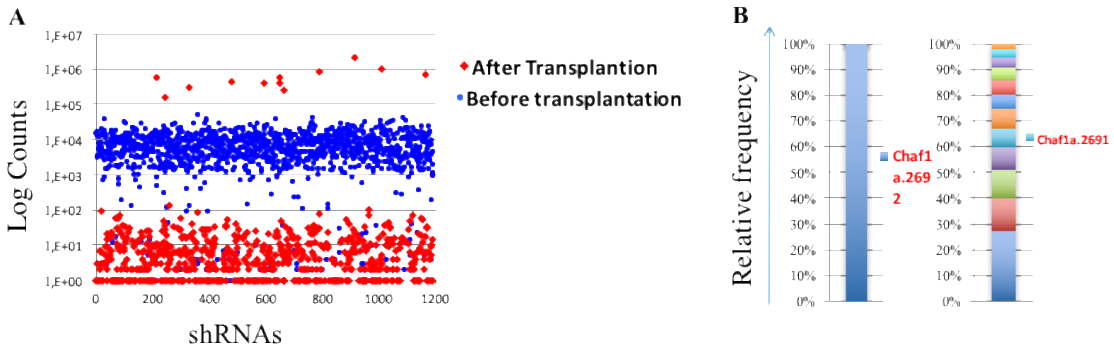


Figure 4-3. **Identification of positive hits in the regenerated mammary glands**

A. Log counts of the shRNAs in a screen. Blue dots: shRNA present in the progenitor cells before injection for regeneration assay; Red dots: upper red dots are the shRNAs identified from sequencing of the pooled-regenerated glands, red dots in bottom are the depleted shRNAs; **B.** Relative frequency of the identified shRNAs in two individual regenerated mammary glands.

38 hits/genes were identified in 6 independent experiments, in which 914,000 total infected cells were screened by regeneration assays. The hits were selected based on the relative frequency of the identified shRNA (>10%). The highest frequency of the shRNAs in the infected cells before injection was 0.6%.

4.1.2 DNA and histone modifiers are enriched and some known inhibitors of iPSC reprogramming are among the positive hits

Most of the hits belong to the DNA and histone modifier categories: 6 hits from DNA methylation, 9 from Histone methylation/demathylation and 6 from acetylation/deacetylation (Fig.4-4).

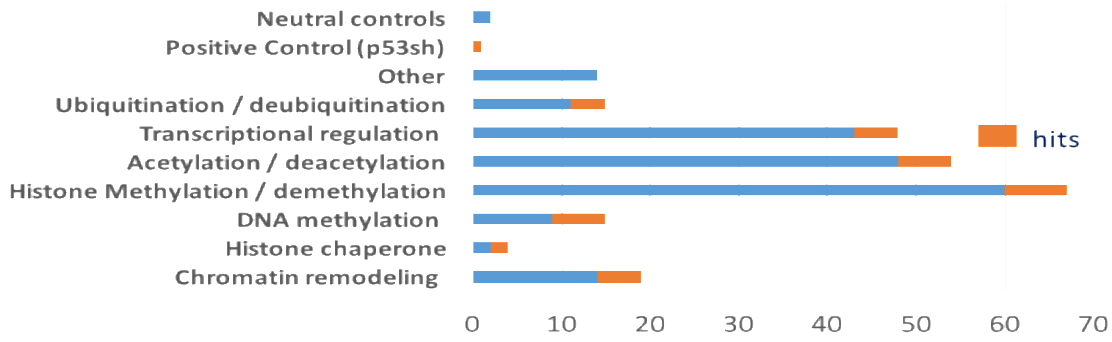


Figure 4-4. **Hits distribution in functional categories**

Notably, some of the hits identified in the shRNA screens, such as Chaf1a (Cheloufi, Elling et al. 2015), Dot1L (Onder, Kara et al. 2012) and p53 (Hong, Takahashi et al. 2009, Utikal, Polo et al. 2009), were recently published as inhibitors of iPSCs reprogramming, hence providing a first level of validation of the screens.

We selected 21 hits for validation, based on the combination of: (i) frequency of the shRNA identified in the screens; (ii) identification of multiple shRNAs inhibiting the same gene; (iii) identification of a hit in more than 1 screening experiment.

4.2 Testing Functional Assays to Validate the Hits

Validation of candidates from the screens was based on *in vivo* and *in vitro* experimental approaches aimed at assessing whether specific shRNAs are able to convert mammary progenitor cells into SCs: i) *in vivo* transplantation assays, to test tissue-reconstitution ability and ii) *in vitro* mammosphere assay, to assess self-renewal capacity (Fig. 4-5).

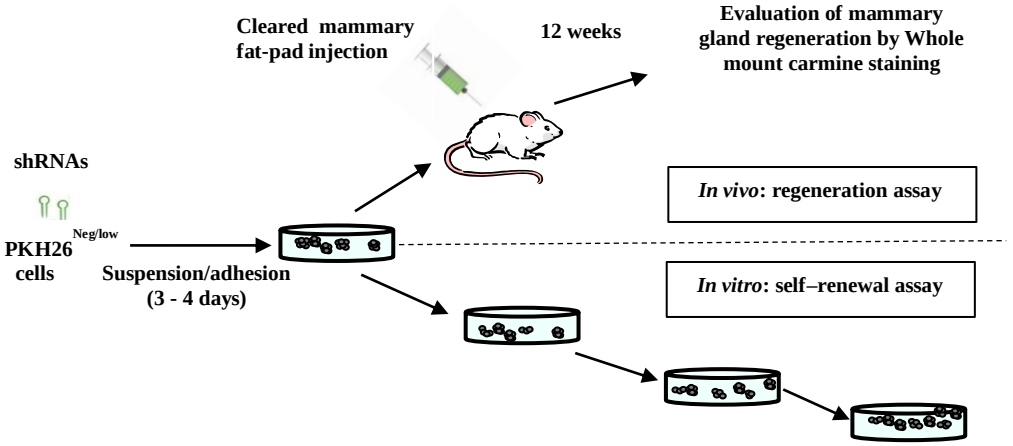


Figure 4-5. ***In vivo and in vitro* validation scheme**

Primary mouse mammary progenitors purified as PKH26$^{Neg/low}$ cells, were infected with individual or 2 pooled shRNAs targeting the same gene. The cells were then either injected into cleared mammary fad pads of 3-weeks old FVB mice (*in vivo*) for regeneration assay or plated as a mammospherses to test the self-renewal ability of the reprogrammed SCs by serial passaging (*in vitro*);

4.2.1 Cloning of the individual shRNAs in the pRSI17 lentiviral vector for validation

The shRNAs identified in the screens were cloned in the same lentiviral vector (pRSI17-U6- (sh)-UbiC-GFP-2A-Puro) in which the libraries were generated (4-2B). Since most of the hits were identified in the screens by two or more shRNAs, we cloned at least two different shRNAs per hit. In total, we have cloned 45 shRNAs into pRSI17, and verified the correct DNA sequence of each shRNA prior to *in vivo* or *in vitro* validations. Before testing the cloned shRNAs biologically, we measured by RT-qPCR the ability of each shRNA to silence its target mRNA, upon infection and puromycin selection in the NMuMG normal mouse mammary epithelial cell line. Most of the shRNAs down-regulate their target mRNA relative expression level from 50% to 90%, as two examples are shown in Figure 4-6.

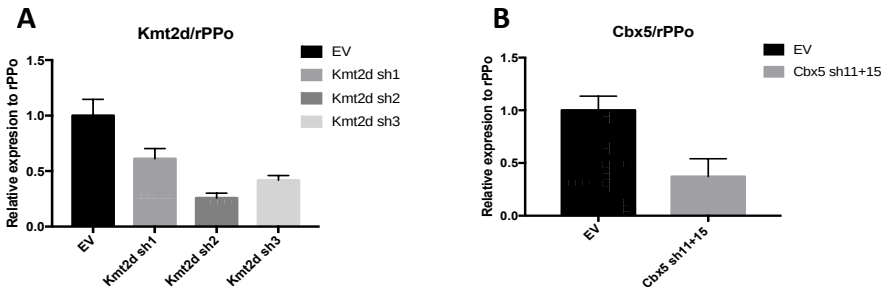


Figure 4- 6. **Kmt2d and Cbx5 knockdown by shRNAs in NMuMG cell line**

A. RT-qPCR analysis testing the efficiencies of Kmt2d down-regulation by three different shRNAs. B. RT-qPCR analysis testing the efficiencies of Cbx5 down-regulation by two shRNAs co-infection. Standard deviation is calculated from technical triplicates;

4.2.2 *In vivo* validations of the hits

In vivo validation experiments were performed following the experimental scheme shown in Figure 4-5. In each validation experiment, mouse primary mammary progenitors ($PKH26^{Neg/low}$) were purified by FACS sorting and infected with a 1:1 mixture of 2 different shRNAs to silence each target gene. As a positive control, in each experiment we included the pWPI-MycER construct, that had been already shown to induce reprogramming of mammary progenitors into mammary SCs (MaSCs) (Pasi, Dereli-Oz et al. 2011, Santoro, Vlachou et al. in press). As a negative control, we used the empty

pRSI17 lentivirus (EV) or the same vector expressing the Luciferase sh (Luc sh). 12 weeks after injection of transduced cells into cleared fat pad, mice were sacrificed and the mammary fat pads were subjected to carmine-alum staining to visualize the regenerated mammary epithelial tissue.

We tested 21 hits/genes in a total of 12 transplantation experiments. Genes were considered validated if scoring positive in at least two independent transplantation experiments, in total 7 hits/genes were scored as validated (see Table 4-1 for two validated genes). Among the validated hits, we selected Kmt2d and Cbx5 for further investigation, based on the relatively high frequency of MaSC generation from mammary progenitors, as estimated by Extreme Limiting Dilution Analysis software (ELDA). In fact, the inhibition of Kmt2d or Cbx5 induced SCs with a frequency of 1/38,450 and 1/71,120 cells, respectively, which was significantly higher than the other validated genes or the MycER positive control (PC, frequency of SCs is 1/97,400 cells) (Table 4-1). Representative images of regenerated mammary glands from all 7 hits are shown Figure 4-7.

Table 4-1. **Summary of the *in vivo* validations**

Controls and Hits	**N of exp**	**Tot infected cell N**	**N of Positive transplantation/ total**	**SC frequency (ELDA)**		
				Lower	**Estimate**	**Upper**
pRSI 17 EV	**12**	**1,101,400**	**1/74**	**9,602,365**	**1,346,770**	**188,890**
pWPI MycER	12	833,000	8/62	196,720	**97,400**	48,200
Dnmt1	2	232,800	1/16	1,624,718	226,950	31,702
Hdac1	2	147,100	1/12	1,078,152	145,160	19,543
Kdm5b	2	182,980	4/10	97,542	35,750	13,102
Cbx5	2	244,000	3/15	220,230	71,120	23,000
Kmt2d	2	130,800	3/10	119,200	38,450	12,400
Sin3b	3	120,420	8/19	43,021	20,670	9,950
Mbd4	4	214,850	2/20	573,672	137,830	33,115

$PKH26^{Neg/low}$ cells infected with shRNAs targeting the hits identified in the screens, negative and positive controls. Stem cell frequency was estimated using the Extreme Limiting Dilution Analysis (ELDA) web tool (http://bioinf.wehi.edu.au/software/elda/) (Hu and Smyth 2009).

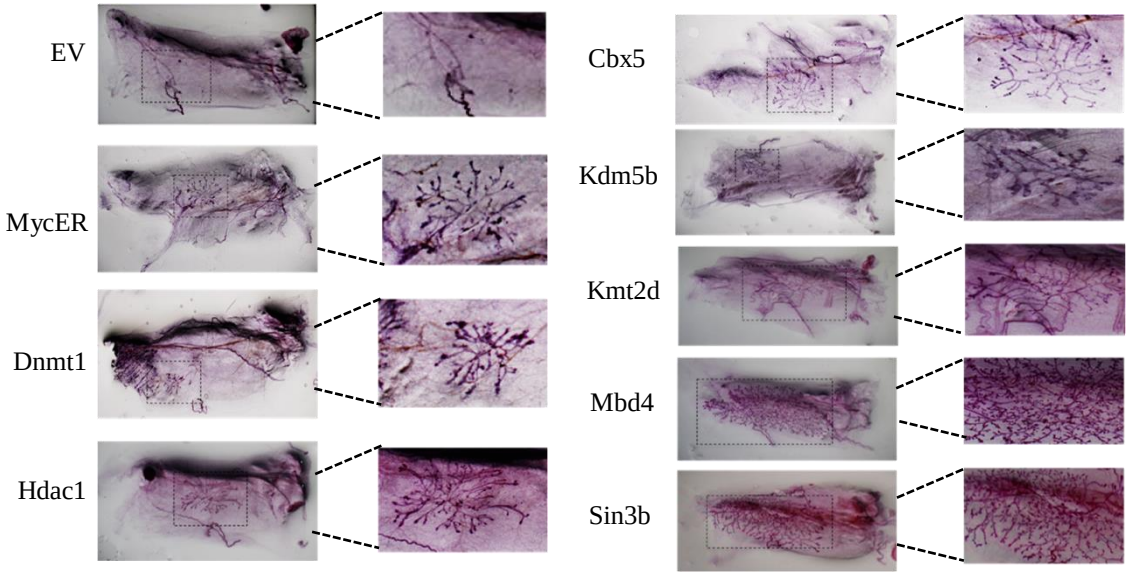


Figure 4-7. **Representative Pictures of regenerated mammary glands upon transplantation of shRNAs infected cells**

Mammary gland regeneration from PKH26$^{Neg/low}$ cells infected with shRNAs targeting 7 different genes assessed by whole-mount Carmine-Alum staining. Empty vector (EV) as negative control; MycER expression used as positive control; 1-7 are the hits we have validated from the screens; The morphology of the glands should be radial, the direction of branches from the injected site are the evaluation criteria for the regenerated mammary gland.

4.2.3 *In vitro* validation of the hits

4.2.3.1 *In vitro* mammosphere assay

To assess whether the *in vivo* validated shRNAs could confer self-renewal ability to mammary progenitors, we employed the *in vitro* mammosphere assay (experimental scheme is shown in Fig. 4-5, lower part). In this assay, cells are cultured in suspension where the progenitor cells die of anoikis while SCs survive and generate spheroids. Mammary progenitors were purified as PKH26$^{Neg/low}$ and infected with the shRNA constructs, as for the *in vivo* validations, and serially passaged in suspension culture every 7 days (Dontu, Abdallah et al. 2003, Cicalese, Bonizzi et al. 2009).

Representative results of two independent *in vitro* mammosphere experiments are shown in Figure 4-8A and B. As expected, the data shows that the PKH26$^{Neg/low}$ cells infected with EV (our negative control) did not grow in suspension and the culture became exhausted after 4-5 passages, while our positive control, i.e. PKH26$^{Neg/low}$ cells transduced with MycER, showed extended growth and self-renewal ability in both 2 independent experiments. However, with the exception of shRNAs targeting Kdm5b which conferred transient self-renewal ability at early passages (Fig. 4-8A), all other

shRNAs did not generate growing mammospheres (Fig. 4-8A and B). Thus, none of the *in vivo* validated shRNAs conferred extended self-renewal ability to mammary progenitor cells through serial passages in suspension culture.

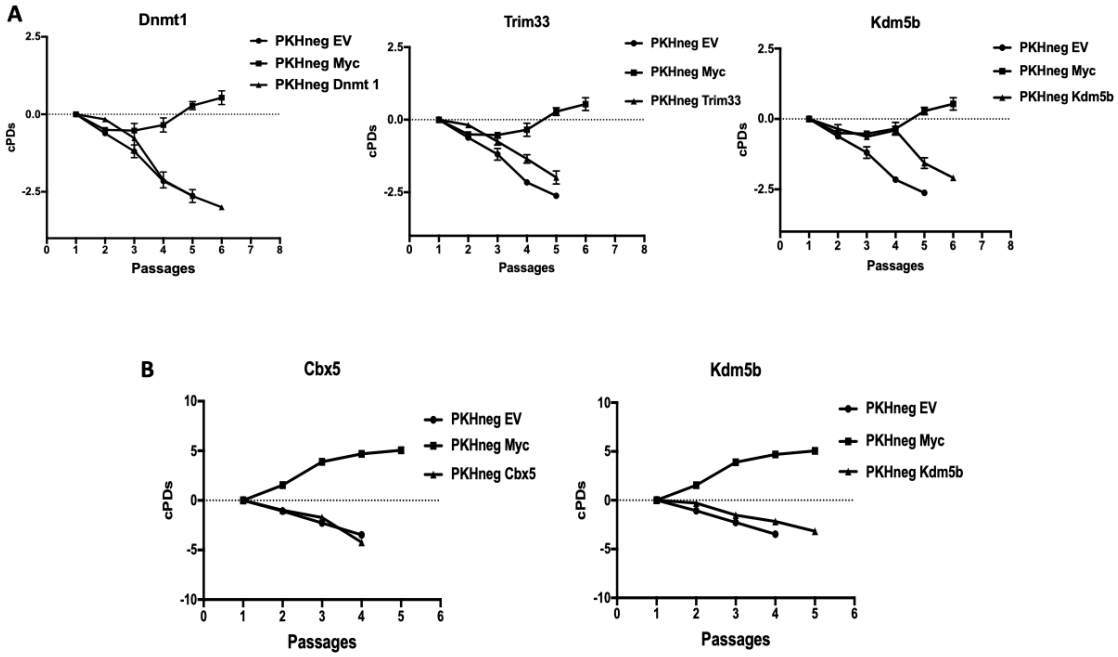


Figure 4- 8. **Growth Curve of Mammospheres from PKH26$^{Neg/low}$ cells infected with shRNAs identified in the Screens**

Two independent *in vitro* experiments, A and B. Briefly, PKH26$^{Neg/low}$ sorted cells infected with shRNAs, targeting 4 different genes separately, were tested in mammosphere growth assay by serial re-plating.

Possible explanation of these results include: a) mammary progenitors were not truly reprogrammed into MaSCs able to self-renew in suspension culture, b) longer time might be required for progenitors to be reprogrammed, and most of them may die by anoikis before being reprogrammed; c) Complete reprogramming may require growth factors or microenvironmental signals present in the mammary fat pad niche, but absent in the suspension culture; d) *in vitro* reprogramming ability by the shRNAs might requires cell proliferation, as given by Myc or YAP/TAZ expression.

4.2.3.2 *In vitro* organoid generation and self-renewal assay

Assessing self-renewal ability is crucial for further characterization of the mechanism by which progenitors are reprogrammed in to SCs. Thus, we tested alternative self-renewal assay, such as the recently reported organoid formation assay (Panciera, Azzolin et al. 2016), used to show reprogramming of mammary Luminal differentiated cells (LD) to

MaSCs by YAP/TAZ transient expression, which leads to generate compact organoids. The organoid self-renewal assay was performed as previously described by Panciera et al. (2016). Briefly, we purified mouse mammary Luminal cells (LUM, Lin$^-$CD24^{hi}CD49f$^+$) and double-negative stromal cells (Lin$^-$CD24$^-$CD49f$^-$) (Fig. 4-9), two mammary gland sub-populations that are deprived of any repopulating ability (Shackleton, Vaillant et al. 2006, Stingl, Eirew et al. 2006). Purified Mammary /Mammary Repopulating Unit (MaSCs/MRU, Lin$^-$CD24$^+$CD49f^{hi}) (Fig. 4-9) were used as an internal positive control. These three purified populations were cultured for 7 days in adhesion on collagen-coated plates, then seeded at a density of 10,000 cells/ well in 6-well ultralow attachment plates in 5% Matrigel mammary colony medium (see section 3.5.2.2). Colonies were counted at the end of each passage (14 days) by Digital image analysis plugin (DIA, Image J; object threshold 100µm), dissociated to obtain single-cells for serial re-plating.

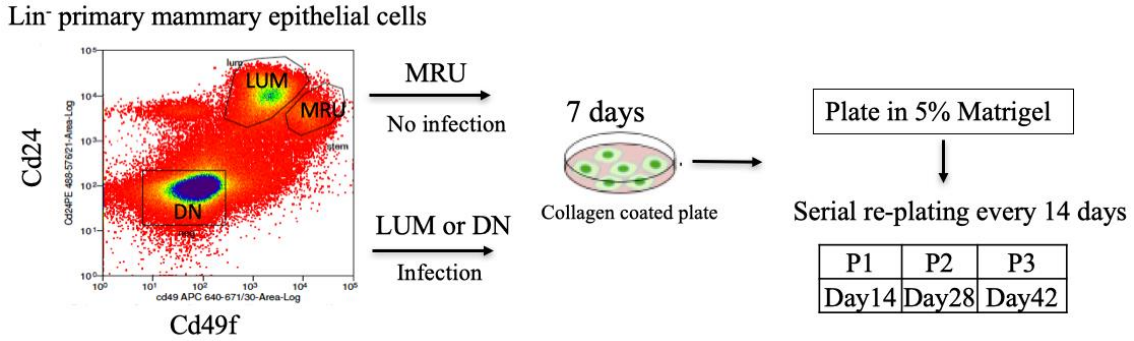


Figure 4-9. **Purification of freshly isolated mammary epithelial cells from mouse mammary glands and schematic representation of the organoid formation assay**

We first tested the effects of MycER over-expression in LPs, as a positive control of the mammary progenitors reprogramming, using the empty vector (EV) as a negative control. We successfully demonstrated that MycER over-expression in the lineage restricted luminal cells (Lin$^-$CD24^{hi}CD49f$^+$) can reprogram them into MaSCs, which generated compact organoids in 5% Matrigel. Moreover, these organoids could be serially passaged, demonstrating self-renewal ability of the MYC-reprogrammed SCs.

As expected, the organoid formation assay showed that MRUs form both compact and acinar organoids, that can be serially passaged, albeit decreasing in number at each passage (Fig. 4-10). Up to passage 2, the colonies were predominantly compact in morphology (Fig.4-10, Pictures in Upper and lower, respectively). Strikingly, we could confirm the presence of functional myoepithelial cells in these colonies by observing contractile movements of compact organoids (link of Videos, (Santoro, Vlachou et al. in press).

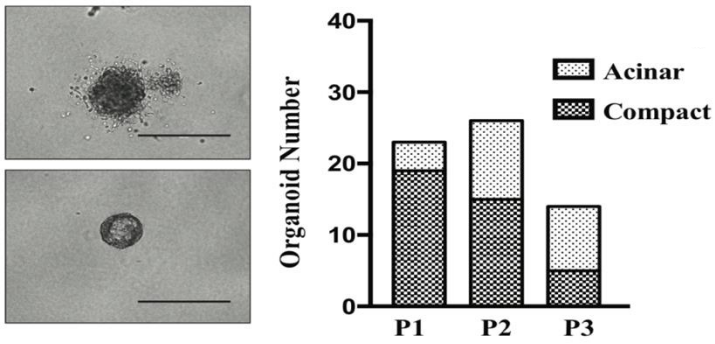


Figure 4-10. **Basal mammary (MRU, Lin$^-$Cd24$^+$Cd49f^{hi}) generate compact organoids and showed limited self-renewal ability**

Representative images of organoids formed by purified MRU (Lin$^-$Cd24$^+$Cd49f^{hi}) in the left; The number of compact organoids decreased by passages, the bar graph on the right, Scale bar: 400 μm. Number of big (diameter ≥100 μm) compact, acinar organoids counted by DIA for 3 consecutive passages (P1, P2 and P3) in 5% matrigel cultures.

MycER over-expression in Luminal progenitors (Lin$^-$Cd24^{hi}Cd49f$^+$) induces the generation of compact and contractile organoid that increase in number through serial passaging. In line with previous reports obtained with immortalized human mammary cells (Poli, Fagnocchi et al. 2018), MycER over-expression bestowed LUM cells with enhanced self-renewal ability (Fig. 4-11B) and differentiation capacity, as shown by the formation of predominantly compact (Fig. 4-11A) and contractile organoids (link of Videos, Santoro, Vlachou et al. in press). Nevertheless, MycER expression had no effect on the morphology of double negative (DN) stromal cells, as compared to control cells infection EV (Fig. 4-12).

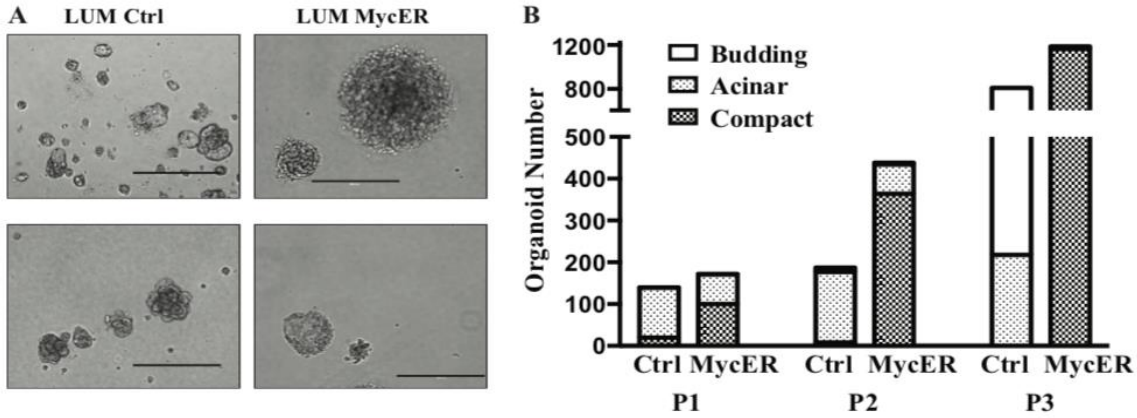


Figure 4-11. **MycER over-expression reprograms luminal cells (Lin$^-$Cd24^{hi}Cd49f$^+$) into MaSCs that generate compact and contractile organoids whose number increases over 3 passages**

A. LUM cells (Lin$^-$CD24^{hi}CD49f$^+$) infected with LUM Ctrl and LUM MycER cells, respectively. Scale bar: 400 µm. B. Number of big (diameter ≥100 µm) compact, acinar and budding organoids counted by DIA for 3 consecutive passages (P1, P2 and P3) in 5% matrigel cultures.

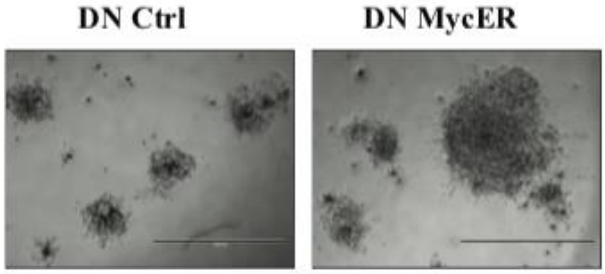


Figure 4-12. **MycER over-expression has no effect on stromal cells (DN)**

Representative images of branched structure formed by purified DN (Lin$^-$CD24$^-$CD49f$^-$); Scale bar: 400 µm.

All together, these data demonstrate that constitutive MycER expression reprograms primary murine lineage- restricted luminal progenitors into MaSCs.

Notably, this approach allowed us to demonstrate *in vitro* reprogramming of freshly isolated, uncultured mouse mammary progenitors into MaSCs, purified by cell surface markers, independently from the PKH26 label retaining assay. This alternative approach is now established in the lab and can be used in the future for *in vitro* validation with the hits obtained from the screens.

4.3 Validation of Cbx5/HP1α as an Inhibitor of Reprogramming of Mammary Progenitors into Mammary Stem Cells

In the shRNA screens, we identified three different shRNAs targeting the Cbx5/HP1α gene in the regenerated mammary glands from two independent screening experiments. Therefore, we validated Cbx5/HP1α as an inhibitor of reprogramming of mammary progenitors into MaSCs.

4.3.1 Cbx5/ HP1α down-regulation induce the generation of mammary repopulating units in PKH26$^{Neg/low}$ progenitor cells

For *in vivo* and *in vitro* validation experiments (Table 4-1 and Fig. 4-8B), primary mammary progenitor cells purified as PKH26$^{Neg/low}$ were co-infected with lentivirus expressing two Cbx5-specific shRNAs in two independent *in vivo* experiments and transplanted into cleared mammary fat pads of pre-pubertal female mice. After 12 weeks, we sacrificed the mice, the mammary glands were proceeded to visualization by carmine-alum on whole mount (section 3.1.3).

In the first two *in vivo* experiments (#1 and #2), the infected PKH26$^{Neg/low}$ cells were cultured in adhesion 3-5 days before injection for *in vivo* regeneration assay, in order to avoid massive cell death occurring in suspension culture, and to allow time for cell reprogramming. The results of experiment #1 showed no mammary gland formation with the EV vector (negative control), while PKH26$^{Neg/low}$ cells co-infected with the two Cbx5 shRNAs-vectors or with the MycER-vector were able to form mammary glands (frequency of engraftment and glands pictures are shown in Table 4-2 and regenerated gland pictures are in Fig. 4-13, respectively).

Table 4-2. ***In vivo* validation experiment #1 of Cbx5 shRNAs in PKH26$^{Neg/low}$ cells**
Cells were cultured in adhesion prior to injection;

PKH26$^{Neg/low}$	Tot N of infected-cells transplanted	N of Infected cells injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
				Lower	Estimate	
Empty pRSI17	144.000	36.000	0 / 4	Inf	**Inf**	84.120
	108.000	18.000	0 /6			
pWPI MycER	112.000	28.000	1/ 4	298.886	**76.144**	19.398
	70.000	14.000	1/5			
pRSI17 Cbx5 sh11+15	70.800	23.600	1/ 3	246.737	**61.479**	15.319
	70.800	11,800	1 / 6			

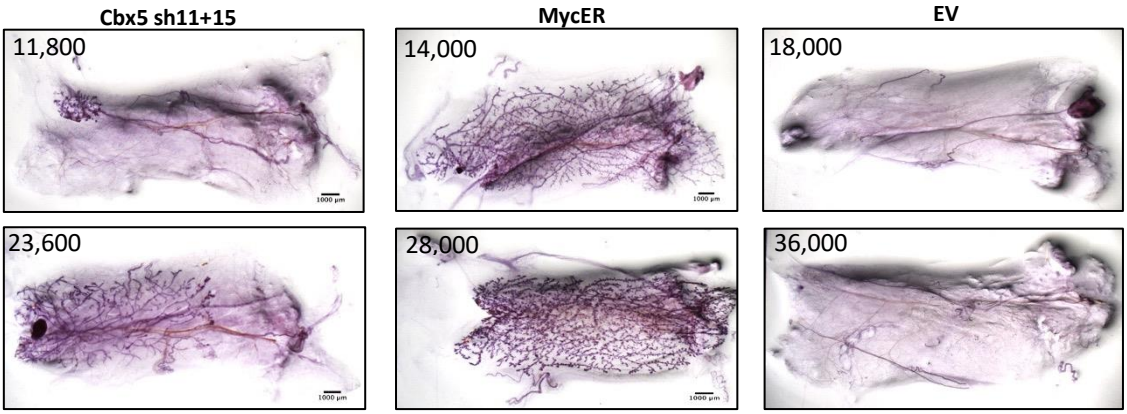


Figure 4-13. **Regenerated mammary glands from the Cbx5 *in vivo* validation experiment #1**

Two mammary glands regenerated from progenitors co-infected with Cbx5 sh11+15, 2 with MycER and 0 with EV infected cells

Experiment #2 confirmed that co-infection of Cbx5 shRNAs is able to reprogram PKH26$^{Neg/low}$ cells into SCs, as shown by regeneration of mammary gland upon transplantation. Notably, the gland size was comparable to the one obtained by injection of PKH26$^{Neg/low}$ cells infected with the positive control MycER-vector. As for the negative control, while 3 out of 4 mammary fat pads were injected with EV-infected cell did not lead to any outgrowths, as expected, in one case we observed a very small rudimentary outgrowth (see The 4-3 and Fig. 4-14).

Table 4-3. ***In vivo* validation experiment #2 of Cbx5 shRNAs in PKH26$^{Neg/low}$ cells**
Cells were cultured in adhesion prior to injection.

PKH26$^{Neg/low}$	tot N of infected-cells transplanted	N of Infected cells/ injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
				Lower	Estimate	Upper
Empty pRSI17	95,800	23,949	1 / 4	595.021	83.252	11.648
pWPI MycER	27.120	13,560	0 / 2	281.734	37.187	4.908
	13.560	6,780	1 / 2			
pRSI17 Cbx5 sh11+15	68.256	22,752	1 / 3	632.236	90.532	12.964
	34.128	11,376	0 / 3			

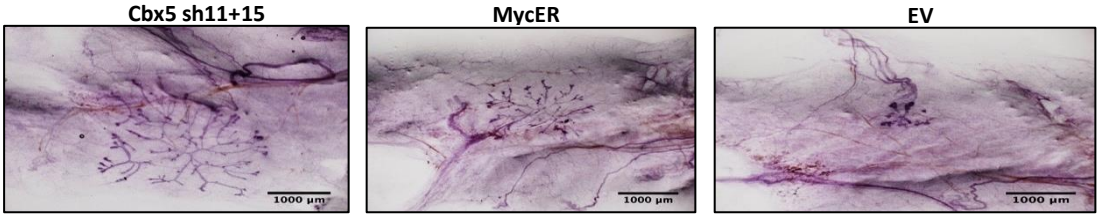


Figure 4-14. **Regenerated mammary glands from the Cbx5 *in vivo* validation experiment #2**

One mammary gland regenerated from progenitors co-infected with Cbx5 sh11+15, 1 with MycER and 1 small outgrowth with EV infected cells

Over all, two independent *in vivo* validation-experiments showed that Cbx5 down-regulation by shRNAs increases mammary repopulating frequency of mammary progenitors defined by $PKH26^{Neg/low}$ cells (Table 4-4).

Table 4-4. **Summary of experiments #1 and #2 for *in vivo* validation of Cbx5 with $PKH26^{Neg/low}$ cells**

$PKH26^{Neg/low}$	Total N of infected cells tested	Total N of Injection	N of regenerated mammary glands	SC frequency (ELDA)		
				Lower	Estimate	Upper
pRSI17 - Cbx5 sh11+15	**243,986**	15	3/15	220,232	**71,116**	22,995
pWPI - MycER	**222,680**	13	3/12	207,900	**65,664**	20,739
pRSI17 - EV	**347,800**	14	1/14	2,395,005	**335,679**	47,048

4.3.2 Cbx5/ HP1α down-regulation in primary mouse mammary luminal cells reprograms them into stem-like cells with regeneration ability

In addition to the PKH26-label retaining approach, we also tested the ability of Cbx5 shRNAs to reprogram freshly-isolated and uncultured primary mouse mammary luminal cells (LUM), purified by FACS-sorting from mammary epithelial cells as $Lin^{-}Cd24^{hi}Cd49f^{+}$ cells (Fig. 4-8). Purified Luminal cells (LUM, $Lin^{-}CD24^{hi}CD49f^{+}$) (Fig. 4-9) were infected with two Cbx5 shRNAs pooled or sh Luc as a negative control, and cultured for 3 days in adhesion on collagen-coated plates, prior to injection into cleared mammary fat pad of pre-pubertal female mice.

As expected, LUM cells infected with the control sh Luc did not show reconstitution ability, while the LUM cells infected with two Cbx5 shRNAs pooled regenerated a mammary gland tissue out of 8 injections, suggesting that Cbx5 down- regulation is indeed able to reprogram purified mammary luminal cells into SCs, though at a lower frequency (1 out of 106,342 cells, ELDA estimation), as compared to that observed with $PKH26^{Neg/low}$ progenitors (1 out of 71,116 cells, ELDA estimation, see Table 4-4). Further experiments will allow us to estimate more accurately the reprogramming

efficiency of LUM cells by Cbx5 down-regulation. The results are summarized in Table 4-5 and Figure 4-15.

Table 4-5. ***In vivo* validation experiment #3 of Cbx5 shRNAs in Luminal cells (Lin-CD24hiCD49f+)**
Cells were cultured in adhesion in collagen-coated plate prior to injection;

Luminal(Lin-$CD24^{hi}CD49f^{+}$)	**tot transplanted infected-cells**	**Infected cells/ injection side**	**Takes**	**SC frequency (ELDA)**		
				Lower	**Estimate**	Upper
Cbx5 sh11+sh15	**113.600**	**14.200**	1 / 8	756.030	**106.342**	14.958
Luc sh	**136.800**	**17.100**	0 / 8	-	-	45.665

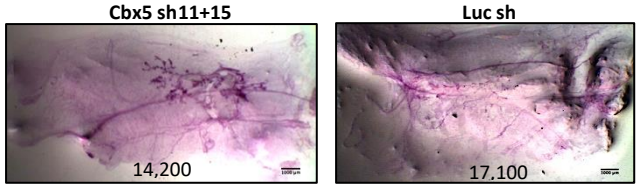


Figure 4-15. **Regenerated mammary gland from luminal progenitor cells infected with Cbx5 shRNAs, experiment #3**

One mammary gland regenerated from luminal cells co-infected with Cbx5 sh11+15, and one representative of EV infected negative control.

4.3.3 Cbx5 shRNAs independent effect on reprogramming of $PKH26^{Neg/low}$ progenitor cells

So far, the *in vivo* validation experiments of Cbx5 were performed as $PKH26^{Neg/low}$ cells or Luminal cells (LUM, $Lin^{-}CD24^{hi}CD49f^{+}$) were infected with pooled sh11+sh15 targeting Cbx5 (Table 4-4 and 5). Therefore, next, in order to rule out off-target effect of shRNAs, we performed two more *in vivo* validation experiments with $PKH26^{Neg/low}$ cells, infected individually with two shRNAs (Cbx5 sh11 or sh15). In this setting, we checked the down-regulation efficiency of shRNAs in both mRNA and protein level in the experiment #4. As shown in Figure 4-16, both Cbx5 sh constructs (sh11 and sh15) achieved similar Cbx5 mRNA (10% residual relative expression) and protein down-regulation.

As a result of *in vivo*, the Cbx5 sh15-infected $PKH26^{Neg/low}$ cells were able to regenerate mammary gland in 1 out of 8 injections, while Cbx5 sh11 did not (0/8). In this experiment we observed a small outgrowth from Luc sh-infected cells in 1/8 injections, as shown in Figure 4-17 and the experiment summary in Table 4-6.

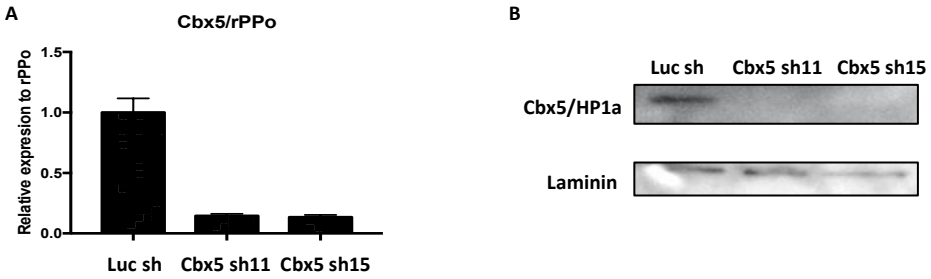


Figure 4-16. **Down-regulation of Cbx5 expression in primary MECs, by individual shRNA constructs**

A. RT-qPCR shows that both Cbx5 shRNAs achieve 90% Cbx5 mRNA reduction; B. Western blot analysis of the same cells;

Table 4-6. ***In vivo* validation experiment #4 of Cbx5 shRNAs individually infected to PKH26$^{Neg/low}$ cells**
Cells were cultured in suspension 3 days after infection;

PKH26$^{Neg/low}$	tot N of infected-cells transplanted	N of Infected cells/ injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
				Lower	**Estimate**	Upper
Empty pRSI17	140,428	17,531	**1/8**	933.378	**131.288**	18.467
pRSI17 Cbx5 sh11	142,760	17,845	**0/8**	Inf	**Inf**	47.655
pRSI17 Cbx5 sh15	185.845	23,231	**1/8**	1.236.855	**173.975**	24.471

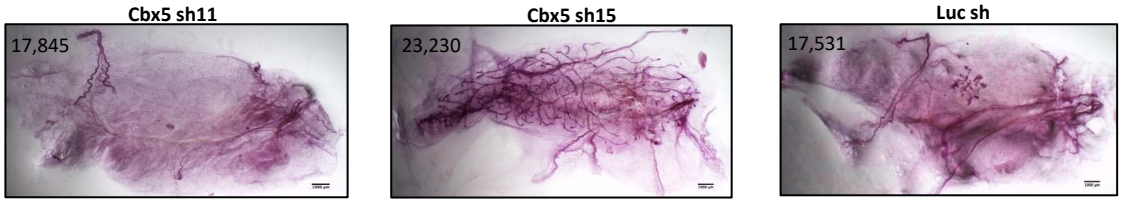


Figure 4-17. **Regenerated mammary glands from the Cbx5 *in vivo* validation experiment #4**

One mammary gland regenerated from Cbx5 sh15 infected progenitors, 0 with Cbx5 sh11 and 1 small outgrowth from EV infected cells

In the second experiment (experiment #5), we obtained an efficient reprogramming activity of Cbx5 sh15 in PKH26$^{Neg/low}$ cells: 3 out 8 regenerated mammary glands with Cbx5 sh15 and 0 out of 8 either with Cbx5 sh11 or Luc sh. Results are shown in Table 4-7 and Figure 4-18.

Table 4-7. ***In vivo* validation experiment #5 of Cbx5 shRNAs individually infected to PKH26$^{Neg/low}$ cells**
Cells were cultured in suspension 3 days after infection;

PKH26$^{Neg/low}$	tot N of infected-cells transplanted	N of Infected cells/ injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
				Lower	**Estimate**	**Upper**
Empty pRSI17	78,360	9,795	**0/8**	Inf	**Inf**	26.157
pRSI17 Cbx5 sh11	64,500	8,062	**0/ 8**	Inf	**Inf**	21.529
pRSI17 Cbx5 sh15	74,640	9,330	**3/ 8**	62.195	**19.851**	6.336

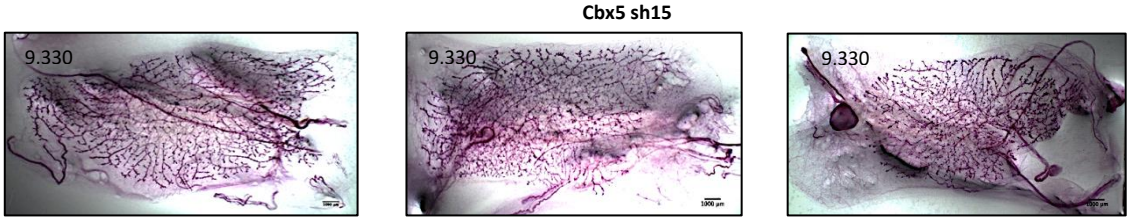


Figure 4-18. **Regenerated mammary gland regenerated from the Cbx5 *in vivo* validation experiment #5**

3 mammary glands regenerated from progenitor cells infected with Cbx5 sh15

In summary, Cbx5 sh15 was clearly able to increase the repopulating frequency of PKH26$^{Neg/low}$ cells, whereas Cbx5 sh11 did not (Table 4-8). Further experiments are ongoing to expand the number of injections with cells infected with Cbx5-shRNA.

In conclusion, down-regulation of Cbx5 by shRNAs in primary mouse mammary progenitors, purified by two different approaches, activates a reprogramming program, that is able to regenerate mammary gland upon transplantation.

Table 4-8. **Summary of experiments #4 and #5 for *in vivo* validation of individual Cbx5 shRNAs in PKH26$^{Neg/low}$ cells**

PKH26$^{Neg/low}$	Total N of infected cells tested	Total N of Injection	N of regenerated mammary glands	SC frequency (ELDA)		
				Lower	Estimate	Upper
pRSI17-Cbx5 sh11	207,300	16	**0/16**	Inf	**Inf**	69,184
pRSI17-Cbx5 sh15	260,500	16	**4/16**	160,387	**58,434**	21,290
pRSI17- Luc	218,800	1	**1/16**	1479901	**209,720**	29720

4.3.4 Cbx5/ HP1α down-regulation increases the conversion of Sca-1^{low} to Sca-1^{high} phenotype in CommaDβ cell line

CommaDβ cells, a normal murine mammary epithelial cell line, includes two different sub-populations in dynamic equilibrium: one expressing high levels the stem cell antigen-1 (Sca-1) and properties of mammary stem/progenitor cells (Sca-1^{high} cells), and the other expressing lower level of Sca-1 and considered as differentiated/luminal cells (Sca-1^{low} cells) (Deugnier, Faraldo et al. 2006). It was also reported that Sca-1^{low} cells can spontaneously convert into Sca-1^{high} cells in culture, thus providing a model to study plasticity of mammary progenitors. We investigated whether the Cbx5 and other genes identified in the shRNA screens could accelerate the spontaneous conversion of the Sca-1^{low} cells into Sca-1^{high} cells.

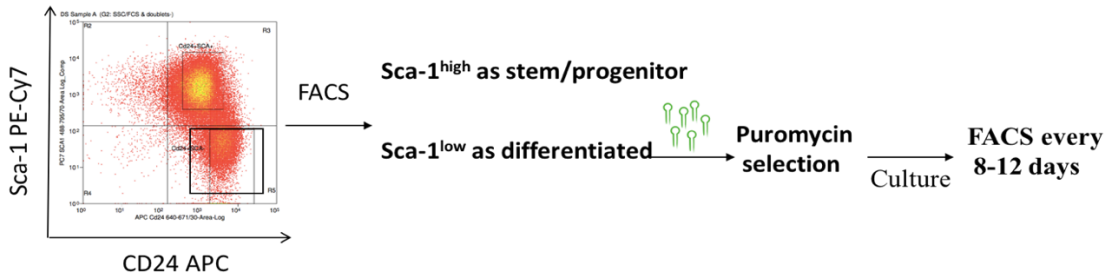


Figure 4-19. **Purification of CommaDβ Sca-1^{high} and Sca-1^{low} cells and experimental scheme of conversion assay from Sca-1^{low} to Sca-1^{high} phenotype**

To do so, we purified the Sca-1^{low} cells fraction by FACS sorting, as shown in Figure 4-19, and were infected with two Cbx5 shRNAs pooled or EV as a negative control. Infected cells were cultured in adhesion and selected with Puromycin, and their Sca-1 phenotype was checked post-infection at day 8 and 21 by FACS analysis, to monitor conversion into the Sca-1^{+} phenotype. The results showed an accelerated conversion of the Sca-1^{low} cells into Sca-1^{high} at two tome points we checked, as compared to controls (non-infected, NI, or infected with empty vector, EV; Fig. 4-20A). FACS analysis of Comma-Dβ Sca-1 positivity/negativity of the samples shown in Figure 4-20A, B, and down-regulation of Cbx5 mRNA is shown in Figure 4-20C.

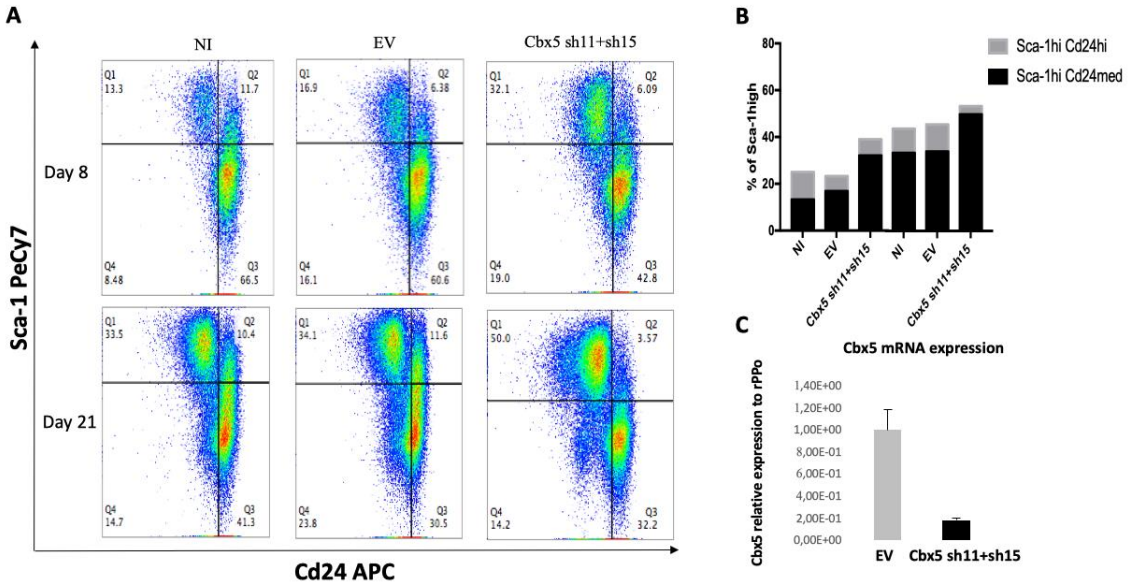


Figure 4-20. **Cbx5 down-regulation accelerates the conversion of Comma-Dβ cells from the Sca-1^{low} to the Sca-1^{high} phenotype**

A. FACS analysis of the Sca-1^{low} cells cultured as Not infected (NI), infected with Empty vector (EV) or infected with Cbx5 sh1+sh2; B. Bar-graph representation of the FACS data; C. RT-qPCR analysis of Cbx5 mRNA relative expression in Sca-1^{low} cells infected with Cbx5 sh1+sh2, or EV. RNA was prepared 8 days after infection and puromycin selection. Standard deviation is calculated from technical triplicates;

4.3.5 Transcriptomic regulation upon Cbx5 down-regulation in PKH26$^{Neg/low}$ cells, by RNA-seq analysis

In order to investigate the effect of Cbx5 down-regulation in PKH26$^{Neg/low}$ at transcriptomic level, while the *in vivo* validation experiment #4 and #5 were ongoing (before knowing the result), we performed gene expression profiling of PKH26$^{Neg/low}$ mammary progenitors infected with Cbx5 sh11 and sh15, respectively, both at 24h and 72h post-infection. Parallel samples infected with Luc sh were used as controls. Two biological replicates were prepared for each sample. RNA preparation and sequencing was performed as described in Material and Method (3.6.2). The pair-end sequenced data, after standard quality control checks, were analyzed with the bioinformatics procedures described in the Material and Method (3.8.1).

First, to validate the down-regulation efficiency of the two Cbx5 shRNAs (sh11 and sh15) in the RNA-seq experiments, we looked at the expression level of Cbx5 mRNA in tall samples at both at the 24h and 72h two-points. The two Cbx5 shRNAs showed comparable down-regulation efficiencies, as shown in Figure 4-21.

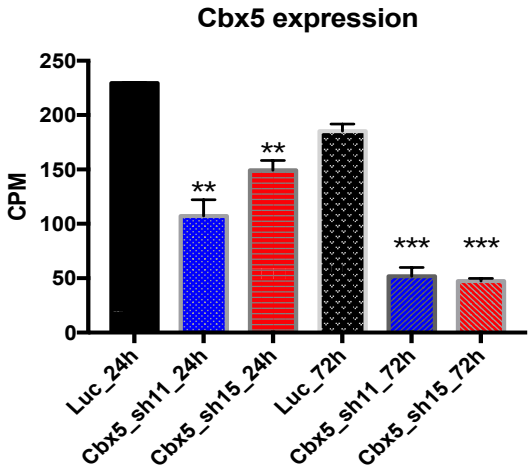


Figure 4-21. **Cbx5 expression in RNA-seq samples**

Error bars indicate the standard deviation of two biological replicates. Statistical significance is obtained by pairwise t-test of each Cbx5 interfered samples versus Luc sh within each time point is indicated, *P<0.05, **P<0.005, ***P<0.0005.

Differential expression analyses showed a small group of Differentially Expressed Genes (DEGs) in the Cbx5-shRNA11 datasets: 9 genes up-regulated (UP) and 7 genes down-regulated (DW) at 24h; 7 genes UP and 4 genes DW at 72h, with a *P*-value threshold set

below or equal to 0.05. Instead, a higher number of genes was regulated by Cbx5-shRNA15 at 72h (36 genes UP and 133 genes DW), while their number at 24h was comparable to those regulated by sh11 (10 UP and 5 DW; $p<0.05$). The number of DEGs concordantly regulated by both Cbx5 shRNAs was also relatively small: 5 DEGs UP and 3 DW at 24h, 1UP and 2DW at 72h (Fig. 4-22), suggesting different effects of these two shRNAs.

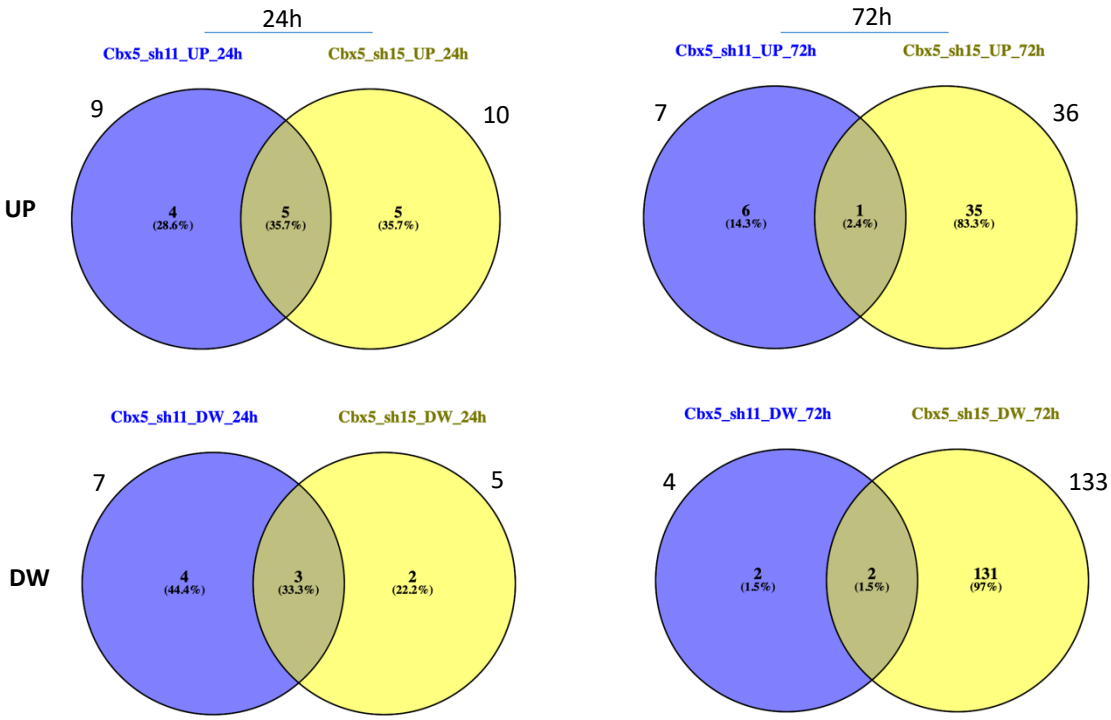


Figure 4-22. **Venn diagrams of genes regulated by Cbx5 sh11 and sh15 both at 24h and 72h**

Since the two shRNAs showed different biological effect in the *in vivo* validation experiments #4 and #5, we decided to focus our analysis on Cbx5 sh15, which achieves the most potent *in vivo* reprogramming efficiency (Table 4-8, and Fig. 4-18).

In the attempt to describe, at a global level, cellular states after Cbx5 down-regulation by "trends" of transcriptomic changes, we performed a Gene Set Enrichment Analysis (GSEA) comparing the transcriptional signature obtained in PKH26$^{Neg/low}$ cells infected with Cbx5 shRNA15 versus Luc sh control (see section 3.8.1).

Among the group of gene-sets available at the GSEA website (www.broadinstitute.org/gsea) in the Molecular signature database (MSigDB v5.0), the

"Hallmark" collection (50 gene-sets) summarizes and represents specific well-defined biological states or processes that display coherent expression. Challenging the gene regulated by Cbx5 sh15 (versus the Luc sh) with this collection, the GSEA analysis uncovered statistically significant enriched gene-sets both at 24h and 72h (Fig. 4-23). Notably, among them we found the "interferon alpha and gamma responses", "Oxidative phosphorylation" and "Myc targets" gene-sets, that are consistently enriched from 24h to 72h (Fig. 4-23A and D). All of them represent critical pathways which deserve further elaboration since they may provide insights about the molecular mechanisms involved in the reprogramming induced by Cbx5 down-regulation.

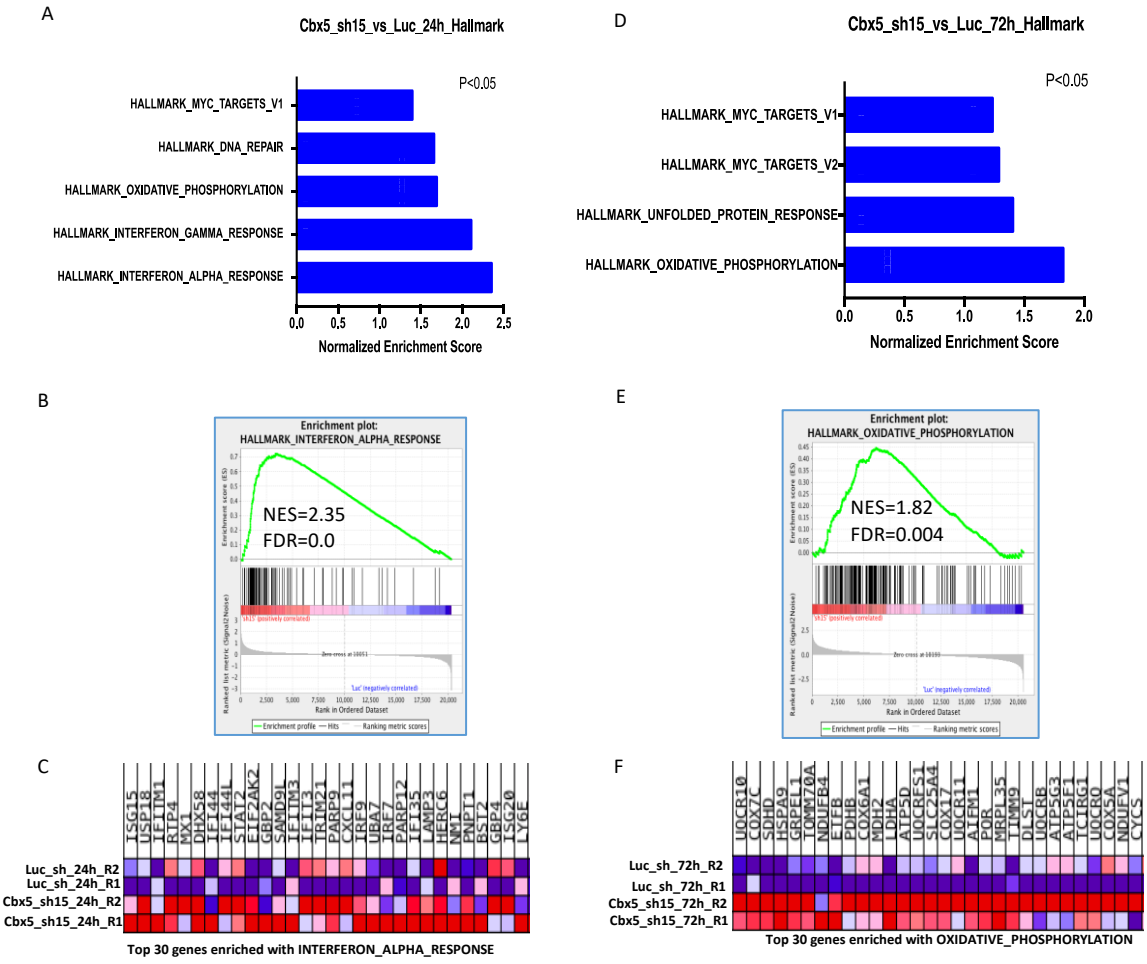


Figure 4-23. **Gene set enrichment analyses of RNA-seq data of PKH26$^{Neg/low}$ cells infected with Cbx5 sh15 or Luc sh, at both 24h or 72h after infection**

A and D. Gene-sets from the MSigDB's Hallmark collection, found significantly enriched in the GSEA analysis comparing Cbx5 sh15 sh versus Luc sh control samples and their corresponding NES, at 24h and 72h respectively; B and E. Enrichment plots showing the gene-sets most significantly enriched within the Cbx5 sh15_vs_Luc regulated genes, both at 24h and 72h; C and F. Heat maps of the corresponding top 30 leading edge genes. The average normalized expression values are indicated.

For each gene-set considered, the GSEA analysis also provides a heatmap of the "leading edge" subset of genes (Fig. 4-23 C and F), i.e. those contributing most to the enrichment to the gene-set. The expression of the leading edge genes in the two Cbx5 sh15 biological replicates showed good correlation at both 24h and 72h. We then performed GSEA analyses using selected published mammary stem cell gene-sets (e.g. LIM_MAMMARY_STEM_UP or SOADY_MAMMARY_STEM_UP gene-sets), and also with all stem cell gene-sets available in the MSigDB database. Notably, none of them was significantly enriched in our sample, suggesting that mammary progenitors have not yet acquired a SC signature at the two time points tested (24h and 72h) upon Cbx5 down-regulation.

Nonetheless, the coherent pattern of gene-set enrichment observed with Cbx5 sh15 from 24h to 72h, might provide mechanistic insights on the results we obtained from the *in vivo* validation experiments. We then asked whether Cbx5 sh11 could also similarly induce transcriptional regulation from 24h to 72h, and performed GSEA analysis of the Cbx5 sh11 transcriptome with the same gene-sets used for the Cbx5 sh15, and did not identify consistently enriched gene-sets (Fig. 4-24A and D). Moreover, the regulated genes from the two biological replicates of Cbx5 sh11 did not show a good correlation at either of the two time points analyzed (see corresponding heat maps in Fig. 4-24C and F). Finally, by checking other enriched gene-sets within the "Hallmark" collection (not significantly enriched ones) we concluded that the global effect of Cbx5 sh11is not coherent from 24h to 72h.

Collectively, transcriptomic analyses of Cbx5-silenced cells at 24h and 72h post-infection revealed regulation of a few pathways ("interferon alpha and gamma responses", "Oxidative phosphorylation" and "Myc targets" gene-sets) that deserve further investigations, and might be critical for the mechanism through which Cbx5 down-regulation endows mammary progenitors with SC properties.

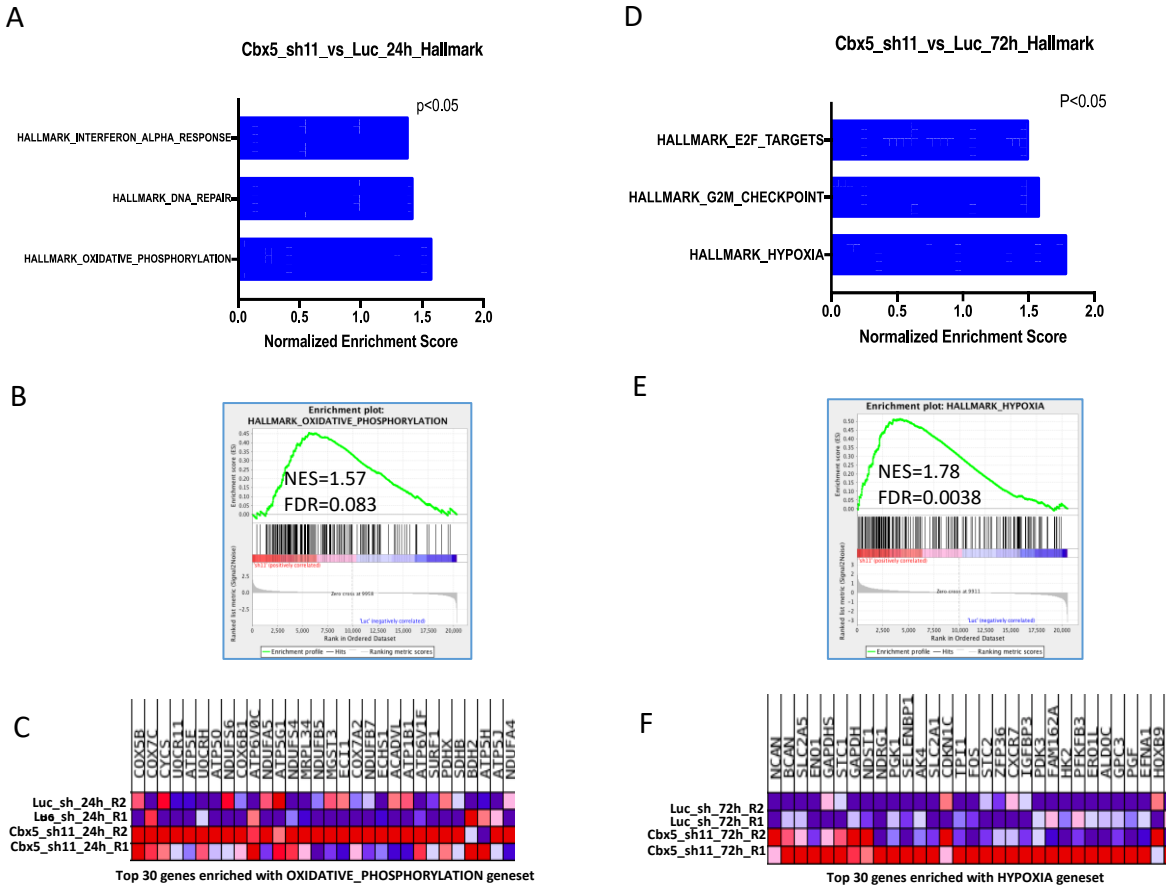


Figure 4-24. **Gene set enrichment analyses of RNA-seq data from PKH26$^{Neg/low}$ cells upon infection with Cbx5 sh15 and Luc sh, at both 24h or 72h**

A and D. Gene-sets from the MSigDB's Hallmark collection, found significantly enriched in the GSEA analysis comparing Cbx5 sh11 sh versus Luc sh control samples, and their corresponding NES, at 24h and 72h respectively; B and E. Enrichment plots showing the gene-sets most significantly enriched within the Cbx5 sh11_vs_Luc regulated genes, both at 24h and 72h; C and F. Heat maps of the corresponding top 30 leading edge genes, the average normalized expression values are indicated.

4.4 Validation of Kmt2d as an Inhibitor of Reprogramming of Mammary Progenitors into Mammary Stem Cells

In the mEpi2 shRNA library screens, one specific shRNA targeting Kmt2d was identified at high frequency (23% of total counts), in one transplanted gland. The identification of Kmt2d in the screen is in line with our working hypothesis that genes inhibiting reprogramming of differentiated cells might function as tumor suppressors. The down-regulation of Kmt2d expression induced reprogramming at high efficiency, according to the results of the *in vivo* validation experiments (Table 4-1).

For the first *in vivo* validation and *in vitro* self-renewal experiments (Table 4-9 and Fig. 4-26), we cloned and tested same shRNA identified in the screening, and 3 additional shRNAs selected from the 10 shRNAs included in the library for further validation experiments. Thus, to rule out off-target effects, we performed three independent *in vivo* validation experiments using purified primary mammary progenitor cells ($PKH26^{Neg/low}$) and three different shRNAs. As for all other validation experiments, infected mammary progenitors were transplanted into cleared mammary fat pads to evaluate their regeneration potential.

4.4.1 Kmt2d down-regulation induces the generation of mammary repopulating units from $PKH26^{Neg/low}$ progenitor cells

In the validations, we observed that the frequency of reprogramming was strongly affected by the culture conditions used to maintain progenitor cells before the *in vivo* regeneration assay. Thus, infected $PKH26^{Neg/low}$ progenitors were cultured either in adhesion or suspension conditions for 3 days prior to injection into cleared mammary fat pads.

After 12 weeks, we obtained positive transplants from Kmt2d sh1 infected progenitors, both cultured in both adhesion and suspension conditions, in the absence of any background outgrowth from the negative control (pRSI17 EV). Notably, the glands regenerated from the cells cultured in adhesion conditions were smaller than those obtained with the suspension cultures. Results are shown in Table 4-9 and Figure 4-25.

Table 4-9. **Summary of *in vivo* validation experiment #1 of Kmt2d shRNA in PKH26$^{Neg/low}$ cells**
Cells were cultured either in adhesion or suspension prior to injection;

PKH26$^{Neg/low}$	tot N of infected-cells transplanted	N of Infected cells/ injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
				Lower	Estimate	Upper
3 days adhesion culture prior to injection						
pRSI17 E.V.	139,200	19,900	0 / 6	-	-	46.642
		9,900	0 / 2			
pRSI17 Kmt2d sh1	57,320	14,327	**1 / 2**	342.675	**49.748**	7.222
		7,163	0 / 4			
pWPI MycER	183,280	14,139	0 / 3	-	-	18.879
		7,069	0 / 2			
3 days suspension culture prior to injection						
pRSI17 E.V.	6,800	3,400	0 / 2	-	-	2.270
pRSI17 Kmt2d sh1	8,360	4,180	**1 / 2**	44.747	**6.059**	821
pWPI MycER	14,500	7,225	**2 / 2**	28.546	**1**	1

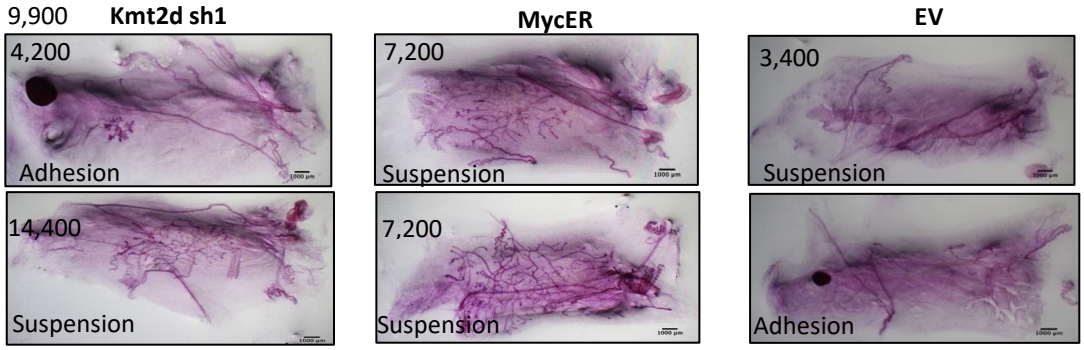


Figure 4-25. **Regenerated mammary gland from the Kmt2d *in vivo* validation experiment #1**

Cells were cultured in either adhesion or suspension prior to injection. 2 regenerated mammary glands from progenitor cells infected with Kmt2d sh1, 2 with MycER and 0 with EV.

4.4.2 Kmt2d down-regulation endows PKH26$^{Neg/low}$ cells with the ability to form self-renewing mammospheres

In order to test whether Kmt2d down-regulation induces self-renewal in PKH26$^{Neg/low}$ cells, we cultured cells in suspension as mammospheres, and re-plated them every 7 days after sphere disaggregation. To avoid immediate exposure to the harsh environment of suspension culture, and allow time for reprogramming we optimized the protocol as follows: after lentivirus infection in suspension conditions (16-17h), cells were plated in 5% Matrigel in the standard stem-cell medium for 7 days, cells were then recovered from Matrigel to start a standard self-renewal assay in suspension conditions (mammosphere assay) by serial re-plating of disaggregated mammospheres every 7 days.

This assay showed that Kmt2d sh1 enables PKH26$^{Neg/low}$ cells to grow as mammospheres in suspension, that can be serially passaged, as shown in Figure 4-26A, C, suggesting that

the PKH26$^{Neg/low}$ cells reprogrammed by Kmt2d sh1 are able to grow in suspension cultures and self-renew.

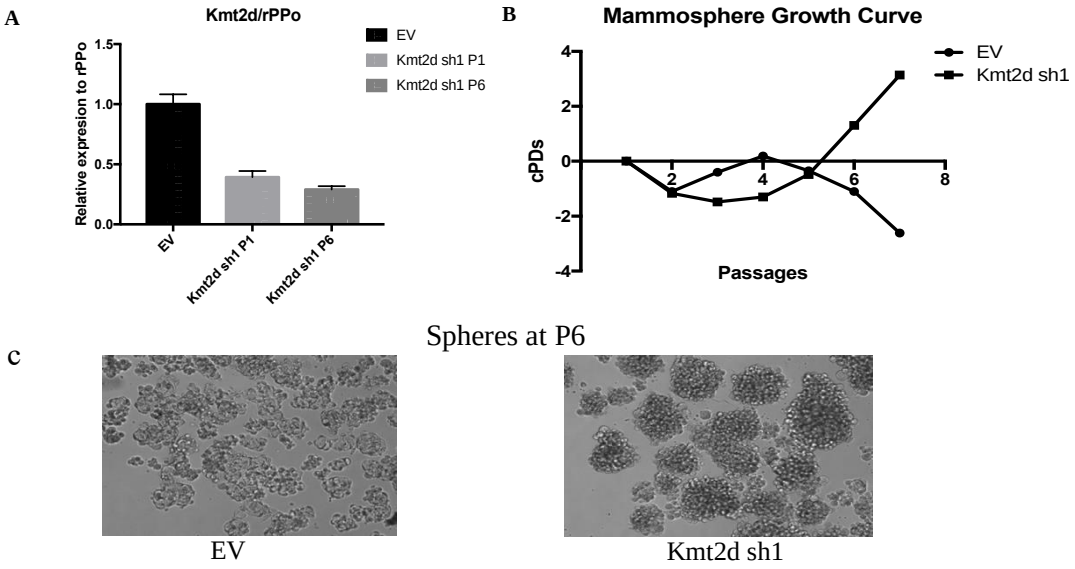


Figure 4-26. **Kmt2d sh1 induces mouse primary mammosphere self-renewal over serial passages**

A. Relative expression of Kmt2d in EV or sh-infected mammosphere at Passage1 and Passage 6 by RT-QPCR, standard deviation is calculated from technical triplicates; B. Cumulative population doublings of PKH26$^{Neg/low}$ progenitors infected with Kmt2d sh1 or EV by mammosphere assay; C. Representative pictures of Mammospheres in suspension culture at P6;

These first round of *in vivo* and *in vitro* experiments suggest that Kmt2d is a putative inhibitor of reprogramming of mammary progenitors into SCs.

4.4.3 Kmt2d shRNA's independent effect on reprogramming of PKH26$^{Neg/low}$ progenitor cells

In order to confirm the function of Kmt2d in SC reprogramming of mammary progenitors and rule out possible off-target effects of individual shRNAs, we tested additional Kmt2d shRNAs (sh2 and sh3) from the library. First, we tested the down-regulation efficiency of Kmt2d sh2 and sh3 in NMuMG cells, to select the most effective ones (Fig. 4-6A). Then, we performed a new *in vivo* validation experiment (experiment #2) using 2 shRNAs (sh1, the same one used in validation #1, and the new sh2 hairpin) each infected individually. The infected cells were cultured 3 days in suspension prior to injection. The result obtained from the *in vivo* validation experiment #2 confirmed that both shRNAs increase the frequency of mammary repopulating units as shown in the Table 4-10 and Figure 4-27. In this experiment, also the negative control (PKH26$^{Neg/low}$ cells infected with EV) generated one small outgrowth out of 9 injections.

Table 4-10. **Summary of *in vivo* validation experiment #2 of Kmt2d shRNAs in PKH26$^{Neg/low}$ cells**
Cells were kept in suspension prior to injection;

PKH26$^{Neg/low}$	tot N of infected-cells transplanted	N of Infected cells/ injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
				Lower	**Estimate**	Upper
pRSI17 E.V.	**97.000**	9.500	**1/9**	568,953	**80,054**	11,264
pRSI17 Kmt2d sh1	**68.400**	6.500	**3/10**	57,518	**18,440**	5,912
pRSI17 Kmt2d sh2	**66.750**	6.500	**2/10**	115,154	**28,717**	7,161

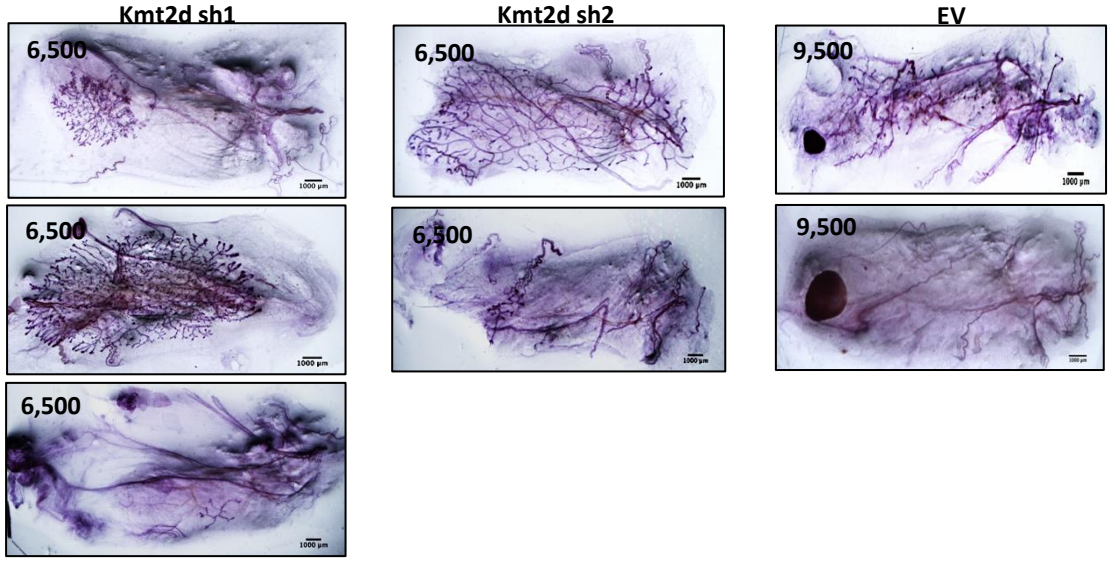


Figure 4-27. **Regenerated mammary gland from the Kmt2d *in vivo* validation experiment #2**

3 Regenerated mammary glands from progenitors infected with Kmt2d sh1 in PKH26$^{Neg/low}$ cells, 2 (1 is small) with Kmt2d sh2 and 1 small outgrowth with EV infected cells

This *in vivo* validations experiment with two different shRNAs targeting Kmt2d, further confirmed that Kmt2d down-regulation reprograms mammary progenitors (PKH26$^{Neg/low}$) into MRUs able to regenerate normal mammary glands.

During the validation of other hits, we observed that when injecting the day after infection, shRNA-infected PKH26$^{Neg/low}$ cells scored higher regeneration rates, as compared to the cells kept in short-term cultures (3 days), in either adhesion culture or suspension. Therefore, we performed one more *in vivo* validation experiment (#3) with Kmt2d shRNAs, to assess whether the length of pre-culture prior to injection affects reconstitution efficiency.

In the *in vivo* validation experiment #3, purified PKH26$^{Neg/low}$ cells were infected with Kmt2d sh2 or another new shRNA (sh3) targeting Kmt2d, or the negative control Luc sh. The *in vivo* regeneration assay was performed at 1 or 3 days after infection. In this experiment, both Kmt2d sh2 and sh3 conferred *in vivo* regeneration ability to mammary progenitors, and confirmed our previous observation that regeneration rate is higher if

cells are injected the day after infection, as shown in in Figure 4-28 and Table 4-11. As in the previous experiment (experiment #2), we obtained also a very small outgrowth from Luc sh infection (Fig. 4-28).

Table 4-11. **Summary of *in vivo* validation experiment #3 of Kmt2d shRNAs in PKH26$^{Neg/low}$ cells**
Cells were cultured in suspension prior to injection;

PKH26$^{Neg/low}$	Infection	tot N of infected-cells transplanted	N of Infected cells/ injected/side	N of takes/ Tot injection	SC frequency (ELDA)		
					Lower	Estimate	Upper
Injection at Day1	pRSI17 Luc	108.000	13.500	**1/8**	718.761	**101.100**	14.221
	pRSI17 Kmt2d sh2	99.600	12.450	**2/8**	173.870	**43.277**	10.772
	pRSI17 Kmt2d sh3	102.000	12.750	**1/8**	678.830	**95.483**	13.431
Injection at Day3	pRSI17 Luc	70.200	17.550	0/4	-	-	23.433
	pRSI17 Kmt2d sh2	55.100	13.800	**1/4**	342.305	**47.893**	6.701
	pRSI17 Kmt2d sh3	68.000	17.000	0/4	-	-	22.699

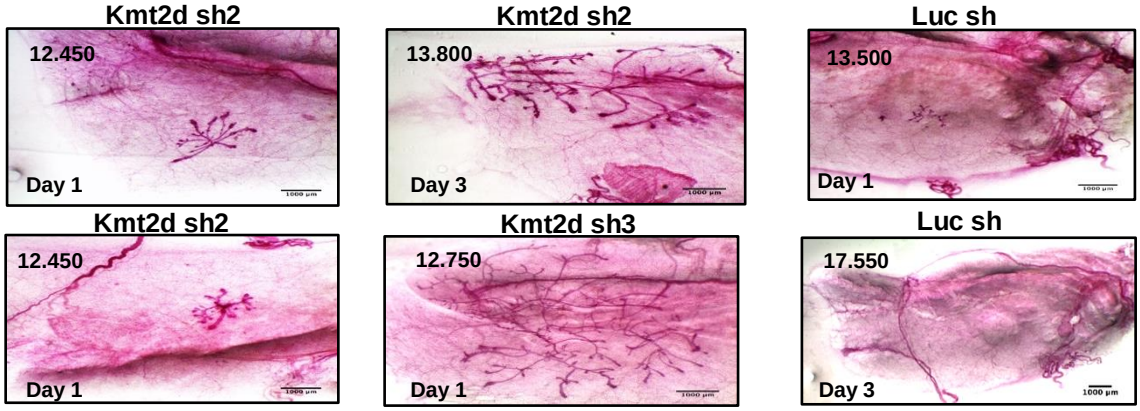


Figure 4-28. **Regenerated mammary glands from the Kmt2d *in vivo* validation experiment #3**

3 mammary glands regenerated from progenitors infected with Kmt2d sh2 in PKH26$^{Neg/low}$ cells as indicated day, 1 with Kmt2d sh2, and 1 very small outgrowth with Luc infected cells

In conclusion, we confirmed that Kmt2d down-regulation-achieved by three different individual shRNAs-induced the generation of mammary repopulating units from mammary progenitors (PKH26$^{Neg/low}$). A summary of all the *in vivo* validations of Kmt2d is shown in Table 4-12.

Table 4-12. **Summary of *in vivo* validations experiments #1, #2 and #3 of Kmt2d shRNAs with PKH26$^{Neg/low}$ cells**

PKH26$^{Neg/low}$	Total N of infected cells tested	Total N of Injection	N of regenerated mammary glands	SC frequency (ELDA)		
				Lower	Estimate	Upper
pRSI17-EV or Luc	439,100	31	2/31	655,420	**202,303**	62,443
pRSI17-Kmt2d sh1	198,750	18	5/18	53,813	**22,200**	9,158
pRSI17-Kmt2d sh2	221,600	22	5/22	92,749	**38,357**	15,862
pRSI17-Kmt2d sh3	170,000	12	1/12	1,167,582	**163,542**	22,907

4.4.4 Kmt2d down-regulation increases the conversion of Sca-1^{low} to Sca-1^{high} phenotype in CommaDβ cell line

Then, we tested whether Kmt2d inhibition could increase cell plasticity in the mouse mammary epithelial CommaDβ cell line, measured as the conversion of the Sca-1^{neg} to Sca-1^{high} phenotype, as described in chapter 4.4.3. FACS sorted Sca-1^{neg} CommaDβ cells were infected with Kmt2d sh1, selected with Puromycin and their Sca-1 phenotype was assessed by FACS analysis at different time points. Indeed, Kmt2d down-regulation accelerates the conversion of Sca-1^{neg} to Sca-1^{high} phenotype. The FACS analysis is shown in Figure 4-29A and B and the RT-qPCR analysis showing Kmt2d mRNA down-regulation in the same experiment is shown in Figure 4-29C

Altogether, these results on the effect of Kmt2d inhibition on the plasticity of normal mouse mammary cells suggest that Kmt2d plays a function in the control of cellular identity in physiological conditions. Therefore, we decided to further investigate the mechanisms through which Kmt2d exerts as an inhibitor of cell plasticity in physiological conditions.

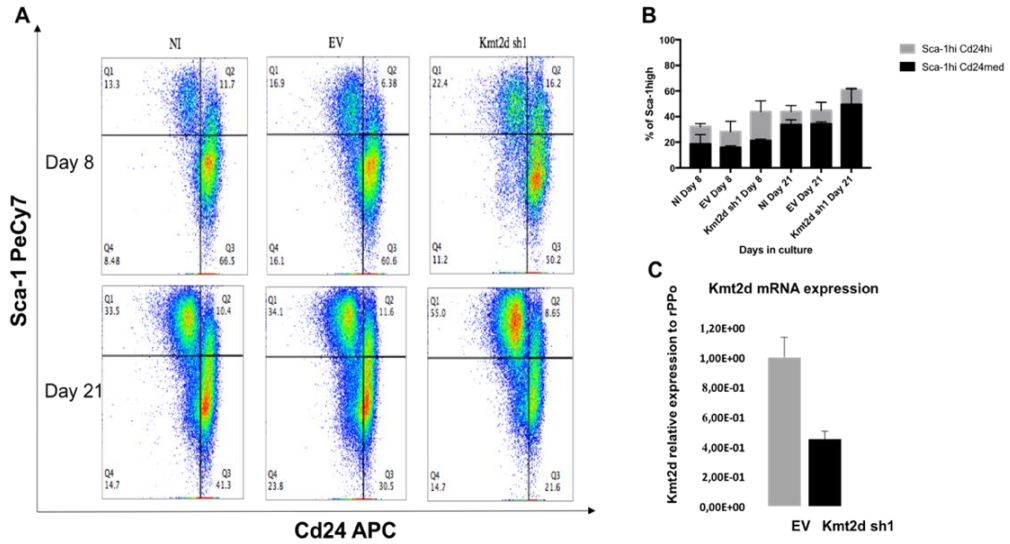


Figure 4-29. **Kmt2d down-regulation accelerates the conversion of Comma-Dβ cells from the Sca-1^{low} to the Sca-1^{high} phenotype.**

A. FACS analysis of Sca-1^{low} cells cultured as Not infected (NI), infected with Empty vector (EV) or infected with Kmt2d sh, one representative experiment is shown; B. Bar-graph showing the fraction of Sca-1^{high} cells, at day 8 and 21after infection and selection with the indicated constructs, standard deviation derived from two independent experiments. C. RT-qPCR analysis of Kmt2d mRNA relative expression in Sca-1^{low} cells infected with Kmt2d sh1, EV, or Kmt2d shRNA. RNA was prepared 8 days after infection and puromycin selection. Standard deviation is calculated from technical triplicates;

4.4.5 Transcriptomic regulation upon Kmt2d knockdown in PKH26$^{Neg/low}$ cells, by RNA-seq analysis

In order to uncover the molecular mechanisms involved in the control of cellular plasticity by Kmt2d, we further investigated the transcriptomic changes induced PKH26$^{Neg/low}$ mammary progenitor cells, upon Kmt2d down-regulation. We infected PKH26$^{Neg/low}$ cells with Kmt2d sh1 or sh2, and Luc sh as control, and performed RNA-seq at 24h and 72h after infection. The DEG analysis showed a strong correlation between the genes regulated by the two Kmt2d shRNAs. In fact, 40%-60% of the significantly regulated genes are concordantly regulated by the two shRNAs at each time point (Fig. 4-30). Surprisingly, the Kmt2d sh1 showed a lower down-regulation efficiency as compared to sh2, as shown in Figure 4-31. This maybe due to the fact that Kmt2d mRNA is one of the longest transcripts in the genome, thus led to artifacts during the RNA-seq procedure or bioinformatics analyses, since we have showed extended knockdown efficiencies of these two shRNAs by RT-qPCR analyses (Fig. 4-6 and 4-26A).

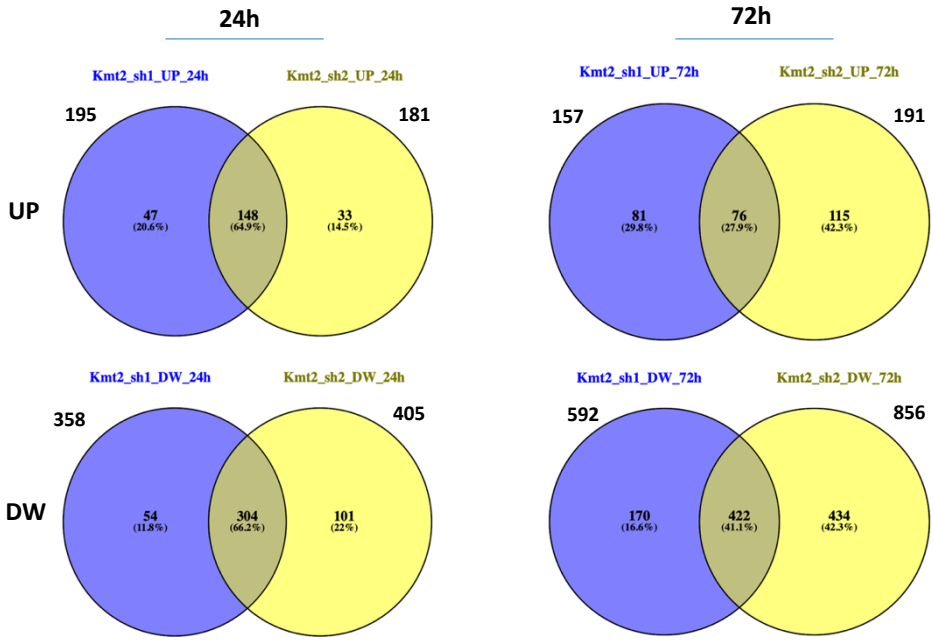


Figure 4-30. **Venn diagrams of genes regulated by Kmt2d sh1 or sh2, both at 24h or 72h**

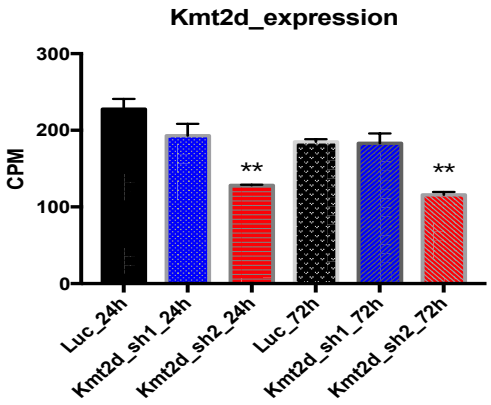


Figure 4-31. **Kmt2d expression in RNA-seq samples**

Error bars indicate standard deviation between the two biological replicates. Statistical significance is obtained by pairwise t-test of each Kmt2d interfered samples versus Luc sh. **P*<0.05, ***P*<0.005.

Since the two Kmt2d shRNAs seem to share highly overlapping gene regulation, we decided to go ahead with the DEG analysis for both time points (24h and 72h), considering the RNA-seq data generated by both Kmt2d shRNAs as one treatment (as 4 replicas of the same sample) and comparing them to Luc sh. Following this approach and using as cutoff p<=0.05 we found: 300 genes UP and 650 genes DW at 24h; 301 genes UP and 905 genes DW at 72h, upon Kmt2d interference, as shown in Table 4-13.

Table 4-13. **Number of genes differentially expressed upon Kmt2d down-regulation in PKH26$^{Neg/low}$ cells**

DEGs at 24h (P-value <= 0.05)		DEGs at 72h (P-value <= 0.05)	
N of UP-regulated genes	**N of Down-regulated genes**	**N of UP-regulated genes**	**N of Down-regulated genes**
300	**650**	**301**	**905**

To describe, at a global level, the cellular states induced Kmt2d down-regulation in PKH26$^{Neg/low}$ cells by “trends” of transcriptomic changes, we performed a GSEA by challenging the transcriptional signature obtained in PKH26$^{Neg/low}$ infected with Kmt2d shRNAs versus Luc sh control with the “Hallmark collection”, as we previously described for Cbx5 (4.3.5). The results are shown in Figure 4-32.

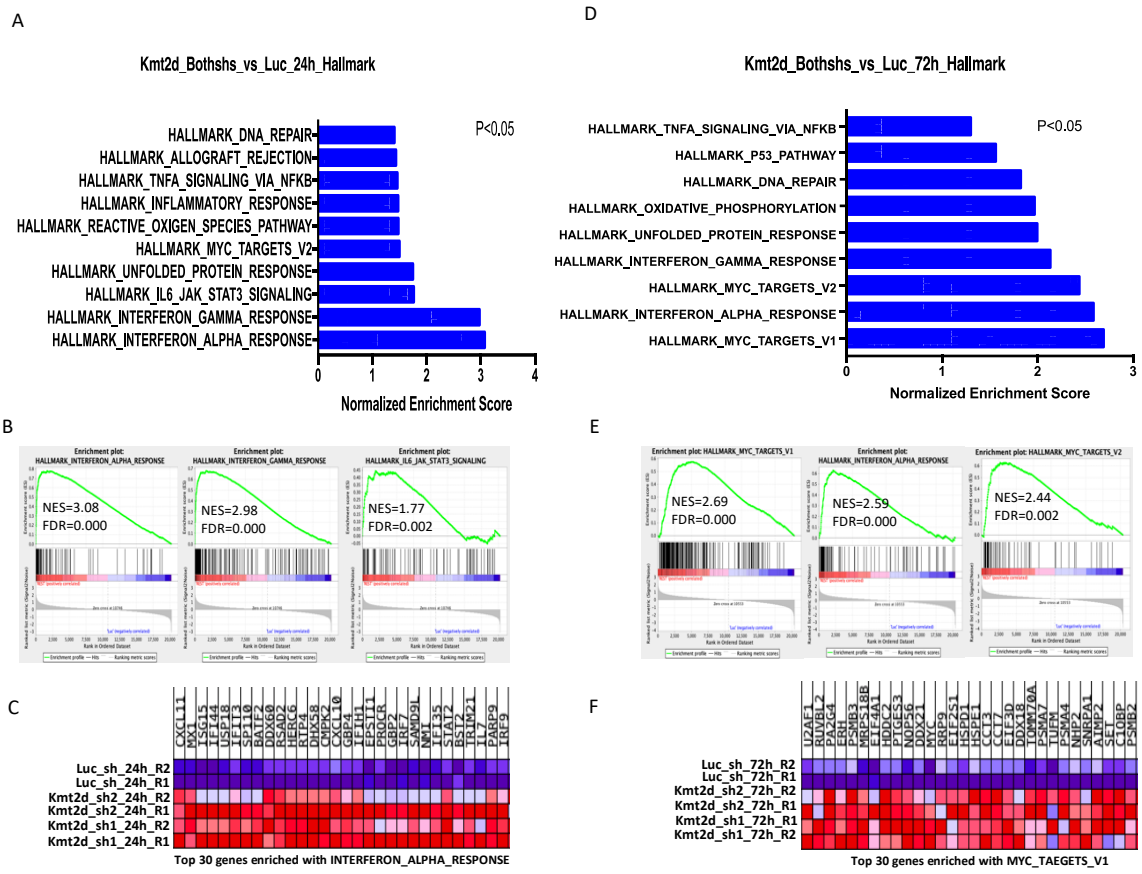


Figure 4-32. **Gene set enrichment analyses of RNA-seq data from PKH26$^{Neg/low}$ cells upon infection with Kmt2d two shRNAs and Luc sh, at both 24h or 72h**

A and D. Gene-sets from the MsigDB's Hallmark collection, found significantly enriched in the GSEA analysis comparing Kmt2d both shRNAs together versus Luc sh control samples and their corresponding NES, at 24h and 72h, respectively; B and E. Enrichment plots of the most significantly enriched gene-sets with the regulated genes in Kmt2d_bothshs_vs_Luc, both at 24h and 72h, respectively; C and F. Heat maps of the corresponding top 30 leading edge genes, the average normalized expression values are indicated.

Surprisingly, the most significantly enriched gene-sets in Kmt2d shRNAs *vs* luc were largely overlapping with those identified in Cbx5 sh15 *vs* sh luc GSEA (Fig. 4-23). In particular, the interferon alpha and gamma responses (24h) and myc targets v1, v2 (72h) and oxidative phosphorylation (72h). These results might suggest the induction of a common transcriptomic responses in the early phases of reprogramming, by the Kmt2d and Cbx5 shRNAs. This hypothesis will be addressed by further experiments.

We then checked whether Kmt2d shRNA expression in PKH26$^{Neg/low}$ cells might affect also Cbx5 mRNA level. Indeed, in the RNA-seq data, Cbx5 expression was significantly down-regulated at 72h, and also, though non-significantly, at 24h after Kmt2d sh

infections, as shown in Figure 4-33. This preliminary observation suggests us that Cbx5 might be a downstream target of Kmt2d.

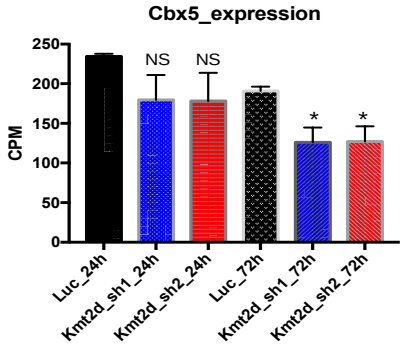


Figure 4-33. **Cbx5 mRNA level in the RNA-seq samples from Kmt2d shRNAs expressing cells**

Error bars indicate standard deviation of two biological replicates. Statistical significance is obtained by pairwise t-test of each Kmt2d interfered samples versus Luc sh. **P*<0.05;

These observations led us to search for common genes regulated by Kmt2d and Cbx5, by leading-edge analysis. This analysis allows to define, within the enriched gene-sets, the non-redundant common genes which most contributed to the enrichment ($p<0.05$). We set stringent criteria (the gene should be present at least in three gene-sets or more), to obtain reliable gene lists that could suggest possible mechanisms underlying cell reprogramming by Kmt2d and Cbx5 down-regulation. Interestingly, this analysis revealed a concordant regulation of several chemokines and cytokines (Cxcl11, Cxcl10, Cxcl9, Ccl5, Ccl7, Il6, Il7, Il15ra, Il2rb), the IL-6 signaling pathway and Myc target genes, by Kmt2d or Cbx5 down-regulation. These genes and their secreted protein products will be further investigated to understand their possible involvement in a common control mechanism of cellular plasticity, shared by Kmt2d and Cbx5.

To validate the RNA-seq data, we checked IL-6 mRNA expression upon Kmt2d down-regulation, in the normal mammary epithelial cell lines CommaDß (mouse) and MCF10A (Human). Indeed, in both cell lines, Kmt2d down-regulation leads to increased IL-6 mRNA, as shown in Figure 4-34. Notably, IL-6 was reported to be involved in reprogramming of fibroblasts into iPSCs (Brady, Li et al. 2013)).

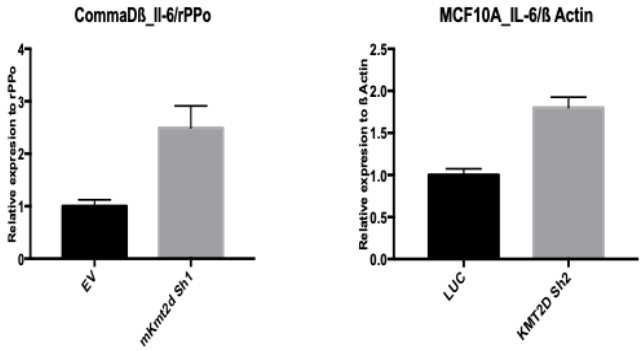


Figure 4-34**. IL-6 mRNA induction upon Km2d down-regulation, in mammary epithelial cell lines CommaDß (mouse) and in MCF10A (Human)**

Furthermore, since the down-regulation of either Cbx5 or Kmt2d led to significant enrichment of Myc_Targets_v1 and v2, we investigated whether some DEGs are shared with those regulated during Myc-induced reprogramming of PKH26$^{Neg/low}$ cells (Santoro, Vlachou et al. in press). To do so, we took advantage of the DEG dataset of Myc-regulated genes generated in our lab (Santoro, Vlachou et al. in press), in mammary PKH26$^{Neg/low}$ cells isolated form the mice expressing an inducible MycER transgene from the Rosa26 locus. Briefly, in this experiment, RNA-seq was performed 10 days after induction of MycER expression, and led to the identification of 2722 Up regulated and 2666 Down-regulated genes ($p<0.05$), called PKH26neg_MycER_DEGs_10days dataset. Interestingly, we found significant overlapping between the genes concordantly regulated by Kmt2d and MYC, as shown in Figure 4-35, suggesting that shared gene subsets may be involved in reprogramming of PKH26$^{Neg/low}$ cells into SCs.

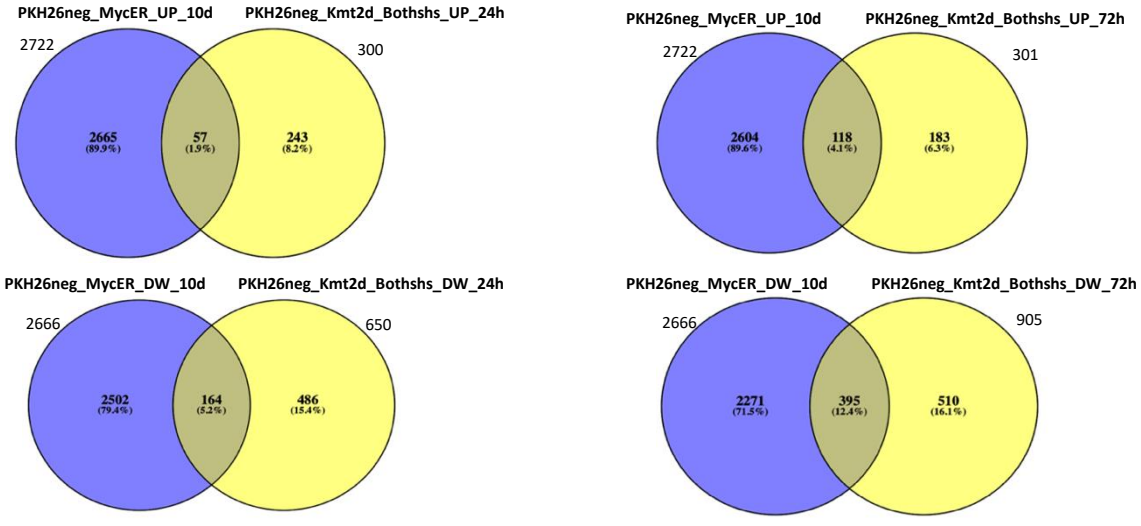


Figure 4-35. **Venn diagrams of differentially expressed gene number in the indicated samples**

We then performed GSEA analysis using the transcriptional signatures obtained from PKH26$^{Neg/low}$ cells infected with Kmt2d shRNAs (versus Luc sh control) and the PKH26neg_MycER_DEGs_10day dataset as a target gene-set. This analysis showed a coherent significantly enrichment of Myc-regulated genes, both at 24h and 72h, as shown in Figure 4-36. We then performed the same analysis to identify enrichment of Myc reprogramming targets within the Cbx5 sh15-regulated transcriptome. Also Cbx5-regulated genes showed a significant enrichment for MYC-regulated genes identified during reprogramming of mammary progenitors.

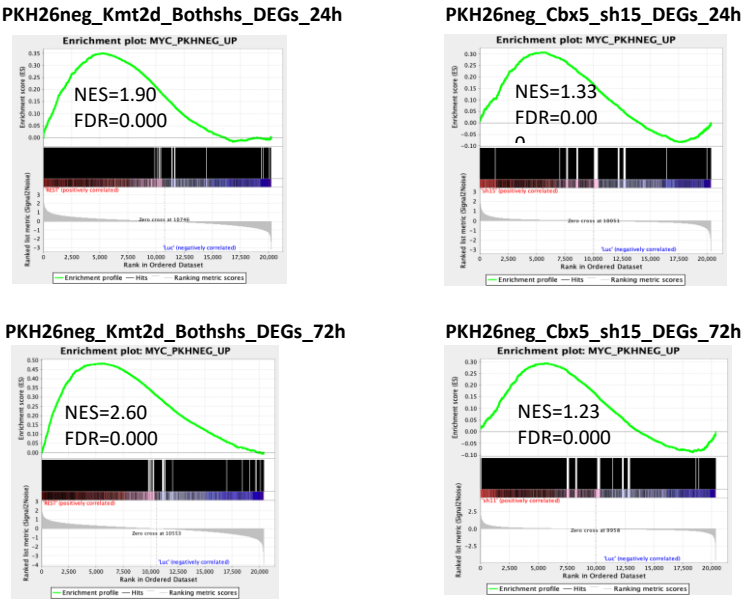


Figure 4-36. **Enrichment analysis of Kmt2d shs or Cbx5 sh15 infected progenitors showing enrichment of Myc-reprogramming signature**

Enrichment plots of Myc-reprogramming signature with PKH26$^{Neg/low}$ cells expressing Kmt2d shs (left panel), or Cbx5 sh15 (right panel) at the indicated time points. Normalized enrichment score (NES) and False discovery rate (FDR) are shown in each panel.

Collectively, analyses of transcriptomes during mammary progenitor reprogramming into stem-like cells suggest a common early transcriptomic responses of PKH26$^{Neg/low}$ cells to the down-regulation of Kmt2d or Cbx5, also largely shared with the transcriptomic changes induced during reprogramming by de-regulated MYC expression. To better understand and characterize the reprogramming events in PKH26$^{Neg/low}$ cells induced by either Kmt2d or Cbx5 down-regulation, we will investigate transcriptome regulation at later time points.

4.4.6 KMT2D down-regulation increases sphere forming efficiency in MCF10A cells

The promising result obtained from both primary mouse mammary epithelial cells and cell lines prompted us to investigate the function of KMT2D in human mammary epithelial cells. We took advantage of a well characterized human normal mammary epithelial immortalized cell line, MCF10A, to test whether KMT2D down-regulation by shRNAs could increase sphere forming efficiencies, suggesting induction of a stem cell phenotype. We performed two independent experiments of mammosphere assay. In the first experiment, we tested 2 shRNAs targeting KMT2D or CBX5, and pWPI_MycER over-expression and TET2 shRNAs (Song, Poliseno et al. 2013) as positive controls, Luc and pWPI_EV as negative controls. Infected-MCF10A cells were plated at a density of 1500cells/well in 1ml of stem medium in 24-well low attachment-plates, and every 5-7 days spheres were counted and re-plated as single cells for following passages. Spheres at P2 are shown in Figure 4-37.

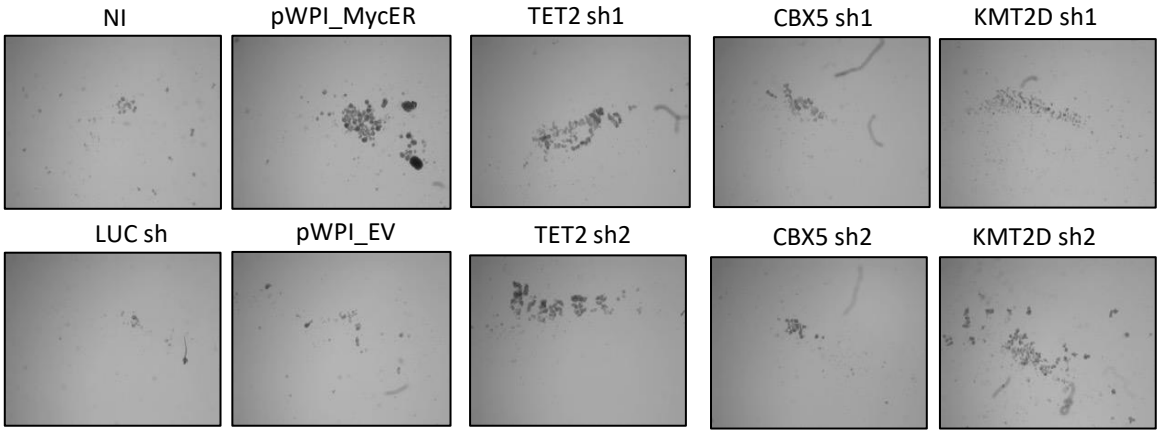


Figure 4-37. **Mammospheres formed by MCF10A cells**

Representative pictures of MCF10A mammospheres infected by different shRNAs targeting the indicated genes, at P2, 2x magnification.

The results are shown in Figure 4-38A. As expected, both MycER over-expression and TET2 shRNAs led to increased sphere numbers throughout serial passages, whereas both negative controls (Luc sh and pWPI_EV) did not. Both KMT2D shRNAs (sh1 and sh2) led to increased sphere numbers through passages, but the sh2 was more efficiently increasing sphere numbers as compared to sh1. Notably, this well correlates with the

corresponding efficiencies of KMT2D down-regulation assessed by RT-qPCR (Fig. 4-38B). Either CBX5 shRNAs were unable to significantly increase the sphere numbers that were comparable to pWPI_EV- infected control.

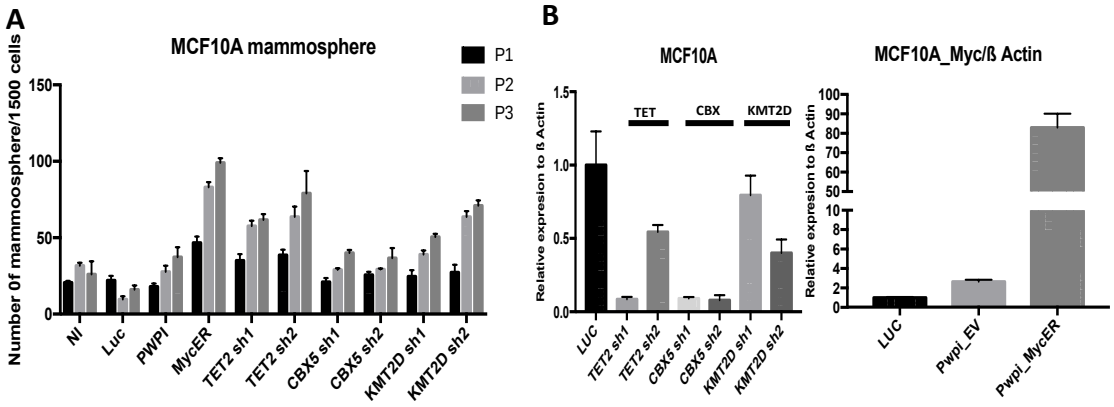


Figure 4-38. **1st MCF10A mammosphere assay upon down-regulation of shRNA-targeted genes and MycER over-expression**

A. Number of MCF10A mammospheres through three serial passages; B. RT-QPCR analyses of relative expression of indicated genes targeted by shRNAs and MycER over expression in MCF10A;

We repeated the mammosphere assay for confirmation in a 2nd experiment, plating the cells in 12-well plates this time. In 24-well plates, cells tend to grow in the periphery of the wells, where suspension condition is somehow altered due to the proximity to the walls of wells. The results obtained from the 2nd experiment are shown in figure 4-39. The sphere formation efficiencies were slightly lower than in the 1st experiment. This could be due to the lower cell density at plating in a larger well, and / or the lower formation of spheres in periphery in 12-well plate. Still, both KMT2D shRNAs showed higher sphere forming efficiencies than the NI and Luc controls, and a slight increase through passages. KMT2D Sh2 might be more efficient than sh1 in this assay, but more experiments would be required to confirm this observation. Again, neither of the two CBX5 shRNAs tested increased sphere formation in MCF10A.

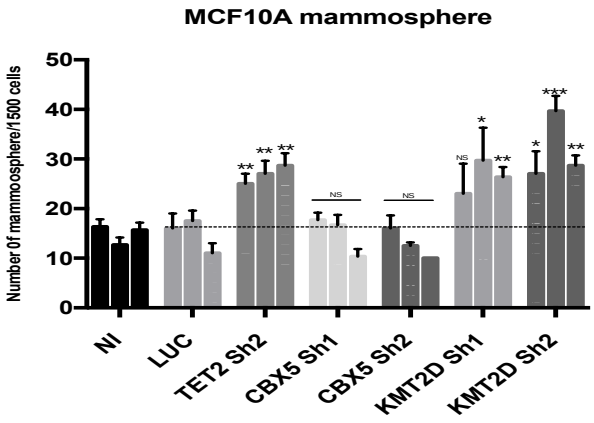


Figure 4-39. **2nd MCF10A mammosphere assay**

A. MCF10A mammospheres by serial re-plating until 3rd passages; The data are represented as mean +_SD from n=3 experimental replicates. *P* value by multiple t-test to LUC sh, *P<0.05, **P<0.005, ***P<0.0005.

These two experiments confirmed that KMT2D down-regulation increases MCF10A sphere initiating cells, similar to its effect in primary mouse mammary progenitors. Therefore, we are planning to use this cell line to perform ChIP-seq on KMT2D and histone modifications (H3K4me3, H3K4me1, H3k27ac), upon the expression of KMT2D sh2 or Luc sh control, to integrate the binding and transcriptomic data, to uncover the mechanism underlying the function of KMT2D inhibiting the cellular plasticity.

4.4.7 KMT2D down-regulation induces an EMT-like phenotype in the MCF10A cell line

In addition to the induction of a self-renewal phenotype, KMT2D down-regulation in MCF10A induced phenotypic changes when cells were cultured in adhesion. MCF10A cells slowed down their proliferation upon infection with KMT2D shRNA, and the morphology of cells in adherence as shown in Figure 4-40A suggests that cell-cell contacts were lost, as compared to Luc infected cells, consistent with the acquisition of an EMT-like phenotype. We then tested the expression of some EMT markers, by RT-qPCR. KMT2D shRNA infected cells showed increased ZEB1 expression, while E-Cadherin mRNA was slightly down-regulated, as shown in Figure 4-40B.

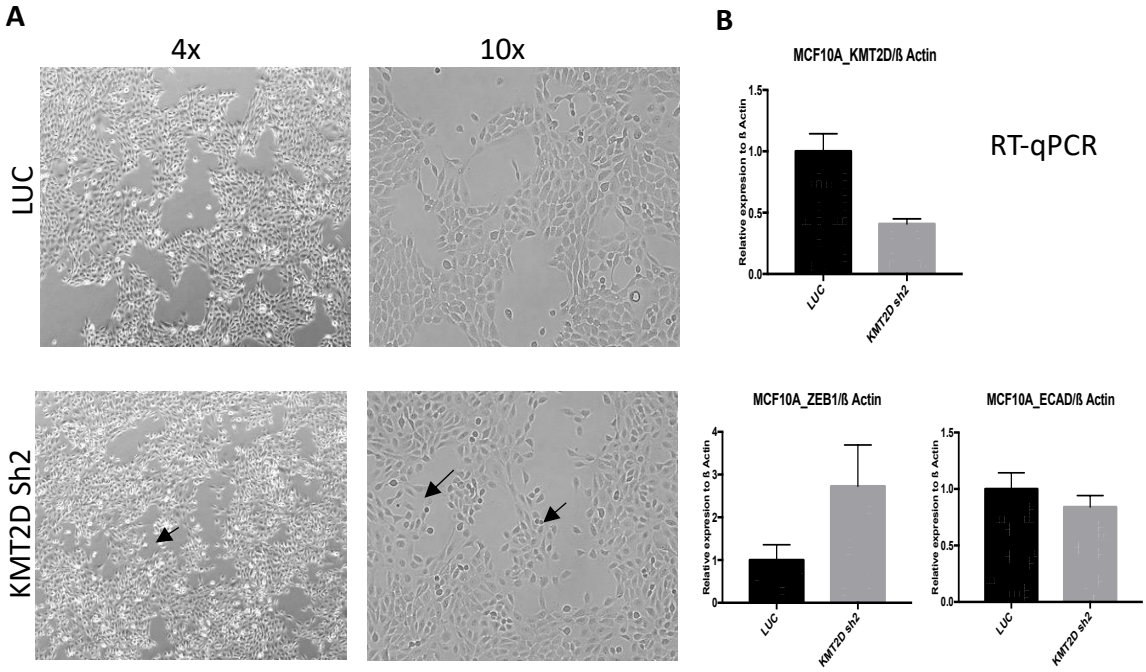


Figure 4-40. **KMT2D sh2 induces an EMT-like phenotype in MCF10A cells**

A. MCF10A cells infected with LUC sh or KMT2D sh2 cultured in adhesion, black arrow shows the areas of mesenchymal-like phenotype; B. RT-QPCR analyses of relative expression of corresponding genes expression in LUC or KMT2D sh2 infected cells;

All together, these preliminary results in MCF10A suggest that this cell line could be a proper model system to further investigate the function of Kmt2d in mammary epithelial cell plasticity.

4.5 Single Cell Transcriptome Analysis of Mammospheres

To better characterize the epithelial subpopulations present in cultured mammospheres, that were employed in the shRNA screens and validations, we decided to investigate their transcriptional heterogeneity. Thus we performed single cell RNA-sequencing (ScRNA-seq) in cells from the entire mammosphere culture (Bulk), and from mammosphere sub-populations differing for the intensity of PKH26 staining (e.g. with different proliferative histories): i. SC-enriched quiescent or slowly proliferating ($PKH26^{High}$); ii. highly proliferating mammary progenitors ($PKH26^{Neg/low}$); and iii. Cells with intermediate proliferative histories ($PKH26^{Med}$) (Cicalese, Bonizzi et al. 2009).

Briefly, either PKH26-labeled or not-labelled (Bulk) primary mouse mammary cells were cultured as mammospheres in suspension for 7 days. Then, PKH26-labeled mammospheres were disaggregated into single cells, and purified by FACS sorting, based on the retained PKH26 label intensity ($PKH26^{High}$, $PKH26^{Med}$ and $PKH26^{Neg/low}$ sub-populations). Also, the Bulk unlabelled cells were subjected to FACS, to separate alive single-cells from dead cells and debris that impair the scRNA sequencing quality. We prepared 5,000 Bulk, 3,000 $PKH26^{Neg/low}$, 2,000 $PKH26^{Med}$ and 2000 $PKH26^{High}$ cells which were then processed for scRNA-seq, using the 10X Chromium platform for single cell capture and library preparation, and sequenced with NovaSeq equipment at a depth of 50,000reads/cell. The experimental scheme is shown in Figure 4-41.

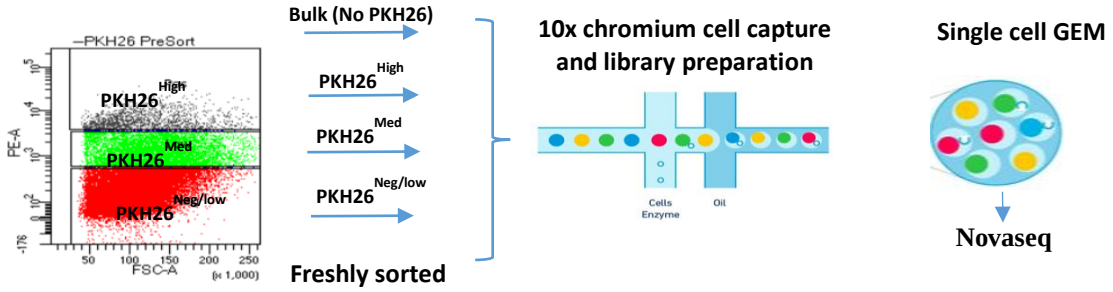


Figure 4-41. **Sorting strategy of primary mouse mammospheres and schematic representation of 10X Chromium single-cell RNA sequencing**

4.5.1 Single-cell RNA sequencing identifies 20 clusters within primary mouse mammospheres

Bioinformatics analyses of scRNA-seq data (in collaboration with Roman Hillje) allowed identification of a total 17,510 individual cells: 7,011_Bulk, 4,367_PKH26$^{Neg/low}$, 2,767_PKH26^{Med}, 3,365_PKH26^{High} cells. Cluster analyses allowed identification of 20 distinct cell populations within the primary mammospheres.

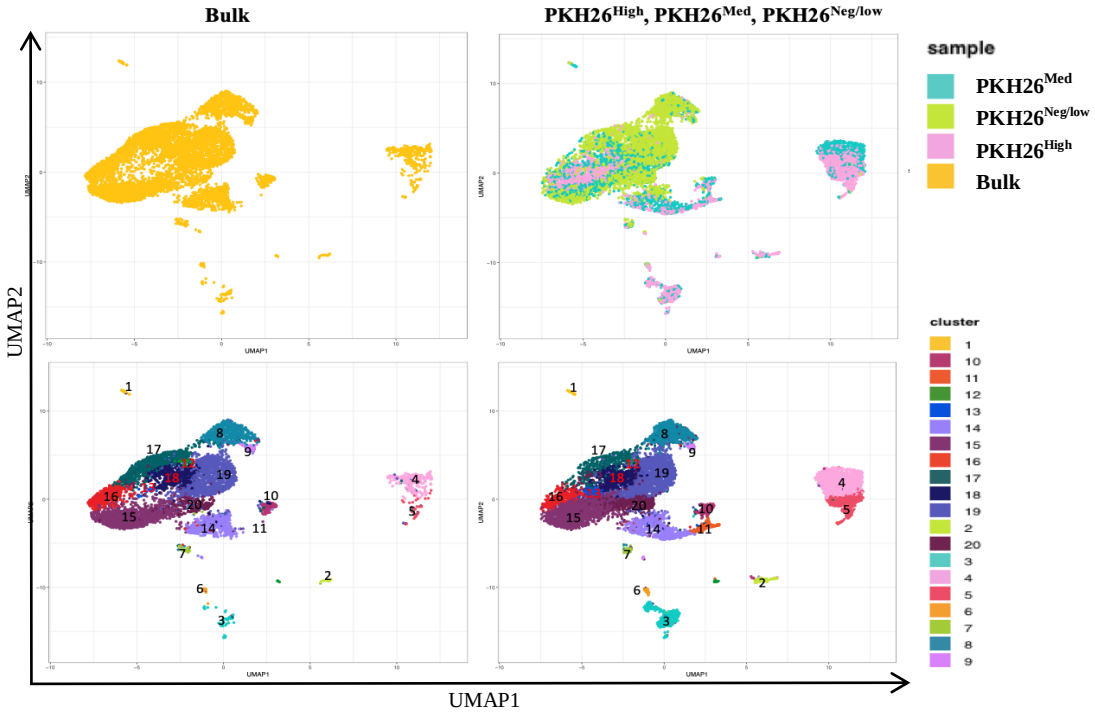


Figure 4-42. **UMAP plots by samples and clusters**

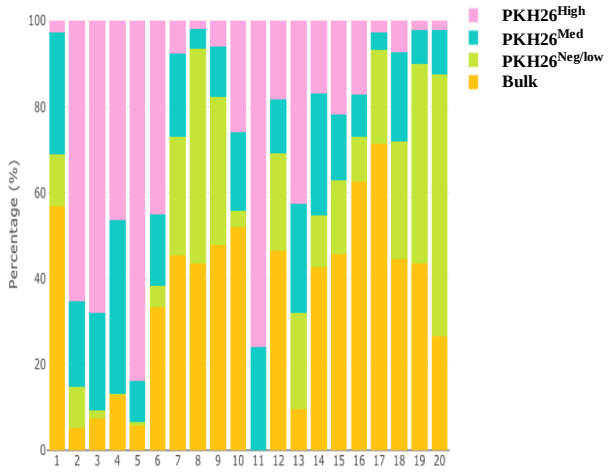
Transcriptome of total mammosphere cells from mouse mammary glands (n=12 mice) generated using the 10X Genomics Chromium platform.

Data were visually represented as UMAPs, by either samples (A and B, upper panel) and clusters (C and D, lower panel), using Uniform Manifold Approximation and Projection (UMAP) algorithm (Fig. 4-42). We obtained a total of 20 Clusters out of 7,011 Bulk cells and 10,499 cells from the three PKH26^{High}, PKH26^{Med} and PKH26$^{Neg/low}$ subpopulations. As expected, the transcriptional patterns derived from PKH26 subpopulations overlapped with the ones derived from Bulk mammosphere cells (Fig. 4-42). The number of genes expressed per cell varied among the samples, showing a correlation with the population's cell-cycle status. In detail, PKH26High was associated with the lowest number of expressed genes per cell and PKH26Neg/low with the highest (Table. 4-14).

Table 4-14. **Summary of samples in single-cell RNA-seq experiment**

Samples	N of cells sequenced	Median of expressed gene number per cell
Bulk	7,011	3,317
PKH26^{High}	3,365	2,199
PKH26^{Med}	2,767	2,641
PKH26$^{Neg/low}$	4,367	4,014

In the representation of cluster composition within samples, Bulk mammosphere cells are distributed in all clusters, consistent with the fact they contain all three subpopulations. Among the clusters, PKH26^{High} clusters composition contain PKH26^{Med} cells as they are intermediate population and closer to PKH26^{High}, as shown in in Figure 4-43.



Cluster	Tot N of cell	Bulk	PKH26Neg	PKH26Med	PKH26Pos
1	74	42	9	21	2
2	170	9	16	34	111
3	541	40	11	122	368
4	1560	199	5	632	724
5	554	32	4	54	464
6	102	34	5	17	46
7	134	61	37	26	10
8	1458	633	732	67	26
9	119	57	41	14	7
10	283	147	11	52	73
11	211	0	0	51	160
12	110	51	25	14	20
13	169	16	38	43	72
14	1332	568	161	379	224
15	3085	1407	534	472	672
16	917	574	96	90	157
17	1334	952	290	56	36
18	1438	642	394	295	107
19	2984	1300	1388	231	65
20	935	247	570	97	21

Figure 4-43. **The distribution of samples within the 20 Clusters**

Left panel shows the bar graph representation of clusters composition by percentage while the right panel shows the number of cells in each cluster.

4.5.2 Identification of MaSCs clusters and putative cell surface markers

In order to identify the MaSC enriched clusters among the 20 identified (regardless of samples), we checked the expression of gene lists that have shown to be up-regulated in MaSCs: i) LIM_MAMMARY_STEM_UP (Lim, Wu et al. 2010); ii) SOADY_MAMMARY_STEM_UP (Soady, Kendrick et al. 2015), and iii) MaSC_UP (Santoro, Vlachou et al. in press). Results showed that Clusters 2, 3 and 6 have a higher

expression of these MaSC-related genes, with higher statistical power for the MaSC_UP signature (Fig. 4-44). Notably, Clusters 2, 3 and 6 within the PKH26^{High} Clusters (C2, 3, 4, 5, 6, 11).

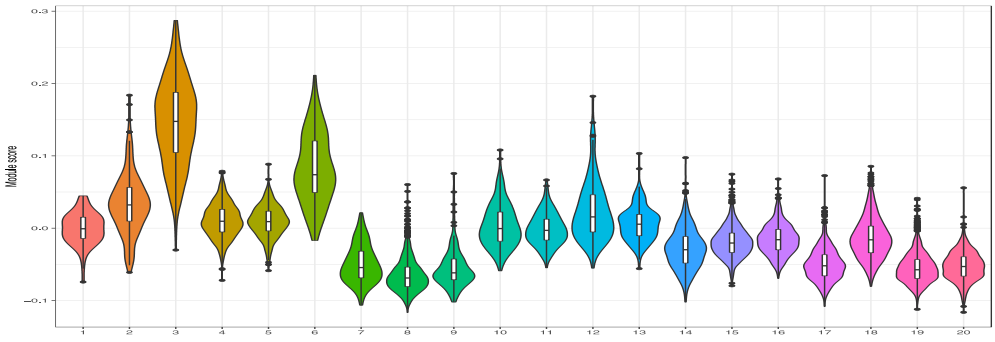


Figure 4-44. **Violin plot of mammary stem cell signature expression within 20 identified clusters**

Cluster 2, 3, 6 have a higher expression of MaSC-related genes, with higher statistical power for the MaSC_UP signature. Conversely, the clusters mainly composed by PKH26Neg/low cells are negatively correlated, such as C7, 8, 9. 17, 19, 20. Y-axis is the average expression of all the genes of the gene list, the value shows z-score.

Then, taking advantage of this powerful technology, we tried to identify putative MaSCs markers genes within the C2, 3 6. In these analyses, each sample/cluster was compared to all other samples/clusters together. Genes were selected based on a positive average log-fold change of at least 0.25 (e.g. the over-expressed genes) and expression in at least 70% of the cells of each single sample/cluster. Statistical analyses were performed using a classical t-test (previously shown to be very accurate in scRNA-seq data-set analyses). In the end, we identified 90 marker genes in the PKH26^{High} samples, as compared with the other samples, 172 marker genes in the C3 and 80 marker genes in the C6, as compared to all other clusters.

In order to find putative cell surface markers that could be possibly used for purification of the MaSCs in the mammospheres, as an alternative to PKH26 label retaining assay, we searched for glycoproteins among the identified marker genes. Notably, we found Cd36 (glycoprotein, collagen type I receptor) gene is exclusively expressed only in the PKH26^{High} population (LogFC=1.75 compared to other samples) and in the corresponding clusters, as shown in Figure 4-45.

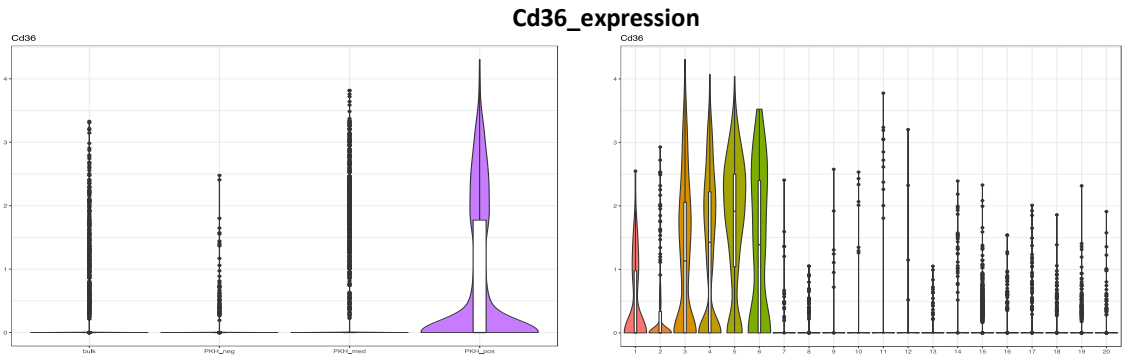


Figure 4-45. **Violin plot of Cd36 expression in samples (left) and clusters (right)**

Cd36 is exclusively expressed in the PKH26^{High} sample, and in the clusters most enriched with PKH26^{High} cells;

Though our bioinformatics search for cell-surface SC-marker is still ongoing, we are currently validating Cd36 in the purification of MaSCs from the mouse primary mammary epithelial mamommsphere culture.

4.5.3 Cbx5 and Kmt2d expressions are lower in PKH26^{High} population and putative Stem cell clusters

Consistently, scRNA-seq results showed that Cbx5 and Kmt2d are expressed at a lower level in PKH26^{High} population, and in the putative stem cell clusters C2, 3, 6 (Fig. 4-46 and 4-47, respectively).

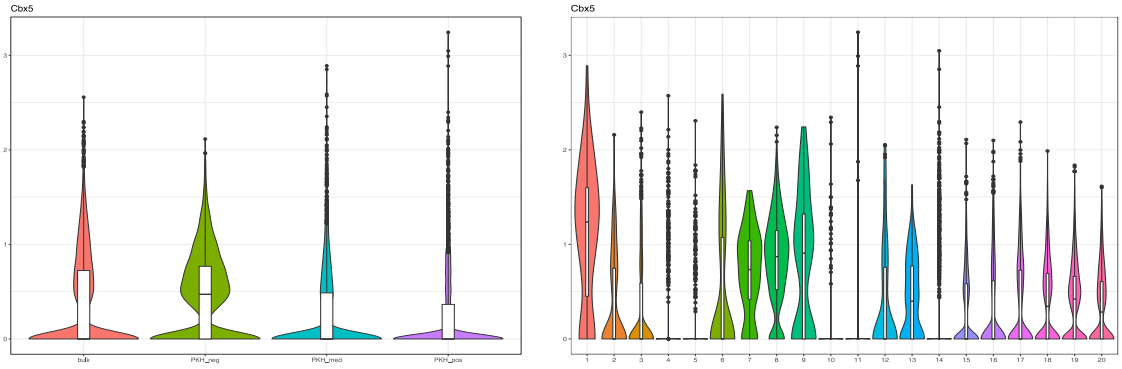


Figure 4-46. **Cbx5 expression is low in PKH26^{High} sample (left) and the putative stem cell clusters (right) by scRNA-seq**

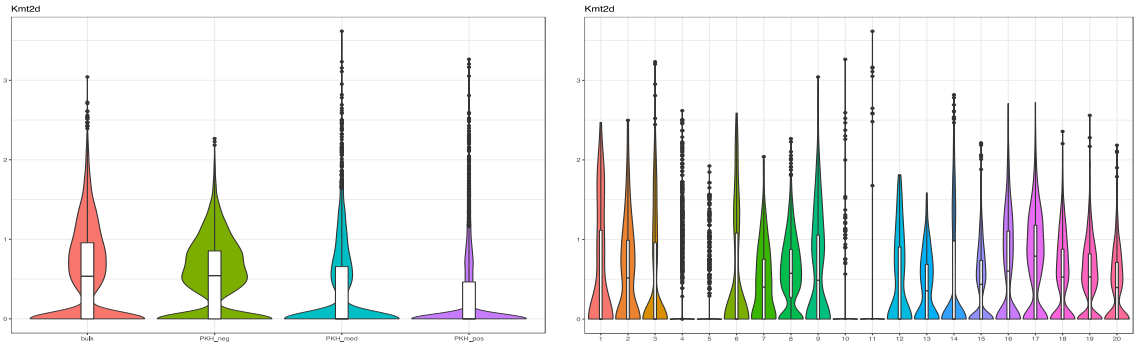


Figure 4-47. **Kmt2d expression is low in PKH26^{High} samples (left) and the putative stem cell clusters (right) by scRNA-seq**

Overall, scRNA-seq data set we generated is helping us to resolve the heterogeneity of MaSCs and progenitors at single cell resolution. It also confirmed the coherence of our validated genes (Cbx5 and Kmt2d) expression at a single cell resolution.

5 Discussion

5.1 Validation of the Genes Identified through shRNA Screening for Epigenetic Inhibitors of Cellular Plasticity

Recent studies suggest that cell plasticity is an important physiological trait, involved in tissue homeostasis maintenance and in tissue regeneration after severe damage, which, also contributes significantly to tumor initiation, tumor cell heterogeneity, metastasization and drug resistance (Varga and Greten 2017). Not surprisingly, these multiple and opposing functions are controlled by both genetic and epigenetic mechanisms. Epigenetic modifiers are known to play an important function in cell identity and cell differentiation during development though much less is known on their function in the control of cell plasticity in adult tissues (Paksa and Rajagopal 2017). In this study, we addressed the function of epigenetic regulation in the inhibition of physiological cellular plasticity, as a mechanism to prevent unscheduled cell fate changes, in particular de-differentiation. To this end, we performed pooled shRNA screens to identify epigenetic inhibitors of the conversion of mammary epithelial progenitors into mammary stem cells. The mouse mammary gland represents a suitable model to study cell plasticity, since it undergoes extensive remodelling, expansion and involution during development, pregnancy and lactation. Moreover, a robust *in vivo* assay for stemness is available in the mouse mammary gland: a single mammary SC injected into a de-epithelized (cleared) mammary fat-pad, can regenerate a fully differentiated mammary gland (Shackleton, Vaillant et al. 2006, Stingl, Eirew et al. 2006), while more differentiated progenitors cannot. In our genetic screenings, we employed lentiviral shRNA libraries to inhibit 234 epigenetic modifiers in mouse primary mammary progenitor cells, each targeted by 10 different shRNAs. In 6 independent screening experiments, we identified shRNAs targeting 38 potentially inhibitors of cell plasticity, out of the 234 genes targeted by our library. Notably, some of the identified genes were

recently identified as inhibitors of iPSCs reprogramming, such as p53 (Utikal, Polo et al. 2009), Dot1L (Onder, Kara et al. 2012), Chaf1 (Cheloufi, Elling et al. 2015), thus providing a first level of validation for our screening strategy. Moreover, also the other identified hits are known to play important functions in development, cell identity, and cancer.

The candidates were then individually subjected to *in vivo* validation experiments, by the same transplantation assay employed in the screenings. We considered as validated hits those scoring positive in at least two independent experiments, testing altogether at least 100,000 shRNA-infected cells. We achieved a good rate of validation, since 7 genes were validated out of 21 tested, in a total of 12 independent experiments. This high validation rate confirmed the value of our pooled *in vivo* screening strategy to identify regulators of cell plasticity. Additional hits' validation is still ongoing, with the final aim of obtaining a more complete picture of the epigenetic regulation of physiological cell plasticity.

Nevertheless, most of the *in vivo* validated shRNAs failed to confer *in vitro* self-renewal ability to mammary progenitors as assessed by serial re-plating of mammosphere cultures, except for Kmt2dsh that conferred self-renewal ability. The main reason for the discrepancy between the results of *in vivo* and *in vitro* validation experiments, could be the high stringency of the in vitro self-renewal assay, that is based on the resistance of SCs to anoikis–induced apoptosis. Mammary progenitors may indeed die in suspension before completing the process of reprogramming into SCs, and become resistant to anoikis. The *in vivo* environment of the mammary niche is likely to be more permissive, thus allowing progenitor survival and reprogramming. Moreover, niche microenvironmental signals and growth factors might be required for reprogramming. Importantly, Myc or YAP/TAZ expression increases proliferation and induces efficient self-renewal *in vitro* (Panciera, Azzolin et al. 2016, Santoro, Vlachou et al. in press), suggesting a requirement for cell proliferation to induce *in vitro* self-renewal, a property that may not be shared by our hits.

Finally, we should not ignore the fact that full reprogramming is a very rare event. In fact, even the reprogramming of fibroblasts into iPSCs by the four "Yamanaka factors" - Oct4, Sox2, Klf4, Myc-is a relatively inefficient process, strongly affected by cell culture conditions and epigenetic alterations (Maherail and Hochdllinger,). Importantly, this process is strongly stimulated by expression of the oncogene Myc, which might in part affect it through induction of self-renewal and cell proliferation.

In conclusion, our data suggest that the *in vivo* mammary gland regeneration assay is more efficient than the *in vitro* mammosphere assay, to score the induction of stemness phenotypes in reprogrammed mammary progenitors. Still, the *in vitro* self-renewal assay is critical because it provides a first hint on the cellular mechanisms of reprogramming. Hence, taking those aforementioned possible reasons into account, we successfully demonstrated that lineage restricted luminal cells can be reprogrammed into MaSCs by Myc expression, using an organoid formation and self-renewal assay (Panciera, Azzolin et al. 2016).

In this assay, freshly isolated mouse mammary luminal cells infected with a MycER expressing construct were able to form compact, contractile organoids, showing extended self-renewal ability through serial passages, as compared to the SC-enriched MRU subpopulation (Santoro, Vlachou et al. in press). The advantages of using the organoid assay are: i) progenitor cells are infected and initially maintained in adhesion on collagen-coated plates, which is a less stressful condition to anchorage-independent cultures; ii) in addition, morphology of colonies formed in Matrigel provides information not only regarding the self-renewal ability of the reprogrammed cells plated, but also on their differentiation potential; iii) the size of the colonies also provides information on the proliferative statues of the cells. Overall, by optimizing this approach one could potentially efficiently characterize and dissect the mechanism of reprogramming at all levels. Thus, this assay may allow us to assess induction of self-renewal *in vitro*, by the *in vivo* validated shRNAs.

Taken together, we provided convincing evidence of the stringency and robustness of the *in vivo* strategy we used to validate the identified genes, and tested possible approaches to assess the induction of self-renewal ability by *in vitro*.

5.2 Cbx5 and Its Function as an Inhibitor of Cellular Plasticity

Here we have reported that Cbx5 down-regulation reprograms mammary progenitors into MaSCs, as assessed by *in vivo* transplantation assay. Target progenitors were purified as PKH26$^{Neg/low}$ cells or Lin-Cd24^{hi}Cd49f+ uncultured mammary luminal cells (LUM). We also showed that Cbx5 down-regulation accelerates the conversion of CommaDβ mouse mammary cells from the Sca-1^{neg} to Sca-1^{high} phenotype. These results strongly suggest that HP1α is an inhibitor of cellular plasticity, at least in mammary cells. Consistently, the HP1 protein family, in particular HP1γ, was recently reported to inhibit reprogramming of fibroblasts into iPSCs. In fact, HP1γ deletion increases the reprogramming efficiencies, while HP1α, or HP1β depletion had a milder and less reproducible effect (Sridharan, Tchieu et al. 2009, Zaidan, Walker et al. 2018). Our results are in line with these findings, further suggesting that HP1α could play a physiological function in cell identity maintenance, acting as a barrier against cell fate transitions, in particular in the generation of SCs from their more differentiated progeny.

The implication of HP1α in the regulation of cell plasticity suggests it might be involved in the generation of CSCs. Indeed, HP1α is down-regulated in metastatic colon, thyroid and breast cancers (De Lange, Burtscher et al. 2001, Wasenius, Hemmer et al. 2003, Norwood, Moss et al. 2006), hinting to its possible function of metastasis - suppressor. In breast cancer, HP1α is the only HP1 family member that is down-regulated both at the mRNA and protein level, in the highly invasive breast cancer cell lines MDA-MB-231 and HS578T, as compared to the non-invasive T47D and MCF7 cell lines (Kirschmann, Lininger et al. 2000, Thomsen, Christensen et al. 2011). Moreover, HP1α is involved in the invasive phenotype of the MCF7 and MDA-MB-231 breast cancer cell lines, *in vitro*.

Cbx5 down-regulation in MCF7 cells leads to a 40% increase of their invasiveness *in vitro*, while the opposite effect is elicited by Cbx5 over-expression in MDA-MB-231 cells (Kirschmann, Lininger et al. 2000, Norwood, Moss et al. 2006). Finally, a recent study suggested a possible function of HP1α in EMT. During the initial steps of TGFβ-induced EMT, HP1α is transiently released from the major satellite repeat sequences located in pericentric heterochromatin, to allow heterochromatin reorganization (Millanes-Romero, Herranz et al. 2013). Collectively, these studies suggest a tumor suppressive function for Cbx5, possibly elicited through the inhibition of EMT and invasive and metastatic potential. Taken together, our result further suggest that HP1α might exert a tumor suppressive function by inhibiting cell plasticity, in particular *de novo* cancer stem cell formation.

In the initial *in vivo* validation experiments, we co-infecetd two shRNAs targeting Cbx5. We then tested them individually for their ability to induce reprogramming, and found that, although both shRNAs down-regulate Cbx5 mRNA and HP1α protein levels with equal efficiency, only Cbx5 sh15 was able to induce the generation of mammary repopulating units. These results might suggest off-target effects. Therefore, we will further investigate this issue, testing the effects of other Cbx5 shRNAs first *in vitro*, in the CommaDB cell line, and then their reprogramming efficiency *in vivo*. In parallel, we will test whether Cbx5 sh11 and sh15 could target different Cbx5 transcript variants, which may account for their different phenotypic outputs. Indeed, a recent study revealed the existence of several novel Cbx5 transcripts in breast cancer cell lines, and described a novel *Cbx5* transcriptional isoform, *STET*, whose expression inversely correlates with HP1α coding mRNA (Vad-Nielsen, Jakobsen et al. 2016).

Consistent with their different reprogramming ability, the two Cbx5 shRNAs induce different effects on the transcriptome, as shown by our RNA-seq studies. We investigated the trancriptomic changes induced in $PKH26^{Neg/low}$ cells upon Cbx5 down-regulation, at 24h and 72h after shRNA infection. The GSEA analysis of the RNA-seq data revealed

both commonly and differently regulated pathways. Cbx5 sh15 induces a consistent enrichment of the oxidative phosphorylation and Myc targets gene sets from 24h to 72h., while the interferon alpha and gamma responses are significantly enriched only at 24h. Instead, the gene sets enriched at 24h and 72h after Cbx5 sh11 infection are different, and more importantly, they share little overlap with those enriched after Cbx5 sh15 infection, especially at 72h. As we mentioned earlier, we will address the possible reasons of the different biological and trancriptional effects of the two Cbx5 shRNAs, testing additional Cbx5 shRNAs.

Still, these RNA-seq data provide us with important indications about the early responses during reprogramming by Cbx5 down-regulation. In addition, they suggested different transcriptomic effects of the two shRNAs employed, that may explain the different biological effects we obtained in the *in vivo* regeneration experiments.

In conclusion, we obtained promising results that Cbx5 validated as a putative inhibitor of mammary epithelial cell plasticity *in vivo*. We still need to address possible off-target effects of the employed shRNAs. After resolving this issue, we will validate the obtained RNA-seq data, further explore the transcriptional changes by performing additional RNAsq experiments, both at the same and longer time points, to uncover the transcriptomic changes during the reprogramming event by Cbx5 down-regulation.

5.3 Kmt2d and Its Function as an Inhibitor of Cellular Plasticity

Kmt2d plays important functions in regulating gene transcription and is frequently mutated in a variety of cancers, including lymphoma (Morin, Mendez-Lago et al. 2011), medulloblastoma (Jones, Jager et al. 2012), and gastric cancer (Zang, Cutcutache et al. 2012). Kmt2d is also mutated in other human diseases, such as Kabuki syndrome (Paulussen, Stegmann et al. 2011). The identified Kmt2d mutations belong to the loss-of-function type, suggesting that its function as a tumor suppressor in various tissues (Zaidi, Choi et al. 2013, Rao and Dou 2015). Therefore, the identification of Kmt2d, a known

tumor suppressor, as an inhibitor of reprogramming in our screens is in line with our initial hypothesis, i.e. that physiological inhibitors of cellular plasticity could play as a tumor suppressor. The function of Kmt2d in maintenance of cell identity or suppression of fate transition has not been specifically addressed, except for a recent report showing that Kmt2d is required for mouse ESCs differentiation, and activates cell-identity genes by priming their enhancers (Wang, Lee et al. 2016).

Here we reported that Kmt2d down-regulation by three different shRNAs reprograms the $PKH^{Neg/low}$ progenitors into SCs able to regenerate mammary glands upon transplantation. This strongly suggests that Kmt2d inhibits unscheduled de-differentiation of mammary progenitors to MaSCs, and argues against off-target effects, since multiple shRNAs targeting Kmt2d are able to induce this phenotype *in vivo*. Moreover, we obtained preliminary evidence that Kmt2d down-regulation endows mammary progenitors with the ability to grow in suspension as mammospheres and self-renewal through serial passages. In this experiment, we tried to improve survival and viability of sorted and infected mammary progenitors, by culturing them in 5% matrigel containing stem medium for 7 days, prior to the mammosphere assay in suspension, to prevent anoikis-induced cell death in suspension culture before cell reprogramming could occur. We will further confirm the inhibitory function of Kmt2d on self-renewal, both in suspension culture and in organoid formation and self-renewal assays. The latter approach will also allow to investigate the effects of Kmt2d inhibition on organoid differentiation *in vitro*.

We further validated Kmt2d as an inhibitor of cell plasticity using the CommaDβ mouse mammary epithelial cell line, in which Kmt2d shRNAs expression accelerates the spontaneous conversion from the $Sca\text{-}1^{neg}$ to $Sca\text{-}1^{high}$ phenotype. In the MCF10A human mammary epithelial cell line, we obtained preliminary results showing that KMT2D down-regulation increased sphere formation efficiency over at least three passages in suspension culture. Moreover, we noticed the induction of an EMT-like phenotype by KMT2D down-regulation in adhesion culture, and accordingly we found increased

expression of ZEB1 mRNA. Altogether, we validated Kmt2d as an inhibitor of cell plasticity in normal mammary epithelial tissue. These results suggest that Kmt2d loss of function in cancer may also induce cell plasticity, possibly leading to the generation of CSCs from more differentiated non-cancer SCs. This question will be specifically addressed in mouse models of mammary tumor and breast cancer patient derived xenografts.

Finally, we performed transcriptomic analyses in PKH26$^{Neg/low}$ mammary progenitors, infected with two different validated shRNAs individually targeting Kmt2d in PKH26$^{Neg/low}$ cells, at 24h and 72h after infection. GSEA analysis of the RNA-seq data showed a significant enrichment of IFN-α, IFN- γ responses, Il-6-Jak-Stat3 inflammatory signaling pathways, at both time points, suggesting an early response, that is maintained at least for 72 hours. Other significantly enriched gene-sets are Myc targets v1 and v2, oxidative phosphorylation, and DNA-repair. Interestingly, these early transcriptomic responses to Kmt2d down-regulation are shared with those obtained upon Cbx5 down-regulation by sh15. Coherently, Cbx5 mRNA level is significantly decreased, 72h after Kmt2d down-regulation, suggesting that Cbx5 might be a downstream target of Kmt2d during reprogramming.

5.4 Kmt2d and Cbx5 Downregulation may Induce Cellular Plasticity through Pro-inflammatory Cytokines

Based on our RNA-seq analyses of the progenitors infected with validated shRNAs, at early two time points, we propose to define the shared trancriptomic inflammatory signature as early transcriptional reprogramming response (ETRR) of mammary progenitor cells to SCs. Leading edge analysis of the most significantly enriched pathways in the ETRR revealed several chemokines and cytokines commonly induced by down-regulation of either Cbx5 or Kmt2d, in PKH26$^{Neg/low}$ cells. These include the inflammatory cytokines: Cxcl11,Cxcl10, Cxcl9, Ccl5, Ccl7, Il6, and Il7 (Yamamura 1992, Cassatella, Meda et al. 1993, Jung, Eckmann et al. 1995), suggesting their possible

function in the reprogramming processes of mammary progenitors. In line with this, we experimentally found induction of Il6 mRNA upon Kmt2d down-regulation, both in mouse and human mammary epithelial cell lines, namely in CommaDβ and MFC10A. Consistently, it was previously shown that transient over-expression of IL-6 at the early stage of MEF reprogramming by the "Yamanaka factors" OSKM, promotes the generation of iPSCs and can even substitute for Myc (Brady, Li et al. 2013). In breast cancer, IL-6 mediated inflammatory signaling was shown to lead to cancer stem cell expansion (Iliopoulos, Hirsch et al. 2011, Korkaya, Kim et al. 2012). These studies suggest that inflammation can stimulate cell reprogramming although it is not yet known which aspects of cell plasticity and what are the precise mechanisms involved.

Moreover, we found a significant overlap between the signature induced by Kmt2d or Cbx5 knock-down, and the MycER over-expression signature in $PKH26^{Neg/low}$ cells (Fig. 4-40). Down regulation of Cbx5 or Kmt2d in $PKH26^{Neg/low}$ cells induce Myc targets at 72h. This evidence reinforces the function of Myc targets in reprogramming of mammary progenitors into SCs, that was recently shown in our group (Santoro, Vlachou et al. in press).

Based on this evidence, we speculate that the proinflammatory cytokines in our ETRR signature may contribute to priming the environment in favor of reprogramming, and Myc targets, together with other pathways, might be subsequently induced to achieve full reprogramming. This hypothesis can be directly tested by supplying a different combination of above mentioned cytokines in the cell culture medium during reprogramming of $PKH26^{Neg/low}$ cells. These experiments should help us to understand the function of those cytokines in the cell plasticity under physiological condition. Eventually, as future work, we can directly check whether these cytokines might induce metastasis through cellular plasticity. Indeed, several recent studies have shown a function of inflammation in metastasis through EMT in breast cancer and other cancers (Cohen, Gao et al. 2015, Qian 2017, Singh, Mishra et al. 2018).

5.5 Single Cell Transcriptomic profile of Mouse Primary Mammopsheres

We reported here some preliminary analyses on scRNA-seq datasets generated from mouse primary mammosphere subpopulations. As expected, we provided the first evidence, at single - cell level, that the transcriptome of quiescent or slowly proliferating PKH^{high} cells are clearly distinguishable from those of their highly proliferating $PKH26^{Neg/low}$ progeny, since they are almost completely separated in the UMAP plot (Fig. 4-42). Moreover, we found a considerable heterogeneity within the mouse primary mammospheres, making still challenging, thus, the definition of the cell types solely on the basis of their transcriptomes. However, thanks to the large amount of studies published on mammary stem cell characterization, we were able to define putative mammary stem cell clusters, taking in consideration all clusters in an unbiased way, regardless of their sample origin. Consistently, we found that those putative mammary stem cell clusters are mainly composed by $PKH26^{High}$ cells.

Then, taking advantage of this powerful technology, we searched for putative markers in the identified SC – enriched clusters identified. In particular, we focused on surface markers, to purify a SC – enriched subpopulation as a possible alternative to the PKH26 label retaining assay. We found the Cd36 glycoprotein, exclusively expressed in the clusters that are mainly composed of $PKH26^{High}$ cells. Consistent with its expression in putative SCs, CD36 was shown as a marker of glioblastoma CSCs and recently shown as a novel marker for fibroblast quiescence in lung fibrosis (Hale, Otvos et al. 2014, Heinzelmann, Lehmann et al. 2018). This suggests that Cd36 could be associated with SCs properties, in particular quiescence (label – retaining cells) also in mammospheres, and in mammary SCs. Therefore, we may investigate the expression of CD36 in freshly purified mammary SC-enriched subpopulation, *in situ* by immunofluorescence, and in mammary organoids.

Taken together, these coherent and promising results we obtained in the scRNA-seq of mammosphere subpopulations, prompt us to further mine the data with the help of a novel and comprehensive tool, Single Cell Browser (developed by Roman Hillje, manuscript in preparation), to display and analyse single cell RNA-seq data. The availability of this platform will hopefully help us in identifying the relevant genes involved in cell plasticity. In fact, we are planning to perform scRNA-seq on $PKH26^{Neg/low}$ cells infected with selected validated shRNAs, at different time points, to investigate possible intermediate reprogramming states by pseudotime analysis of the obtained data (Campbell and Yau 2016, Reid and Wernisch 2016, Street, Risso et al. 2018).

Appendix

During the course of my PhD studies I contributed to the successful completion of the revision process of an original research article titled as “P53-loss in breast cancer leads to Myc activation, increased cell plasticity and expression of a mitotic signature with prognostic value”, which has been recently accepted for publication in Cell Reports (Santoro, Vlachou et al. in press). The results I provided and have been incorporated in the final version of the manuscript to address specific Reviewer comments are shown in the thesis section 4.2.3.2, Figures 4-10, 11, and 12. The title list of authors and abstract are provided in the following pages.

P53-loss in breast cancer leads to Myc activation, increased cell plasticity and expression of a mitotic signature with prognostic value

Angela Santoro[1^], Thalia Vlachou[1^], Lucilla Luzi[1], Giorgio Melloni[2,3], Luca Mazzarella[1], Errico D'Elia[1], **Xieraili Aobuli[1]**, Cristina Elisabetta Pasi[1], Linsey Reavie[4], Paola Bonetti[2], Simona Punzi[1], Lucia Casoli[2], Arianna Sabò[2], Maria Cristina Moroni[1], Gaetano Ivan Dellino[1,5], Bruno Amati[1,2], Francesco Nicassio[2], Luisa Lanfrancone[1] and Pier Giuseppe Pelicci[1,5]*

[1]Department of Experimental Oncology, European Institute of Oncology, Via Adamello 16, 20139 Milan, Italy.

[2]Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, Via Adamello 16, 20139 Milan, Italy.

[3]Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115, USA.

[4]Friedrich Miescher Institute, Maulbeerstr 66, 4058 Basel, Switzerland.

[5]Department of Oncology and Hemato-Ongology, University of Milan, Via Santa Sofia 9, 20142 Milan, Italy.

^ Equally contributing *Correspondence to Pier Giuseppe Pelicci: piergiuseppe.pelicci@ieo.it

Abstract

Loss of p53 function is invariably associated with cancer. Its role in tumor growth was recently linked to its effects on cancer SCs (CSCs), though underlying molecular mechanisms remain unknown. Here we show that c-myc is a transcriptional target of p53 in mammary stem cells (MaSCs) and is activated in breast tumors as a consequence of p53-loss. Constitutive Myc expression in normal mammary cells leads to increased frequency of SC symmetric divisions, extended SC replicative-potential and SC-reprogramming of progenitors, while Myc activation in breast cancer is necessary and sufficient to maintain the expanding pool of CSCs and tumor growth. Concomitant p53-loss and Myc activation trigger the expression of 189 mitotic genes, which identify patients at high risk of mortality and relapse, independently of other risk factors. Our data indicate that de-regulation of the p53:Myc axis in mammary tumors increases CSC content and plasticity, and is a critical determinant of tumor growth and clinical aggressiveness.

References

Anders, S., P. T. Pyl and W. Huber (2015). "HTSeq--a Python framework to work with high-throughput sequencing data." Bioinformatics **31**(2): 166-169.

Barker, N., R. A. Ridgway, J. H. van Es, M. van de Wetering, H. Begthel, M. van den Born, E. Danenberg, A. R. Clarke, O. J. Sansom and H. Clevers (2009). "Crypt stem cells as the cells-of-origin of intestinal cancer." Nature **457**(7229): 608-611.

Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**(6): 1513-1523.

Blanpain, C. and E. Fuchs (2014). "Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration." Science **344**(6189): 1242281.

Brabletz, T., A. Jung, S. Spaderna, F. Hlubek and T. Kirchner (2005). "Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression." Nat Rev Cancer **5**(9): 744-749.

Brady, J. J., M. Li, S. Suthram, H. Jiang, W. H. Wong and H. M. Blau (2013). "Early role for IL-6 signalling during generation of induced pluripotent stem cells revealed by heterokaryon RNA-Seq." Nat Cell Biol **15**(10): 1244-1252.

Britschgi, A., S. Duss, S. Kim, J. P. Couto, H. Brinkhaus, S. Koren, D. De Silva, K. D. Mertz, D. Kaup, Z. Varga, H. Voshol, A. Vissieres, C. Leroy, T. Roloff, M. B. Stadler, C. H. Scheel, L. J. Miraglia, A. P. Orth, G. M. Bonamy, V. A. Reddy and M. Bentires-Alj (2017). "The Hippo kinases LATS1 and 2 control human breast cell fate via crosstalk with ERalpha." Nature **541**(7638): 541-545.

Campbell, K. R. and C. Yau (2016). "Order Under Uncertainty: Robust Differential Expression Analysis Using Probabilistic Models for Pseudotime Inference." PLoS Comput Biol **12**(11): e1005212.

Cardiff, R. D. and S. R. Wellings (1999). "The comparative pathology of human and mouse mammary glands." J Mammary Gland Biol Neoplasia **4**(1): 105-122.

Cassatella, M. A., L. Meda, S. Bonora, M. Ceska and G. Constantin (1993). "Interleukin 10 (IL-10) inhibits the release of proinflammatory cytokines from human polymorphonuclear leukocytes. Evidence for an autocrine role of tumor necrosis factor and IL-1 beta in mediating the production of IL-8 triggered by lipopolysaccharide." J Exp Med **178**(6): 2207-2211.

Celia-Terrassa, T. (2018). "Mammary Stem Cells and Breast Cancer Stem Cells: Molecular Connections and Clinical Implications." Biomedicines **6**(2).

Chaffer, C. L., I. Brueckmann, C. Scheel, A. J. Kaestli, P. A. Wiggins, L. O. Rodrigues, M. Brooks, F. Reinhardt, Y. Su, K. Polyak, L. M. Arendt, C. Kuperwasser, B. Bierie and R. A. Weinberg (2011). "Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state." Proc Natl Acad Sci U S A **108**(19): 7950-7955.

Chaffer, C. L., N. D. Marjanovic, T. Lee, G. Bell, C. G. Kleer, F. Reinhardt, A. C. D'Alessio, R. A. Young and R. A. Weinberg (2013). "Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity." Cell **154**(1): 61-74.

Chakrabarti, R., J. Hwang, M. Andres Blanco, Y. Wei, M. Lukacisin, R. A. Romano, K. Smalley, S. Liu, Q. Yang, T. Ibrahim, L. Mercatali, D. Amadori, B. G. Haffty, S. Sinha and Y. Kang (2012). "Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2." Nat Cell Biol **14**(11): 1212-1222.

Cheloufi, S., U. Elling, B. Hopfgartner, Y. L. Jung, J. Murn, M. Ninova, M. Hubmann, A. I. Badeaux, C. Euong Ang, D. Tenen, D. J. Wesche, N. Abazova, M. Hogue, N. Tasdemir, J. Brumbaugh, P. Rathert, J. Jude, F. Ferrari, A. Blanco, M. Fellner, D. Wenzel, M. Zinner, S. E. Vidal, O. Bell, M. Stadtfeld, H. Y. Chang, G. Almouzni, S. W. Lowe, J. Rinn, M. Wernig, A. Aravin, Y. Shi, P. J. Park, J. M. Penninger, J. Zuber and K. Hochedlinger (2015). "The histone chaperone CAF-1 safeguards somatic cell identity." Nature **528**(7581): 218-224.

Cheng, L., S. Bao and J. N. Rich (2010). "Potential therapeutic implications of cancer stem cells in glioblastoma." Biochem Pharmacol **80**(5): 654-665.

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

Cohen, E. N., H. Gao, S. Anfossi, M. Mego, N. G. Reddy, B. Debeb, A. Giordano, S. Tin, Q. Wu, R. J. Garza, M. Cristofanilli, S. A. Mani, D. A. Croix, N. T. Ueno, W. A. Woodward, R. Luthra, S. Krishnamurthy and J. M. Reuben (2015). "Inflammation Mediated Metastasis: Immune Induced Epithelial-To-Mesenchymal Transition in Inflammatory Breast Cancer Cells." PLoS One **10**(7): e0132710.

Cordenonsi, M., F. Zanconato, L. Azzolin, M. Forcato, A. Rosato, C. Frasson, M. Inui, M. Montagner, A. R. Parenti, A. Poletti, M. G. Daidone, S. Dupont, G. Basso, S. Bicciato and S. Piccolo (2011). "The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells." Cell **147**(4): 759-772.

De Koning, L., A. Savignoni, C. Boumendil, H. Rehman, B. Asselain, X. Sastre-Garau and G. Almouzni (2009). "Heterochromatin protein 1alpha: a hallmark of cell proliferation relevant to clinical oncology." EMBO Mol Med **1**(3): 178-191.

De Lange, R., H. Burtscher, M. Jarsch and U. H. Weidle (2001). "Identification of metastasis-associated genes by transcriptional profiling of metastatic versus non-metastatic colon cancer cell lines." Anticancer Res **21**(4A): 2329-2339.

De Sousa, E. M. F., X. Wang, M. Jansen, E. Fessler, A. Trinh, L. P. de Rooij, J. H. de Jong, O. J. de Boer, R. van Leersum, M. F. Bijlsma, H. Rodermond, M. van der Heijden, C. J. van Noesel, J. B. Tuynman, E. Dekker, F. Markowetz, J. P. Medema and L. Vermeulen (2013). "Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions." Nat Med **19**(5): 614-618.

Deugnier, M. A., M. M. Faraldo, J. Teuliere, J. P. Thiery, D. Medina and M. A. Glukhova (2006). "Isolation of mouse mammary epithelial progenitor cells with basal characteristics from the Comma-Dbeta cell line." Dev Biol **293**(2): 414-425.

Do, J. T. and H. R. Scholer (2004). "Nuclei of embryonic stem cells reprogram somatic cells." Stem Cells **22**(6): 941-949.

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

dos Santos, C. O., C. Rebbeck, E. Rozhkova, A. Valentine, A. Samuels, L. R. Kadiri, P. Osten, E. Y. Harris, P. J. Uren, A. D. Smith and G. J. Hannon (2013). "Molecular hierarchy of mammary differentiation yields refined markers of mammary stem cells." Proc Natl Acad Sci U S A **110**(18): 7123-7130.

Dravis, C., C. Y. Chung, N. K. Lytle, J. Herrera-Valdez, G. Luna, C. L. Trejo, T. Reya and G. M. Wahl (2018). "Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity." Cancer Cell **34**(3): 466-482 e466.

Eissenberg, J. C. and S. C. Elgin (2014). "HP1a: a structural chromosomal protein regulating transcription." Trends Genet **30**(3): 103-110.

Fan, B., Y. Malato, D. F. Calvisi, S. Naqvi, N. Razumilava, S. Ribback, G. J. Gores, F. Dombrowski, M. Evert, X. Chen and H. Willenbring (2012). "Cholangiocarcinomas can originate from hepatocytes in mice." J Clin Invest **122**(8): 2911-2915.

Fu, N. Y., A. C. Rios, B. Pal, C. W. Law, P. Jamieson, R. Liu, F. Vaillant, F. Jackling, K. H. Liu, G. K. Smyth, G. J. Lindeman, M. E. Ritchie and J. E. Visvader (2017). "Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive." Nat Cell Biol **19**(3): 164-176.

Grivennikov, S. I., F. R. Greten and M. Karin (2010). "Immunity, inflammation, and cancer." Cell **140**(6): 883-899.

Gu, B., K. Watanabe, P. Sun, M. Fallahi and X. Dai (2013). "Chromatin effector Pygo2 mediates Wnt-notch crosstalk to suppress luminal/alveolar potential of mammary stem and basal cells." Cell Stem Cell **13**(1): 48-61.

Guo, W., Z. Keckesova, J. L. Donaher, T. Shibue, V. Tischler, F. Reinhardt, S. Itzkovitz, A. Noske, U. Zurrer-Hardi, G. Bell, W. L. Tam, S. A. Mani, A. van Oudenaarden and R. A. Weinberg (2012). "Slug and Sox9 cooperatively determine the mammary stem cell state." Cell **148**(5): 1015-1028.

Hale, J. S., B. Otvos, M. Sinyuk, A. G. Alvarado, M. Hitomi, K. Stoltz, Q. Wu, W. Flavahan, B. Levison, M. L. Johansen, D. Schmitt, J. M. Neltner, P. Huang, B. Ren, A. E. Sloan, R. L. Silverstein, C. L. Gladson, J. A. DiDonato, J. M. Brown, T. McIntyre, S. L. Hazen, C. Horbinski, J. N. Rich and J. D. Lathia (2014). "Cancer stem cell-specific scavenger receptor CD36 drives glioblastoma progression." Stem Cells **32**(7): 1746-1758.

Hay, E. D. (2005). "The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it." Dev Dyn **233**(3): 706-720.

Hayakawa, Y., H. Ariyama, J. Stancikova, K. Sakitani, S. Asfaha, B. W. Renz, Z. A. Dubeykovskaya, W. Shibata, H. Wang, C. B. Westphalen, X. Chen, Y. Takemoto, W. Kim, S. S. Khurana, Y. Tailor, K. Nagar, H. Tomita, A. Hara, A. R. Sepulveda, W. Setlik, M. D. Gershon, S. Saha, L. Ding, Z. Shen, J. G. Fox, R. A. Friedman, S. F. Konieczny, D. L. Worthley, V. Korinek and T. C. Wang (2015). "Mist1 Expressing Gastric Stem Cells Maintain the Normal and Neoplastic Gastric Epithelium and Are Supported by a Perivascular Stem Cell Niche." Cancer Cell **28**(6): 800-814.

Heinzelmann, K., M. Lehmann, M. Gerckens, N. Noskovicova, M. Frankenberger, M. Lindner, R. Hatz, J. Behr, A. Hilgendorff, M. Konigshoff and O. Eickelberg (2018). "Cell-surface phenotyping identifies CD36 and CD97 as novel markers of fibroblast quiescence in lung fibrosis." Am J Physiol Lung Cell Mol Physiol.

Herz, H. M., M. Mohan, A. S. Garruss, K. Liang, Y. H. Takahashi, K. Mickey, O. Voets, C. P. Verrijzer and A. Shilatifard (2012). "Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4." Genes Dev **26**(23): 2604-2620.

Holliday, H., L. A. Baker, S. R. Junankar, S. J. Clark and A. Swarbrick (2018). "Epigenomics of mammary gland development." Breast Cancer Res **20**(1): 100.

Hong, H., K. Takahashi, T. Ichisaka, T. Aoi, O. Kanagawa, M. Nakagawa, K. Okita and S. Yamanaka (2009). "Suppression of induced pluripotent stem cell generation by the p53-p21 pathway." Nature **460**(7259): 1132-1135.

Hu, Y. and G. K. Smyth (2009). "ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays." J Immunol Methods **347**(1-2): 70-78.

Humphreys, R. C., M. Krajewska, S. Krnacik, R. Jaeger, H. Weiher, S. Krajewski, J. C. Reed and J. M. Rosen (1996). "Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis." Development **122**(12): 4013-4022.

Iliopoulos, D., H. A. Hirsch, G. Wang and K. Struhl (2011). "Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion." Proc Natl Acad Sci U S A **108**(4): 1397-1402.

Inman, J. L., C. Robertson, J. D. Mott and M. J. Bissell (2015). "Mammary gland development: cell fate specification, stem cells and the microenvironment." Development **142**(6): 1028-1042.

Ito, M., Y. Liu, Z. Yang, J. Nguyen, F. Liang, R. J. Morris and G. Cotsarelis (2005). "Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis." Nat Med **11**(12): 1351-1354.

Jones, D. T., N. Jager, M. Kool, T. Zichner, B. Hutter, M. Sultan, Y. J. Cho, T. J. Pugh, V. Hovestadt, A. M. Stutz, T. Rausch, H. J. Warnatz, M. Ryzhova, S. Bender, D. Sturm, S. Pleier, H. Cin, E. Pfaff, L. Sieber, A. Wittmann, M. Remke, H. Witt, S. Hutter, T. Tzaridis, J. Weischenfeldt, B. Raeder, M. Avci, V. Amstislavskiy, M. Zapatka, U. D. Weber, Q. Wang, B. Lasitschka, C. C. Bartholomae, M. Schmidt, C. von Kalle, V. Ast,

C. Lawerenz, J. Eils, R. Kabbe, V. Benes, P. van Sluis, J. Koster, R. Volckmann, D. Shih, M. J. Betts, R. B. Russell, S. Coco, G. P. Tonini, U. Schuller, V. Hans, N. Graf, Y. J. Kim, C. Monoranu, W. Roggendorf, A. Unterberg, C. Herold-Mende, T. Milde, A. E. Kulozik, A. von Deimling, O. Witt, E. Maass, J. Rossler, M. Ebinger, M. U. Schuhmann, M. C. Fruhwald, M. Hasselblatt, N. Jabado, S. Rutkowski, A. O. von Bueren, D. Williamson, S. C. Clifford, M. G. McCabe, V. P. Collins, S. Wolf, S. Wiemann, H. Lehrach, B. Brors, W. Scheurlen, J. Felsberg, G. Reifenberger, P. A. Northcott, M. D. Taylor, M. Meyerson, S. L. Pomeroy, M. L. Yaspo, J. O. Korbel, A. Korshunov, R. Eils, S. M. Pfister and P. Lichter (2012). "Dissecting the genomic complexity underlying medulloblastoma." Nature **488**(7409): 100-105.

Jung, H. C., L. Eckmann, S. K. Yang, A. Panja, J. Fierer, E. Morzycka-Wroblewska and M. F. Kagnoff (1995). "A distinct array of proinflammatory cytokines is expressed in human colon epithelial cells in response to bacterial invasion." J Clin Invest **95**(1): 55-65.

Kirschmann, D. A., R. A. Lininger, L. M. Gardner, E. A. Seftor, V. A. Odero, A. M. Ainsztein, W. C. Earnshaw, L. L. Wallrath and M. J. Hendrix (2000). "Down-regulation of HP1Hsalpha expression is associated with the metastatic phenotype in breast cancer." Cancer Res **60**(13): 3359-3363.

Kiselev, V. Y., K. Kirschner, M. T. Schaub, T. Andrews, A. Yiu, T. Chandra, K. N. Natarajan, W. Reik, M. Barahona, A. R. Green and M. Hemberg (2017). "SC3: consensus clustering of single-cell RNA-seq data." Nat Methods **14**(5): 483-486.

Kopp, J. L., M. Grompe and M. Sander (2016). "Stem cells versus plasticity in liver and pancreas regeneration." Nat Cell Biol **18**(3): 238-245.

Kordon, E. C. and G. H. Smith (1998). "An entire functional mammary gland may comprise the progeny from a single cell." Development **125**(10): 1921-1930.

Koren, S. and M. Bentires-Alj (2015). "Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy." Mol Cell **60**(4): 537-546.

Koren, S., L. Reavie, J. P. Couto, D. De Silva, M. B. Stadler, T. Roloff, A. Britschgi, T. Eichlisberger, H. Kohler, O. Aina, R. D. Cardiff and M. Bentires-Alj (2015). "PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours." Nature **525**(7567): 114-118.

Korkaya, H., G. I. Kim, A. Davis, F. Malik, N. L. Henry, S. Ithimakin, A. A. Quraishi, N. Tawakkol, R. D'Angelo, A. K. Paulson, S. Chung, T. Luther, H. J. Paholak, S. Liu, K. A. Hassan, Q. Zen, S. G. Clouthier and M. S. Wicha (2012). "Activation of an IL6 inflammatory loop mediates trastuzumab resistance in HER2+ breast cancer by expanding the cancer stem cell population." Mol Cell **47**(4): 570-584.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Lee, H. J., R. A. Hinshelwood, T. Bouras, D. Gallego-Ortega, F. Valdes-Mora, K. Blazek, J. E. Visvader, S. J. Clark and C. J. Ormandy (2011). "Lineage specific methylation of the Elf5 promoter in mammary epithelial cells." Stem Cells **29**(10): 1611-1619.

Liang, G. and Y. Zhang (2013). "Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective." Cell Res **23**(1): 49-69.

Lim, E., D. Wu, B. Pal, T. Bouras, M. L. Asselin-Labat, F. Vaillant, H. Yagita, G. J. Lindeman, G. K. Smyth and J. E. Visvader (2010). "Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways." Breast Cancer Res **12**(2): R21.

Lin, B., P. Srikanth, A. C. Castle, S. Nigwekar, R. Malhotra, J. L. Galloway, D. B. Sykes and J. Rajagopal (2018). "Modulating Cell Fate as a Therapeutic Strategy." Cell Stem Cell **23**(3): 329-341.

Maherali, N. and K. Hochedlinger (2008). "Guidelines and techniques for the generation of induced pluripotent stem cells." Cell Stem Cell **3**(6): 595-605.

Mani, S. A., W. Guo, M. J. Liao, E. N. Eaton, A. Ayyanan, A. Y. Zhou, M. Brooks, F. Reinhard, C. C. Zhang, M. Shipitsin, L. L. Campbell, K. Polyak, C. Brisken, J. Yang and R. A. Weinberg (2008). "The epithelial-mesenchymal transition generates cells with properties of stem cells." Cell **133**(4): 704-715.

Margueron, R. and D. Reinberg (2011). "The Polycomb complex PRC2 and its mark in life." Nature **469**(7330): 343-349.

Marjanovic, N. D., R. A. Weinberg and C. L. Chaffer (2013). "Cell plasticity and heterogeneity in cancer." Clin Chem **59**(1): 168-179.

McNally, S. and T. Stein (2017). "Overview of Mammary Gland Development: A Comparison of Mouse and Human." Methods Mol Biol **1501**: 1-17.

Merrell, A. J. and B. Z. Stanger (2016). "Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style." Nat Rev Mol Cell Biol **17**(7): 413-425.

Michalak, E. M., K. Nacerddine, A. Pietersen, V. Beuger, I. Pawlitzky, P. Cornelissen-Steijger, E. Wientjens, E. Tanger, J. Seibler, M. van Lohuizen and J. Jonkers (2013). "Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal progenitor pool." Stem Cells **31**(9): 1910-1920.

Millanes-Romero, A., N. Herranz, V. Perrera, A. Iturbide, J. Loubat-Casanovas, J. Gil, T. Jenuwein, A. Garcia de Herreros and S. Peiro (2013). "Regulation of heterochromatin transcription by Snail1/LOXL2 during epithelial-to-mesenchymal transition." Mol Cell **52**(5): 746-757.

Molyneux, G., F. C. Geyer, F. A. Magnay, A. McCarthy, H. Kendrick, R. Natrajan, A. Mackay, A. Grigoriadis, A. Tutt, A. Ashworth, J. S. Reis-Filho and M. J. Smalley (2010). "BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells." Cell Stem Cell **7**(3): 403-417.

Morin, R. D., M. Mendez-Lago, A. J. Mungall, R. Goya, K. L. Mungall, R. D. Corbett, N. A. Johnson, T. M. Severson, R. Chiu, M. Field, S. Jackman, M. Krzywinski, D. W. Scott, D. L. Trinh, J. Tamura-Wells, S. Li, M. R. Firme, S. Rogic, M. Griffith, S. Chan, O. Yakovenko, I. M. Meyer, E. Y. Zhao, D. Smailus, M. Moksa, S. Chittaranjan, L. Rimsza, A. Brooks-Wilson, J. J. Spinelli, S. Ben-Neriah, B. Meissner, B. Woolcock, M. Boyle, H. McDonald, A. Tam, Y. Zhao, A. Delaney, T. Zeng, K. Tse, Y. Butterfield, I.

Birol, R. Holt, J. Schein, D. E. Horsman, R. Moore, S. J. Jones, J. M. Connors, M. Hirst, R. D. Gascoyne and M. A. Marra (2011). "Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma." Nature **476**(7360): 298-303.

Nasser, W., A. Amitai-Lange, D. Soteriou, R. Hanna, B. Tiosano, Y. Fuchs and R. Shalom-Feuerstein (2018). "Corneal-Committed Cells Restore the Stem Cell Pool and Tissue Boundary following Injury." Cell Rep **22**(2): 323-331.

Norwood, L. E., T. J. Moss, N. V. Margaryan, S. L. Cook, L. Wright, E. A. Seftor, M. J. Hendrix, D. A. Kirschmann and L. L. Wallrath (2006). "A requirement for dimerization of HP1Hsalpha in suppression of breast cancer invasion." J Biol Chem **281**(27): 18668-18676.

Okita, K., T. Ichisaka and S. Yamanaka (2007). "Generation of germline-competent induced pluripotent stem cells." Nature **448**(7151): 313-317.

Onder, T. T., N. Kara, A. Cherry, A. U. Sinha, N. Zhu, K. M. Bernt, P. Cahan, B. O. Marcarci, J. Unternaehrer, P. B. Gupta, E. S. Lander, S. A. Armstrong and G. Q. Daley (2012). "Chromatin-modifying enzymes as modulators of reprogramming." Nature **483**(7391): 598-602.

Paksa, A. and J. Rajagopal (2017). "The epigenetic basis of cellular plasticity." Curr Opin Cell Biol **49**: 116-122.

Pal, B., T. Bouras, W. Shi, F. Vaillant, J. M. Sheridan, N. Fu, K. Breslin, K. Jiang, M. E. Ritchie, M. Young, G. J. Lindeman, G. K. Smyth and J. E. Visvader (2013). "Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2." Cell Rep **3**(2): 411-426.

Panciera, T., L. Azzolin, A. Fujimura, D. Di Biagio, C. Frasson, S. Bresolin, S. Soligo, G. Basso, S. Bicciato, A. Rosato, M. Cordenonsi and S. Piccolo (2016). "Induction of Expandable Tissue-Specific Stem/Progenitor Cells through Transient Expression of YAP/TAZ." Cell Stem Cell **19**(6): 725-737.

Pasi, C. E., A. Dereli-Oz, S. Negrini, M. Friedli, G. Fragola, A. Lombardo, G. Van Houwe, L. Naldini, S. Casola, G. Testa, D. Trono, P. G. Pelicci and T. D. Halazonetis (2011). "Genomic instability in induced stem cells." Cell Death Differ **18**(5): 745-753.

Pathania, R., S. Ramachandran, S. Elangovan, R. Padia, P. Yang, S. Cinghu, R. Veeranan-Karmegam, P. Arjunan, J. P. Gnana-Prakasam, F. Sadanand, L. Pei, C. S. Chang, J. H. Choi, H. Shi, S. Manicassamy, P. D. Prasad, S. Sharma, V. Ganapathy, R. Jothi and M. Thangaraju (2015). "DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis." Nat Commun **6**: 6910.

Paulussen, A. D., A. P. Stegmann, M. J. Blok, D. Tserpelis, C. Posma-Velter, Y. Detisch, E. E. Smeets, A. Wagemans, J. J. Schrander, M. J. van den Boogaard, J. van der Smagt, A. van Haeringen, I. Stolte-Dijkstra, W. S. Kerstjens-Frederikse, G. M. Mancini, M. W. Wessels, R. C. Hennekam, M. Vreeburg, J. Geraedts, T. de Ravel, J. P. Fryns, H. J. Smeets, K. Devriendt and C. T. Schrander-Stumpel (2011). "MLL2 mutation spectrum in 45 patients with Kabuki syndrome." Hum Mutat **32**(2): E2018-2025.

Pietersen, A. M., B. Evers, A. A. Prasad, E. Tanger, P. Cornelissen-Steijger, J. Jonkers and M. van Lohuizen (2008). "Bmi1 regulates stem cells and proliferation and differentiation of committed cells in mammary epithelium." Curr Biol **18**(14): 1094-1099.

Poli, V., L. Fagnocchi, A. Fasciani, A. Cherubini, S. Mazzoleni, S. Ferrillo, A. Miluzio, G. Gaudioso, V. Vaira, A. Turdo, M. Gaggianesi, A. Chinnici, E. Lipari, S. Bicciato, S. Bosari, M. Todaro and A. Zippo (2018). "MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state." Nat Commun **9**(1): 1024.

Qian, B. Z. (2017). "Inflammation fires up cancer metastasis." Semin Cancer Biol **47**: 170-176.

Qin, H., A. Diaz, L. Blouin, R. J. Lebbink, W. Patena, P. Tanbun, E. M. LeProust, M. T. McManus, J. S. Song and M. Ramalho-Santos (2014). "Systematic identification of barriers to human iPSC generation." Cell **158**(2): 449-461.

Rais, Y., A. Zviran, S. Geula, O. Gafni, E. Chomsky, S. Viukov, A. A. Mansour, I. Caspi, V. Krupalnik, M. Zerbib, I. Maza, N. Mor, D. Baran, L. Weinberger, D. A. Jaitin, D. Lara-Astiaso, R. Blecher-Gonen, Z. Shipony, Z. Mukamel, T. Hagai, S. Gilad, D. Amann-Zalcenstein, A. Tanay, I. Amit, N. Novershtern and J. H. Hanna (2013). "Deterministic direct reprogramming of somatic cells to pluripotency." Nature **502**(7469): 65-70.

Rao, R. C. and Y. Dou (2015). "Hijacked in cancer: the KMT2 (MLL) family of methyltransferases." Nat Rev Cancer **15**(6): 334-346.

Reid, J. E. and L. Wernisch (2016). "Pseudotime estimation: deconfounding single cell time series." Bioinformatics **32**(19): 2973-2980.

Rios, A. C., N. Y. Fu, G. J. Lindeman and J. E. Visvader (2014). "In situ identification of bipotent stem cells in the mammary gland." Nature **506**(7488): 322-327.

Robinson, M. D. and A. Oshlack (2010). "A scaling normalization method for differential expression analysis of RNA-seq data." Genome Biol **11**(3): R25.

Sancho, R., R. Gruber, G. Gu and A. Behrens (2014). "Loss of Fbw7 reprograms adult pancreatic ductal cells into alpha, delta, and beta cells." Cell Stem Cell **15**(2): 139-153.

Santoro, A., T. Vlachou, L. Luzi, G. Melloni, L. Mazzarella, E. D'Elia, X. Aobuli, C. E. Pasi, L. R. , P. Bonetti, S. Punzi, L. Casoli, A. Sabò, M. C. Moroni, G. I. Dellino, B. Amati, F. Nicassio, L. Lanfrancone and P. G. Pelicci (in press). P53-loss in breast cancer leads to Myc activation, increased cell plasticity and expression of a mitotic signature with prognostic value.

Savagner, P. (2015). "Epithelial-mesenchymal transitions: from cell plasticity to concept elasticity." Curr Top Dev Biol **112**: 273-300.

Scheel, C., E. N. Eaton, S. H. Li, C. L. Chaffer, F. Reinhardt, K. J. Kah, G. Bell, W. Guo, J. Rubin, A. L. Richardson and R. A. Weinberg (2011). "Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast." Cell **145**(6): 926-940.

Schwitalla, S., A. A. Fingerle, P. Cammareri, T. Nebelsiek, S. I. Goktuna, P. K. Ziegler, O. Canli, J. Heijmans, D. J. Huels, G. Moreaux, R. A. Rupec, M. Gerhard, R. Schmid, N.

Barker, H. Clevers, R. Lang, J. Neumann, T. Kirchner, M. M. Taketo, G. R. van den Brink, O. J. Sansom, M. C. Arkan and F. R. Greten (2013). "Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties." Cell **152**(1-2): 25-38.

Shackleton, M., F. Vaillant, K. J. Simpson, J. Stingl, G. K. Smyth, M. L. Asselin-Labat, L. Wu, G. J. Lindeman and J. E. Visvader (2006). "Generation of a functional mammary gland from a single stem cell." Nature **439**(7072): 84-88.

Sharma, S. and G. Gurudutta (2016). "Epigenetic Regulation of Hematopoietic Stem Cells." Int J Stem Cells **9**(1): 36-43.

Shibue, T. and R. A. Weinberg (2017). "EMT, CSCs, and drug resistance: the mechanistic link and clinical implications." Nat Rev Clin Oncol **14**(10): 611-629.

Singh, S. K., M. K. Mishra, I. A. Eltoum, S. Bae, J. W. Lillard, Jr. and R. Singh (2018). "CCR5/CCL5 axis interaction promotes migratory and invasiveness of pancreatic cancer cells." Sci Rep **8**(1): 1323.

Sleeman, K. E., H. Kendrick, A. Ashworth, C. M. Isacke and M. J. Smalley (2006). "CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells." Breast Cancer Res **8**(1): R7.

Soady, K. J., H. Kendrick, Q. Gao, A. Tutt, M. Zvelebil, L. D. Ordonez, J. Quist, D. W. Tan, C. M. Isacke, A. Grigoriadis and M. J. Smalley (2015). "Mouse mammary stem cells express prognostic markers for triple-negative breast cancer." Breast Cancer Res **17**: 31.

Song, S. J., L. Poliseno, M. S. Song, U. Ala, K. Webster, C. Ng, G. Beringer, N. J. Brikbak, X. Yuan, L. C. Cantley, A. L. Richardson and P. P. Pandolfi (2013). "MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling." Cell **154**(2): 311-324.

Sridharan, R., J. Tchieu, M. J. Mason, R. Yachechko, E. Kuoy, S. Horvath, Q. Zhou and K. Plath (2009). "Role of the murine reprogramming factors in the induction of pluripotency." Cell **136**(2): 364-377.

Stange, D. E., B. K. Koo, M. Huch, G. Sibbel, O. Basak, A. Lyubimova, P. Kujala, S. Bartfeld, J. Koster, J. H. Geahlen, P. J. Peters, J. H. van Es, M. van de Wetering, J. C. Mills and H. Clevers (2013). "Differentiated Troy+ chief cells act as reserve stem cells to generate all lineages of the stomach epithelium." Cell **155**(2): 357-368.

Stingl, J., P. Eirew, I. Ricketson, M. Shackleton, F. Vaillant, D. Choi, H. I. Li and C. J. Eaves (2006). "Purification and unique properties of mammary epithelial stem cells." Nature **439**(7079): 993-997.

Street, K., D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom and S. Dudoit (2018). "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics." BMC Genomics **19**(1): 477.

Sun, X., J. C. Chuang, M. Kanchwala, L. Wu, C. Celen, L. Li, H. Liang, S. Zhang, T. Maples, L. H. Nguyen, S. C. Wang, R. A. Signer, M. Sorouri, I. Nassour, X. Liu, J. Xu, M. Wu, Y. Zhao, Y. C. Kuo, Z. Wang, C. Xing and H. Zhu (2016). "Suppression of the

SWI/SNF Component Arid1a Promotes Mammalian Regeneration." Cell Stem Cell **18**(4): 456-466.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Tata, P. R., H. Mou, A. Pardo-Saganta, R. Zhao, M. Prabhu, B. M. Law, V. Vinarsky, J. L. Cho, S. Breton, A. Sahay, B. D. Medoff and J. Rajagopal (2013). "Dedifferentiation of committed epithelial cells into stem cells in vivo." Nature **503**(7475): 218-223.

Tetteh, P. W., H. F. Farin and H. Clevers (2015). "Plasticity within stem cell hierarchies in mammalian epithelia." Trends Cell Biol **25**(2): 100-108.

Thiery, J. P. (2003). "Epithelial-mesenchymal transitions in development and pathologies." Curr Opin Cell Biol **15**(6): 740-746.

Thomsen, R., D. B. Christensen, S. Rosborg, T. E. Linnet, J. Blechingberg and A. L. Nielsen (2011). "Analysis of HP1alpha regulation in human breast cancer cells." Mol Carcinog **50**(8): 601-613.

Thorel, F., N. Damond, S. Chera, A. Wiederkehr, B. Thorens, P. Meda, C. B. Wollheim and P. L. Herrera (2011). "Normal glucagon signaling and beta-cell function after near-total alpha-cell ablation in adult mice." Diabetes **60**(11): 2872-2882.

Tian, H., B. Biehs, S. Warming, K. G. Leong, L. Rangell, O. D. Klein and F. J. de Sauvage (2011). "A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable." Nature **478**(7368): 255-259.

Trapnell, C., L. Pachter and S. L. Salzberg (2009). "TopHat: discovering splice junctions with RNA-Seq." Bioinformatics **25**(9): 1105-1111.

Utikal, J., J. M. Polo, M. Stadtfeld, N. Maherali, W. Kulalert, R. M. Walsh, A. Khalil, J. G. Rheinwald and K. Hochedlinger (2009). "Immortalization eliminates a roadblock during cellular reprogramming into iPS cells." Nature **460**(7259): 1145-1148.

Vad-Nielsen, J., K. R. Jakobsen, T. F. Daugaard, R. Thomsen, A. Brugmann, B. S. Sorensen and A. L. Nielsen (2016). "Regulatory dissection of the CBX5 and hnRNPA1 bi-directional promoter in human breast cancer cells reveals novel transcript variants differentially associated with HP1alpha down-regulation in metastatic cells." BMC Cancer **16**: 32.

Vad-Nielsen, J. and A. L. Nielsen (2015). "Beyond the histone tale: HP1alpha deregulation in breast cancer epigenetics." Cancer Biol Ther **16**(2): 189-200.

van Amerongen, R., A. N. Bowman and R. Nusse (2012). "Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland." Cell Stem Cell **11**(3): 387-400.

van Es, J. H., T. Sato, M. van de Wetering, A. Lyubimova, A. N. Yee Nee, A. Gregorieff, N. Sasaki, L. Zeinstra, M. van den Born, J. Korving, A. C. M. Martens, N. Barker, A. van Oudenaarden and H. Clevers (2012). "Dll1+ secretory progenitor cells revert to stem cells upon crypt damage." Nat Cell Biol **14**(10): 1099-1104.

Van Keymeulen, A., M. Y. Lee, M. Ousset, S. Brohee, S. Rorive, R. R. Giraddi, A. Wuidart, G. Bouvencourt, C. Dubois, I. Salmon, C. Sotiriou, W. A. Phillips and C. Blanpain (2015). "Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity." Nature **525**(7567): 119-123.

Van Keymeulen, A., A. S. Rocha, M. Ousset, B. Beck, G. Bouvencourt, J. Rock, N. Sharma, S. Dekoninck and C. Blanpain (2011). "Distinct stem cells contribute to mammary gland development and maintenance." Nature **479**(7372): 189-193.

Varga, J. and F. R. Greten (2017). "Cell plasticity in epithelial homeostasis and tumorigenesis." Nat Cell Biol **19**(10): 1133-1141.

Visvader, J. E. (2009). "Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis." Genes Dev **23**(22): 2563-2577.

Visvader, J. E. and H. Clevers (2016). "Tissue-specific designs of stem cell hierarchies." Nat Cell Biol **18**(4): 349-355.

Visvader, J. E. and J. Stingl (2014). "Mammary stem cells and the differentiation hierarchy: current status and perspectives." Genes Dev **28**(11): 1143-1158.

Wahl, G. M. and B. T. Spike (2017). "Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity." NPJ Breast Cancer **3**: 14.

Wang, C., J. E. Lee, B. Lai, T. S. Macfarlan, S. Xu, L. Zhuang, C. Liu, W. Peng and K. Ge (2016). "Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition." Proc Natl Acad Sci U S A **113**(42): 11871-11876.

Wang, D., C. Cai, X. Dong, Q. C. Yu, X. O. Zhang, L. Yang and Y. A. Zeng (2015). "Identification of multipotent mammary stem cells by protein C receptor expression." Nature **517**(7532): 81-84.

Wang, L., S. Wang and W. Li (2012). "RSeQC: quality control of RNA-seq experiments." Bioinformatics **28**(16): 2184-2185.

Wang, R., K. Chadalavada, J. Wilshire, U. Kowalik, K. E. Hovinga, A. Geber, B. Fligelman, M. Leversha, C. Brennan and V. Tabar (2010). "Glioblastoma stem-like cells give rise to tumour endothelium." Nature **468**(7325): 829-833.

Wasenius, V. M., S. Hemmer, E. Kettunen, S. Knuutila, K. Franssila and H. Joensuu (2003). "Hepatocyte growth factor receptor, matrix metalloproteinase-11, tissue inhibitor of metalloproteinase-1, and fibronectin are up-regulated in papillary thyroid carcinoma: a cDNA and tissue microarray study." Clin Cancer Res **9**(1): 68-75.

Wuestefeld, T., M. Pesic, R. Rudalska, D. Dauch, T. Longerich, T. W. Kang, T. Yevsa, F. Heinzmann, L. Hoenicke, A. Hohmeyer, A. Potapova, I. Rittelmeier, M. Jarek, R. Geffers, M. Scharfe, F. Klawonn, P. Schirmacher, N. P. Malek, M. Ott, A. Nordheim, A. Vogel, M. P. Manns and L. Zender (2013). "A Direct in vivo RNAi screen identifies MKK4 as a key regulator of liver regeneration." Cell **153**(2): 389-401.

Yamamura, M. (1992). "Defining protective responses to pathogens: cytokine profiles in leprosy lesions." Science **255**(5040): 12.

Yang, M. H., M. Z. Wu, S. H. Chiou, P. M. Chen, S. Y. Chang, C. J. Liu, S. C. Teng and K. J. Wu (2008). "Direct regulation of TWIST by HIF-1alpha promotes metastasis." Nat Cell Biol **10**(3): 295-305.

Yanger, K., Y. Zong, L. R. Maggs, S. N. Shapira, R. Maddipati, N. M. Aiello, S. N. Thung, R. G. Wells, L. E. Greenbaum and B. Z. Stanger (2013). "Robust cellular reprogramming occurs spontaneously during liver regeneration." Genes Dev **27**(7): 719-724.

Ye, X. and R. A. Weinberg (2015). "Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression." Trends Cell Biol **25**(11): 675-686.

Yoo, K. H., S. Oh, K. Kang, C. Wang, G. W. Robinson, K. Ge and L. Hennighausen (2016). "Histone Demethylase KDM6A Controls the Mammary Luminal Lineage through Enzyme-Independent Mechanisms." Mol Cell Biol **36**(16): 2108-2120.

Yu, Y. H., G. Y. Chiou, P. I. Huang, W. L. Lo, C. Y. Wang, K. H. Lu, C. C. Yu, G. Alterovitz, W. C. Huang, J. F. Lo, H. S. Hsu and S. H. Chiou (2012). "Network biology of tumor stem-like cells identified a regulatory role of CBX5 in lung cancer." Sci Rep **2**: 584.

Zaidan, N. Z., K. J. Walker, J. E. Brown, L. V. Schaffer, M. Scalf, M. R. Shortreed, G. Iyer, L. M. Smith and R. Sridharan (2018). "Compartmentalization of HP1 Proteins in Pluripotency Acquisition and Maintenance." Stem Cell Reports **10**(2): 627-641.

Zaidi, S., M. Choi, H. Wakimoto, L. Ma, J. Jiang, J. D. Overton, A. Romano-Adesman, R. D. Bjornson, R. E. Breitbart, K. K. Brown, N. J. Carriero, Y. H. Cheung, J. Deanfield, S. DePalma, K. A. Fakhro, J. Glessner, H. Hakonarson, M. J. Italia, J. R. Kaltman, J. Kaski, R. Kim, J. K. Kline, T. Lee, J. Leipzig, A. Lopez, S. M. Mane, L. E. Mitchell, J. W. Newburger, M. Parfenov, I. Pe'er, G. Porter, A. E. Roberts, R. Sachidanandam, S. J. Sanders, H. S. Seiden, M. W. State, S. Subramanian, I. R. Tikhonova, W. Wang, D. Warburton, P. S. White, I. A. Williams, H. Zhao, J. G. Seidman, M. Brueckner, W. K. Chung, B. D. Gelb, E. Goldmuntz, C. E. Seidman and R. P. Lifton (2013). "De novo mutations in histone-modifying genes in congenital heart disease." Nature **498**(7453): 220-223.

Zanconato, F., G. Battilana, M. Cordenonsi and S. Piccolo (2016). "YAP/TAZ as therapeutic targets in cancer." Curr Opin Pharmacol **29**: 26-33.

Zang, Z. J., I. Cutcutache, S. L. Poon, S. L. Zhang, J. R. McPherson, J. Tao, V. Rajasegaran, H. L. Heng, N. Deng, A. Gan, K. H. Lim, C. K. Ong, D. Huang, S. Y. Chin, I. B. Tan, C. C. Ng, W. Yu, Y. Wu, M. Lee, J. Wu, D. Poh, W. K. Wan, S. Y. Rha, J. So, M. Salto-Tellez, K. G. Yeoh, W. K. Wong, Y. J. Zhu, P. A. Futreal, B. Pang, Y. Ruan, A. M. Hillmer, D. Bertrand, N. Nagarajan, S. Rozen, B. T. Teh and P. Tan (2012). "Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes." Nat Genet **44**(5): 570-574.

Zou, M. R., J. Cao, Z. Liu, S. J. Huh, K. Polyak and Q. Yan (2014). "Histone demethylase jumonji AT-rich interactive domain 1B (JARID1B) controls mammary gland development by regulating key developmental and lineage specification genes." J Biol Chem **289**(25): 17620-17633.

ياشلىق، ئۇگەن

لۇتپۇللا مۇتەللىپ

ياشلىق، تىرىش-تىرماش ھارماي ئۇگەن،
ئىلىمنىڭ دېڭىزى قاينام ھەم چوڭقۇر.
نادانلىق ئۇ ، ھەممىنى بىلدىم دېگەن،
كۆرۈپ... ۋاراقلاپلا بىر قۇر.

ئىلىم مۇجىزىلەرنىڭ سىرىن ئاچقۇچى،
بىر كۈن ياشاش ئۈچۈن ئوقى ئون يىل!
كۆزلا ئەمەس ئالدىڭغا نۇر چاچقۇچى،
قاراڭغۇنى يورىتىدۇ يورۇق دىل.

ھەربىر ھادىسە يوشۇرۇن سىر ساقلايدۇ،
توقۇلمىلىرى ئۆزنىڭ ئاجايىپ.
ئىلىم ھارماس مەرت ئادەمنى ياقلايدۇ،
نادانىنىڭ كۆزىدىن ئۇ ھەرقاچان غايىپ.

راھىپ بول ئىلىمنىڭ نازاكەت چۆلىدە،
مۇجىزىلەر گۈمبىزىگە «شەيخ»
كۈن-تۈن زىكرى قىل شۇلار تۆۋدە،
ئىجتىھات سېنى دېگۈزمەس دەرىخ.

ياشلىق، چاقماقسەن، چېقىپ ئۆچسەن،
ئەۋرىشىم چېغىڭ مۇشۇ، ئۇگەنگىن، ئۇگەن!
قاراڭغۇ دىل چۆلىگە بىر كۈن چۆكسەن،
ئادەم بولۇش دۇلدۇلىغا سالمىساڭ يۈگەن!

1943-يىلى ئاۋغۇست، ئۈرۈمچى