Genetics

http://www.aimspress.com/journal/Genetics

Review

Uncovering the stem cell hierachy by genetic lineage tracing in the mammary gland

Liliana Osório[1,2], Fei Long[1,2] and Zhongjun Zhou[1,2,]*

[1] School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong Hong Kong, China
[2] Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen, China

* Correspondence: Email: zhongjun@hku.hk; Tel: +852-3917-9542.

Abstract: The mammary gland is the distinct feature that gives the name to the class of mammals and distinguishes them from other animals. Functionally, the mammary gland is a secretory organ which main role is to produce milk to nourish the offspring. Organogenesis of the mammary gland starts during embryogenesis but occurs mainly after birth at puberty under the influence of hormonal cues. Throughout the adult life as well as during pregnancy, the mammary gland shows a remarkable regenerative ability, thus constituting an excellent model for studying stem cell biology. Although the mammary gland consists of a relatively simple epithelial structure with a luminal and a basal cell layers, these are indeed composed by distinct subsets of mammary epithelial cells. Flow cytometry and transplantation assay have identified several subpopulations of stem and/or progenitor cells in the mammary gland. Yet, physiological and developmental relevant information can only be obtained when investigating the stem cell hierarchy in the intact mammary gland. Genetic lineage tracing studies have offered unprecedented levels of information regarding the organization of the stem cell compartment and possible role of resident stem and/or progenitor cells at different stages of the mammary gland organogenesis. These studies, although creating a passionate debate, highlight the existence of heterogeneous stem cell compartment, where bipotent as well as unipotent mammary stem cells seems to co-exist. Genetic lineage tracing experiments provide relevant information on stem cells that are key for understanding both normal development as well as associated pathologies in human. It holds the promise of providing new insights into the cell-of-origin and heterogeneity of breast tumorigenesis.

Keywords: Mammary gland; mammary stem cell hierarchy; genetic lineage tracing; mouse

Abbreviations

MEC: mammary epithelial cells; FACS: fluorescence-activated cell sorting.

1. Introduction

The mammary gland is a secretory organ, which main function is to produce milk that provides nutrition and immunological protection to the offspring. It is the distinct feature that gives the name to the class of mammals and distinguishes mammals from the other animals. Mammary glands are epidermal appendages derived from the ventral skin. Organogenesis of the mammary gland begins during embryogenesis but occurs mainly after birth [1]. In the mouse, mammary gland development starts at mid-gestation with the formation of the milk line, followed by generation of the mammary placodes and mammary buds. The mammary buds generate a rudimentary ductal structure that remains quiescent until puberty. At around three to four weeks in the mouse, under influence of the ovarian hormones such as estrogen, the ends of the rudimentary ducts proliferate and form distinct multilayered epithelial structures known as the terminal end buds. These undergo successive rounds of elongation and bifurcation. Lateral branching occurs concomitantly. The mammary gland grows until it invades the entire fat pad and forms an elaborated epithelial tree, which can be observed at around nine to ten weeks in the mouse. In the adults, the mammary epithelium undergoes recurring rounds of cell proliferation and cell apoptosis, with each ovarian estrous cycle. Fully development of the mammary gland only takes place after pregnancy. During alveologenesis, there is a dramatic expansion of the epithelial tree and formation of the alveoli. Suckling stimulates lactogenic differentiation, resulting in milk secretion into the lumen of the alveoli. When the stimuli for milk production are lost at the weaning, a process called involution occurs. Involution removes the milk-producing cells after lactation and remodels the mammary gland back to its pre-pregnancy state. Pregnancy, lactation and involution constitute a remarkable example of the remodeling ability of the mammary gland. It is widely acknowledged that such extensive regenerative capacity is maintained by stem cells, but their nature and location is still a subject of passionate debate. A better understanding of the stem and progenitor cell populations provides useful insight into the mechanisms that sustain tissue turnover and repair upon injury as well as into human diseases such as cancer.

2. Mammary epithelial cell types

The mammary gland is an epithelial tree embedded in the mammary fat pad. The fat pad is a complex stroma containing a variety of cell types, such as adipocytes, fibroblasts, immune and vascular cells. As in other glandular tissues, they work together to generate and maintain a functional organ. The importance of the stromal cells to mammary gland organogenesis has been reviewed elsewhere [10]. The mammary gland itself is a relatively simple epithelial structure traditionally described as being composed by the apically orientated luminal layer that lines the ducts, surrounded by an outer layer of basal myoepithelial cells in contact with the basement membrane [1]. Luminal cells express keratin 8 (K8) and keratin 18 (K18) while basal cells are positive for keratin 5 (K5) and keratin 14 (K14) as well as smooth muscle actin (SMA). Alveolar cells located in the luminal layer express wheat acidic protein (WAP) and beta-casein (CSN2). There is increasingly evidence that the

mammary epithelium is indeed more complex than initially thought. The first evidence that several distinct cell types compose the mammary epithelium came from ultrastructural studies in which five different subtypes of mammary epithelial cells (MEC) were described: classical luminal and myoepithelial cells, undifferentiated and differentiated large light cells and small light cells [2]. Additional work has provided genetic evidence that MEC are constituted by at least four independent lineages: the classical luminal and myoepithelial cells, L-cell lineage and S-cell lineage [3]. S-cells are important for the spatial placement of the tertiary branches, while L-cells coordinate formation of alveolar lumen and spatial organization of alveolar clusters. Over the years, fluorescence-activated cell sorting (FACS) has allowed characterization of distinct immunophenotypes of MEC [4]. These experiments provided great evidence for the existence of several epithelial cell subsets, which express a particular combination of cell surface markers, in the mammary gland. For instance, the luminal cell population seems to be composed by distinct stem or progenitor subpopulations as well as mature cells that can be distinguished by markers such as CD61, SCA1 and CD14. A heterogeneous cell population is also found in the basal myoepithelial cell layer. Although the cellular complexity of the mammary epithelium is now broadly accepted, there are still a few pertinent questions that remain open. Of particular interest, the different approaches have revealed distinct MEC subtypes but their relationship across the different studies has not yet been established. Does the small light cells described by Chepko and Smith correspond to the S-cell lineage identified by Sale and colleagues? Whether there is an overlap between the several subtypes of MEC waits to be demonstrated experimentally.

3.   Mammary stem cells in transplantation assay

The existence of stem cells in the mammary gland has long been appreciated. The gold standard method that has been extensively used for assessing stem cell activity in the mammary gland is the transplantation assay. A seminal paper published by DeOme and colleagues, more than half-century ago, provided the first indication of the presence of stem cells in the mammary gland. In their work, they show that transplantation of small fragments of mammary gland into de-epithelized (cleared) mammary fat pad gives rise to the entire epithelial tree [5]. The progeny of the transplanted cells could be serially transplanted, supporting the existence of MEC that have self-renewing properties and multi-lineage potential. Subsequent work has showed that transplantation of dissociated MEC at limiting dilutions gives rise to ductal, lobular or mixed ductal and lobular outgrowths [6], suggesting the presence of lineage-restricted progenitors in the mammary epithelium. Clonal analysis of retrovirally-labeled MEC suggested that a single multipotent stem cell was capable of reconstituting the entire mammary epithelium and might account for the renewal of the mammary epithelium over several transplant generations [7]. Subsequent studies using FACS provided functional evidence that a single cell from a mature mammary gland could indeed reconstitute the mammary epithelium when transplanted into a cleared mammary fat pad [8,9]. The cells with this ability are referred as mammary repopulation units (MRU) and they constitute a subpopulation of the basal cells. Following studies aimed at refining the surface markers of the MRU leading to the identification of several of these populations [10]. These show a great variability in the cell-surface markers they express as well as their frequency. This seems to be due to differences in both the experimental design, such as the donor age and transplant conditions, as well as in the methods of MEC isolation [11]. Altogether, these studies highlight the remarkable regenerative capacity of the MRU

but whether they contribute to normal mammary gland development remains extensively debated. There is a growing concern that transplantation, which involves loss of the mammary tissue architecture and wound generation, might lead transplanted MEC to behave in a way that do not reflect their physiological developmental fate [12-15]. On the other hand, some may argue that the full potential of a given cell is only manifested at a very specific developmental time point in vivo but that its complete lineage potential is unleashed in the transplantation assay [4].

4. Genetic lineage tracing

Studying stem cell activity in physiological developmental contexts is of paramount importance in any field, but particularly relevant for epithelial tissues such as mammary gland. The methods that allow stem cell studies in intact tissues or organisms have been commonly referred as lineage tracing (or lineage cell fate) [16]. Lineage tracing involves labeling of a single cell with a heritable, permanent molecular 'mark' that allows the identification of all its progeny and thus track of its fate. Lineage tracing has been extensively used in studies of developmental biology and its origin dates back to the nineteenth-century. Any lineage 'mark' should not change the properties of the marked cell, its progeny and its neighbors; it should also be passed to all progeny of the founder cell and be retained over time, without being transferred to unrelated, neighbor cells. Genetic lineage tracing uses genetic recombination to label the cells of interest and it has been widely used in the mammary gland. A recombinase enzyme is expressed in a cell-specific manner to activate the expression of a conditional reporter gene. Upon cell division, the labeled cells transmit the reporter gene to their daughter cells in a way that all the progeny of the marked cell is permanently labeled. One of the biggest strengths of this method is the considerable amount of promoters readily available to drive expression of the recombinase enzyme, in combination with the tremendous amount of conditional reporters that can be used.

4.1. Constitutive Cre-LoxP recombination system

The Cre-LoxP recombination system, derived from the bacteriophage P1, has been the preferred choice for most of the genetic lineage tracing studies in mice. Cre recombinase is an enzyme that recognizes specific genetic sequences known as LoxP sites, excises the DNA sequence that lies between the two LoxP sites and rejoins their exposed ends [17]. The first step to the use this system for lineage tracing studies involves the generation of a mouse line in which Cre is expressed under the control of a cell-specific promoter. This mouse line is then crossed with a second mouse line in which a reporter is flanked by a LoxP-STOP-LoxP sequence. In animals expressing both constructs, Cre specifically activates the reporter in cells that express the promoter by excising the STOP sequence (Figure 1). Reporter constructs are usually expressed from the *Rosa26* locus. The first reporter used was beta-galactosidase [18] but a great amount of fluorescent reporters, such as EGFP or YFP, are now available. Although used less frequently, another recombination system derived from *Saccharomyces cerevisiae* that uses FLP recombinase and FRT sites is available [19]. Since its first application in mice, the Cre-LoxP system has seen many variations that are further detailed in the next sections.
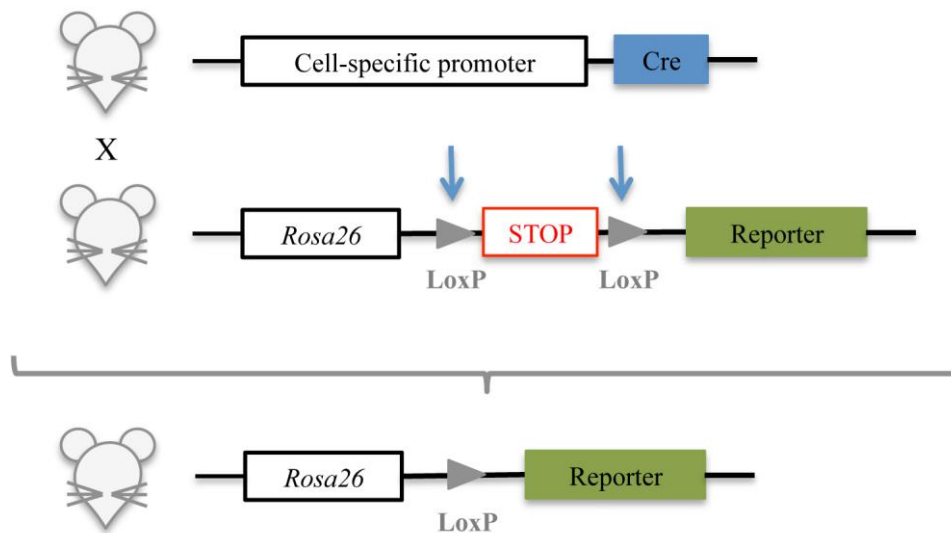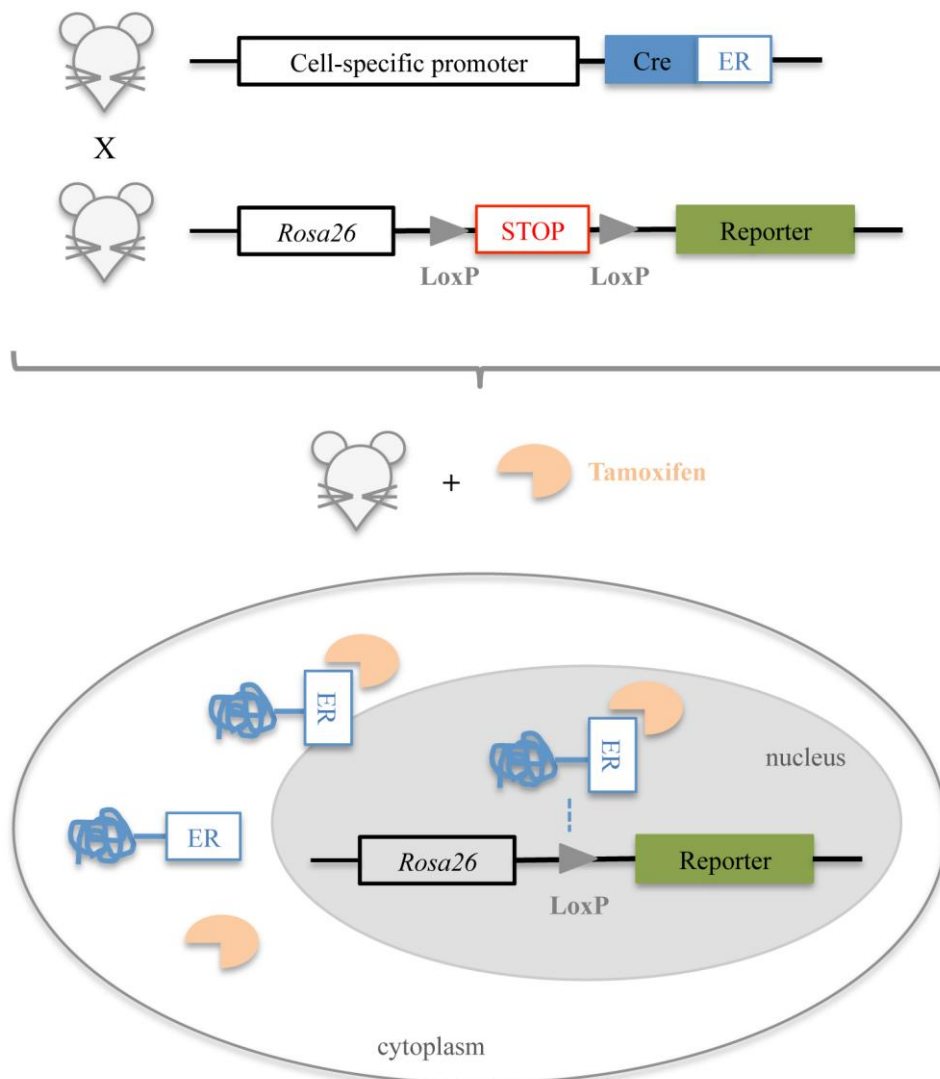
Figure 1. Cre-LoxP recombination system. Cre recombinase is an enzyme that recognizes specific genetic sequences known as LoxP sites. Cre excises the DNA sequence that lies between the two LoxP sites and rejoins their exposed ends. A mouse line in which Cre is expressed under the control of a cell-specific promoter is crossed with a second mouse line in which a reporter, flanked by a LoxP-STOP-LoxP sequence, has been inserted into the *Rosa26* locus. In animals expressing both constructs, Cre specifically activates the reporter in cells that express the promoter by excising the STOP sequence.

4.2. Inducible Cre-LoxP recombination systems

The use of a specific promoter that drives constitutive expression of Cre and thus of the reporter may lead to misinterpretation and inconclusive results. One can easily understand that, if distinct cell populations sequentially express the same promoter, the cells positive for the reporter not only derive from the initial population but also represent *de novo* genetic recombination in unrelated cells. Hence, temporal and spatial control of the Cre activity is fundamental in studies of genetic lineage tracing. Several inducible Cre-LoxP systems have been developed for this purpose. With these systems and as an example, activation of Cre can be induced selectively in the adult mice via a promoter that is also expressed during embryonic development. In the most commonly used system, Cre recombinase has been fused to the human estrogen receptor (ER). In the absence of ligands (estrogen 17beta-estradiol, anti-estrogen tamoxifen or its active metabolite 4-OHT), the CreET is kept in the cytoplasm. Binding of the ligand to the ER causes its conformational change followed by its translocation to the nucleus, where Cre can recombine the LoxP sites [20] (Figure 2).

Since only exposure to tamoxifen will induce Cre activity, this system allows both spatial and temporal control of Cre expression. CreER^{T2} is an improved version of the original construct, in which a mutated ER allows for increased sensitivity to low doses of tamoxifen while displaying reduced activation by endogenous mouse 17beta-estradiol [21]. These two parameters should be carefully taken into consideration when designing any experiment using the inducible Cre-LoxP system. The dose of tamoxifen needs to be optimized for each Cre mouse line and this is particularly relevant for the studies in the mammary gland, as estrogen is indispensable for proper mammary gland development [22]. It is

thus essential to use transient and low doses of tamoxifen since higher doses impair mammary gland development [23]. It is also of great importance to evaluate the 'leakiness' (extent of non-specific activation) that occurs in the absence of tamoxifen in these systems.

Figure 2. Inducible Cre-LoxP recombination system. Cre recombinase has been fused to the human estrogen receptor (ER). In the absence of ligands, the CreET is kept in the cytoplasm. After administration of tamoxifen, tamoxifen binds to ER, causing its conformational change followed by translocation of the CreER fusion protein to the nucleus. Cre recombine the LoxP sites and activates the expression of the reporter.

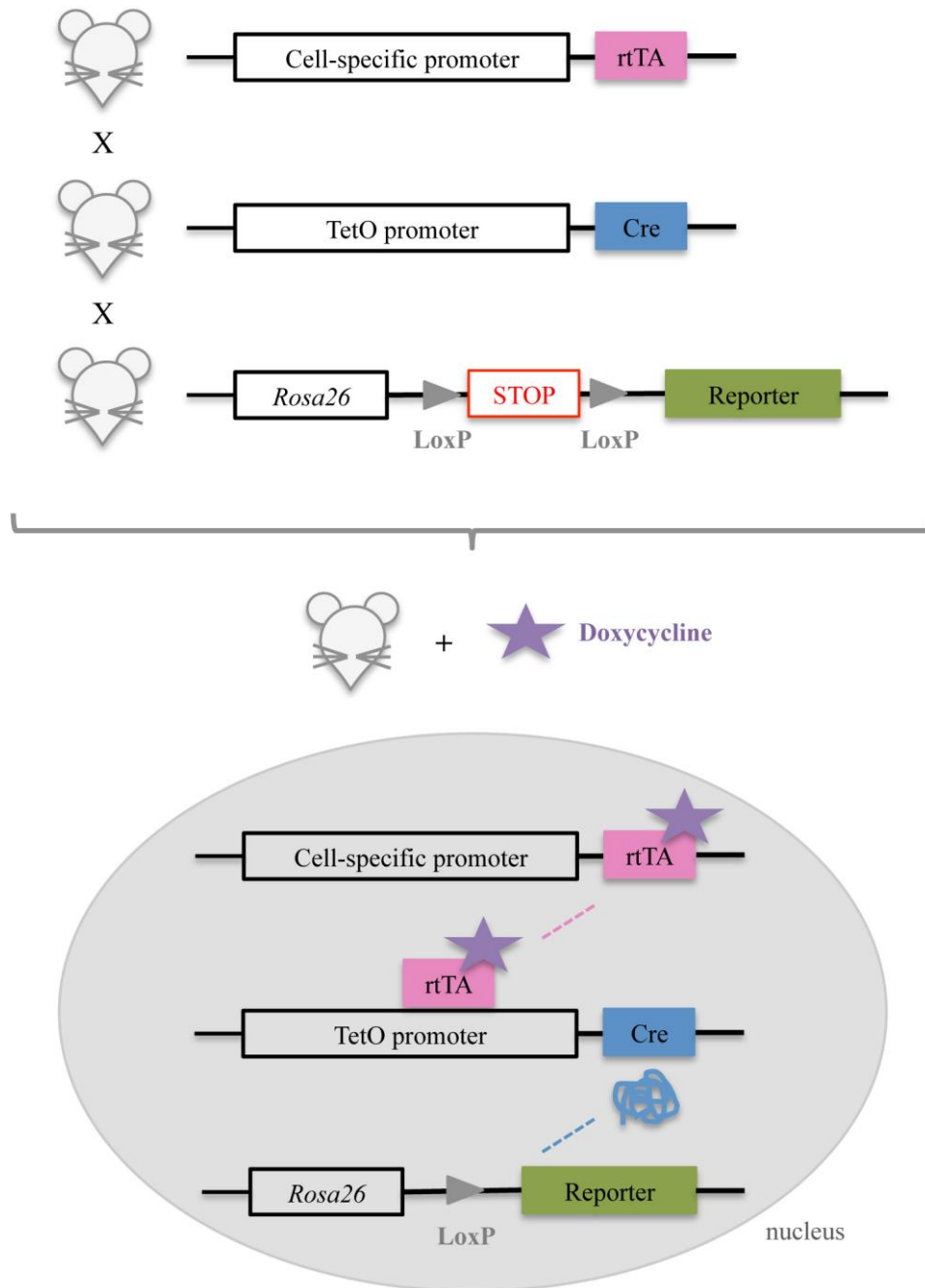## 4.3. Inducible and reversible Tet recombination systems

The 'Tet' inducible systems are based on the tetracycline bacterial resistance gene operon, TetO. In these systems, administration of tetracycline (or its analogue doxycycline) either allows or prevents gene expression, known as 'Tet-ON' or 'Tet-OFF', respectively [24,25]. We will focus on the Tet-ON system, which has been used in genetic lineage tracing experiments in the mammary

gland. The first step involves the creation of a mouse line in which a cell-specific promoter drives the expression of a reverse tetracycline-controlled activator (rtTA). In a second mouse line, the rtTA-dependent TetO promoter controls Cre expression. Upon ligation by doxycycline, the rtTA will bind to the tetracycline response elements (TRE) of the TetO promoter, which drives Cre expression. Cre will then excise the STOP sequence and allow expression of the reporter (Figure 3). Improved Tet systems that include transactivators with less toxicity, improved ligand sensitivity and reduced leakiness, such as rtTA^2 and rtTA^3, have also been generated [26]. An important hallmark of the Tet systems is that allow transient and reversible gene expression: Cre expression only occurs in the presence of doxycycline and it is reversed after doxycycline withdrawal. Reminiscent of the inducible Cre-LoxP system, the dose of doxycycline to be administrated to the mice as well as the leakiness of the system are key factors to consider when designing experiments using the Tet systems.

4.4. The choice of the reporter

One of the major strengths of the genetic lineage tracing approaches is related to the incredible variety of reporter mouse lines that are currently available. The first reporter mouse line created uses the *Escherichia coli* LacZ gene that encodes beta-galactosidase [18]. Beta-galactosidase produces an intense blue precipitate when incubated with the substrate analogue X-gal. In addition to the technical difficulties related to X-gal staining and production of variable results, this approach cannot be used in live cells. Development of the fluorescent reporter mouse lines has revolutionized the field of genetic lineage tracing and they are now the norm. Their main advantages lie on the facts that they can be directly detected by epifluorescence, with no need for antibody-based methods, and be used in live imaging experiments. Common used fluorescent reporters include EGFP [27], EYFP and ECFP [28] and tdTomato [29].

The choice of the appropriate fluorescent reporter depends on the type of the analyses to be performed. For example, tdTomato is one of the brightest fluorescent proteins currently available. Its strong epifluorescence can be either an advantage or a limitation: it may be useful when tracking a small number of labeled cells within a tissue but it may be problematic if used in FACS experiments. Multicolor reporter mouse lines have become popular as a tool to examine how different cell types contribute to the development and homeostasis of a given tissue. The first constructs employed two colors, such as EGFP and LacZ [30] or tdTomato and EGFP [31]. The most recent multicolor reporter mouse line, the *Rosa26R*-Confetti mice, which is derived from the 'Brainbow' and 'Rainbow' mice [32], enables the combined expression of four fluorescent proteins in a stochastic manner; after random recombination, up to ten colors can be generated using the Confetti reporter [33]. This reporter mouse may be ideal for stem cell lineage analysis as it allows the labeling of distinct stem cells within the same tissue with distinct colors and monitor their progeny.

Figure 3. Inducible and reversible Tet recombination system. The reverse tetracycline-controlled activator (rtTA) is expressed under the control of a cell-specific promoter in one mouse line. In a second mouse line, the rtTA-dependent TetO promoter controls Cre expression. Upon ligation by doxycycline, the rtTA will bind to the TetO promoter, which in turn drives Cre expression. Cre will then excise the STOP sequence and allow expression of the reporter.

5. Genetic lineage tracing in the mammary gland: Cre and rtTA mouse lines

In any genetic lineage tracing study, a major challenge is to engineer mice that express Cre/rtTA under the control of a promoter that is active only in the cell of interest. A list of the mouse strains

that have been used for lineage tracing in the mammary gland is summarized in Table 1. The choice of the promoters used to generate these mouse lines has been based on several criteria: markers of the major cell lineages of the mammary epithelium, such as K5 [12,23,24], K8 [12], K18 [35], K14 [36-38], ACTA2 [39], Elf5 [23] and WAP [40]; markers of stem cells that have been identified in other tissues, such as Lgr5 [30] and Procr [41]; members of the signaling pathways Notch and WNT, such as Notch2, Notch3 [42] and Axin2 [13], which are known to be important for stem cell activity and lineage specification in other tissues. A careful characterization of the promoter expression pattern is a pre-requisite for an accurate interpretation of the lineage tracing results. Of interest, one should keep in mind that the same promoter can be expressed by distinct MEC subsets and/or, it can switch on and off in a developmental stage-dependent manner in a particular MEC subset. Another important aspect is to determine how faithfully the expression of the promoter is reproduced by Cre activity, which can be influenced by several factors. Among these, the strategy used to generate the mouse Cre lines as well as the dose and the timing of drug administration, when using the inducible systems, seem to be of critical importance in regulating Cre activity and are discussed in further detail below. As summarized in Table 1, a wide variety of strategies have been used to generate the different mouse Cre lines: some of these are transgenic or BAC transgenic whereas others are knock-in. All have advantages and limitations that should be considered. Generation of knock-in mice is expensive and time-consuming, but knock-in alleles are more likely to reproduce the endogenous gene expression pattern. Transgenic mice are generated much faster and cheaper but normally only include minimal promoter regions and both positional effects and transgene copy number are likely to influence Cre activity. BAC transgenes normally contain large segments of genomic DNA that presumably include most of the regulatory elements necessary to recapitulate the endogenous gene expression pattern driven by the promoter of interest, but often show positional and/or copy number effects. The development of novel genome-engineering technologies, such as CRISPR/Cas, may be useful in overcoming some of the limitations of these traditionally used methods. Functioning as the RNA-based adaptive immune system in bacteria and archaea [43,44], CRISPR (clustered regularly interspaced short palindromic repeat) and Cas (CRISPR-associated) proteins have become a popular tool for mouse genome editing [45,46]. The Cas9 protein binds to specific genomic regions, under the guidance of a synthetic single-guide RNA, where it generates targeted double-strand breaks that facilitate gene targeting. Mouse knock-in lines carrying reporter constructs in endogenous locus, which are important tools in lineage tracing experiments, have been successfully generated using this system [47]. Importantly, when compared to the traditional strategy of making knock-in reporter mice, both cost and time are greatly reduced. One should thus expect that lineage tracing studies in the mammary gland employing mice generated by CRISPR/Cas in the near future. In addition to the strategy used to generate the mouse Cre lines, the dose and the timing of tamoxifen or doxycycline administration are particularly relevant when using the inducible systems. Low or sub-optimal doses label the population of interest at clonal density while high doses maximize labeling of the entire population, including low frequency putative stem cell populations, though it may have adverse effects. In addition, it is essential to determine the extent of Cre 'leakiness' in the absence of tamoxifen or doxycycline. At this point, it is worth mentioning that all these aspects, taken together, may have a significant impact in the lineage tracing studies. In the mammary gland, different results have been obtained when different Cre lines employing the same promoter are used, fuelling the passionate debate on the mammary stem cell hierarchy models.

Table 1. Cre and rtTA mouse lines used in lineage tracing in mammary gland.

Mouse strain	Type	Promoter	Original citation	Commercial source
WAP-Cre	Transgenic	2.6 Kb mouse WAP promoter	Wagner et al., 1997	Jax Stock 003552
K14-Cre	Transgenic	2.1 Kb human K14 promoter	Vasioukhin et al., 1999	Jax Stock 005107
K14-CreER^{T2}	Transgenic	2.0 Kb human K14 promoter	Indra et al., 1999	IMSR
K14-rtTA	Transgenic	2.1 Kb human K14 promoter	Nguyen et al., 2006	Jax Stock 008099
Lgr5-GFP-IRES-CreER^{T2}	Knock-in, null allele	Lgr5 endogenous promoter	Barker et al., 2007	Jax Stock 008875
K5-CreER^{T2}	Transgenic	6.0 Kb human K5 promoter	Rock et al., 2009	Not available
K18-CreER^{T2}	Transgenic	2.5 Kb human K18 promoter	Van Keymeulen et al., 2009	Jax Stock 017948
Acta2-CreER^{T2}	BAC transgenic	BAC Acta2 promoter	Wendling et al., 2009	Not available
K5-IRES-CreER^{T2}	Knock-in, 3'UTR	K5 endogenous promoter	Van Keymeulen et al., 2011	Not available
K8-CreER^{T2}	Transgenic	3.5 Kb mouse K8 promoter	Van Keymeulen et al., 2011	Jax Stock 017947
Notch2-CreER^{T2}	Knock-in, null allele	Notch2 endogenous promoter	Fre et al., 2011	Not available
Notch3-CreER^{T2}	Knock-in, null allele	Notch3 endogenous promoter	Fre et al., 2011	Not available
Axin2-CreER^{T2}	Knock-in, null allele	Axin2 endogenous promoter	Van Amerogen et al., 2012	Jax Stock 009120
K5-rtTA^{3}-IRES-GFP	Transgenic	5.2 Kb bovine K5 promoter	Rios et al., 2014	Not available
Elf5-rtTA^{3}-IRES-GFP	BAC transgenic	BAC Elf5 promoter	Rios et al., 2014	Not available
Procr-CreER^{T2}	Knock-in, null allele	Procr endogenous promoter	Wang et al., 2014	Not available

6. Mammary stem cell hierarchy revealed by genetic lineage tracing

Genetic lineage tracing studies have offered unprecedented levels of information regarding clonal analyses, organization of the stem cell compartment and possible role of resident stem and/or progenitor cells at different stages of the mammary gland organogenesis. The first study in the mammary gland was done over a decade ago using transgenic mice carrying Cre recombinase under the control of WAP, a pregnancy-specific promoter [40]. In this study, a MEC subpopulation that functions as self-renewing alveolar progenitors, called 'parity-induced' MEC, has been identified. Since then, innumerous studies have shed light into additional MEC subpopulations that are important for mammary gland development and homeostasis. Lineage tracing of K14-expressing cells in the rudimentary mammary gland at late embryogenesis revealed the existence of bipotent mammary stem cells that give rise to both myoepithelial and luminal cells [12,13]. The intense debate on the nature and location of the mammary stem cells relates to postnatal stages and there are currently two proposed models for the mammary stem cell hierarchy. Despite the controversy, there is compiling evidence that supports heterogeneity within the stem cell compartment in the mammary gland. One model suggests the existence of distinct pools of unipotent stem cells that progressively replace the fetal mammary stem cells [3,12,14,48]. In the pubertal gland, K14-, K5- or Lgr5-positive cells give rise to only basal cells with no contribution to the luminal cell lineage, whereas K8- and K18-positive cells contribute only to the luminal and alveolar and not basal cells [12,14]. Lgr5 labels a small subpopulation of basal cells [14]. Axin2 is a direct transcriptional target of the Wnt/Beta-catenin signaling pathway [49] and a defined stem cell marker [50]. Lineage tracing of Axin2-positive cells during puberty has revealed that two MEC lineages arise in a consecutive order that independently give rise to basal cells and luminal and alveolar cell lineages [13]. Further, a highly clonogenic and transiently quiescent luminal and alveolar population has been identified when tracing the Notch3-positive cells in the pubertal gland [48]. All these studies support the existence of unipotent, committed stem/progenitors in both the myoepithelial and luminal cell lineages. It remains unknown, however, whether there is any overlap between the distinct stem/progenitor cell populations identified in these studies. The other model for the mammary stem cell hierarchy proposes the existence of a bipotent mammary stem cell population in the pubertal gland that contributes to the major stages of morphogenesis of the mammary gland [23,41]. The *in situ* identification of population was experimentally demonstrated by using an inducible Tet-ON system with a multicolor Confetti reporter in combination with a novel 3D imaging technique [23]. In this study, cells labeled for K14, K5 or Lgr5 in the pre-pubertal mammary gland give rise to clonal patches of cells from both the myoepithelial and luminal and alveolar lineages, while luminal Elf5-positive cells give rise to only luminal and alveolar cells. Lineage tracing of K5-positive cells over long periods of time revealed a shift from mixed to predominantly luminal clones, suggesting that bipotent progenitors give rise to a population of luminal progenitors important for ductal maintenance. The existence of a multipotent basal cell population in the pubertal gland is supported by results from another study performing genetic lineage tracing of Procr-positive cells. These localize to the basal layer but express low levels of K5 and K14 while showing characteristics of the epithelial-to-mesenchyme transition [41]. Both these studies suggest that the basal cell population is heterogeneous, likely composed by bipotent basal stem cells as well as basal-restricted progenitor cells. It is worth mentioning that there are clear discrepancies between the different cell lineages

tracing studies using keratin gene promoters to drive expression in the basal compartment [12,13]. These are likely due to inherent differences between the mouse models, as previously discussed, as well as differences in imaging techniques and/or experimental design. Although most of the studies use CreER/rtTA systems, there is a considerable variation in how the experiments have been performed. The most critical aspects seem to be related to induction of Cre activity: when, how and for how long the inducing drugs are administered. While some of the studies employing the K5-CreER, K14-CreER and Lgr5-GFP-IRES-CreER mice used a single intraperitoneal injection of 1.5 mg of tamoxifen [23], others used three intraperitoneal injections of 5 mg of tamoxifen every other day [12,14]. Similarly, when using doxycycline, some studies opted for an oral administration of doxycycline food diet for 5 days or one month [12] whereas others used either one or two intraperitoneal injections of doxycycline [14]. These differences in the experimental design are likely to contribute to the differences observed in the several studies.

7.   Genetic lineage tracing in breast cancer

Breast cancer is a disease that comprises distinct pathological and molecular subtypes, which are relevant for the treatment efficiency and clinical outcome of the patients. Inter-tumor heterogeneity is believed to reflect the cell-of-origin as well as their genetic mutational profiles. There is a great promise that uncovering the stem cell hierarchy in the mammary gland will have a profound impact in the understanding of the mechanisms of breast tumorigenesis. Thus, genetic lineage tracing studies, in which specific subsets of MEC are transformed, labeled and traced in the intact mammary gland, are expected to boom in the coming years. The first of these studies used intravital analyses of clonal dynamics in MMTV-PyMT model of breast cancer [51,52]. In this work, compound MMTV-PyMT/R26-CreER^{T2}/R26R-Confetti mice were treated with tamoxifen at different time points to induce random expression of one of the four Confetti colors. The labeled cells as well as their progeny were then followed in live animals. Such analyses provide evidence of not only the existence of cancer stem cells in unperturbed tumors but also the high degree of plasticity of these cells. Recently, genetic lineage tracing was used to define the cellular origin of PIK3CA-derived tumors and the impact of this mutation on tumor heterogeneity [53]. PIK3CA, together with TP53, is one of the most frequently mutated genes and is associated with different types of human breast cancer. In this study, oncogenic PI3KCA^{H1047R} mutant was expressed specifically in basal or luminal cells, using K5-CreER^{T2} and K8-CreER^{T2} mouse lines respectively. Interestingly, expression in unipotent basal cells give rise to luminal-like tumors while its expression in unipotent luminal cells gave rise to luminal or basal-like tumors. These results indicate that PI3KCA^{H1047R} activates a multipotent genetic program in normally lineage-restricted populations at the early stage of tumor initiation, contributing to the tumor heterogeneity. This study is a great example on how genetic lineage tracing experiments are likely to provide fundamental insight on breast tumorigenesis.

8.   Conclusion

The mammary gland has been traditionally described as a relatively simple epithelial structure, composed of an inner luminal cell layer and surrounded by an outer basal layer. Over the years, it has become clear that the mammary epithelium is indeed more complex than initially thought, where

distinct subsets of MEC are present. Studies using FACS and transplantation assay have identified several subpopulations of stem and/or progenitor cells in the mammary gland. Nevertheless, only investigation of the stem cell hierarchy in the intact mammary gland can provide relevant information in a physiological and developmental context. Genetic lineage tracing is the approach that it has been used for this purpose and has provided exciting knowledge of mammary gland organogenesis and homeostasis that provides unique information on how tissue turnover and repair is sustained. These studies highlight the existence of heterogeneous stem cell compartment, where bipotent as well as unipotent mammary stem cells seems to co-exist. Many questions still remain though, namely about the nature and location of these cells. Some challenge will be to reconcile the results obtained using different lineage tracing strategies or to determine the overlap between the various MEC subpopulations identified by different approaches. Uncovering the stem cell hierarchy in the mammary gland is likely to have a profound impact in the understanding of breast cancer. Genetic lineage tracing experiments hold the promise of providing unprecedented insight into the cell-of-origin and heterogeneity of breast tumorigenesis.

Acknowledgment

Conflict of interest

All the authors declare no conflicts of interest in this paper.

References

1. Macias H, Hinck L (2012) Mammary gland development. *Wiley Interdiscip Rev Dev Biol* 1: 533-557.
2. Chepko G, Smith GH (1997) Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue Cell* 29: 239-253.
3. Šale S, Lafkas D, Artavanis-Tsakonas S (2013) Notch2 genetic fate mapping reveals two previously unrecognized mammary epithelial lineages. *Nat Cell Biol* 15: 451-460.
4. Visvader JE, Stingl J (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* 28: 1143-1158.
5. Deome KB, Faulkin LJ, Bern HA, Blair PB (1959) Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res* 19: 515-520.
6. Smith GH (1996) Experimental mammary epithelial morphogenesis in an in vivo model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res Treat* 39: 21-31.
7. Kordon EC, Smith GH (1998) An entire functional mammary gland may comprise the progeny from a single cell. *Development* 125: 1921-1930.
8. Shackleton M, Vaillant F, Simpson KJ, et al. (2006) Generation of a functional mammary gland from a single stem cell. *Nature* 439: 84-88.

9. Stingl J, Eirew P, Ricketson I, et al. (2006) Purification and unique properties of mammary epithelial stem cells. *Nature* 439: 993-997.

10. Inman JL, Robertson C, Mott JD, et al. (2015) Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* 142: 1028-1042.

11. Hines WC, Su Y, Kuhn I, et al. (2014) Sorting out the FACS: a devil in the details. *Cell Rep* 6: 779-781.

12. Van Keymeulen A, Rocha AS, Ousset M, et al. (2011) Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479: 189-193.

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

14. Plaks V, Brenot A, Lawson DA, et al. (2013) Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep* 3: 70-78.

15. Prater MD, Petit V, Alasdair Russell I, et al. (2014) Mammary stem cells have myoepithelial cell properties. *Nat Cell Biol* 16: 942–50–1–7.

16. Kretzschmar K, Watt FM (2012) Lineage tracing. *Cell* 148: 33-45.

17. Hoess RH, Wierzbicki A, Abremski K (1986) The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res* 14: 2287-2300.

18. Soriano P (1999) Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* 21: 70-71.

19. Zhu XD, Pan G, Luetke K, et al. (1995) Homology requirements for ligation and strand exchange by the FLP recombinase. *J Biol Chem* 270: 11646-11653.

20. Metzger D, Clifford J, Chiba H, et al. (1995) Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc Natl Acad Sci USA* 92: 6991-6995.

21. Feil R, Wagner J, Metzger D, et al. (1997) Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* 237: 752-757.

22. Asselin-Labat M-L, Vaillant F, Sheridan JM, et al. (2010) Control of mammary stem cell function by steroid hormone signalling. *Nature* 465: 798-802.

23. Rios AC, Fu NY, Lindeman GJ, et al. (2014) In situ identification of bipotent stem cells in the mammary gland. *Nature* 506: 322-327.

24. Gossen M, Freundlieb S, Bender G, et al. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268: 1766-1769.

25. Kistner A, Gossen M, Zimmermann F, et al. (1996) Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc Natl Acad Sci USA* 93: 10933-10938.

26. Urlinger S, Baron U, Thellmann M, et al. (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci USA* 97: 7963-7968.

27. Mao X, Fujiwara Y, Chapdelaine A, et al. (2001) Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. *Blood* 97: 324-326.

28. Srinivas S, Watanabe T, Lin CS, et al. (2001) Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* 1: 4.

29. Madisen L, Zwingman TA, Sunkin SM, et al. (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* 13: 133-140.

30. Barker N, van Es JH, Kuipers J, et al. (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449: 1003-1007.

31. Muzumdar MD, Tasic B, Miyamichi K, et al. (2007) A global double-fluorescent Cre reporter mouse. *Genesis* 45: 593-605.

32. Livet J, Weissman TA, Kang H, et al. (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450: 56-62.

33. Snippert HJ, van der Flier LG, Sato T, et al. (2010) Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* 143: 134-144.

34. Rock JR, Onaitis MW, Rawlins EL, et al. (2009) Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci USA* 106: 12771-12775.

35. Van Keymeulen A, Mascre G, Youseff KK, et al. (2009) Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis. *J Cell Biol* 187: 91-100.

36. Vasioukhin V, Degenstein L, Wise B, et al. (1999) The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc Natl Acad Sci USA* 96: 8551-8556.

37. Indra AK, Warot X, Brocard J, et al. (1999) Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Res* 27: 4324-4327.

38. Nguyen H, Rendl M, Fuchs E (2006) Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* 127: 171-183.

39. Wendling O, Bornert JM, Chambon P, et al. (2009) Efficient temporally-controlled targeted mutagenesis in smooth muscle cells of the adult mouse. *Genesis* 47: 14-18.

40. Wagner K-U, Boulanger CA, Henry MD, et al. (2002) An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* 129: 1377-1386.

41. Wang D, Cai C, Dong X, et al. (2015) Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* 517: 81-84.

42. Fre S, Hannezo E, Šale S, et al. (2011) Notch lineages and activity in intestinal stem cells determined by a new set of knock-in mice. *PLoS ONE* 6: e25785.

43. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria nd archaea. *Science* 327: 167-170.

44. Weidenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482: 331-338.

45. Cong L, Ran FA, Cox D, et al. (2013) Multiplex genome engineering using CRISPR/Cas system. *Science* 339: 819-823.

46. Wang H, Yang H, Shivalila CS, et al. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153: 910-918.

47. Yang H, Wang H, Shivalila CS, et al. (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154: 1370-1379.

48. Lafkas D, Rodilla V, Huyghe M, et al. (2013) Notch3 marks clonogenic mammary luminal progenitor cells in vivo. *J Cell Biol* 203: 47-56.

49. Lustig B, Jerchow B, Sachs M, et al. (2002) Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol Cell Biol* 22: 1184-1193.

50. Zeng YA, Nusse R (2010) Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* 6: 568-577.

51. Zomer A, Ellenbroek SIJ, Ritsma L, et al. (2013) Intravital imaging of cancer stem cell plasticity in mammary tumors. *Stem Cells* 31: 602-606.

52. Guy CT, Cardiff RD, Muller WJ (1992) Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol* 12: 954-961.

53. Van Keymeulen A, Lee MY, Ousset M, et al. (2015) Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* 525: 119-123.