doi:10.1038/nature13851

Identification of multipotent mammary stem cells by protein C receptor expression

Daisong Wang[1]*, Cheguo Cai[1]*, Xiaobing Dong[1], Qing Cissy Yu[1], Xiao-Ou Zhang[2], Li Yang[2] & Yi Arial Zeng[1]

The mammary gland is composed of multiple types of epithelial cells, which are generated by mammary stem cells (MaSCs) residing at the top of the hierarchy[1,2]. However, the existence of these multipotent MaSCs remains controversial and the nature of such cells is unknown[3,4]. Here we demonstrate that protein C receptor (Procr), a novel Wnt target in the mammary gland, marks a unique population of multipotent mouse MaSCs. Procr-positive cells localize to the basal layer, exhibit epithelial-to-mesenchymal transition characteristics, and express low levels of basal keratins. Procr-expressing cells have a high regenerative capacity in transplantation assays and differentiate into all lineages of the mammary epithelium by lineage tracing. These results define a novel multipotent mammary stem cell population that could be important in the initiation of breast cancer.

The mammary gland is an epithelial organ consisting of myoepithelial (basal) cells and luminal cells. During pregnancy, the luminal cells at side branches undergo terminal differentiation and form alveolar cells. Previous studies using the surface markers Lin$^-$, CD24$^+$ and CD29^{hi}, and transplantation assays, indicate that MaSCs reside in the basal layer of the epithelium[1,2]. This population is heterogeneous, including MaSCs, differentiated basal cells and potential intermediate progenitors. Until now, no marker specific for MaSCs has been identified. On the other hand, the existence of multipotent MaSCs in adults remains in debate as lineage-tracing studies using the pan-basal markers keratin 5 (K5; also known as Krt5) and K14 generate controversial results[3,4]. Multipotent MaSCs may have been missed in other basal subpopulation lineage tracing studies using *Lgr5* or *Axin2*, and the rare occurrence of clones containing both lineages (bi-lineage) could be due to the periodic luminal expression of these genes[3,5,6]. Here we show that Procr, a novel Wnt target in the mammary gland, marks a unique population of multipotent MaSCs.

Wnt signalling is instrumental for MaSC self-renewal[5,7,8]. Our previous work demonstrated that the Wnt3A protein can expand MaSCs in three-dimensional Matrigel culture and maintain their stem cell properties[7]. Taking advantage of this *in vitro* system, we performed microarray analysis of the cultured MaSCs in an attempt to identify Wnt targets specifically expressed in MaSCs (Extended Data Fig. 1a). Among the candidates whose expression was increased in the presence of Wnt3A, we identified *Procr* (Extended Data Fig. 1a). Quantitative polymerase chain reaction (qPCR) confirmed that the gene is upregulated by Wnt3A treatment (Extended Data Fig. 1b).

Procr is a single-pass transmembrane protein originally recognized as protein C receptor through its roles in anticoagulation, inflammation and haematopoiesis[9–14]. We investigated whether Procr is normally expressed in the mammary epithelium. We isolated basal (Lin$^-$ CD24$^+$ CD29^{hi}) and luminal (Lin$^-$ CD24$^+$ CD29^{lo}) cells from 8-week-old virgin mammary glands (Fig. 1a), and found that *Procr* is expressed at higher levels in basal cells (Fig. 1b). Furthermore, fluorescence-activated cell sorting (FACS) analysis indicated that Procr labels 3–7% of basal cells depending on the genetic background (about 2.9 $\pm$ 0.5% in CD1 and 7 $\pm$ 1.5% in B6), while Procr$^+$ cells were not found among luminal cells (Fig. 1c and Extended Data Fig. 1c–g). Procr$^+$ cells were also detected in the stromal cell compartment (Extended Data Fig. 1c–g). Notably, the Procr expression patterns were similar throughout development (Extended Data Fig. 1c–g). Immunostaining confirmed that a subpopulation of basal cells expresses Procr (Fig. 1d). Intriguingly, Procr$^+$ cells appeared to express less K14 in comparison to their neighbouring Procr$^-$ basal cells (Fig. 1d). Next, we isolated Procr$^+$ and Procr$^-$ cells from the basal cell population and performed RNA-sequencing (RNA-seq) analysis. We found that basal Procr$^+$ cells exhibit features of epithelial-to-mesenchymal transition (EMT), with lower expression of epithelial signatures, for example, Epcam, E-cadherin and claudins, and with increased expression of mesenchymal signature genes, for example, *Vim*, N-cadherin (also known as *Cdh2*), *Foxc2*, *Zeb1* and *Zeb2* (Fig. 1e). Of note, the basal keratins K5 and K14 were expressed at lower levels in Procr$^+$ cells compared with Procr$^-$ basal cells (Fig. 1e). These observations were confirmed by qPCR analysis (Fig. 1f).

We next examined the behaviours of Procr$^+$ basal cells *in vitro* and in transplantation assays. We isolated total basal cells (CD24$^+$ CD29^{hi}), Procr$^+$ basal cells (Procr$^+$ CD24$^+$ CD29^{hi}) and Procr$^-$ basal cells (Procr$^-$ CD24$^+$ CD29^{hi}), and compared their colony-forming ability in three-dimensional Matrigel culture as previously described[7] (Fig. 2a, b). We found that the enrichment of Procr$^+$ cells increased colony-forming efficiency by fivefold when compared to the total basal cell group. One colony formed out of 15 plated total basal cells, while one colony formed out of three plated Procr$^+$ basal cells (Fig. 2b and Extended Data Fig. 2a). Colony sizes were indistinguishable between the two groups (Fig. 2b). In striking contrast, Procr$^-$ basal cells were not able to form colonies in Matrigel culture, suggesting that MaSCs that have colony-forming abilities were absent from this group.

To assess their mammary gland reconstitution capacity, the three groups of isolated cells were transplanted into cleared fat pads. We found that Procr$^+$ basal cells generate the mammary gland more efficiently (repopulating frequency of 1/12) than total basal cells (1/68) (Fig. 2e). The outgrowths displayed normal morphology and marker expression (Fig. 2c). When recipient mice were in late pregnancy, the mammary gland resulting from the transplanted Procr$^+$ basal cells consisted of a dense ductal system ending in clusters of milk-producing alveoli (Fig. 2d). In contrast, Procr$^-$ basal cells showed markedly lower stem cell frequency (1/2,084) (Fig. 2e). These findings demonstrate that the CD24$^+$ CD29^{hi} basal population can be further enriched for MaSCs using the marker Procr.

We found that Lgr5$^+$ cells fell into the Procr$^-$ population that has drastically reduced regenerative capability (Extended Data Fig. 3a, b), raising the question as to whether Lgr5$^+$ cells are enriched for MaSCs. To address this, we isolated the three subpopulations of basal cells, Procr$^+$ Lgr5$^-$, Procr$^-$ Lgr5$^+$ and Procr$^-$ Lgr5$^-$ and examined their regenerative capacities. Consistent with our earlier results, the Procr$^+$ Lgr5$^-$ cells efficiently formed colonies *in vitro* and readily reconstituted mammary gland in transplantation (repopulating frequency of 1/14) (Extended Data Fig. 3c, d). Procr$^-$ Lgr5$^+$ cells were not able to form colonies *in vitro*.

[1]The State Key Laboratory of Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [2]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
*These authors contributed equally to this work.

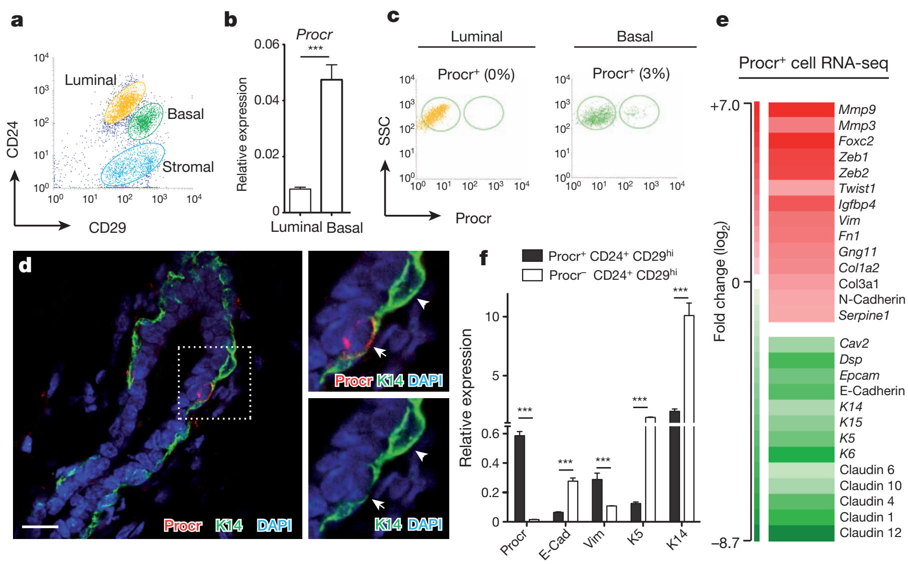
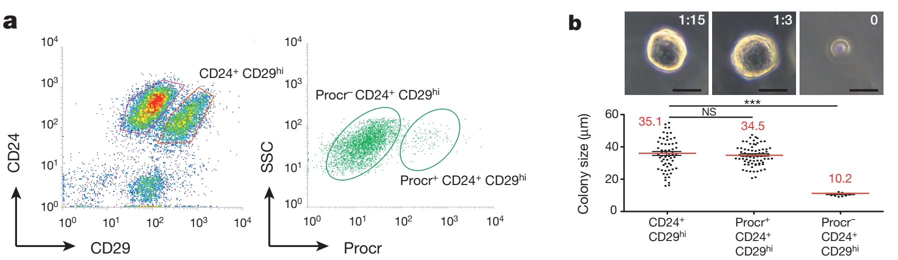


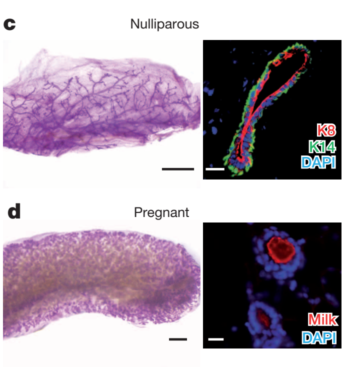
Figure 1 | Procr$^+$ basal cells express lower levels of basal keratin. **a**, **b**, Basal and luminal cells were FACS-isolated and analysed for *Procr* expression by qPCR. **c**, FACS analysis of Procr expression in 8-week-old CD1 mammary epithelial cells. **d**, Immunohistochemistry indicating the expression of Procr in a subpopulation of basal cells (arrows). Ninety-four per cent of Procr$^+$ basal cells ($n = 206$) expressed less K14 compared with the neighbouring basal cells (arrowheads). DAPI, 4′,6-diamidino-2-phenylindole. Scale bar, 20 μm. **e**, Expression of EMT-related genes in Procr$^+$ basal cells. **f**, qPCR analysis of Procr$^+$ basal cells and Procr$^-$ basal cells. Quantification was performed on three independent experiments. E-Cad, E-cadherin. **b**, **f**, Data are presented as mean ± standard deviation (s.d.). ***$P < 0.01$.

Interestingly, they were able to reconstitute *in vivo* by transplantation, although with a significantly lower repopulating frequency (1/165) (Extended Data Fig. 3c, d). Considering a repopulating frequency of 1/68 for total basal cells, our results indicated that Lgr5 expression was not enriched in regenerative MaSCs. This conclusion is different from a previous report[15], yet is consistent with two other studies[4,6]. Finally, the Procr$^-$ Lgr5$^-$ cells were depleted of MaSCs and failed to regenerate *in vitro* or *in vivo* (Extended Data Fig. 3c, d).

We next investigated whether the population of Procr$^+$ cells behave as multipotent MaSCs under physiological conditions. To this end, we generated a knock-in allele of Procr by integrating a *CreERT2-IRES-tdTomato* cassette at the first ATG codon (Fig. 3a and Extended Data Fig. 2b, c). Heterozygous mice were healthy and fertile. Homozygotes died before embryonic day (E)10.5 (Extended Data Fig. 2d, e), resembling the *Procr*-null mutant mice[16]. Confocal imaging of histological sections indicated that tdTomato$^+$ cells resided in the basal layer yet expressed lower levels of K5 and K14 compared with tdTomato$^-$ basal cells (Fig. 3b). FACS analysis indicated that 3% of basal cells were tdTomato$^+$ and no tdTomato$^+$ cells were found in luminal cells. Reminiscent of the expression of Procr itself, some tdTomato$^+$ cells were present in stromal cells (Fig. 3c). We isolated tdTomato$^+$ and tdTomato$^-$ cells from the basal group and assessed their colony formation capability. We found that tdTomato$^+$ cells form colonies efficiently *in vitro*, whereas tdTomato$^-$ cells cannot (Fig. 3d). These results demonstrate that the *Procr$^{CreERT2\text{-}IRES\text{-}tdTomato}$* allele faithfully recapitulates endogenous Procr expression.

The generation of the *Procr$^{CreERT2\text{-}IRES\text{-}tdTomato}$* mouse allowed us to examine the expression of Procr$^+$ cells in detail. Using whole-mount confocal imaging analysis, Procr$^+$ cells were identified as being sparsely located in E18.5 and newborn (postnatal day (P)1.5) mammary gland. At these stages, before the formation of the terminal end buds (TEBs), Procr$^+$ cells could be detected in the middle or at the tip of the mammary ducts (Extended Data Fig. 4a, b). In puberty, dispersed individual Procr$^+$ cells were predominantly present in the mammary ducts, whereas no Procr$^+$ cells were detected in the TEBs (Extended Data Fig. 4c, d). In the mature mammary gland, individual Procr$^+$ cells were also located over the ducts (Extended Data Fig. 4e). As the TEB is the most proliferative structure in the pubertal gland, our observations suggest that Procr$^+$ cells are not the major proliferative force, consistent with the properties of stem cells rather than transient amplifying





e

	Number of cells injected per fat pad	Number of outgrowths	Fat pad filled	Repopulating frequency
CD24$^+$ CD29^{hi}	100	7/8	●×6 ◕×1 ○×1	1/68.3 (41.3–113.0)
	50	4/8	●×4 ○×4	
	25	2/12	◕×1 ◔×1 ○×10	
	13	2/12	◔×2 ○×10	
	1	1/22	◔×1 ○×21	
Procr$^+$ CD24$^+$ CD29^{hi}	100	8/8	●×8	1/11.9*** (7.3–19.4)
	50	6/6	●×6	
	25	9/12	●×7 ◕×1 ◑×1 ○×3	
	13	5/8	●×2 ◑×3 ○×3	
	1	5/18	◔×5 ○×13	
Procr$^-$ CD24$^+$ CD29^{hi}	500	2/8	●×2 ○×6	1/2084.2*** (675.3–6,432.1)
	250	1/8	◑×1 ○×7	
	100	0/6	○×6	
	50	0/6	○×6	

Figure 2 | Procr$^+$ cells are enriched for mammary stem cells with regenerative capabilities. **a**, Isolation of total basal, Procr$^+$ basal and Procr$^-$ basal populations. **b**, Colony-formation efficiency and the colony sizes in Matrigel culture. Scale bars, 20 μm. ***$P < 0.01$. NS, not significant. See also Extended Data Fig. 2a. **c**, **d**, Whole-mount and section images of an outgrowth derived from the transplantation of Procr$^+$ basal cells in nulliparous and late pregnant mammary tissues. Scale bars, 2 mm in whole mount; 20 μm in section. **e**, Transplantation of sorted cells in limiting dilution. Data are pooled from four independent experiments. ***$P < 0.01$.

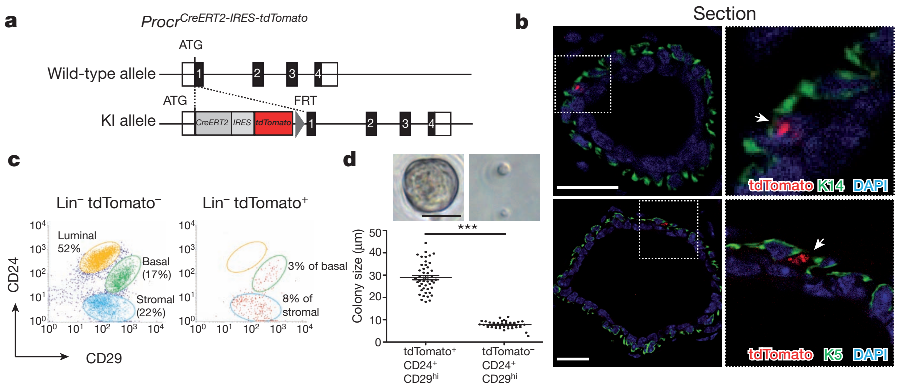


Figure 3 | $Procr^{CreERT2\text{-}IRES\text{-}tdTomato}$ knock-in mouse recapitulates the Procr expression pattern and labelled cell behaviour. a, Targeting strategy to generate the $Procr^{CreERT2\text{-}IRES\text{-}tdTomato}$ knock-in (KI) mouse. See also Extended Data Fig. 2b. **b,** Immunostaining analysis of the knock-in mammary sections. Scale bar, 20 μm. **c,** FACS analysis indicating that tdTomato^{+} cells are located in 3% of basal and 8% of stromal cells. $n = 4$ mice. One of four similar experiments is shown. **d,** Colony formation of tdTomato^{+} and tdTomato^{-} basal cells in Matrigel culture. Scale bars, 20 μm. $***P < 0.01$.

cells. By 5-ethymyl-2′-deoxyuridine (EdU) incorporation assays, we found that in pubertal or mature mammary glands, the majority of Procr^{+} cells indeed enter the cell cycle (Extended Data Fig. 4f–i). The seemingly higher percentage of EdU^{+} Procr^{+} population cells in mature ducts is probably due to a lower number of EdU^{+} cells at this stage (Extended Data Fig. 4i). Our data suggest that Procr^{+} cells are proliferative cells residing in the mammary ducts.

To trace the fate of Procr^{+} cells, we crossed the $Procr^{CreERT2\text{-}IRES\text{-}tdTomato/+}$ allele with the $Rosa26^{mTmG/+}$ ($R26^{mTmG/+}$) reporter strain[17] (Fig. 4a). We first tracked the developmental fate of Procr^{+} cells in postnatal mammary glands by administering tamoxifen to $Procr^{CreERT2/+}$;$R26^{mTmG/+}$ pubertal mice (5 weeks old) and analysing the contribution of labelled cells to the mature epithelial network once the mice had reached adulthood. Expression of green fluorescent protein (GFP) was not observed in un-induced mice (data not shown). Short time tracing (48 h) and confocal whole-mount imaging allowed us to visualize single elongated cells initially labelled by GFP (Fig. 4b, e). Immunostaining in tissue sections confirmed that the initially GFP^{+} cells were basal cells (Fig. 4f). Quantification of the labelling events by FACS analysis indicated that no luminal cells are labelled at the beginning of the analysis (Fig. 4c, d). There were some labelled stromal fibroblasts, which also express Procr (Fig. 4c, d). After 3 weeks of tracing, the GFP^{+} cells expanded in number (Fig. 4g–i). In addition, their pattern extended to include luminal cells. Clonal expansion of GFP^{+} cells was visualized by whole-mount imaging, and the clones consisted of both elongated and columnar cells (Fig. 4j). Immunostaining confirmed that GFP^{+} cells are present in both basal and luminal layers (Fig. 4k). Clonal analysis revealed that the majority of the clones (72%) are bi-lineage, giving rise to basal and luminal cells. Notably, about 13% of the clones were single basal cells that had not entered division since labelling; no luminal clones were found (Extended Data Fig. 5a, b). Importantly, the majority of two-cell clones (65%) comprised one luminal cell and one basal cell (Extended Data Fig. 5c). When the tracing was prolonged to 6 weeks, the average clone sizes increased over time, and the proportion of the bi-lineage clone also increased (from 72% to 93%), indicating that more basal cells had differentiated into luminal cells (Extended Data Fig. 5d, e). The percentage of bi-lineage two-cell clones increased (from 65% to 85%), suggesting that more initially GFP^{+} labelled cells had asymmetrically divided to become luminal cells (Extended Data Fig. 5f). During pregnancy, GFP^{+} cells contributed to alveolus formation (Fig. 4l–o). One alveolus could consist solely of GFP^{+} cells or harbour both GFP^{+} and mTomato^{+} cells (Fig. 4p), indicating that alveoli can originate from one or more progenitor(s), which is consistent with previous reports[4,5]. GFP^{+} cells were maintained at similar percentages across multiple pregnancies, showing that Procr^{+} cells are capable of long-term self-renewal (Extended Data Fig. 6a–f). Notably, GFP^{+} stromal cells did not expand over tracing, suggesting a less proliferative nature of Procr^{+} fibroblasts (Fig. 4c, h, m and Extended Data Fig. 6b, d).

The multipotency of Procr^{+} basal cells was examined by initiating the labelling in 8-week-old adult mice (Extended Data Fig. 7a). After 3 weeks of tracing, the majority of labelled cells differentiated into bi-lineage clones (74%) (Extended Data Fig. 7b–g), and the percentage increased to 94% by 6 weeks (Extended Data Fig. 7h–j). From 3 weeks to 6 weeks, the percentage of bi-lineage two-cell clones also increased

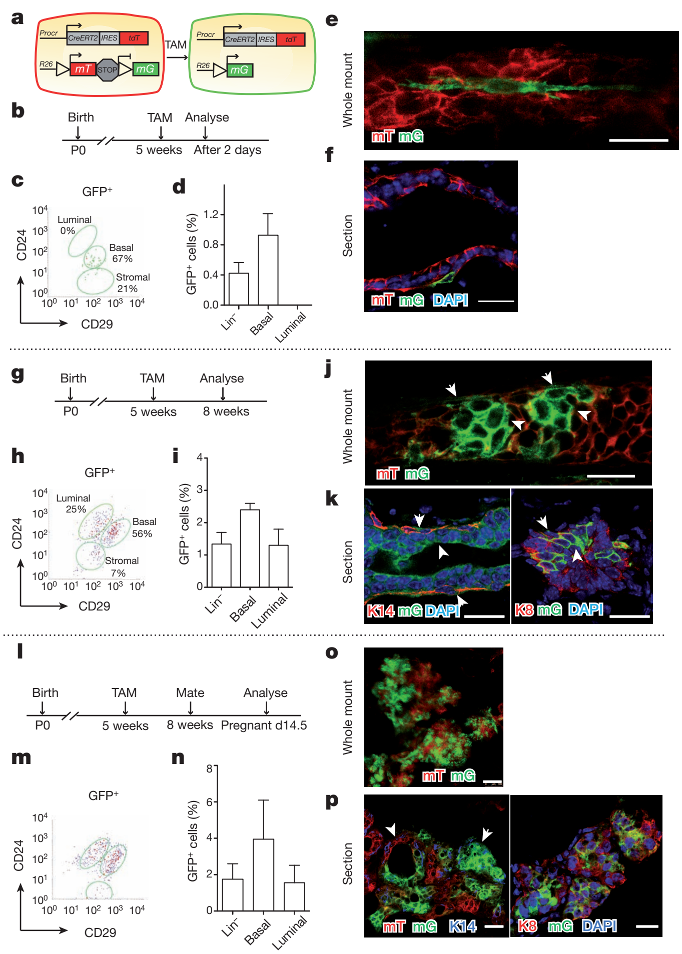


Figure 4 | Procr labels multipotent adult mammary stem cells. a, Illustration of lineage tracing strategy. *tdT*, *tdTomato*. **b,** Experimental setup used in short-term tracing. **c, d,** FACS analysis indicating that GFP^{+} cells were restricted to the basal cells at 48 h after tamoxifen (TAM) administration. **e,** Whole-mount confocal microscopy showing an elongated GFP^{+} basal cell. **f,** Section imaging indicating the basal location of the GFP^{+} cell. **g–k,** FACS and imaging analysis of GFP^{+} cell distribution after 3 weeks of tracing. **l–p,** FACS and imaging analysis of GFP^{+} cells distribution at pregnant day 14.5 after 5 weeks of tracing. Scale bars, 20 μm. **d, i, n,** Data are presented as mean ± s.d. $n = 3$ mice.

(from 70% to 90%). Upon pregnancy, GFP^+ cells differentiated to form alveoli (Extended Data Fig. 7k–n).

We next investigated the contribution of $Procr^+$ cells to early mammary development by initiating the labelling in E18.5, P0.5 and prepubescent mice (2 weeks old) (Extended Data Figs 8, 9) and analysing the contribution of GFP^+ cells in mature adults. By FACS analysis and immunostaining, GFP^+ cells were found in both basal and luminal populations. Eight-week tracing from late embryo or at birth predominantly led to bi-lineage clones in adults (98% and 99%) (Extended Data Fig. 8f, m). Consistently, 6-week tracing of $Procr^+$ cells in prepubescent mice (2 weeks old) mostly resulted in bi-lineage clones (90%) (Extended Data Fig. 9f). During pregnancy, GFP^+ cells contributed to alveoli formation (Extended Data Fig. 9h–k). Taken together, these lineage tracing experiments initiated in the embryonic and various postnatal stages show that $Procr^+$ cells contribute to both basal and luminal cell lineages.

To investigate the physiological requirement of $Procr^+$ cells in mammary gland development, we performed targeted ablation of these cells in developing mammary glands. We generated the $Procr^{CreERT2/+};R26^{DTA/+}$ strain to conditionally express diphtheria toxin (DTA) in $Procr^+$ cells (Extended Data Fig. 10a). We administered tamoxifen in $Procr^{CreERT2/+};R26^{DTA/+}$ pubertal mice at P33 every 3 days, and evaluated the effects of targeted ablation of $Procr^+$ cells 9 days later (Extended Data Fig. 10b). At this stage, both the oil-treated control mammary epithelium and the tamoxifen-treated $R26^{DTA/+}$ control mammary epithelium had grown to the distal edge of the fat pad (Extended Data Fig. 10c, d). In striking contrast, tamoxifen administration in $Procr^{CreERT2/+};R26^{DTA/+}$ mice largely prevented the growth of the epithelium (Extended Data Fig. 10e, f). FACS analysis indicated that the basal $Procr^+$ cells were efficiently ablated (Extended Data Fig. 10g, h). Together, these results suggest that the $Procr^+$ cells are important for the development and maintenance of adult mammary gland.

Our study identifies Procr as a novel Wnt target in the mammary epithelium. $Procr^+$ cells express lower levels of K5/K14 compared with other basal cells. They are also unique in that they are multipotent by lineage tracing, and show the highest repopulation efficiency by transplantation. Such a population of cells has not been described before. Much effort has been devoted to delineating the relationships between different epithelial cell populations in the mammary gland. Our work suggests that $Procr^+$ cells are at the top of the hierarchy, supporting the model that multipotent and unipotent stem cells coexist in the adult mammary gland, reconciling the differences found between previous lineage tracing and transplantation studies (Extended Data Fig. 10i).

EMT has been linked to the stemness properties of cancer cells[18]. As $Procr^+$ MaSCs exhibit EMT signatures in the normal mammary gland, it is tempting to speculate that $Procr^+$ MaSCs represent one of the origins of breast cancer stem cells. Indeed, in human breast cancer, Procr is expressed in the $CD44^+$ (cancer-stem-cell-enriched) group[19]. Procr expression in cancer cell lines promotes tumour formation[20,21] and metastasis[22,23]. More similarities may exist between normal stem cells and malignant stem cells.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 February; accepted 9 September 2014.
Published online 19 October 2014.

1. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439,** 84–88 (2006).
2. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439,** 993–997 (2006).
3. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479,** 189–193 (2011).
4. Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. *In situ* identification of bipotent stem cells in the mammary gland. *Nature* **506,** 322–327 (2014).
5. van Amerongen, R., Bowman, A. N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/β-catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* **11,** 387–400 (2012).
6. de Visser, K. E. *et al.* Developmental stage-specific contribution of $LGR5^+$ cells to basal and luminal epithelial lineages in the postnatal mammary gland. *J. Pathol.* **228,** 300–309 (2012).
7. Zeng, Y. A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6,** 568–577 (2010).
8. Badders, N. M. *et al.* The Wnt receptor, Lrp5, is expressed by mouse mammary stem cells and is required to maintain the basal lineage. *PLoS ONE* **4,** e6594 (2009).
9. Fukudome, K. & Esmon, C. T. Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor. *J. Biol. Chem.* **269,** 26486–26491 (1994).
10. Cheng, T. *et al.* Activated protein C blocks p53-mediated apoptosis in ischemic human brain endothelium and is neuroprotective. *Nature Med.* **9,** 338–342 (2003).
11. Balazs, A. B., Fabian, A. J., Esmon, C. T. & Mulligan, R. C. Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. *Blood* **107,** 2317–2321 (2006).
12. Vetrano, S. *et al.* Unexpected role of anticoagulant protein C in controlling epithelial barrier integrity and intestinal inflammation. *Proc. Natl Acad. Sci. USA* **108,** 19830–19835 (2011).
13. Ivanova, N. B. *et al.* A stem cell molecular signature. *Science* **298,** 601–604 (2002).
14. Bae, J. S., Yang, L., Manithody, C. & Rezaie, A. R. The ligand occupancy of endothelial protein C receptor switches the protease-activated receptor 1-dependent signaling specificity of thrombin from a permeability-enhancing to a barrier-protective response in endothelial cells. *Blood* **110,** 3909–3916 (2007).
15. Plaks, V. *et al.* Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep.* **3,** 70–78 (2013).
16. Gu, J. M. *et al.* Disruption of the endothelial cell protein C receptor gene in mice causes placental thrombosis and early embryonic lethality. *J. Biol. Chem.* **277,** 43335–43343 (2002).
17. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45,** 593–605 (2007).
18. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133,** 704–715 (2008).
19. Shipitsin, M. *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11,** 259–273 (2007).
20. Schaffner, F. *et al.* Endothelial protein C receptor function in murine and human breast cancer development. *PLoS ONE* **8,** e61071 (2013).
21. Hwang-Verslues, W. W. *et al.* Multiple lineages of human breast cancer stem/progenitor cells identified by profiling with stem cell markers. *PLoS ONE* **4,** e8377 (2009).
22. Beaulieu, L. M. & Church, F. C. Activated protein C promotes breast cancer cell migration through interactions with EPCR and PAR-1. *Exp. Cell Res.* **313,** 677–687 (2007).
23. Spek, C. A. & Arruda, V. R. The protein C pathway in cancer metastasis. *Thromb. Res.* **129** (suppl. 1), S80–S84 (2012).

Acknowledgements The screening was initiated in the laboratory of R. Nusse and we are grateful for his support and generosity. We thank Y. Zhang and Biocytogen for assistance in knock-in mouse generation, and Q. Yin and H. Fang for technical support in RNA-seq analysis. We are grateful to C.-C. Hui and E. Verheyen for critical reading of the manuscript. We thank D. Li for helpful discussion. This work is supported by grants from the Ministry of Science and Technology of China (2014CB964800, 2012CB945000 to Y.A.Z., 2014CB910600 to L.Y.), National Natural Science Foundation of China (31171421, 31371500 to Y.A.Z., 31201098 to C.C.), the Chinese Academy of Sciences (XDA01010307, 2010OHTP03 to Y.A.Z., 2012OHTP08 to L.Y.), Shanghai Municipal Science and Technology Commission (12PJ1410100 to Y.A.Z.).

Author Contributions Y.A.Z. conceived the study and designed experiments. D.W. performed transplantation and lineage tracing experiments and carried out data analysis. C.C. and X.D. performed array and screening experiments. D.W., Q.C.Y., X.-O.Z. and L.Y. performed RNA-seq and analysis. Y.A.Z. wrote the manuscript.

Author Information Microarray and RNA-seq data have been deposited in the Gene Expression Omnibus under accession numbers GSE55117 and GSE58088, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.A.Z. (yzeng@sibcb.ac.cn).

METHODS

Experimental animals. To generate mice expressing CreERT2-IRES-tdTomato under control of the endogenous *Procr* promoter, we generated the targeting construct depicted in Fig. 3a and Extended Data Fig. 2b. Female mice of $Rosa26^{mTmG/+}$, $Rosa26^{DTA/+}$ (Jackson Laboratories), $Axin2^{lacZ/+}$ (ref. 24), $Lgr5^{eGFP\text{-}IRES\text{-}CreERT2/+}$ (*Lgr5-GFP*) (ref. 25), CD1, B6 and Nude strains were used in this study. For lineage tracing experiments induced in prepubescent, pubertal and mature adult mice, animals received a single intraperitoneal injection of 4 mg per 25 g body weight of tamoxifen (TAM; Sigma-Aldrich) diluted in sunflower oil. For lineage tracing experiments induced at birth, each mouse received a single injection of 125 μg tamoxifen. To induce recombination in embryos, pregnant mothers at day 18.5 were injected with a single dose of 0.5 mg per 25 g body weight of tamoxifen. For DTA-mediated cell ablation experiments, pubertal mice were injected with 4 mg per 25 g body weight of tamoxifen in sunflower oil every 3 days a total of three times. Experimental procedures were approved by the Animal Care and Use Committee of Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences.

Quantification of lineage-specific cells and the size of clones. A minimum of three different mice were analysed per condition. Dissociated single mammary cells were FACS analysed for the GFP^+ cells proportion in basal and luminal compartments. A minimum of 3 mice were analysed by FACS analysis and a minimum of 20 sections were analysed by immunohistochemistry of K14 and K8 to discern the basal and luminal composition of GFP^+ cells. Representative clones were documented by confocal imaging. For clonal analysis, a minimum of 200 GFP^+ clones were analysed per time point. For each clone, the number of cells and their K14 or K8 expression were scored. The clones were grouped in three classes: one-cell, two-cell, and clones with more than two cells.

Antibodies. Antibodies used were: rat anti-Procr (1:50, eBioscience, catalogue #13-2012, clone 1560), rat anti-K8 (1:250, Developmental Hybridoma Bank, TROMA-I), rabbit anti-K14 (1:1,000, Covance), rabbit anti-K5 (1:1000, Covance), rabbit anti-milk (1:500, Nordic Immunological Laboratories).

Primary cell preparation. Mammary glands from 8- to 12-week-old virgin or an otherwise specified stage of female mice were isolated. The minced tissue was placed in culture medium (RPMI 1640 with 25 mM HEPES, 5% fetal bovine serum, 1% penicillin-streptomycin-glutamine (PSQ), 300 U ml^{-1} collagenase III (Worthington)) and digested for 2 h at 37 °C. After lysis of the red blood cells in NH_4Cl, a single-cell suspension was obtained by sequential incubation with 0.25% trypsin-EDTA at 37 °C for 5 min and 0.1 mg ml^{-1} DNase I (Sigma) for 5 min with gentle pipetting, followed by filtration through 70 μm cell strainers.

Cell labelling and flow cytometry. The following antibodies in 1:200 dilutions were used: biotinylated and FITC-conjugated CD31, CD45, TER119 (BD PharMingen, clone MEC 13.3, 30-F11 and TER-119; catalogue # 553371, #55307, # 553672, # 553372, # 553080 and # 557915), CD24-PE/cy7, CD29-APC (Biolegend, clone M1/69 and HMβ1-1; catalogue #101822 and #102216), Procr-PE (eBioscience, clone 1560, catalogue #12-2012), Streptavidin-V450, and Streptavidin-FITC (BD PharMingen). Antibody incubation was performed on ice for 15 min in HBSS with 10% fetal bovine serum. All sortings were performed using a FCASJazz (Becton Dickinson). The purity of sorted population was routinely checked and ensured to be more than 95%.

***In vitro* colony formation assay.** FACS-sorted cells were resuspended at a density of 4×10^5 cells ml^{-1} in chilled 100% growth-factor-reduced Matrigel (BD Bioscience), and the mixture was allowed to polymerize before covering with culture medium (DMEM/F12, ITS (1:100; Sigma), 50 ng ml^{-1} EGF, plus either vehicle (1% CHAPS in PBS) or 200 ng ml^{-1} Wnt3A[26]). Culture medium was changed every 24 h. Primary colony numbers were scored after 6–7 days in culture. The colonies were mostly spherical. In cases that colonies were oval, the long axis was measured.

RNA extraction, microarray and RNA sequencing. For microarray, total RNA from second-passage MaSC colonies cultured in the presence of vehicle and Wnt3A was extracted with PicoPure (Arcturs) in accordance with the manufacturer's protocol. At the second passage, MaSC colonies in Wnt3A treatment can efficiently reconstitute new mammary glands in transplantation assays, an indication of retaining stemness, while MaSC colonies in vehicle have completely lost the reconstitution capabilities[7]. RNA concentration was determined with NanoDrop ND-1000, and quality was determined using the RNA 6000 Nano assay on the Agilent 2100 Bioanalyzer (Agilent Technologies). Affymetrix microarray analysis, fragmentation of RNA, labelling, hybridization to Mouse Genome 430 2.0 microarrays and scanning were performed in accordance with the manufacturer's protocol (Affymetrix). For RNA-seq, total RNA from freshly isolated Lin^- $CD24^+$ $CD29^{hi}$ $Procr^+$ cells and Lin^- $CD24^+$ $CD29^{hi}$ $Procr^-$ cells were extracted with Trizol. RNA-seq libraries were prepared according to the manufacturer's instructions (Illumina) and then applied to sequencing on Illumina HiSeq 2000 in the CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai. In total, around 32 million 1×100 single reads for each sample were obtained and uniquely mapped to the mm9 mouse genome with more than 70% mapping rate for both samples using TopHat 1.3.3. Differential gene expression analysis was carried out using Cuffdiff 2.0.2, and genes with significant alteration were extracted for a further analysis.

EdU labelling. *In vivo* EdU labelling was accomplished by intraperitoneal injections of EdU (0.2 mg per 10 g body weight) followed by harvest 3 h after injection. Samples were subjected to Click-it chemistry (Invitrogen).

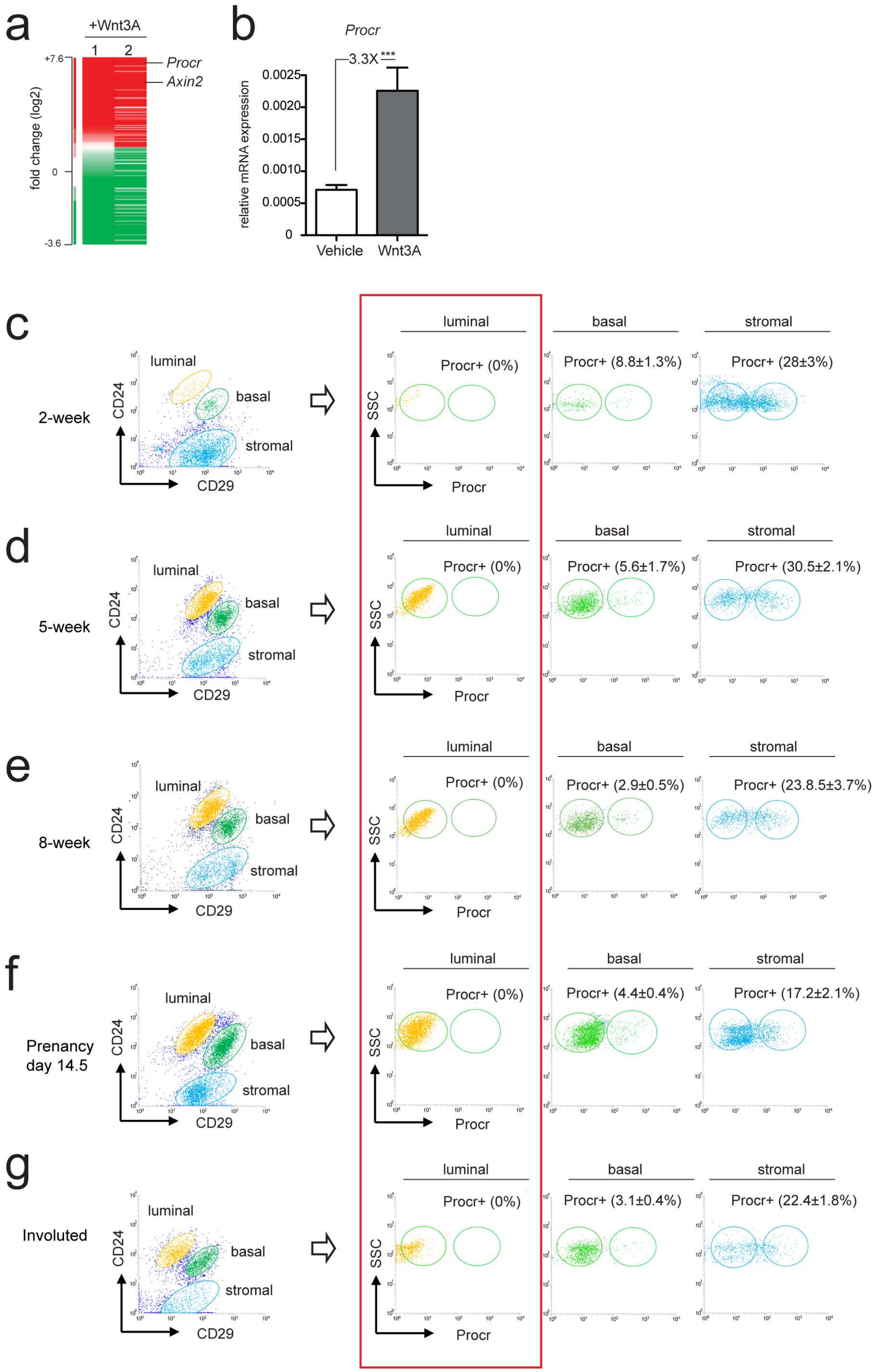
Immunohistochemistry. Whole-mount staining was performed as described previously[4]. Frozen sections were prepared by air-drying and fixation for 1 h in cold MeOH or PFA. Tissue sections were incubated with primary antibodies at 4 °C overnight, followed by washes, incubation with secondary antibodies for 2 h at 25 °C, and counterstaining with DAPI (Vector Laboratories). For all the immunoflourescence staining at least three independent experiments were conducted. Representative images are shown in the figures.

Mammary fat pad transplantation and analysis. Sorted cells were resuspended in 50% Matrigel, PBS with 20% FBS, and 0.04% Trypan Blue (Sigma), and injected in 10 μl volumes into the cleared fat pads of 3-week-old female. Reconstituted mammary glands were harvested after 8–10 weeks post-surgery. Outgrowths were detected by under a dissection microscope (Leica) after Carmine staining. Outgrowths with more than 10% of the host fat pad filled were scored as positive.

Statistical analysis. Student's *t*-test was performed and the *P* value was calculated in Prism on data represented by bar charts, which consisted of results from three independent experiments unless specified otherwise. For all experiments with error bars, the standard deviation (s.d.) was calculated to indicate the variation within each experiment. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Primers used in qPCR analysis. Primers used were as follows. Procr forward, CTCTCTGGGAAAACTCCTGACA; Procr reverse, CAGGGAGCAGCTAACA GTGA; K5 forward, TCTGCCATCACCCCATCTGT; K5 reverse, CCTCCGCC AGAACTGTAGGA; K14 forward, TGACCATGCAGAACCTCAATGA; K14 reverse, ATTGGCATTGTCCACGG; E-cadherin forward, CAGGTCTCCTCAT GGCTTTGC; E-cadherin reverse, CTTCCGAAAAGAAGGCTGTCC; Vim forward, CGTCCACACGCACCTACAG; Vim reverse, GGGGGATGAGGAATA GAGGCT; Lgr5 forward, CCTACTCGAAGACTTACCCAGT; Lgr5 reverse, GC ATTGGGGTGAATGATAGCA; Axin2 forward, AGCCTAAAGGTCTTATGTG GCTA; Axin2 reverse, ACCTACGTGATAAGGATTGACT.

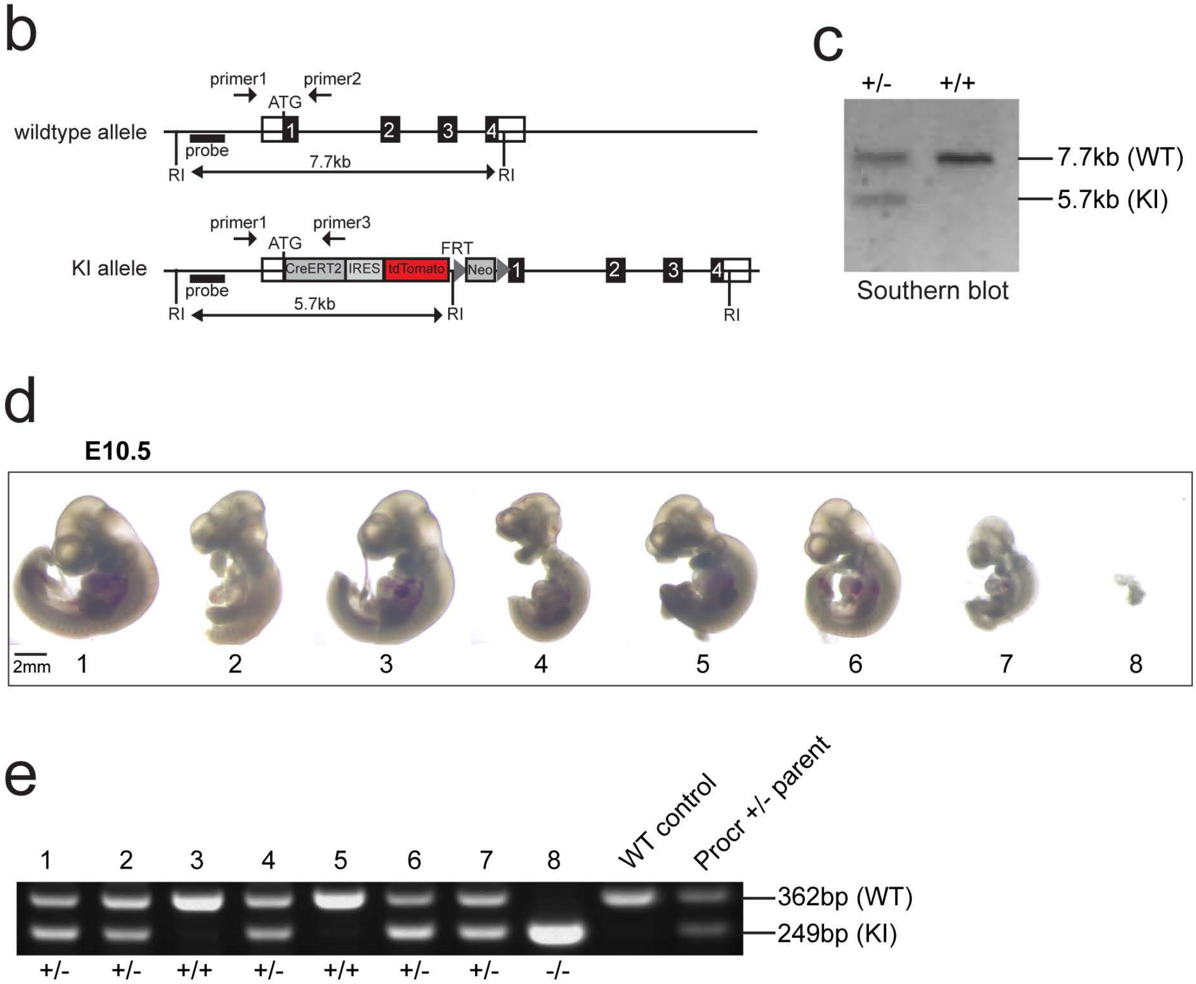
24. Lustig, B. *et al.* Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. *Mol. Cell. Biol.* **22,** 1184–1193 (2002).
25. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
26. Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423,** 448–452 (2003).
27. Shehata, M. *et al.* Phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* **14,** R134 (2012).



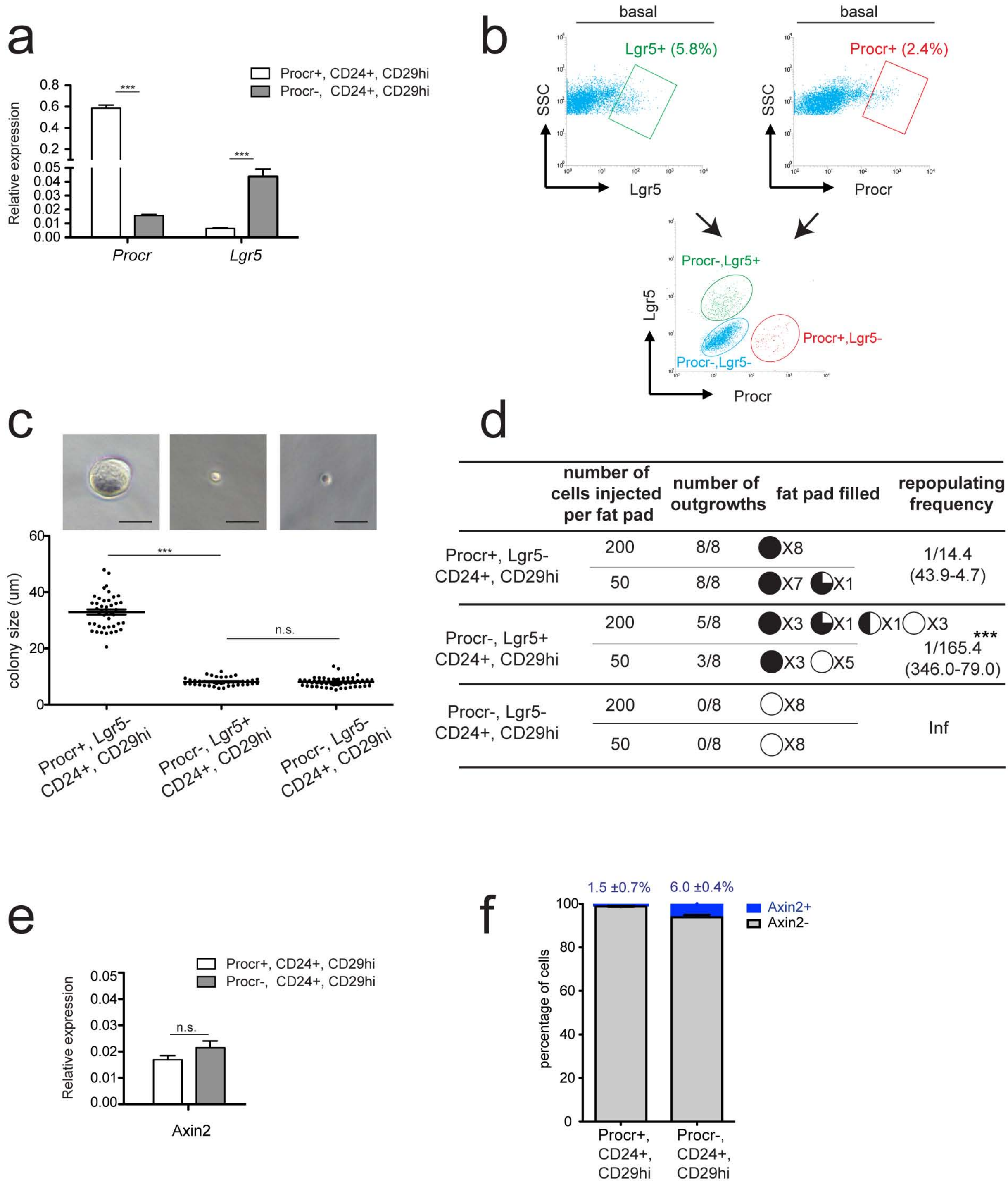
Extended Data Figure 1 | No Procr$^+$ cells are found in luminal cells throughout postnatal development. **a**, **b**, Microarray of three-dimensional cultured basal cells in the presence of Wnt3A versus vehicle. 1 and 2 represent two independent experiments. See Methods for details. qPCR indicated that *Procr* is upregulated in basal cells cultured in the presence of Wnt3A compared with cells grown in the absence of Wnt3A. Data are pooled from three independent experiments. Data are presented as mean ± s.d. ***$P < 0.01$. **c**–**g**, The 4th inguinal mammary glands harvested from 2-week-old (**c**), 5-week-old (**d**), 8-week-old (**e**), pregnant day 14.5 (**f**) and 2 weeks post-weaning (**g**) CD1 mice were analysed by FACS. Procr$^+$ cells were distributed in basal cells (ranging from 2.9% to 8.8%) and stromal fibroblasts (from 17.2% to 30.5%). No Procr$^+$ cells were detected in luminal cells at any postnatal stage.

a

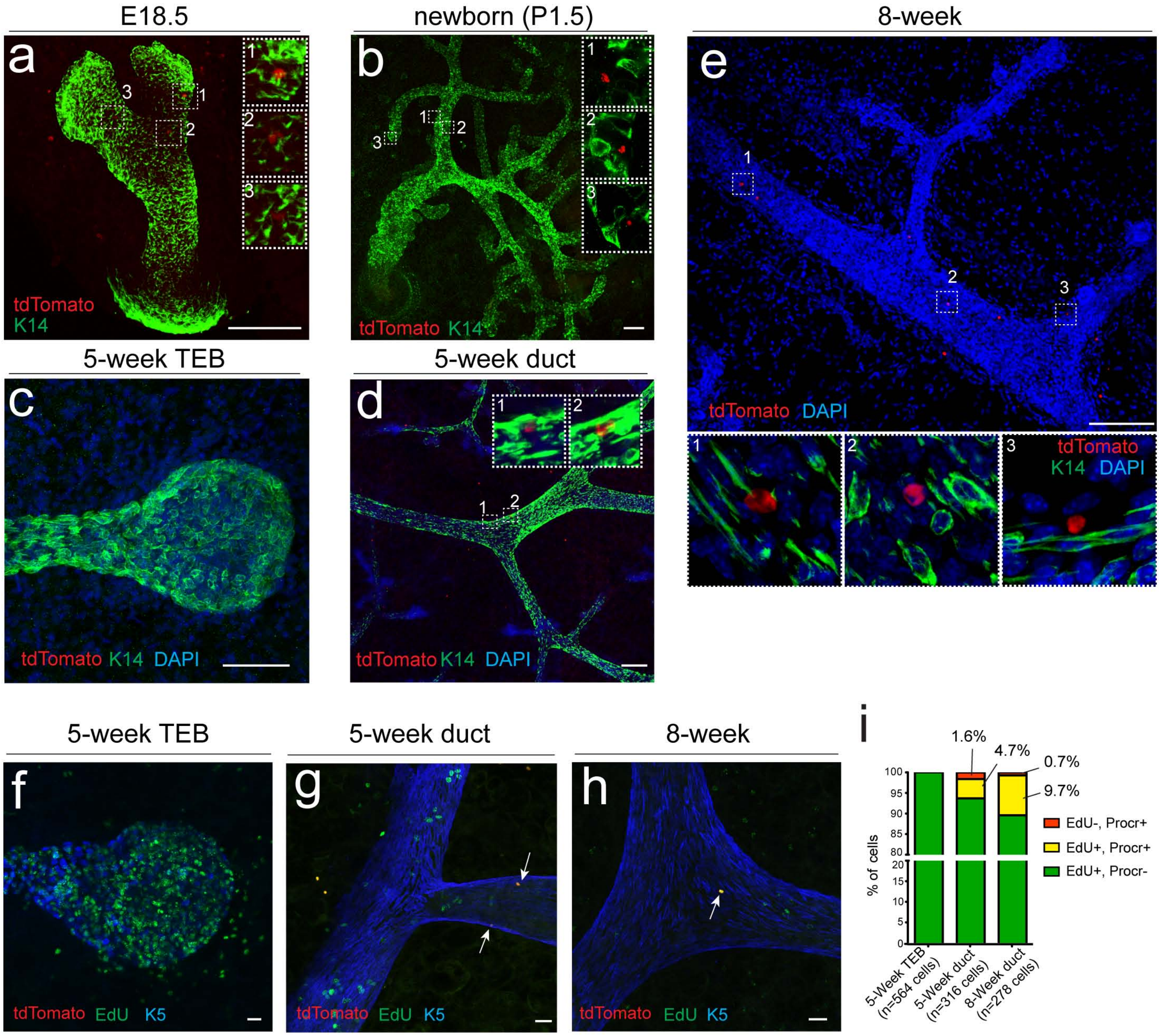
	number of cells plated	number of colonies formed	colony-forming efficiency
Lin-, CD24+, CD29hi	30K	2.7K	7.3 ± 0.6%
	20K	1.5K	
	15K	1K	
	10K	620	
Procr+, Lin-, CD24+, CD29hi	20K	7.8K	38.6 ± 4.1%***
	12K	5.2K	
	10K	2.7K	
	8K	3.6K	
Procr- Lin-, CD24+, CD29hi	30K	0	N/A Inf
	30K	0	
	20K	0	
	8K	0	



Extended Data Figure 2 | Generation of the *Procr*$^{CreERT2\text{-}IRES\text{-}tdTomato}$ knock-in mouse. **a**, Basal (Lin$^-$ CD24$^+$ CD29^{hi}), Procr$^+$ CD24$^+$ CD29^{hi} and Procr$^-$ CD24$^+$ CD29^{hi} cells were FACS-isolated and placed in Matrigel to assess the colony-formation ability. Data are pooled from four independent experiments. ***$P < 0.01$. **b**, Targeting strategy to generate the *Procr*$^{CreERT2\text{-}IRES\text{-}tdTomato/+}$ knock-in (KI) mouse. Designs of Southern blot probe and genotyping primers are as indicated. **c**, Southern blot analysis with a 5′ external probe of EcoRI-digested DNA from mouse embryonic stem cells, showing a 5.7 kb band in addition to the 7.7 kb wild-type (WT) band in clones that have undergone homologous recombination at the *Procr* locus. **d**, **e**, Embryos resulting from a cross of heterozygous male and female mice were dissected at E10.5 (**d**). Genotyping PCR indicated the proper distribution of wild type and heterozygotes as Mendel's law of segregation, and that homozygotes were lethal before this time point as the embryo had mostly been absorbed (**d**, **e**). One of three similar experiments is shown.



Extended Data Figure 3 | Procr^{+} cells and Lgr5^{+} cells are mutually exclusive populations in the mammary gland, while Procr^{+} cells and Axin2^{+} cells are largely non-overlapping in mammary basal cells. **a**, Procr^{+} CD24^{+} CD29^{hi} and Procr^{-} CD24^{+} CD29^{hi} cells were FACS-isolated and analysed by qPCR. Procr^{+} cells expressed significantly lower levels of *Lgr5* compared with Procr^{-} cells. Data are pooled from three independent experiments. $***P < 0.01$. **b**, Cells isolated from Lgr5–GFP mammary gland were analysed for the expression of GFP and Procr. 5.8% of basal cells (Lin^{-} CD24^{+} CD29^{hi}) were Lgr5–GFP^{+} cells, while 2.4% of basal cells were Procr^{+} cells. These two populations were not overlapped. **c**, Three basal subpopulations as indicated were FACS-isolated and cultured in Matrigel for colony formation. Only Procr^{+} Lgr5^{-} cells formed colonies whereas Procr^{-} Lgr5^{+} and Procr^{-} Lgr5^{-} cells could not. Data are presented as mean ± s.d. Scale bars, 20 μm. $***P < 0.01$. **d**, Recipient fat pads were injected with freshly sorted basal subpopulation cells as indicated and harvested at 8 weeks after surgery. Procr^{+} Lgr5^{-} cells efficiently formed new mammary glands (frequency 1/14.4). Comparably, Procr^{-} Lgr5^{+} cells had significantly lower reconstitution efficiency (1/165.4). Procr^{-} Lgr5^{-} cells were not able to reconstitute. **e**, Procr^{+} CD24^{+} CD29^{hi} and Procr^{-} CD24^{+} CD29^{hi} cells were FACS-isolated and analysed by qPCR. No significant difference in *Axin2* level was detected in the two populations. Data are pooled from three independent experiments. $***P < 0.01$. **f**, Procr^{+} CD24^{+} CD29^{hi} and Procr^{-} CD24^{+} CD29^{hi} cells were isolated from *Axin2-lacZ* mammary gland and underwent X-gal staining. About 1.5% of Procr^{+} CD24^{+} CD29^{hi} cells were Axin2-lacZ^{+}, while 6.0% of Procr^{-} CD24^{+} CD29^{hi} cells were Axin2-lacZ^{+}. Data are pooled from two independent experiments (n = 1,085 cells and n = 1,103 cells).



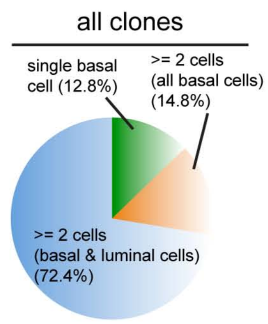
Extended Data Figure 4 | Procr^{+} cells are located in mammary ducts and are proliferative cells. **a**–**e**, The 4th inguinal mammary glands harvested from E18.5 (**a**), P1.5 (**b**), 5-week-old (**c**, **d**) and 8-week-old (**e**) *Procr*$^{CreERT2\text{-}IRES\text{-}tdTomato/+}$ mice were analysed by whole-mount confocal imaging. Individual tdTomato^{+} cells were dispersedly located in all stages of mammary ducts. tdTomato^{+} cells were not found in TEBs of 5-week-old glands (**c**). A minimum of 50 TEBs and 50 ducts were analysed. Scale bars, 100 µm. **f**–**i**, Analysis of proliferative cells in TEBs and ducts of *Procr*$^{CreERT2\text{-}IRES\text{-}tdTomato}$ mice at 3 h after EdU injection. Five-week-old TEBs exhibited abundant EdU^{+} cells (green) but no Procr^{+} cells (red) (**f**). EdU^{+} Procr^{+} cells (yellow) were found in both 5-week-old ducts (**g**) and 8-week-old mammary gland (**h**). Quantification is shown in **i**. Scale bars, 20 µm.

a Puberty labelled, 3-week tracing

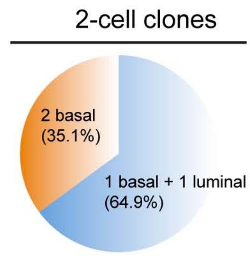
number of luminal cells per clone (columns); number of basal cells per clone (rows)

	0	1	2	3	4	5	6	7	8	10	12	total
1	38	61	36	15	6	6	1	0	1	0	0	164
2	33	23	12	13	5	4	2	2	0	0	1	95
3	5	8	3	2	2	0	1	1	1	1	0	24
4	6	2	4	0	0	0	0	0	1	0	0	13
5	0	1	1	0	0	0	0	0	0	0	0	2
total	82	95	56	30	13	10	4	3	3	1	1	298

b



c

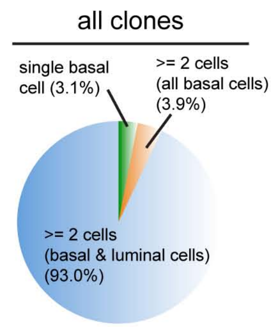


d Puberty labelled, 6-week tracing

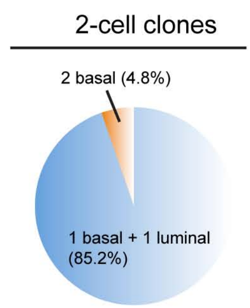
number of luminal cells per clone (columns); number of basal cells per clone (rows)

	0	1	2	3	4	5	6	7	8	9	>10	total
1	7	40	18	18	6	3	0	2	0	0	1	95
2	2	13	12	12	8	5	0	1	2	1	0	56
3	5	8	12	4	1	3	0	1	3	0	0	37
4	1	6	2	4	2	0	1	3	2	0	1	22
5	1	2	1	0	1	0	1	1	0	0	0	7
6	0	1	1	1	1	0	0	0	0	0	1	5
7	0	0	1	0	1	0	0	0	0	0	0	2
8	0	0	0	0	0	0	0	0	1	0	0	1
10	0	0	0	1	0	0	0	0	0	0	0	1
>10	0	0	0	0	0	0	0	0	0	1	0	1
total	16	70	47	40	20	11	2	8	8	2	3	227

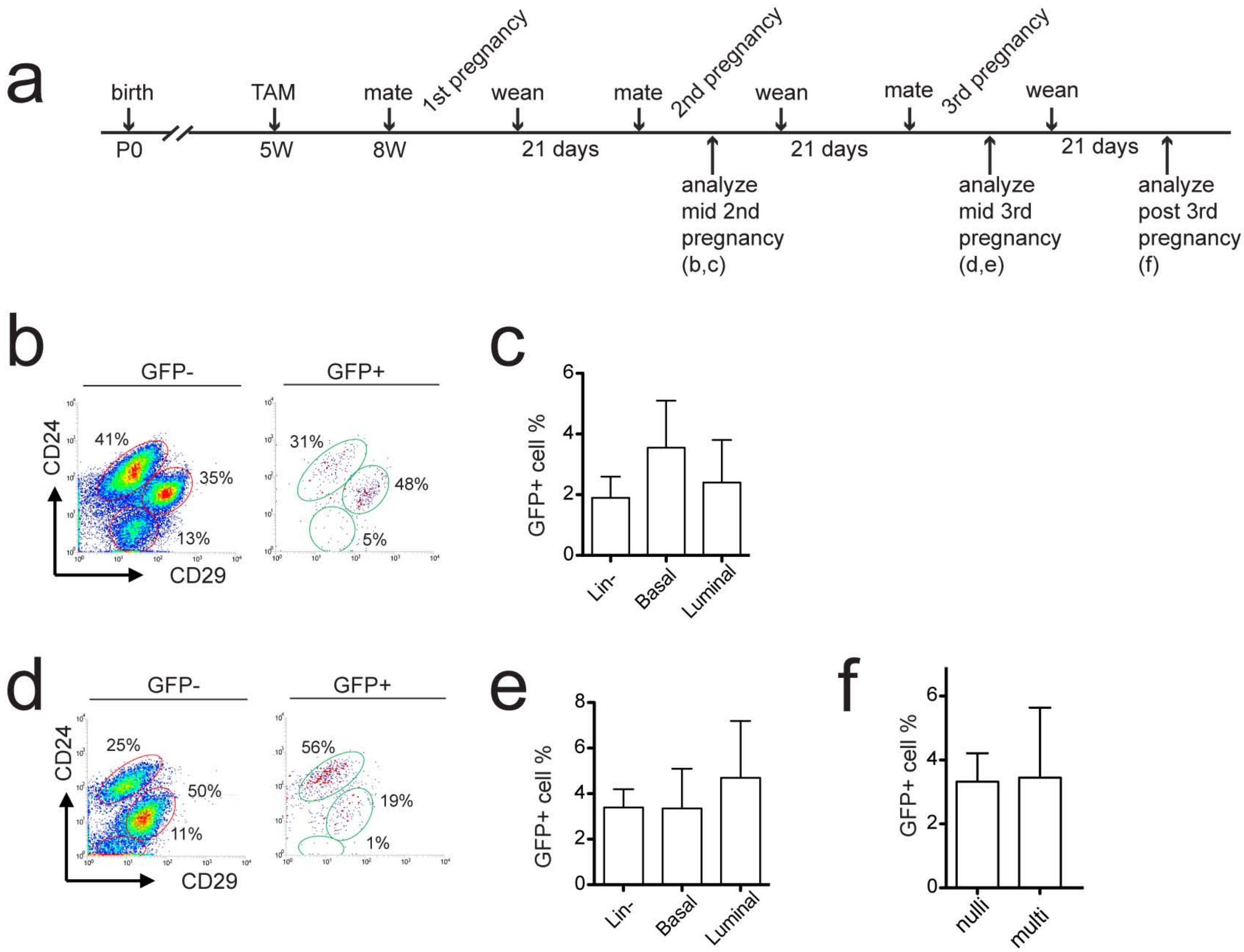
e



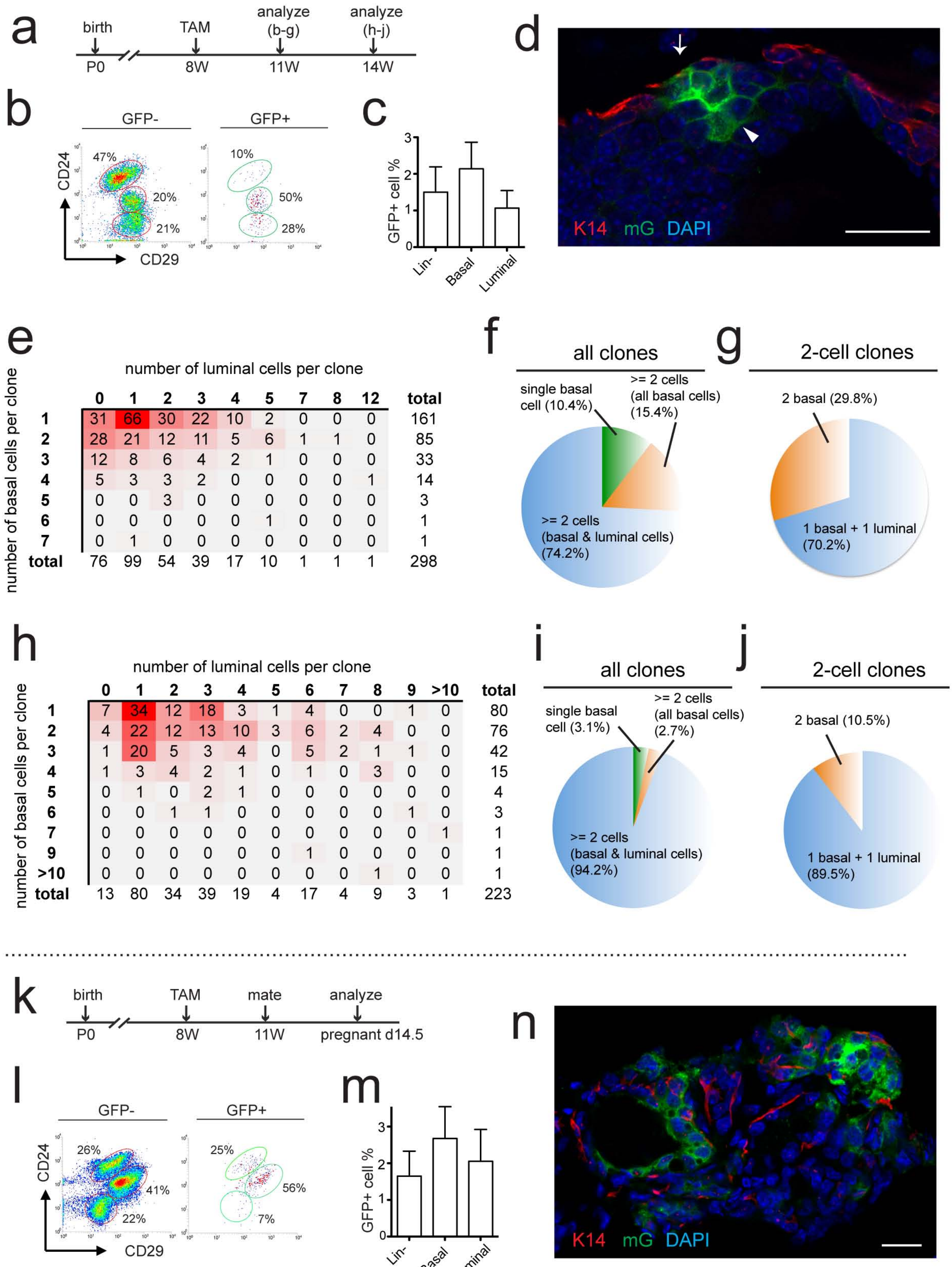
f



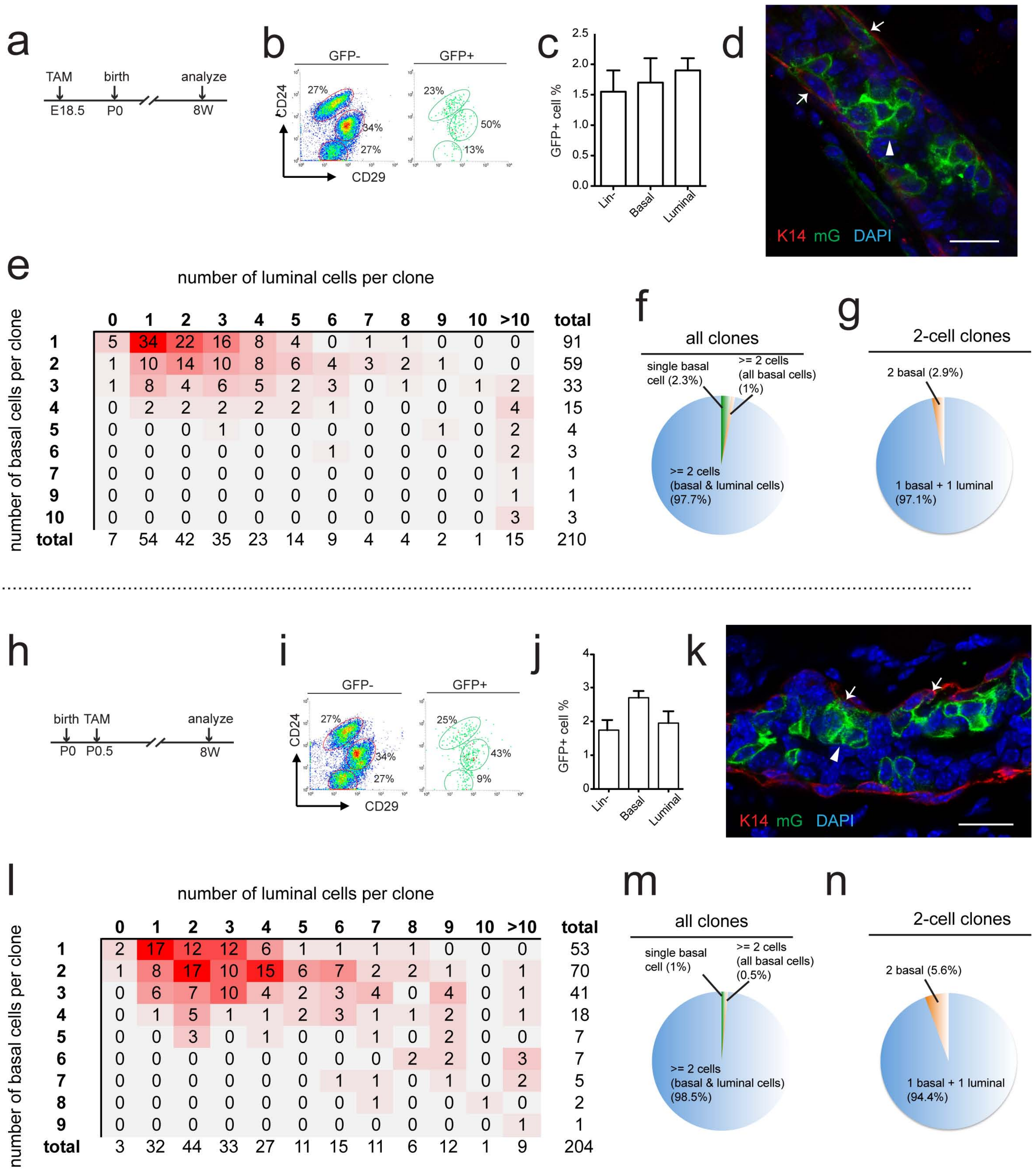
Extended Data Figure 5 | Quantitative clonal analysis of Procr-labelled cells in mammary glands induced in puberty. **a**–**f**, The number of basal and luminal cells in individual GFP$^+$ clones were scored in *Procr*$^{CreERT2/+}$; *R26*$^{mTmG/+}$ mammary glands after 3 weeks (**a**–**c**) or 6 weeks (**d**–**f**) induction. Basal cell numbers are shown along the *y*-axis, and luminal cell numbers are shown along the *x*-axis. Red shading indicates the relative frequency of certain clone composition, with deeper shading indicating higher frequency. **d**, Note that the deeper shading boxes shifted to the right in tracing experiments undertaken for a longer period. **b**, In clones after 3-week tracing, 72.4% were bi-lineage, 14.8% were solely basal cells derived from Procr$^+$ cell division, 12.8% were single basal cells that had not divided. **c**, Among two-cell clones, 64.9% were composed of one basal cell and one luminal cell, while 35.1% consisted of two basal cells. **d**, **e**, In clones after 6-week tracing, the proportion of bi-lineage clone increased to 93%, while the percentage of the other two groups decreased to 3.9% and 3.1%. **f**, In two-cell clones, bi-lineage clones increased to 85.2%, while clones consisting of two basal cells decreased to 4.8%. $n = 4$ mice for 3-week tracing and $n = 3$ mice for 6-week tracing.



Extended Data Figure 6 | Procr$^+$ cells are long-lived multipotent MaSCs retained beyond multiple rounds of pregnancy. **a**–**d**, Tamoxifen (TAM) was administered in 5-week-old *Procr$^{CreERT2/+}$;R26$^{mTmG/+}$* mice. Labelled cell contribution was analysed as illustrated in **a**. **b**–**e**, FACS analysis indicating that GFP$^+$ cells are distributed in both basal and luminal layer in mid-2nd pregnancy (**b**, **c**) and mid-3rd pregnancy (**d**, **e**). **f**, Quantification of GFP$^+$ cells indicating no difference in the percentage of GFP$^+$ basal cells between nulliparous mice and multiparous mice that have gone through three complete cycles of pregnancy and involution. $n = 3$ mice. **c**, **e**, **f**, Data are presented as mean ± s.d.



Extended Data Figure 7 | Procr labels multipotent mammary stem cells in mature adult mice. **a**–**g**, Tamoxifen (TAM) was administered in 8-week old *Procr*$^{CreERT2/+}$;*R26*$^{mTmG/+}$ mice. Labelled cell contribution was analysed after 3-week or 6-week induction. After 3 weeks, FACS analysis indicated that GFP$^+$ cells were distributed in both basal and luminal layers (**b**, **c**). Immunostaining in sections showed the clonal expansion of GFP$^+$ cells and confirmed their distribution in both basal and luminal layers. Basal cells were marked by K14, while cells apical to K14$^+$ cells were luminal cells (arrow and arrowhead in **d**). Clonal analysis indicated that bi-lineage clones are the majority in all clones (74.2%) (**e**, **f**), and in two-cell clones (70.2%) (**g**). **h**–**j**, Clonal analysis of 6-week induction indicating that clone sizes are larger (red shaded boxes shifted to the right) (**h**), bi-lineage clone percentage has also increased to 94% in all clones (**i**) and to 89.5% in two-cell clones (**j**). **k**–**n**, Mammary glands were analysed at pregnant day 14.5 after tamoxifen administration at 8 weeks. GFP$^+$ cells were in both basal and luminal layers as indicated by FACS analysis (**l**, **m**). GFP$^+$ cells contributed to alveologenesis by immunohistochemistry analysis in sections (**n**). $n = 3$ mice for 3-week tracing and $n = 3$ mice for 6-week tracing. Scale bars, 20 μm. **c**, **m**, Data are presented as mean ± s.d.



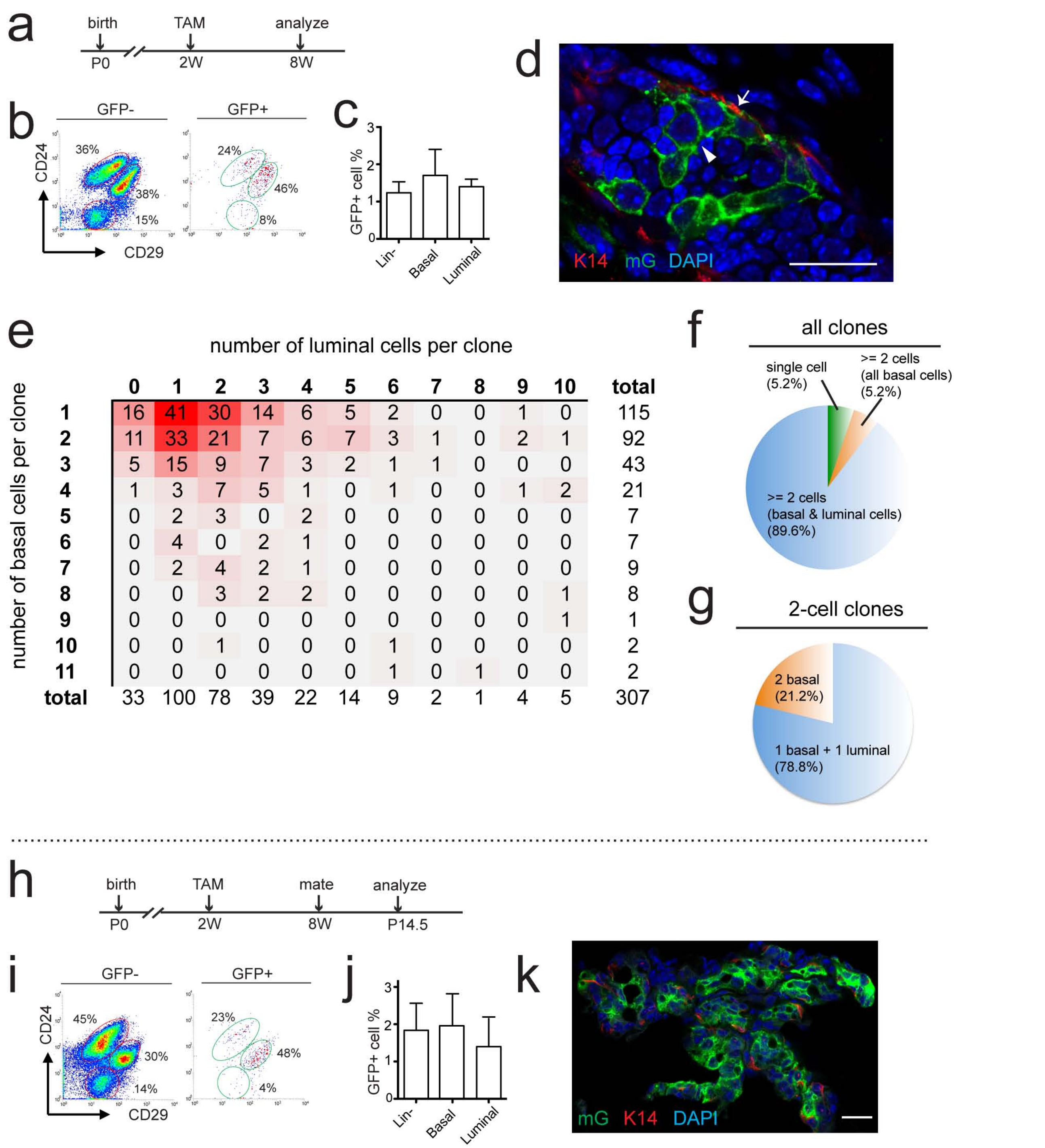
e

	0	1	2	3	4	5	6	7	8	9	10	>10	total
1	5	34	22	16	8	4	0	1	1	0	0	0	91
2	1	10	14	10	8	6	4	3	2	1	0	0	59
3	1	8	4	6	5	2	3	0	1	0	1	2	33
4	0	2	2	2	2	2	1	0	0	0	0	4	15
5	0	0	0	1	0	0	0	0	0	1	0	2	4
6	0	0	0	0	0	0	1	0	0	0	0	2	3
7	0	0	0	0	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	0	0	0	0	1	1
10	0	0	0	0	0	0	0	0	0	0	0	3	3
total	7	54	42	35	23	14	9	4	4	2	1	15	210

l

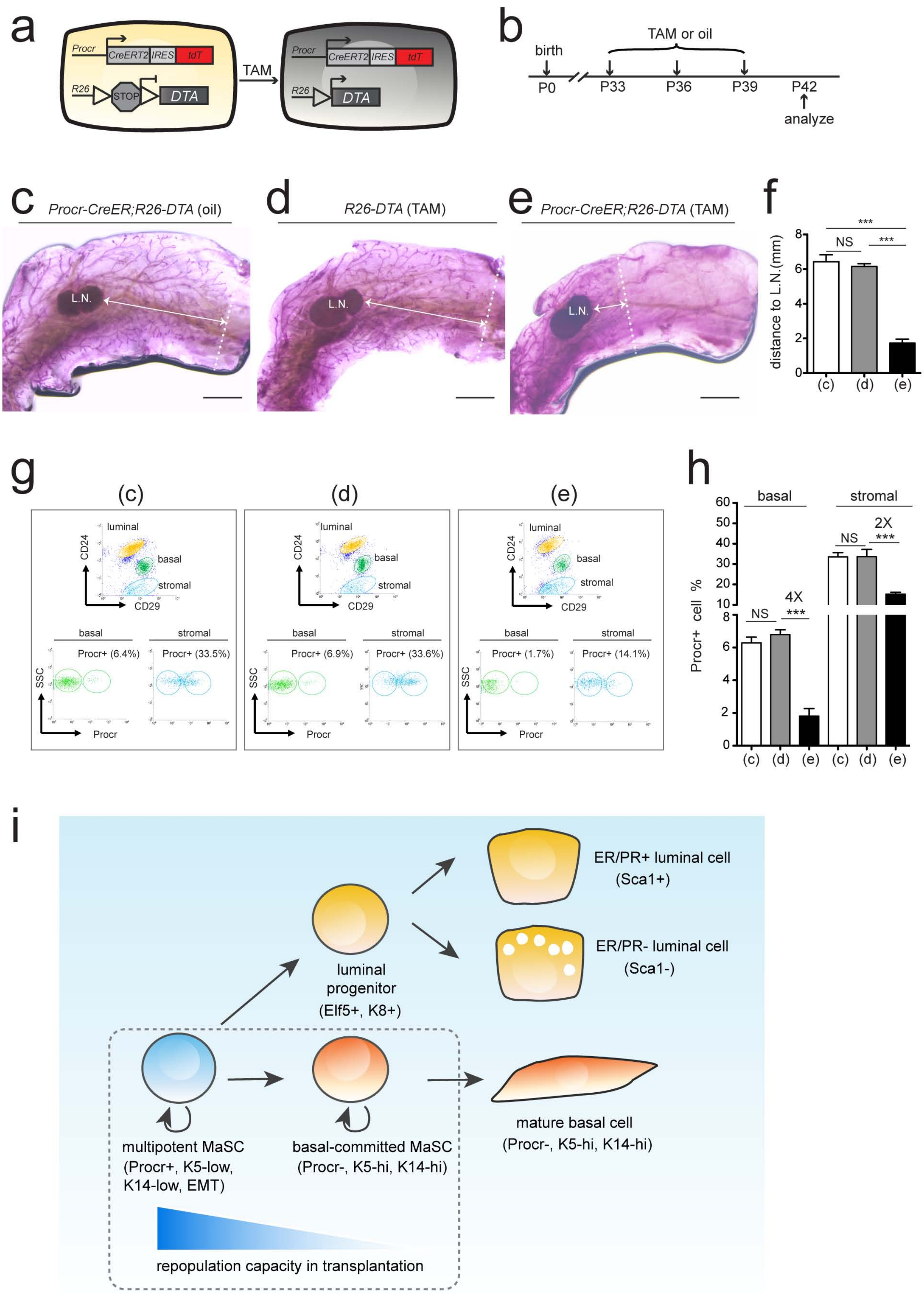
	0	1	2	3	4	5	6	7	8	9	10	>10	total
1	2	17	12	12	6	1	1	1	1	0	0	0	53
2	1	8	17	10	15	6	7	2	2	1	0	1	70
3	0	6	7	10	4	2	3	4	0	4	0	1	41
4	0	1	5	1	1	2	3	1	1	2	0	1	18
5	0	0	3	0	1	0	0	1	0	2	0	0	7
6	0	0	0	0	0	0	0	0	2	2	0	3	7
7	0	0	0	0	0	0	1	1	0	1	0	2	5
8	0	0	0	0	0	0	0	1	0	0	1	0	2
9	0	0	0	0	0	0	0	0	0	0	0	1	1
total	3	32	44	33	27	11	15	11	6	12	1	9	204

Extended Data Figure 8 | Procr$^+$ cells are multipotent MaSCs in embryonic or newborn mammary gland. **a**–**g**, Tamoxifen (TAM) was administered in pregnant day 18.5 mothers bearing *Procr*$^{CreERT2/+}$;*R26*$^{mTmG/+}$ mice. Labelled cell contribution in the pups was analysed after 8-week induction (**a**). FACS analysis indicated that GFP$^+$ cells are distributed in both basal and luminal layers (**b**, **c**). Immunostaining in sections showed the clonal expansion of GFP$^+$ cells and confirmed their distribution in both basal and luminal layers (arrow and arrowhead in **d**). Scale bar, 20 µm. Clonal analysis indicated that bi-lineage clones are the majority in all clones (97.7%) (**e**, **f**), and in two-cell clones (97.1%) (**g**). $n = 5$ mice. **c**, Data are presented as mean ± s.d. **h**–**n**, Tamoxifen was administered in P0.5 *Procr*$^{CreERT2/+}$;*R26*$^{mTmG/+}$ mice. Labelled cell contribution was analysed after 8-week induction (**h**). FACS analysis indicated that GFP$^+$ cells are distributed in both basal and luminal layers (**i**, **j**). Immunostaining in sections showed the clonal expansion of GFP$^+$ cells and confirmed their distribution in both basal and luminal layers. Basal cells were marked by K14, while cells apical to K14$^+$ cells were luminal cells (arrows and arrowhead in **k**). Scale bar, 20 µm. Clonal analysis indicated that bi-lineage clones are the majority in all clones (98.5%) (**l**, **m**), and in two-cell clones (94.4%) (**n**). $n = 5$ mice. **j**, Data are presented as mean ± s.d.



	0	1	2	3	4	5	6	7	8	9	10	total
1	16	41	30	14	6	5	2	0	0	1	0	115
2	11	33	21	7	6	7	3	1	0	2	1	92
3	5	15	9	7	3	2	1	1	0	0	0	43
4	1	3	7	5	1	0	1	0	0	1	2	21
5	0	2	3	0	2	0	0	0	0	0	0	7
6	0	4	0	2	1	0	0	0	0	0	0	7
7	0	2	4	2	1	0	0	0	0	0	0	9
8	0	0	3	2	2	0	0	0	0	0	1	8
9	0	0	0	0	0	0	0	0	0	0	1	1
10	0	0	1	0	0	0	1	0	0	0	0	2
11	0	0	0	0	0	0	1	0	1	0	0	2
total	33	100	78	39	22	14	9	2	1	4	5	307

Extended Data Figure 9 | Procr labels multipotent mammary stem cells in prepubescent mice. **a**–**d**, Tamoxifen (TAM) was administered in 2-week old *Procr*$^{CreERT2/+}$;*R26*$^{mTmG/+}$ mice. Labelled cell contribution was analysed at 8 weeks. FACS analysis indicated that GFP^{+} cells are distributed in both basal and luminal layers (**b**, **c**). Immunostaining in sections showed the clonal expansion of GFP^{+} cells and confirmed their distribution in both basal and luminal layers. Basal cells were marked by K14, while cells apical to K14^{+} cells were luminal cells (arrow and arrowhead in **d**). **e**–**g**, Clonal analysis indicated that bi-lineage clones are the majority in all clones (89.6%) (**e**, **f**), and in two-cell clones (78.8%) (**g**). $n = 4$ mice. **h**–**k**, Mammary glands were analysed at day 14.5 gestation after tamoxifen administration at 2 weeks. GFP^{+} cells were in both basal and luminal layers as indicated by FACS analysis (**i**, **j**). GFP^{+} cells contributed to alveologenesis by immunohistochemistry analysis in sections (**k**). Scale bars, 20 μm. **c**, **j**, Data are presented as mean ± s.d.



a
Procr
CreERT2
IRES
tdT
R26
STOP
DTA
TAM
b
TAM or oil
birth
P0
P33
P36
P39
P42
analyze
c
Procr-CreER;R26-DTA (oil)
d
R26-DTA (TAM)
e
Procr-CreER;R26-DTA (TAM)
L.N.
f
distance to L.N.(mm)
NS
***
(c) (d) (e)
g
(c) (d) (e)
luminal
basal
stromal
CD24
CD29
SSC
Procr
Procr+ (6.4%)
Procr+ (33.5%)
Procr+ (6.9%)
Procr+ (33.6%)
Procr+ (1.7%)
Procr+ (14.1%)
h
Procr+ cell %
basal
stromal
NS
4X
2X
***
i
ER/PR+ luminal cell
(Sca1+)
ER/PR- luminal cell
(Sca1-)
luminal
progenitor
(Elf5+, K8+)
multipotent MaSC
(Procr+, K5-low,
K14-low, EMT)
basal-committed MaSC
(Procr-, K5-hi, K14-hi)
mature basal cell
(Procr-, K5-hi, K14-hi)
repopulation capacity in transplantation

Extended Data Figure 10 | Procr$^+$ cells are important for mammary development. **a**, Schematic illustration of targeted ablation of Procr$^+$ cells using the Procr-CreERT2 model to drive expression of DTA. **b**, Tamoxifen (TAM) was administered every 3 days a total of three times followed by analysing the 4th mammary gland. **c**–**e**, Whole-mount imaging of the mammary epithelium at P42. The lymph node (L.N.) is indicated. Both the oil-treated control mammary epithelium (**c**) and the tamoxifen-treated $R26^{DTA/+}$ control mammary epithelium (**d**) had grown to the distal edge of the fat pad. Tamoxifen administration in $Procr^{CreERT2/+};R26^{DTA/+}$ mice largely prevented the growth of the epithelium: the forefront of the epithelium halted at a position close to where the forefront was at the initiation of cell ablation (slightly past the lymph node) (**e**). Scale bars, 2 mm. **f**, Quantification of the distance from the epithelium forefront to the lymph node indicated that epithelium extension in the Procr$^+$ cell-ablation group is largely compromised (comparing **e** with **d** or **c**). $^{***}P < 0.01$. NS, not significant. **g**, **h**, FACS analysis indicating that Procr$^+$ basal and stromal cells are ablated (fourfold and twofold). $n = 3$ mice. The role of Procr$^+$ stromal cells in this study should also be taken into consideration as they were also affected by the ablation. Nonetheless, the reduction of Procr$^+$ stromal cells was not as pronounced as Procr$^+$ basal cells, probably due to the less proliferative nature of Procr$^+$ fibroblasts, thereby fewer progeny cells were affected. Data are presented as mean ± s.d. $^{***}P < 0.01$. **i**, Multipotent and unipotent MaSCs coexist in the mammary epithelial cell hierarchy. Multipotent MaSCs are characterized by Lin$^-$ CD24$^+$ CD29^{hi} Procr$^+$ K5^{low} K14^{low}, and express EMT features. Multipotent MaSCs generate all differentiated cell types, as determined by lineage tracing, and display the highest repopulation efficiency by transplantation. Basal-committed MaSCs are destined for basal cells in development, yet can repopulate both basal and luminal cells in transplantation, underlining the plasticity of basal-committed MaSCs in response to intervention. Luminal progenitors contribute to only luminal cells in lineage tracing and are not able to repopulate in transplantation. Their markers are as previously reported[3,4,27].