ARTICLES
https://doi.org/10.1038/s41556-018-0095-2

nature cell biology

Corrected: Publisher Correction

Early lineage segregation of multipotent embryonic mammary gland progenitors

Aline Wuidart[1,5], Alejandro Sifrim[2,3,5], Marco Fioramonti[1], Shigeru Matsumura[1], Audrey Brisebarre[1], Daniel Brown[2], Alessia Centonze[1], Anne Dannau[1], Christine Dubois[1], Alexandra Van Keymeulen[1], Thierry Voet[2,3] and Cédric Blanpain[1,4*]

The mammary gland is composed of basal cells and luminal cells. It is generally believed that the mammary gland arises from embryonic multipotent progenitors, but it remains unclear when lineage restriction occurs and what mechanisms are responsible for the switch from multipotency to unipotency during its morphogenesis. Here, we perform multicolour lineage tracing and assess the fate of single progenitors, and demonstrate the existence of a developmental switch from multipotency to unipotency during embryonic mammary gland development. Molecular profiling and single cell RNA-seq revealed that embryonic multipotent progenitors express a unique hybrid basal and luminal signature and the factors associated with the different lineages. Sustained p63 expression in embryonic multipotent progenitors promotes unipotent basal cell fate and was sufficient to reprogram adult luminal cells into basal cells by promoting an intermediate hybrid multipotent-like state. Altogether, this study identifies the timing and the mechanisms mediating early lineage segregation of multipotent progenitors during mammary gland development.

The mammary gland (MG) is a branched epithelium that produces milk during lactation. It is composed of two main lineages: basal cells (BCs), surrounding inner luminal cells (LCs). The LCs can be subdivided into oestrogen receptor (esr1 or ER)-positive and negative ductal cells, and alveolar cells that produce milk[1].

The MG derives from the ectoderm at around embryonic day 10.5 (E10.5). At E13, the MG placodes invaginate to form buds that continue to sprout until E16, when they start to branch. By E18.5, the epithelium forms a rudimentary ductal structure. From E18.5, the MG grows proportionally to the body size until puberty when oestrogen stimulates rapid growth and further branching of the MG. During pregnancy and lactation, MG further develops and gives rise to alveolar LCs that differentiate into milk-producing cells. At the end of lactation, the MG involutes and goes back to its virgin appearance, ready to undergo a new cycle of growth in the next pregnancy[1].

Lineage tracing experiments demonstrate that postnatal pubertal development and adult remodelling are mediated by unipotent basal and luminal progenitors/stem cells[2–12]. Whereas multicolour clonal analysis combined with statistical modelling demonstrates the unipotency of adult BCs and LCs[10–12], such experimental approaches have never been undertaken during MG development. Lineage tracing of the keratin 14 (K14)-expressing cells that compose the embryonic MG at E17 demonstrated that both basal and luminal lineages arise from K14-expressing cells during embryonic development[8], and suggest the existence of embryonic multipotent progenitors (EMPs). However, these experiments cannot discriminate whether the apparent multipotency of embryonic MG arises from the labelling of distinct pools of already pre-committed BCs and LCs or whether EMPs are truly multipotent at the single cell level. In addition, it remains unclear when the basal and luminal lineage segregation occurs and what are the mechanisms responsible for the switch from multipotency to unipotency during MG morphogenesis.

Here, using multicolour clonal analysis in mice, we demonstrate the multipotency of EMPs and the existence of a switch from multipotency to unipotency that occurs during embryonic MG development. Using molecular profiling and single cell RNA sequencing, we demonstrate that multipotency is associated with a hybrid basal and luminal gene expression signature. Finally we show the key role of p63 in promoting BC fate in EMPs.

Results

Clonal analysis demonstrates the switch from multipotency to unipotency during MG development. To assess whether MG arises from early multipotent progenitors or from a mixture of different lineage restricted progenitors, we performed clonal analysis using lineage tracing experiments at the early stages of MG development, when K14 is homogenously expressed in all MG cells (Fig. 1a). To this end, we generated K14rtTA/TetO-Cre/Rosa-Confetti mice (Fig. 1b) and titrated the dose of doxycycline (DOX) that leads to clonal labelling of the MG. Among the four colours of the confetti reporter system, nGFP is much less frequently recombined than the other fluorescent proteins[10,13,14]. Consequently, nuclear GFP (nGFP) can be used to further ensure clonal labelling in lineage tracing experiments. By administrating 1 μg of DOX per g of mouse to pregnant mice at E13 by intravenous injection (i.v.), we found that about 80% of the MGs were not labelled by nGFP (Fig. 1c,d). Similar proportions of MGs (20%) were labelled with nGFP two days after the injection (E15) and at postnatal day 5 (P5) when MG has branched and basal and luminal lineages are clearly separated (Fig. 1d–f).

[1]Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles, Brussels, Belgium. [2]Department of Human Genetics, University of Leuven, KU Leuven, Leuven, Belgium. [3]Sanger Institute—EBI Single-Cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, UK. [4]Walloon Excellence in Life Sciences and Biotechnology (WELBIO), Université Libre de Bruxelles (ULB), Brussels, Belgium. [5]These authors contributed equally: Aline Wuidart and Alejandro Sifrim. *e-mail: Cedric.Blanpain@ulb.ac.be

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

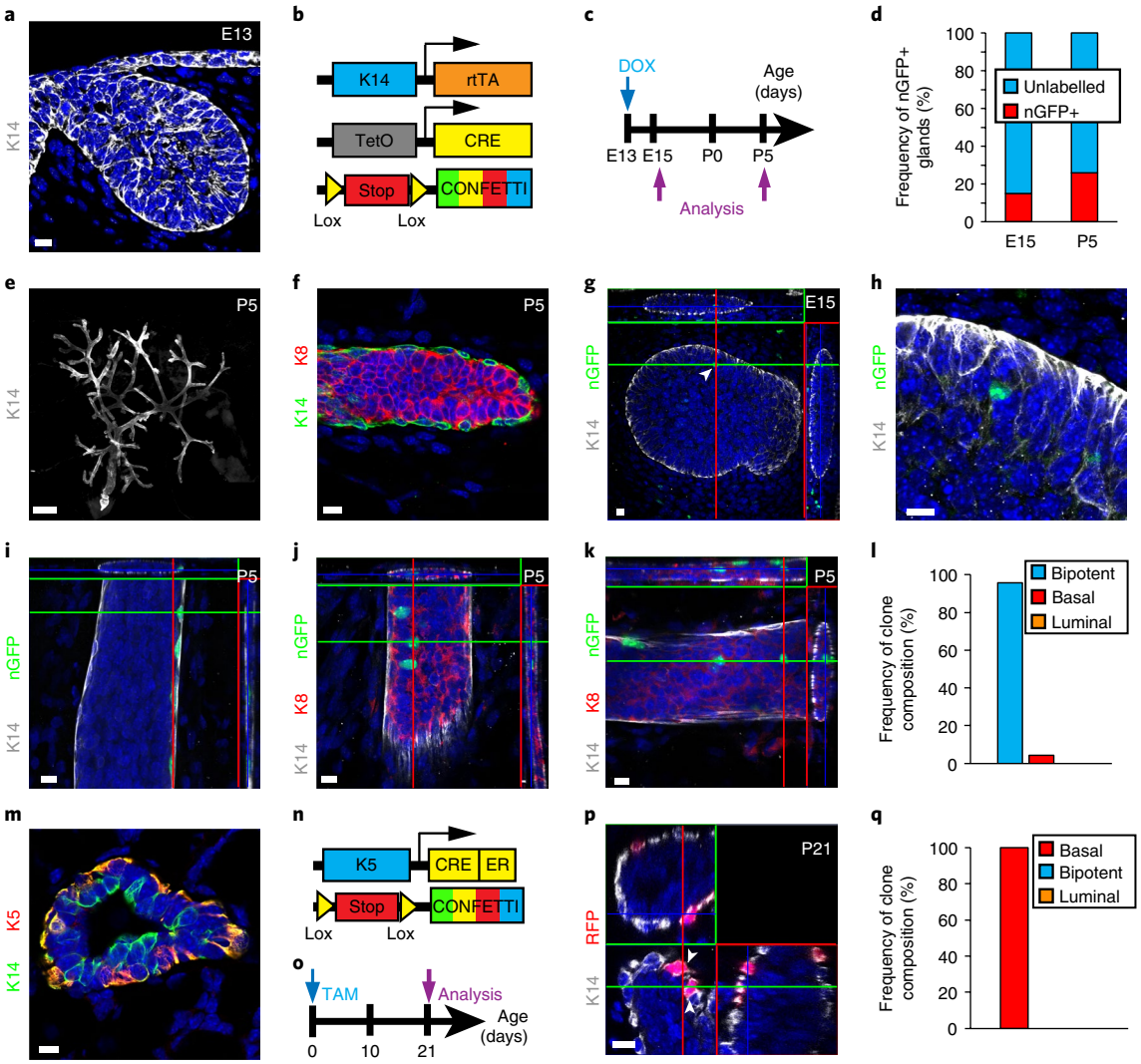

Fig. 1 | Clonal analysis demonstrates the switch from multipotency to unipotency during MG development. a, Confocal imaging of immunostaining of K14 in embryonic MG (E13) (eight embryos). **b**, Genetic strategy used to target Confetti expression in K14-expressing cells. **c**, Protocol used to study the fate of cells targeted during embryogenesis. **d**, Graph representing the fraction of glands containing nGFP+ cells at E15 ($n=73$ from eight embryos) and P5 ($n=85$ from 13 mice). **e**,**f**, Confocal imaging of immunostaining of K14 in P5 postnatal MG at low magnification (**e**) and of K14 and K8 in P5 postnatal MG (**f**) (85 glands from 13 mice). **g**, Confocal imaging of immunostaining of K14 and nGFP in E15 K14rtTA/TetO-Cre/Rosa-Confetti embryo induced at E13 with 1 µg DOX per g shows clonal induction in mammary buds (arrow). **h**, Zoom onto the labelled cell (73 glands from 8 embryos). **i–k**, Confocal imaging of immunostaining of K14, nGFP and/or K8 in P5 K14rtTA/TetO-Cre/Rosa-Confetti glands induced at E13 with 1 µg DOX per g shows the presence of isolated BCs (**i**), isolated LCs (**j**) and adjacent BCs and LCs (**k**) (85 glands from 13 mice). **l**, Graph representing the frequency of nGFP clone composition observed in P5 K14rtTA/TetO-Cre/Rosa-Confetti mice induced at E13 with 1 µg DOX per g ($n=85$ gland from 13 mice). **m**, Confocal imaging of immunostaining of K14 and K5 in P5 MG (three mice). **n**, Genetic strategy used to target Confetti expression in K5-expressing cells. **o**, Protocol used to study the fate of cells targeted at birth. **p**, Confocal imaging of immunostaining of K14 and nGFP in P21 K5CreER/Rosa-Confetti mice induced at birth with 50 µg of TAM (12 glands from three mice). **q**, Graph representing the frequency of clone composition observed in P21 K5CreER/Rosa-Confetti mice induced at birth with 50 µg of TAM ($n=36$ clones from 12 glands from three mice). In **g**, **i**, **j**, **k** and **p** orthogonal projections of three-dimensional (3D) stacks are shown. Scale bars except **e**, 10 µm. Scale bar in **e**, 500 µm. Arrowheads represent labelled nGFP cells at E15 (**g**) and labelled BCs at P21 (**p**). See Supplementary Table 1 for source data for **d**, **l** and **q**.

At E15, MG contained one to three nGFP+ cells spatially close to each other (Fig. 1g,h), consistent with the clonal expression of nGFP. Interestingly, at P5, almost all nGFP clones contained both BCs and LCs, even though BCs and LCs could be relatively distant from one another (Fig. 1i–l). These data clearly show that the MG initially develops through multipotent progenitors.

It has previously been shown that pubertal development and adult MG homeostasis are mediated by distinct lineage-restricted stem cells[2–12]. However, it is still unclear when the lineage restriction between BCs and LCs occurs. At P1, keratin 5 (K5) was specifically expressed in the outer layer of the MG (Fig. 1m). Clonal lineage tracing of BCs using K5CreER^{T2}/Rosa-Confetti mice at P1 lead to the exclusive labelling of BCs (Fig. 1n–q), showing that at P1 all BCs are already unipotent.

Hybrid adult basal and luminal gene expression in EMPs. To understand the molecular mechanisms regulating embryonic multipotency, we developed a strategy to FACS (fluorescence-activated cell sorting) isolate EMPs at E14, enabling us to compare their transcriptome to adult BCs and LCs. To this end, we used Lgr5-GFP

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

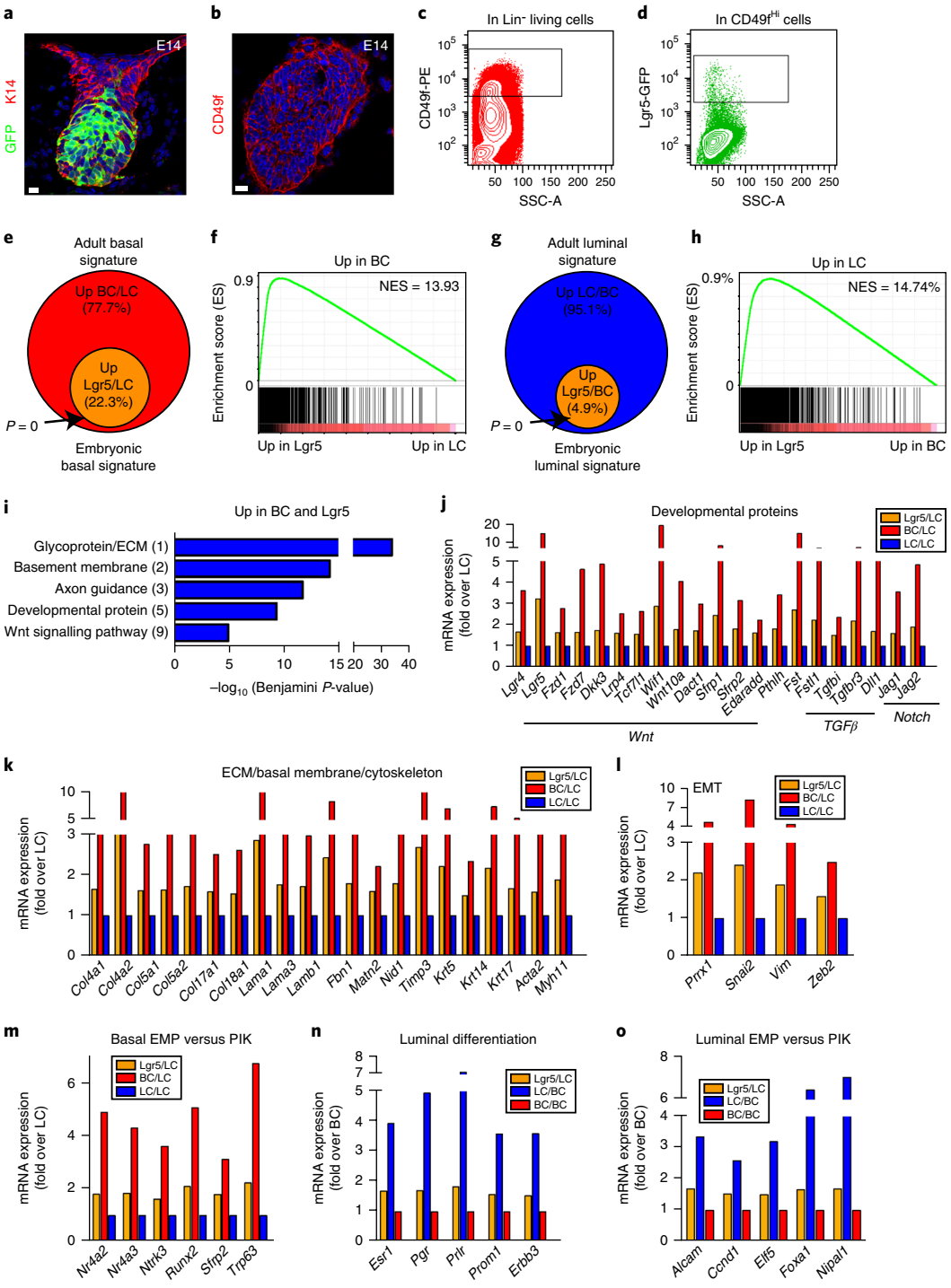

Fig. 2 | Transcriptional profiling of EMPs reveals their compound basal and luminal gene signature. a,**b**, Confocal imaging of immunostaining of K14 and GFP in Lgr5-IRES-GFP embryo at E14 (**a**) and of CD49f in wild-type E14 embryo (**b**) (five embryos analysed). **c**,**d**, FACS analysis of CD49f and GFP expression in Lin$^-$ epithelial cells (**c**) and Lin$^-$ CD49f^{Hi} mammary cells (**d**) in E14 Lgr5-GFP embryos (five embryos analysed). **e**, Venn diagram showing the overlap between the genes upregulated 1.5-fold in BCs compared to LCs (adult basal signature) and in Lgr5 cells compared to LCs (embryonic basal signature). **f**, GSEA of the upregulated genes in BCs (versus LCs) with the genes upregulated in Lgr5 cells (versus LCs), showing the enrichment of the basal signature in Lgr5 cells. Normalized enrichment score (NES) is indicated in figure **g**, Venn diagram showing the overlap between the genes upregulated in LCs compared to BCs (adult luminal signature) and in Lgr5 cells compared to BCs (embryonic luminal signature). **h**, GSEA of the upregulated genes in LCs (versus BCs) with the genes upregulated in Lgr5 cells (versus BCs), showing enrichment of the LC signature in Lgr5 cells. **i**, Gene ontology (GO) analysis of genes upregulated more than 1.5-fold in both BCs and Lgr5 cells compared to LCs. Histograms represent –log$_{10}$ of the Benjamini score. **j**–**m**, Graphs representing mRNA expression measured by microarray analysis of upregulated genes in FACS-isolated BCs and Lgr5 cells (fold over LC), showing the transcriptional priming of BC genes in Lgr5 cells. **n**,**o**, Graphs representing mRNA expression measured by microarray analysis of upregulated genes in FACS-isolated LCs and Lgr5 cells (fold over BC), showing the transcriptional priming of LC genes in Lgr5 cells. PIK, pik3ca (H1047R). Analyses presented in **e**–**o** are derived from the fold change ratio of the mean of Lgr5 microarray data ($n=3$) over the mean of BC ($n=2$) or LC ($n=2$) microarray data. Enrichment P values in **e** and **g** are derived from hypergeometric tests performed with R software without adjustment ($n=2{,}958$ and 3,999 genes, respectively, for BC/LC and LC/BC signatures). Scale bars, 10 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

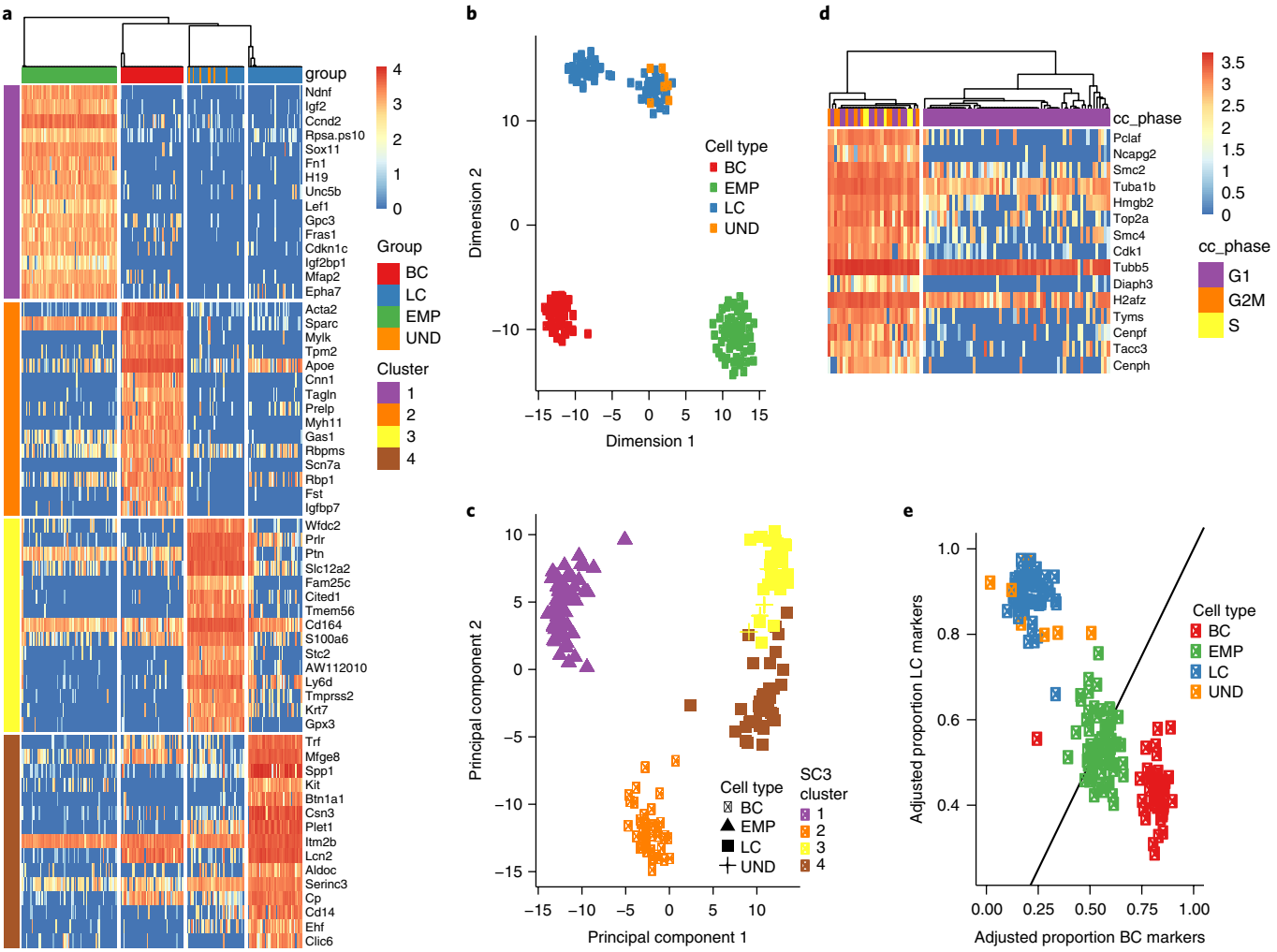

Fig. 3 | Single-cell RNA-seq shows the EMP hybrid gene signature. a, Unsupervised clustering using SC3 on EMPs ($n=68$), adult BC ($n=45$) and LCs (n = 73) using clustering parameter $k=4$. Heatmaps of the top 15 marker genes for each cluster and their corresponding normalized expression (area under the curve (AUC) > 0.8 and Wilcoxon signed rank test false discovery rate (FDR) adjusted P value < 0.01). Columns represent single cells, colour-coded by their respective lineage. UND (undetermined significance) represents few FACS-isolated CD29^{Hi} CD24^{+} cells with LC gene signature, which probably represent errors during cell sorting. **b**,**c**, Dimensionality reduction using t-distributed stochastic neighbour embedding (t-SNE) (**b**) and principal component analysis (PCA) (**c**), where every dot ($n=193$) represents one cell, with the colour representing either cell type or the assigned SC3 cluster represented in **a**, respectively. **d**, SC3 clustering of EMPs ($n=68$) using clustering parameter $k=2$. Heatmap of top 15 marker genes for each cluster and their corresponding normalized expression (AUC > 0.9 and Wilcoxon signed rank test FDR adjusted P value < 0.01). Columns represent single cells, colour-coded by their assigned cell-cycle phase. **e**, Scatter plot, with the x axis representing the adjusted proportion of BC-specific marker genes detected by SC3 ($n=53$) and the y axis LC-specific marker genes ($n=47$). Marker genes were selected to be expressed in at least 75% of the respective cell type and in less than 50% of the opposite cell type. The proportion of expressed markers is computed as the fraction of markers with >0 expression over the total number of markers. Every dot ($n=193$) represents one cell and is colour-coded according to cell type.

reporter mice[15] to specifically isolate EMPs (CD49f^{Hi}/Lgr5-GFP^{Hi}) from the underlying mesenchyme and surrounding epidermis[16] and compared their transcriptome to adult BCs (CD24^{+}CD29^{Hi}) and LCs (CD24+ CD29^{Lo})[17] (Fig. 2a–d and Supplementary Fig. 1). To determine to what extent EMPs resemble adult BCs, we compared the transcriptome of EMPs with LCs and defined the genes upregulated 1.5-fold in EMPs (embryonic basal signature) and assessed which genes upregulated in this signature were also upregulated in adult BCs (when compared to LCs) by more than 1.5-fold (adult basal signature). A high proportion of the genes of the adult basal signature were also upregulated in EMPs when compared to LCs (22.3%) (Fig. 2e). Lgr5-isolated EMPs expressed a number of genes previously shown to be enriched in laser capture embryonic MG[18]. Gene set enrichment analysis (GSEA) confirmed the high enrichment of adult basal genes in EMPs (Fig. 2f). Interestingly, when EMPs were compared with adult BCs (embryonic luminal signature) and assessed for the expression of markers of the adult luminal signature (genes upregulated 1.5-fold in adult LCs when compared to BCs), EMPs also expressed luminal-specific genes (Fig. 2g). GSEA confirmed the enrichment of adult luminal genes in EMPs (Fig. 2h). These data demonstrate that, at the population level at E14, EMPs express genes of basal and luminal lineages.

Gene ontology (GO) analysis of the genes belonging to the basal signature and expressed in EMPs revealed a high enrichment for developmental proteins Wnt, Edar Pth, TGFβ, and Notch pathways (Fig. 2i,j and Supplementary Fig. 2a), which are well known to regulate MG morphogenesis and adult maintenance[3,5,19–27]. Extracellular matrix (ECM) genes were also strongly enriched (Fig. 2i,k and Supplementary Fig. 2a), suggesting that these progenitors contribute to the formation of their own niche. In addition,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

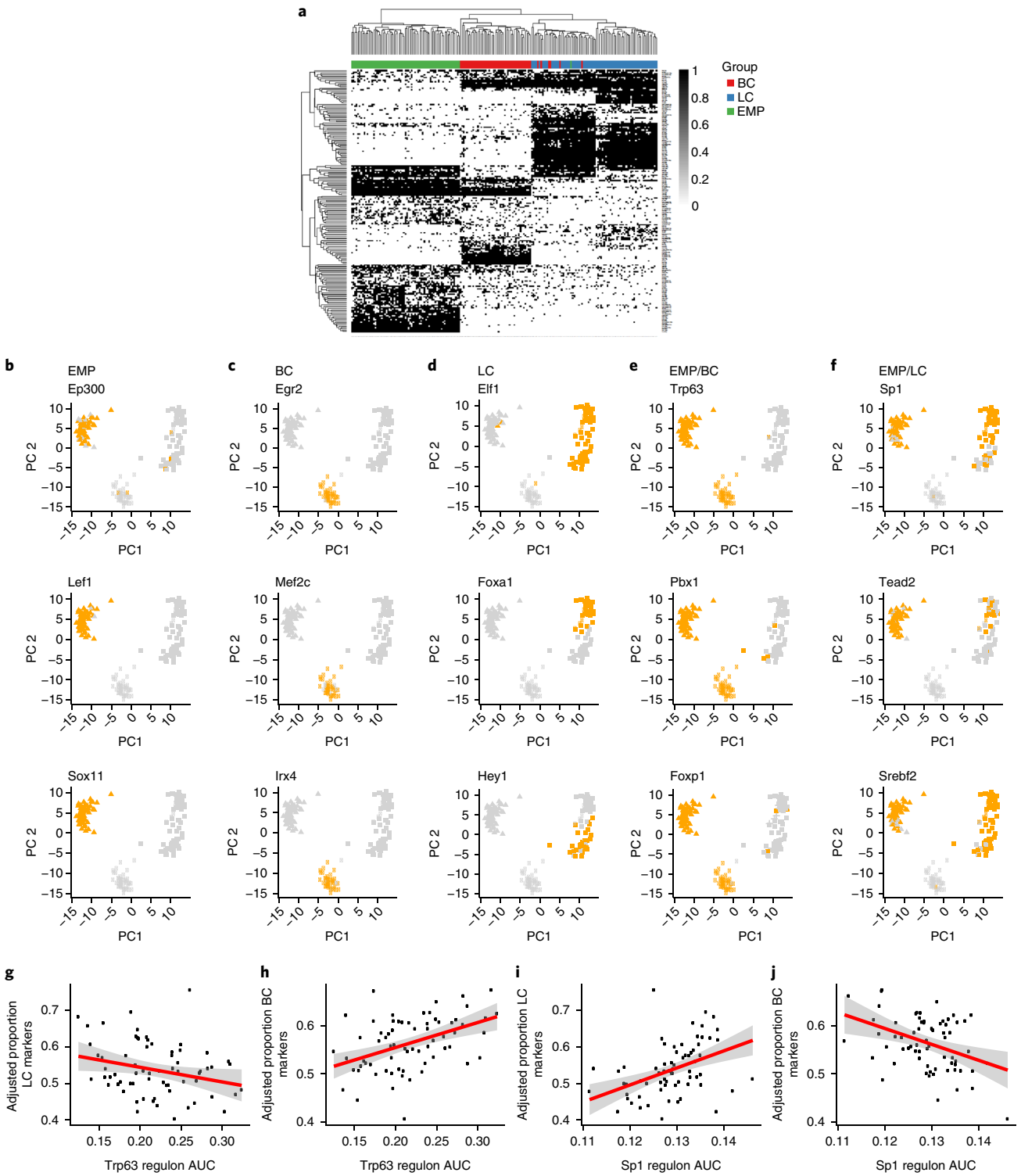

Fig. 4 | SCENIC analysis of EMP, LC and BC scRNA-seq data. a, Binary activity matrix for regulons inferred by SCENIC. Regulons were determined to be active (black) if they exceeded a manually adjusted AUC regulon-specific threshold, or inactive under this threshold (white). Columns represent cells (n=193) colour-coded by cell type; rows represent regulons. Hierarchical clustering was performed and clusters of regulons can be observed specific for each cell population but also shared between the different cell populations. Only regulons with an absolute correlation with any other regulon > 0.3 and at least active in 1% of cells are shown. **b–f**, PCA plots showing the binary activity of regulons inferred by SCENIC. PCA was performed on scRNA-seq data on the normalized expression values of the top 500 most variable genes. Every dot (n=193) represents a single cell whereas the activity of the respective regulon is colour-coded as active (orange) or inactive (light grey) for each cell. Examples of regulons are grouped by their respective populations: only active in EMPs (**b**), only active in BCs (**c**), only active in LCs (**d**), active in both EMPs and BCs but not active in LCs (**e**), active in both EMPs and LCs but not in BCs (**f**). **g–j**, Scatter plots depicting the linear relationship in the EMP population (n=68) between regulon activity and the adjusted proportion of specific LC (**g**,**i**) and BC (**h**,**j**) markers for Trp63 (**g**,**h**) and Sp1 (**i**,**j**) regulons. Regulon activity was measured as the regulon AUC, which is a function of the number of inferred target genes of that regulon being expressed (0 means no genes and 1 means all genes being expressed). The red line represents a linear model fitted using the lm function in R, and the grey area represents the 95% confidence interval.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

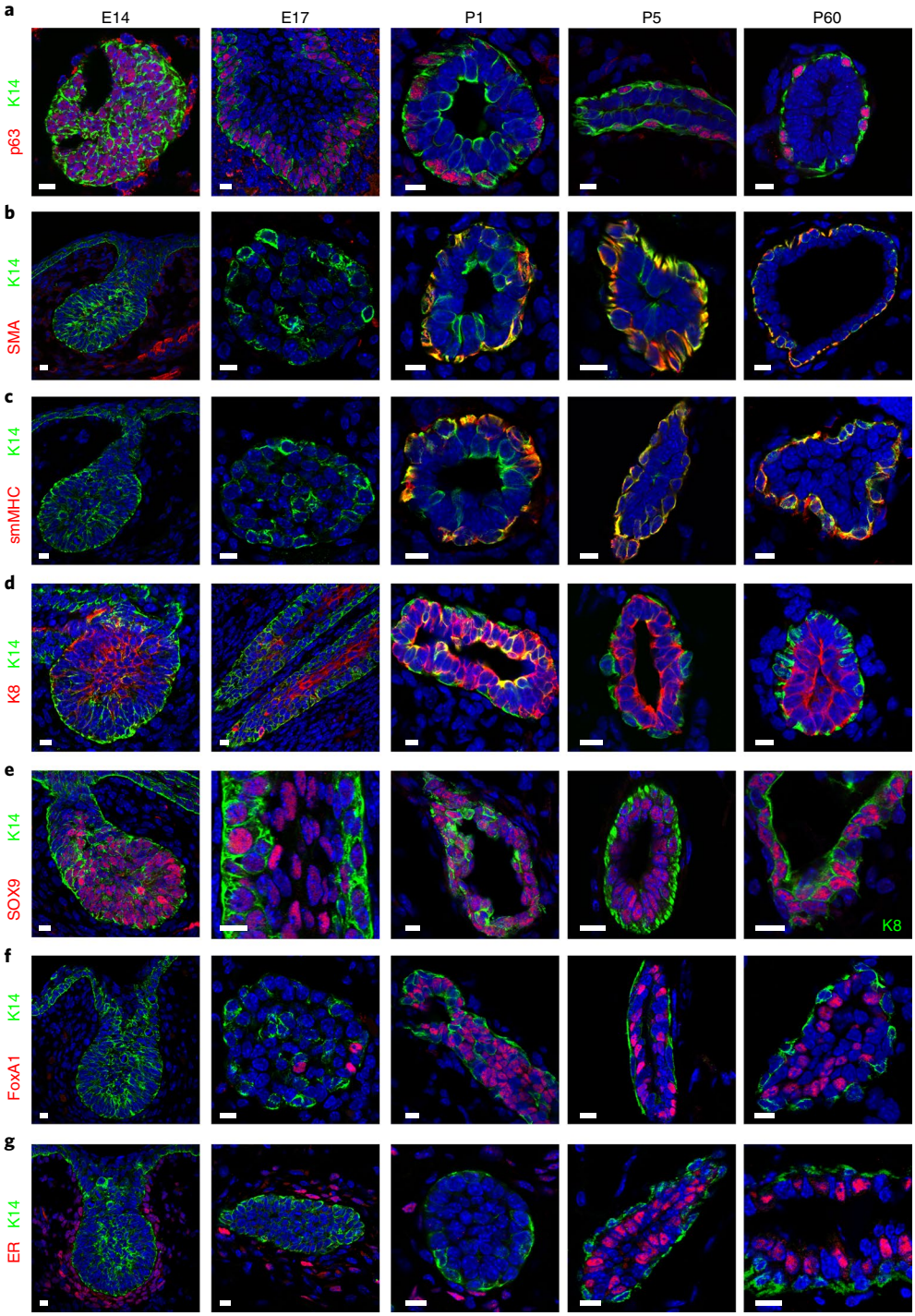

Fig. 5 | Asymmetrical expression of basal and luminal markers marks the early step of MG lineage segregation. **a**–**g**, Confocal imaging of immunostaining of K14 and p63 (**a**), SMA (**b**), smMHC (**c**), K8 (**d**), Sox9 (**e**), FoxA1 (**f**) and ER (**g**) in E14, E17, P1, P5 and P60 MG shows the temporality of cellular heterogeneity during MG development (except in **e**, where the MG at P60 is represented with K8 staining). Representative images from three independent mice were analysed per time point. Scale bars, 10 µm.

EMPs were enriched in developmental proteins, including *Trp63*, a key transcription factor (TF) known to be essential for MG embryonic development and expressed in adult BCs[28–30], and epithelial to mesenchymal transition (EMT) genes (Fig. 2l), a feature associated with embryonic MG[31] and adult BCs[32]. EMPs also expressed cell adhesion and cytoskeleton molecules including basal keratins, members of the planar cell polarity pathway, growth factor binding and axon guidance molecules (Fig. 2i,k and Supplementary Fig. 2a,b), which may regulate the growth and branching of MG. Oncogenic Pik3ca expression in adult BCs or LCs promotes the activation of a multipotent program[17]. Interestingly, the EMP basal signature encompassed a high number of genes (31.2%, 124 out of 398 genes) previously associated with the oncogenic Pik3ca-induced multipotency signature (Fig. 2m–o)[17].

GO analysis of the genes of the adult luminal signature expressed in the EMPs revealed that the EMPs were highly enriched for genes regulating cell cycle and mitosis (Supplementary Fig. 2c–e), and for well-known key regulators promoting MG luminal lineage

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

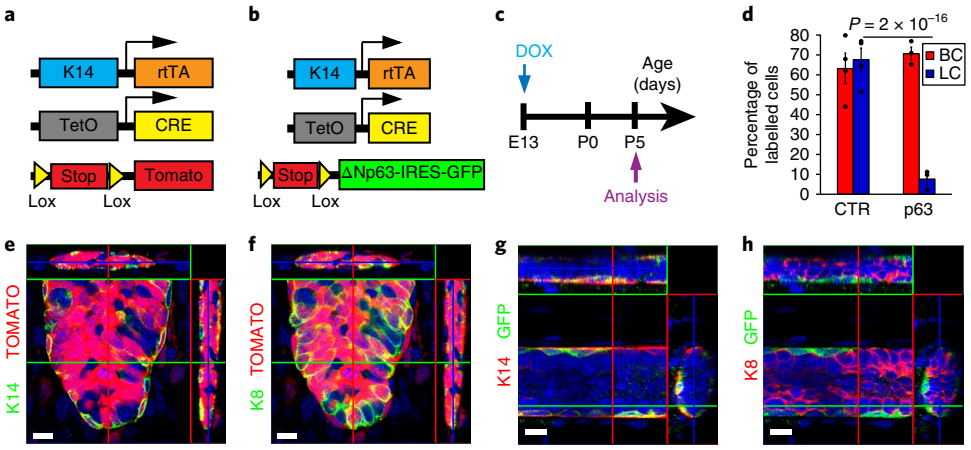

Fig. 6 | p63 promotes unipotent BC fate in EMPs. a,**b**, Scheme summarizing the genetic strategy used to target tdTomato (**a**) or ΔNp63-IRES-GFP (**b**) expression in K14-expressing cells at E13. **c**, Scheme summarizing the protocol used to study the fate of cells targeted during embryogenesis using K14rtTA/TetO-Cre/Rosa-tdTomato or Rosa-ΔNp63-IRES-GFP mice. **d**, Graph representing the mean percentage of labelled BCs and LCs in control versus p63-overexpressing mice. Respectively, $n = 4$ and $n = 3$ independent mice were analysed in K14rtTA/TetO-Cre/Rosa-tdTomato and in Rosa-ΔNp63-IRES-GFP mice. Individual data points are represented as dots. Error bars indicate s.e.m. *P* values are derived from Fisher exact test without adjustment. See Supplementary Table 1 for source data for **d**. **e**,**f**, Confocal imaging of immunostaining of K14 (**e**) or K8 (**f**) and Tomato in K14rtTA/TetO-Cre/Rosa-tdTomato mice induced at E13 with 15 µg DOX per g. Representative images from four mice were analysed. **g**,**h**, Confocal imaging of immunostaining of K14 (**g**) or K8 (**h**) and GFP in K14rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP mice induced at E13 with 15 µg DOX per g. Representative images from three mice are analysed. Images in **e**–**h** represent orthogonal projections of 3D stacks. Scale bars, 10 µm.

specification and maintenance[18,33–38], together with the known ER+ and ER− luminal markers (Fig. 2n) that characterize luminal lineages[9,39]. A significant proportion of the EMP luminal signature was common with the oncogenic Pik3ca-induced multipotent gene signature (14.3%, 20 out of 140 genes) (Fig. 2o).

Single cell RNA-seq uncovers the hybrid EMP signature. To further define the molecular features associated with the multipotency of EMPs at the early stages of MG morphogenesis, we performed single-cell RNA-seq (scRNA-seq) following a SMARTseq2-based approach on FACS-isolated cells. After very stringent quality control, 69 single EMPs at E14, 51 adult BCs and 73 adult LCs were retained for downstream analyses. Unsupervised clustering analysis using the SC3 method[40] showed that these cells could be individualized into four main clusters specific for the EMPs, adult BCs and both *Esr1*+ and *Esr1*− LCs (Fig. 3a and Supplementary Fig. 3). EMPs were specifically enriched for genes that regulate proliferation and signalling and were previously reported to be expressed in breast cancers[41–47], whereas BCs and LCs expressed classical basal and luminal genes, respectively. GO analysis of the single-cell signature of EMPs, similar to the microarray analysis, showed high enrichment for developmental proteins, differentiation, transcription, axon guidance, Wnt signalling and ECM genes (Supplementary Fig. 3). EMPs at E14 did not express *Sox10*, contrasting with its expression in late embryonic development (E18.5) and adult MG[31,38] (Supplementary Fig. 3d). The four main clusters found by SC3 were also found by dimensionality reduction using t-distributed stochastic neighbour embedding (t-SNE) and principal component analysis (PCA) (Fig. 3b,c), demonstrating the robustness of these clusters. The first component in the PCA reflected 21% of total variance and could be attributed to the difference between the embryonic and adult cell types. Consistent with what we found in microarray analysis, BCs more closely resemble EMPs along PC1. The second component of the PCA constitutes 12% of the total variance and represents transcriptional differences between adult LCs and BCs. The third (5% of total variance) and fourth (2% of total variance) components represent variance attributed to the *Esr1*+ and *Esr1*− status of the LCs.

We then assessed whether EMPs constitute a transcriptionally homogenous population expressing a hybrid LC/BC gene signature or are composed of distinct pre-committed subpopulations of cells with distinct gene signatures corresponding to BC or LC progenitors. Only scRNA-seq can discriminate between these two scenarios. We performed SC3 unsupervised clustering on the EMPs only, and found two subclusters (Fig. 3d). One subcluster comprises the majority of EMPs, expressing no significant cluster-specific marker genes. The second subcluster represents a proliferative subpopulation of EMPs expressing marker genes associated with cell cycle. To avoid aberrant clustering, these cycling cells were omitted from further analysis. Very interestingly, only EMPs, and not adult LCs or BCs, expressed a hybrid transcriptional signature comprising markers for both LC and BC lineages (Fig. 3e and Supplementary Fig. 4). Lowering the stringency of the quality controls for the scRNA-seq pipeline, such as the number of expressed genes per cell, may result in the appearance of a few adult LCs and BCs with artefactual hybrid gene expression profiles (Supplementary Fig. 5).

We also found that a few Lgr5-expressing cells at E14 were clustered in a separate mesenchymal cluster distinct from all the other EMPs. Some of these cells did not express any epithelial markers, including K14, K5 or K8, and correspond to stromal cells of the embryonic MG, which were not eliminated by the α6-integrin/CD49f FACS gating strategy that separates embryonic mammary epithelial from mesenchymal Lgr5-expressing cells. However, a few epithelial cells, as demonstrated by their keratin expression, also expressed a higher level of mesenchymal genes, and thus correspond to an EMT hybrid state (Supplementary Fig. 6a).

By looking at the proportion of exclusively expressed lineage-specific markers in EMPs, we observed no clear subpopulations of EMPs expressing a higher proportion of either sets of genes (Fig. 3e and Supplementary Fig. 6), suggesting that EMPs were not committed to the basal or luminal lineage at E14.

To further assess the regulatory mechanisms associated with the different cell states, we performed gene regulatory network analysis using SCENIC[48], which allows the identification of regulatory modules by inferring co-expression between TFs and genes containing the respective TF binding motif in their regulatory regions.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

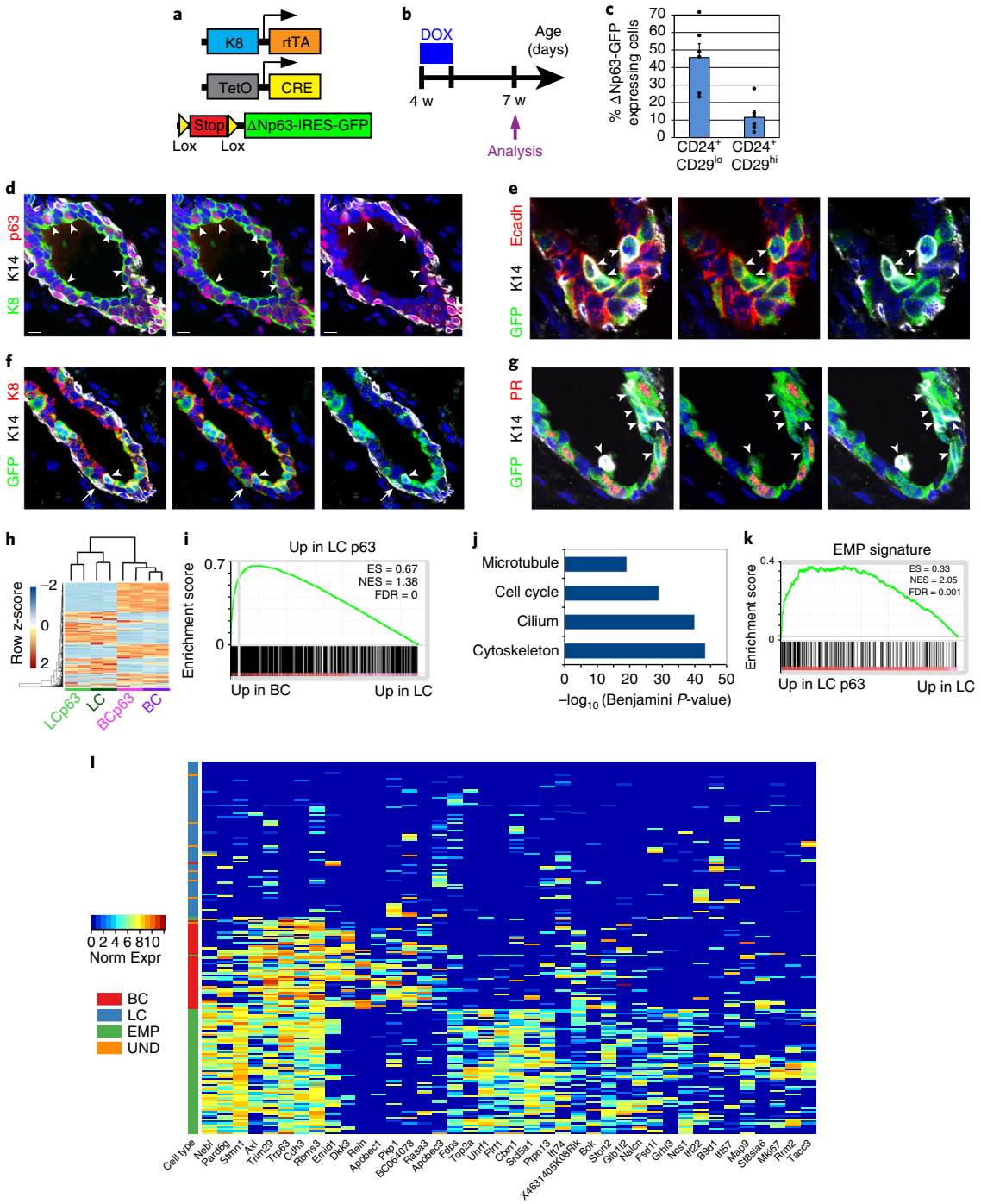

Fig. 7 | In vivo reprogramming of adult LCs into BCs by p63. a, Genetic strategy to target ΔNp63-IRES-GFP expression in LCs. **b**, Protocol used to study the fate of cells targeted using K8rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP mice. **c**, Percentage of GFP-labelled LCs (CD24^{+} CD29^{lo}) and newly formed BCs (CD24^{+} CD29^{Hi}) following expression of ΔNp63-IRES-GFP in adult LCs ($n=6$ mice) (mean+s.e.m.). Dots indicate individual data points. See Supplementary Table 1 for source data. **d**–**g**, Immunofluorescence of p63, K14 and K8 (**d**), GFP, K14 and Ecadh (**e**), GFP, K14 and K8 (**f**) and GFP, K14 and PR (**g**) two weeks following p63-IRES-GFP expression in LCs. Representative images from six mice. Arrows indicate BCs arising from LCs, and arrowheads indicate hybrid LCs expressing basal markers. Scale bars, 10 μm. **h**, Heatmap representing the similarities between the different BC (CD24^{+} CD29^{Hi}) and LC populations (CD24^{+} CD29^{lo}) populations ($n=2$ RNA-seq data for each population). The top 500 most variable genes across the eight samples are plotted in the heatmap. The dendrogram shows hierarchical gene expression clustering of BCs and LCs with or without p63 overexpression. Blue and red correspond to low and high expressed genes, respectively. The two major branches of the tree perfectly discriminate between LCs and BCs and between wild-type (WT) and ΔNp63 cells. **i**, GSEA of the genes upregulated by p63 in LCs (versus WT LCs) with upregulated genes in BCs (versus LCs), showing enrichment of basal genes in p63 upregulated genes in LCs. **j**, Bar chart of Benjamini-corrected enrichment *P* values of the first four functional annotation clusters for the 902 genes overexpressed in ΔNp63 expressing LC compared to WT LCs. **k**, GSEA of the upregulated genes in EMPs (versus LCs) with the genes upregulated by p63 in LCs (versus WT LCs), showing enrichment of the EMP signature in p63 upregulated genes in LCs. **l**, Heatmap representing genes overexpressed in ΔNp63 expressing LC bulk RNA-seq data. The top 200 overexpressed genes are chosen and the 40 genes with the highest variance among the single cells are plotted in the heatmap. Columns represent single cells colour-coded by cell type. Colours in the heatmap represent normalized expression values. RNA-seq analyses in **i**–**k** are derived from the means of two RNA-seq data sets per condition.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

These modules or regulons were then assessed in each cell to ascertain their activity and infer cell-specific states. By hierarchical clustering of the cells according to the binary activity of regulons, we recapitulated the clusters previously found by SC3. We identified regulons active only in EMPs, only in LCs, only in BCs, in both EMPs and LCs, and in EMPs and BCs (Fig. 4a–f). We then investigated the correlation between the degree of regulon activity by the number of their expressed target genes, and basal or luminal marker expression. Interestingly, in the EMP population we identified TFs such as *p63* for BC or *Sp1* for LC, which showed opposite correlations to LC and BC marker expression, suggesting their role as cell lineage regulators (Fig. 4g–j).

Asymmetrical expression of basal and luminal markers marks the early step of lineage segregation. To gain further insight into the mechanisms that regulate cellular heterogeneity and lineage segregation during MG morphogenesis, we assessed by immunofluorescence the temporal expression of basal and luminal markers differentially regulated at the bulk mRNA level in EMPs, adult BCs and LCs. At E14, basal (K14 and p63) and luminal (K8 and Sox9) markers were expressed in all cells of the MG, although lower levels of K8 were already visible along the outer region of the embryonic MG (Fig. 5a–e). At E17, the cells of the outer layer of the embryonic MG expressed p63 and lower levels of K8, whereas the inner cells of the embryonic MG expressed higher levels of K8 and no p63 (Fig. 5a,d). Sox9, which is expressed by a fraction of adult LCs[9,10,49] was still expressed by BCs and LCs at this stage of embryonic development (Fig. 5e). Foxa1 and ER were only observed at the protein level in a fraction of LCs during late embryogenesis for Foxa1 and in the postnatal MG for ER (Fig. 5f,g). Basal markers SMA and smMHC (Fig. 5b,c), which were detected at E14 at the mRNA level, were only detected at the protein level at P1 in outer BCs.

p63 promotes unipotent BC fate in EMPs. To further dissect the molecular mechanisms associated with MG lineage segregation, we assessed whether p63 controls the switch from multipotency to BC unipotency during MG development. p63 deletion during embryonic development leads to the absence of MG formation[28,29], demonstrating the essential early function of p63 during MG specification but preventing study of the role of p63 in MG lineage segregation. Overexpression or ShRNA knockdown of ΔNp63 in primary culture of mouse mammary epithelial cells (MMECs) in vitro showed that ΔNp63 decreases the expression of LC markers and increases BC marker expression, suggesting that ΔNp63, directly or indirectly, regulates BC characteristics. Furthermore, transplantation of MMECs overexpressing ΔNp63 reduced the proportion of LCs in the MG outgrowth, suggesting that downregulation of ΔNp63 is required for proper luminal lineage differentiation[50]. However, it remains unclear whether the spatial restriction of ΔNp63 expression during MG embryonic development is important to promote BC fate.

To address this question, we used a genetic approach allowing p63 expression to be sustained in the outer and inner cells of the embryonic MG[51]. DOX administration to K14rtTA/TetO-Cre/Rosa-tdTomato at E13 led to labelling of the same proportion of BCs (Tomato+/K14+) and LCs (Tomato+/K8+) (70%) at P5. In contrast, DOX administration to K14rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP mice led to the labelling of 70% of BCs (GFP+/K14+) and only 10% of LCs (Fig. 6a–h), demonstrating that most LCs arise from the rare EMPs that did not express the p63 transgene.

Reprogramming of adult LCs into BCs by ΔNp63. We next examined whether ΔNp63 could reprogram adult LCs into the BC lineage, and if so by which molecular mechanisms. To express ΔNp63 in adult LCs, we administered DOX for seven consecutive days to four-week-old K8rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP females (Fig. 7a,b), and analysed the presence of GFP in BCs and LCs by FACS two weeks after the last DOX administration. Interestingly, we found that ΔNp63 expression in adult LCs was sufficient to reprogram a fraction of ΔNp63 expressing LCs into BCs (Fig. 7c), demonstrating the important plasticity of LCs and the master regulatory function of ΔNp63 to promote BC fate in vivo.

To gain further insights into the cellular mechanisms that accompanied LC fate transition upon ectopic p63 expression, we performed immunofluorescence characterization of the MG during the cellular reprogramming. In good accordance with the FACS data, p63 immunostaining showed that p63 was expressed in about half of the LCs (Fig. 7d). Interestingly, we found that some LCs expressing p63-IRES-GFP co-expressed basal and luminal markers (Fig. 7e–g and Supplementary Fig 7). Most of these hybrid cells were still located in the inner cells together with the other LCs, and still presented the round shape of LCs (Fig. 7e–g). More rarely at this stage, some of these p63-targeted LCs were located in the outer part of the mammary epithelium, lost the expression of LC markers, expressed BC markers and presented BC elongated morphology (Fig. 7f). Although p63-IRES-GFP was initially evenly expressed in ER/PR+ and − LCs, the LCs that co-expressed basal and luminal markers did not express Foxa1 or PR (Fig. 7g and Supplementary Fig. 7d), suggesting that LCs need to shut down their hormone receptor differentiation program to undergo cell fate reprogramming upon ΔNp63 expression.

To understand the molecular mechanisms by which ΔNp63 reprograms adult LCs into BCs, we performed bulk RNA-seq of FACS-isolated LCs and BCs upon ΔNp63 expression in adult MGs. Gene clustering demonstrated that the newly generated basal-like cells upon ΔNp63 expression in LCs molecularly resemble WT BCs (Fig. 7h). Venn diagrams and GSEA also showed that a very significant fraction of the LC upregulated genes upon ΔNp63 expression belong to the BC signature and, among the 48 p63 target genes inferred by SCENIC, 18 were overexpressed in LCs expressing ΔNp63 ($P=1.349e-06$ given a random sampling of genes; Fig. 7i and Supplementary Fig. 7), supporting the notion that ΔNp63 expression in LCs induced a transient hybrid multipotent state before the de novo generation of BCs. GO analysis showed that the genes upregulated by ΔNp63 in LCs are highly enriched for cytoskeleton, cilium biogenesis and cell division genes (Fig. 7j). Interestingly, the scRNA-seq EMP signature was also enriched in adult LCs expressing ΔNp63 (Fig. 7k). A subset of the most upregulated genes in adult LCs expressing ΔNp63 versus WT LCs were expressed in the single-cell EMP data but not in BCs, further suggesting that the reprogramming of adult LCs following ΔNp63 expression leads to a multipotent embryonic-like hybrid state before giving rise to fully reprogrammed BCs (Fig. 7l).

Discussion

Here, using clonal analysis during MG development, we formally demonstrate that MG initially develops from EMPs that rapidly switch from multipotency to unipotency during the course of embryonic development. Similar conclusions were made from lineage tracing experiments using Notch1-CreER at different stages of embryogenesis in a related study in ref. [52]. This rapid switch from multipotency to unipotency may explain the very rare and large bipotent clones dispersed from the nipple region to the distal part of the epithelial tree found in lineage tracing induced by random recombination[12].

Microarray and scRNA-seq analyses indicate that EMPs are associated with a hybrid signature that overlaps with both basal and luminal lineages. The greater resemblance of EMPs to BCs may explain why BCs are multipotent in transplantation assays[4,7,8]. The hybrid gene expression of EMPs at E14 sharing similarities with both BCs and LCs is consistent with their multipotent fate at this stage of embryonic development.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Reactivation of multipotency is associated with the early stages of mammary tumour initiation and the development of tumour heterogeneity. Our data showing that EMPs express the same genes as expressed during the reactivation of multipotency and cell fate changes induced by oncogenic Pik3ca expression[17] support the notion that the mechanisms associated with multipotency during tumorigenesis recapitulate at least partially the genetic program that regulates multipotency during embryonic MG development. Moreover, many genes found to be specifically expressed in EMPs by scRNA-seq such as *Sox11*, *Stmn1* and *Mdk* are expressed in human breast cancers with poor prognosis[41–46], further suggesting that the reactivation of an EMP gene expression program during tumorigenesis is essential for tumour growth and invasion.

Temporal in situ characterization reveals that the early signs of lineage segregation visible at the protein level consist of a restricted expression of p63 together with a decreased expression of K8 in the outermost layer of the embryonic MG in contact with the stroma. By sustaining the expression of p63 in EMPs, we found that p63 promotes the differentiation of EMPs into BCs during MG development. Similarly, p63 overexpression in adult MMECs decreases the formation of LCs upon transplantation of these cells into the mammary mesenchyme[50]. Expression of p63 in adult LCs is sufficient to reprogram these cells into BCs, further demonstrating that p63 acts as a master regulator of BC fate. It has been proposed that p63 and Notch signalling act in an opposite manner to promote, respectively, BC and LC fates[50,52]. It is thus possible that the downregulation of K8 and the restricted expression of p63 at the outer layer are the consequences of an inhibition of Notch signalling in these cells.

Ectopic expression of ΔNp63 in adult LCs induces the reprogramming of these cells into basal-like cells. This cell fate change is a progressive process rather than a direct trans-differentiation event that passes through a hybrid state, reminiscent of the EMP. Our molecular analysis identifies a specific gene program regulated by p63 sufficient to convert an adult LC into a BC and that encompasses many genes expressed by EMPs, further suggesting that the cell fate reprogramming of an LC into a BC by p63 reactivates some features of the embryonic multipotent genetic program.

Although our study uncovers the importance of hybrid gene expression in regulating the multipotent state during embryonic development and p63-mediated reprogramming of adult LCs into BCs, and defines the BC molecular program controlled by p63 under these conditions, many important questions remain unanswered. Further studies will be needed to better understand the molecular mechanisms that regulate the switch from multipotency to unipotency during MG development. What are the intrinsic and extrinsic signals that act upstream and sustain the expression of p63 and inhibit the expression of luminal genes in the inner part of the MG at the early stage of lineage segregation? What are the key p63 target genes that regulate BC fate in EMP and the reactivation of multipotency in adult unipotent luminal SCs? What is the role of the genes other than p63 that are specifically expressed by EMPs and regulate multipotency during development? As p63 is also expressed in the basal lineage of other epithelial tissues[53,54], the key role of p63 in regulating the switch from multipotency to unipotency is likely to be conserved across different organs. Defect in lineage segregation and regulation of multipotency by p63 may potentially explain some pathological features of the different developmental syndromes affecting these tissues, including the MG, that are associated with p63 mutations in humans[55].

In conclusion, our study identifies the temporal switch from multipotency to unipotency that occurs during MG development, as well as the molecular mechanisms that regulate cell fate transition and lineage segregation during this process. The paradigm and molecular mechanisms uncovered here have important implications for understanding multipotency, unipotency and lineage segregation in other tissues and during tumorigenesis.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41556-018-0095-2.

Received: 1 November 2017; Accepted: 26 March 2018;
Published online: 21 May 2018

References

1. Watson, C. J. & Khaled, W. T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, 995–1003 (2008).
2. 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).
3. Lafkas, D. et al. Notch3 marks clonogenic mammary luminal progenitor cells in vivo. *J. Cell Biol.* **203**, 47–56 (2013).
4. Prater, M. D. et al. Mammary stem cells have myoepithelial cell properties. *Nat. Cell Biol.* **16**, 942–950 (2014).
5. Rodilla, V. et al. Luminal progenitors restrict their lineage potential during mammary gland development. *PLoS Biol.* **13**, e1002069 (2015).
6. Tao, L., van Bragt, M. P., Laudadio, E. & Li, Z. Lineage tracing of mammary epithelial cells using cell-type-specific cre-expressing adenoviruses. *Stem Cell Rep.* **2**, 770–779 (2014).
7. van Amerongen, R., Bowman, A. N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* **11**, 387–400 (2012).
8. Van Keymeulen, A. et al. Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189–193 (2011).
9. Wang, C., Christin, J. R., Oktay, M. H. & Guo, W. Lineage-biased stem cells maintain estrogen-receptor-positive and -negative mouse mammary luminal lineages. *Cell Rep.* **18**, 2825–2835 (2017).
10. Wuidart, A. et al. Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev.* **30**, 1261–1277 (2016).
11. Scheele, C. L. et al. Identity and dynamics of mammary stem cells during branching morphogenesis. *Nature* **542**, 313–317 (2017).
12. Davis, F. M. et al. Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. *Nat. Commun.* **7**, 13053 (2016).
13. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
14. Lescroart, F. et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat. Cell Biol.* **16**, 829–840 (2014).
15. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
16. Trejo, C. L., Luna, G., Dravis, C., Spike, B. T. & Wahl, G. M. Lgr5 is a marker for fetal mammary stem cells, but is not essential for stem cell activity or tumorigenesis. *NPJ Breast Cancer* **3**, 16 (2017).
17. Van Keymeulen, A. et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119–123 (2015).
18. Wansbury, O. et al. Transcriptome analysis of embryonic mammary cells reveals insights into mammary lineage establishment. *Breast Cancer Res.* **13**, R79 (2011).
19. Biggs, L. C. & Mikkola, M. L. Early inductive events in ectodermal appendage morphogenesis. *Semin. Cell Dev. Biol.* **25–26**, 11–21 (2014).
20. Boras-Granic, K., Chang, H., Grosschedl, R. & Hamel, P. A. Lef1 is required for the transition of Wnt signaling from mesenchymal to epithelial cells in the mouse embryonic mammary gland. *Dev. Biol.* **295**, 219–231 (2006).
21. Chu, E. Y. et al. Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. *Development* **131**, 4819–4829 (2004).
22. Hiremath, M. & Wysolmerski, J. Parathyroid hormone-related protein specifies the mammary mesenchyme and regulates embryonic mammary development. *J. Mammary Gland Biol. Neoplasia* **18**, 171–177 (2013).
23. Howard, B. A. & Lu, P. Stromal regulation of embryonic and postnatal mammary epithelial development and differentiation. *Semin. Cell Dev. Biol.* **25–26**, 43–51 (2014).
24. Wysolmerski, J. J., McCaughern-Carucci, J. F., Daifotis, A. G., Broadus, A. E. & Philbrick, W. M. Overexpression of parathyroid hormone-related protein or parathyroid hormone in transgenic mice impairs branching morphogenesis during mammary gland development. *Development* **121**, 3539–3547 (1995).
25. Raafat, A. et al. Expression of Notch receptors, ligands, and target genes during development of the mouse mammary gland. *J. Cell Physiol.* **226**, 1940–1952 (2011).
26. Sale, S., Lafkas, D. & Artavanis-Tsakonas, S. Notch2 genetic fate mapping reveals two previously unrecognized mammary epithelial lineages. *Nat. Cell Biol.* **15**, 451–460 (2013).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

27. Robinson, G. W. Cooperation of signalling pathways in embryonic mammary gland development. *Nat. Rev. Genet* **8**, 963–972 (2007).
28. Mills, A. A. et al. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).
29. Yang, A. et al. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).
30. Forster, N. Basal cell signaling by p63 controls luminal progenitor function and lactation via NRG1. *Dev. Cell* **28,** 147-160 (2014).
31. Dravis, C. et al. Sox10 regulates stem/progenitor and mesenchymal cell states in mammary epithelial cells. *Cell Rep.* **12**, 2035–2048 (2015).
32. Ye, X. et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
33. Choi, Y. S., Chakrabarti, R., Escamilla-Hernandez, R. & Sinha, S. Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: failure of Stat5 activation and functional differentiation in the absence of Elf5. *Dev. Biol.* **329**, 227–241 (2009).
34. Oakes, S. R., Hilton, H. N. & Ormandy, C. J. The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. *Breast Cancer Res* **8**, 207 (2006).
35. Bernardo, G. M. et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development* **137**, 2045–2054 (2010).
36. Balko, J. M. et al. The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium. *Proc. Natl Acad. Sci. USA* **109**, 221–226 (2012).
37. Kendrick, H. et al. Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* **9**, 591 (2008).
38. Spike, B. T. et al. A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* **10**, 183–197 (2012).
39. Van Keymeulen, A. et al. Lineage-restricted mammary stem cells sustain the development, homeostasis, and regeneration of the estrogen receptor positive lineage. *Cell Rep.* **20**, 1525–1532 (2017).
40. Kiselev, V. Y. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).
41. Ibusuki, M. et al. Midkine in plasma as a novel breast cancer marker. *Cancer Sci.* **100**, 1735–1739 (2009).
42. Kuang, X. Y. et al. Stathmin and phospho-stathmin protein signature is associated with survival outcomes of breast cancer patients. *Oncotarget* **6**, 22227–22238 (2015).
43. Prochazkova, I. et al. Targeted proteomics driven verification of biomarker candidates associated with breast cancer aggressiveness. *Biochim. Biophys. Acta* **1865**, 488–498 (2017).
44. Saal, L. H. et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc. Natl Acad. Sci. USA* **104**, 7564–7569 (2007).
45. Shepherd, J. H. et al. The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget* **7**, 13106–13121 (2016).
46. Zvelebil, M. et al. Embryonic mammary signature subsets are activated in Brca1$^{-/-}$ and basal-like breast cancers. *Breast Cancer Res.* **15**, R25 (2013).
47. Fernandez-Garcia, B. et al. Expression and prognostic significance of fibronectin and matrix metalloproteases in breast cancer metastasis. *Histopathology* **64**, 512–522 (2014).
48. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
49. Malhotra, G. K. et al. The role of Sox9 in mouse mammary gland development and maintenance of mammary stem and luminal progenitor cells. *BMC Dev. Biol.* **14**, 47 (2014).
50. Yalcin-Ozuysal, O. et al. Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. *Cell Death Differ.* **17**, 1600–1612 (2010).
51. Latil, M. et al. Cell-type-specific chromatin states differentially prime squamous cell carcinoma tumor-initiating cells for epithelial to mesenchymal transition. *Cell Stem Cell* **20**, 191–204 (2017).
52. Lilja, A. M. et al. Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nat. Cell Biol.* https://doi.org/10.1038/s41556-018-0108-1 (2018).
53. Signoretti, S. et al. p63 is a prostate basal cell marker and is required for prostate development. *Am. J. Pathol.* **157**, 1769–1775 (2000).
54. Kurita, T., Medina, R. T., Mills, A. A. & Cunha, G. R. Role of p63 and basal cells in the prostate. *Development* **131**, 4955–4964 (2004).
55. van Bokhoven, H. & McKeon, F. Mutations in the p53 homolog p63: allele-specific developmental syndromes in humans. *Trends Mol. Med* **8**, 133–139 (2002).

Acknowledgements

The authors acknowledge the animal house facility from ULB (Erasme campus). Sequencing was performed at the Brussels Interuniversity Genomics High Throughput core (www.brightcore.be) and the Genomics Core Leuven. The authors thank N. Dedoncker for help with single-cell RNA-seq library construction. C.B. is an investigator with WELBIO, A.W. is supported by a FNRS fellowship. M.F. is supported by a Télévie fellowship. A.V.K. is Maître de Recherches of the FNRS. A.S., D.B. and T.V. are supported by KU Leuven (SymBioSys, PFV/10/016), Stichting Tegen Kanker (2015-143) and FWO (postdoctoral fellow number 12W7318N, [PEGASUS]^2 Marie Skłodowska-Curie fellow number 12O5617N). The authors thank colleagues who provided reagents mentioned in the text, and J.-M. Vanderwinden for help with confocal imaging. This work was supported by the FNRS, a research grant from the Fondation Contre le Cancer, the ULB fondation, the Fond Gaston Ithier, the Télévie, the foundation Bettencourt Schueller, the foundation Baillet Latour, and the European Research Council (EXPAND).

Author contributions

A.W. and C.B. designed the experiments and performed data analysis. A.W., S.M., M.F., A.C. and A.V.K performed the biological experiments. A.B. performed GSEA analysis. A.S. and T.V. performed single-cell RNA-seq data analysis and provided the related figures and methods. D.B. and T.V. performed single-cell RNA-seq processing and sequencing, and provided the related methods, A.D. provided technical support. C.D. provided technical support for cell sorting. A.W., A.S., C.B. and A.V.K. prepared the figures. C.B. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0095-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Methods

Mice. Lgr5-EGFP-IRES-CreER[15] and Rosa-tdTomato[56] mice were obtained from the Jackson Laboratory. Rosa-Confetti[13] mice were provided by H. Clevers. K14rtTA transgenic mice[57] were provided by E. Fuchs. TetO-Cre mice[58] were provided by A. Nagy. Rosa26-ΔNp63-IRES-GFP mice[51] were provided by W. Declercq. The generation of K5CreER and K8rtTA has been described previously[8,10]. Mice colonies were maintained in a certified animal facility in accordance with European guidelines. The experiments were approved by the local ethical committee (CEBEA) under protocols #477 and #527. The study is compliant with all relevant ethical regulations regarding animal research. Mice were analysed at embryonic stages E14, E15 and E17, during postnatal development at P1, P5, P21, P60 and 7w and in adult mice (over 8w), as indicated in the figure legends.

Targeting Confetti, tdTomato or ΔNp63-IRES-GFP expression in the MG. For clonal lineage tracing, K14rtTA/TetO-Cre/Rosa-Confetti embryos were induced at E13 by i.v. injection in the tail vein of pregnant mothers with 1 µg doxycycline per g (diluted in sterile PBS, Sigma cat. no. D9891) and at E15 or P5. K5CreER/Rosa-Confetti pups at P1 were induced by intraperitoneal (i.p.) injection of 50 µg tamoxifen (diluted in sunflower seed oil, Sigma cat. no. T5648) and killed 21 days later. For K14rtTA/TetO-Cre/Rosa-tdTomato or K14rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP, embryos were induced at E13 by i.v. injection of pregnant mothers with 15 µg doxycycline per g (diluted in sterile PBS) and killed at P5. K8rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP female mice were induced at 4w with doxycyclin 10 g kg^{-1} in diet (Bio-Serv), 2 g l^{-1} in drinking water (AG Scientific cat. no. D2545) and three intraperitoneal injections of 2 mg in 200 µl PBS over 7 days.

MG whole-mount (WM) processing and immunostaining. For MG processing at E15, whole embryos were collected and fixed in PFA 4% for 2 h at room temperature (RT) or overnight at 4 °C. The following day, the whole skin (of female embryos) containing mammary buds was dissected and stained as detailed in the following. For MG processing at P5, thoracic and inguinal MGs were dissected and fixed in PFA 4% for 2 h at RT or overnight at 4 °C, then stained. For MG processing at P21, inguinal glands were dissected and enzymatically digested in HBSS (GIBCO) + 300 U ml^{-1} collagenase (Sigma cat. no. C0130) + 300 µg ml^{-1} hyaluronidase (Sigma cat. no. 4272) for 20 min at 37 °C under shaking. Glands were fixed in PFA 4% for 2 h at RT. Confetti samples were washed in NH_4Cl (0.5 M in PBS) for 20 min, followed by washes in PBS. For WM immunostaining, all samples were incubated in blocking buffer for 3 h (bovine serum albumin (BSA) 1%, horse serum (HS) 5%, TritonX 0,8% in PBS) at RT. The different primary antibodies were incubated overnight at RT and washed for 1 h at RT with PBS 0.2% Tween20 before incubation with secondary antibodies diluted in blocking buffer at 1:400 for 5 h at RT. The following primary antibodies were used: anti-K14 (rabbit or chicken, 1:1,000, Thermo), anti-K8 (rat, 1:500, Developmental Studies Hybridoma Bank, University of Iowa), anti-GFP (chicken, 1:500, ab13970, Abcam). The following secondary antibodies were used: anti-rabbit, anti-rat, anti-chicken conjugated to AlexaFluor488 (Molecular Probes), Rhodamine Red-X or Cy5 (JacksonImmunoResearch). Nuclei were stained with a Hoechst solution (1:1,000 in PBS 0.2% Tween20) for 30 min and washed for another hour in PBS 0.2% Tween20 before mounting on slides in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

MG immunofluorescence on sections. Whole embryos and newborn pups collected at E13, E14 and E17 and dissected MGs from P5 and adult mice were pre-fixed in PFA 4% for 2 h at RT or directly embedded in optimum cutting temperature (OCT) compound (Sakura) and kept at −80 °C. Pre-fixed tissues were washed in PBS, incubated overnight in 30% sucrose in PBS at 4 °C and embedded in OCT and kept at −80 °C. Sections of 10 µm were cut using a HM560 Microm cryostat (Mikron Instruments). Tissue sections were fixed in PFA 4% for 10 min at RT (for non-pre-fixed sections only) and incubated in blocking buffer (BSA 1%, HS 5%, Triton-X 0.2% in PBS) for 1 h at RT. The different primary antibodies were incubated overnight at 4 °C. Sections were then rinsed in PBS and incubated with the corresponding secondary antibodies diluted at 1:400 in blocking buffer for 1 h at RT. The following primary antibodies were used: anti-GFP (chicken, 1:1,000, ab13970, Abcam), anti-K8 (rat, 1:1,000, Troma-I, Developmental Studies Hybridoma Bank, University of Iowa), anti-K14 (rabbit or chicken, 1:1,000, Thermo), anti-K5 (rabbit, 1:1,000, PRB-160P-0100, Covance), anti-CD49f-PE (rat, 1:100, clone GoH3, eBiosciences), anti-p63 (rabbit, 1:500, clone EPR5701, Abcam), anti-SMA-Cy3 (mouse, 1:500, clone 1A4, Sigma), anti-smMHC (rabbit, 1:100, BT562, Biomedical Technologies), anti-Sox9 (rabbit, 1:5,000, AB5535, Millipore), anti-FoxA1 (rabbit, 1:100, clone EPR10881, Abcam), anti-ER (rabbit, 1:300, sc542, Santa Cruz), anti-PR (rabbit 1/250, MA5-14505, ThermoFisher Scientific) and anti-Ecadherin (rat, 1/500, 14-3249-82, Ebioscience). The following secondary antibodies, diluted 1:400, were used: anti-rabbit (A21206), anti-rat (A21208), anti-chicken (A11039) conjugated to AlexaFluor488 (Molecular Probes), anti-rabbit (711-295-152), anti-rat (712-295-153), anti-chicken (703-295-155) Rhodamine Red-X or anti-rabbit (711-605-152), anti-rat (712-605-153) and anti-chicken (703-605-155) Cy5 (JacksonImmunoResearch). Nuclei were stained with Hoechst solution (1:2,000) and slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Microscope image acquisition. Confocal images were acquired at RT using a Zeiss LSM780 confocal microscope fitted on an Axiovert M200 inverted microscope equipped with a C-Apochromat (×40, NA = 1.2) water immersion objective (Carl Zeiss). Optical sections (1,024 × 1,024 pixels) were collected sequentially for each fluorochrome. The data sets generated were merged and displayed with ZEN 2 software.

Quantification of percent GFP-labelled glands and clone composition. Whole mounts of E15 and P5 K14rtTA/TetO-Cre/Rosa-Confetti MG were analysed by confocal microscopy. For E15 embryos, 73 MGs (n = 8 embryos, 1 litter) were analysed, among which 11 were positive for the presence of nGFP cells (16%). For P5 mice, 85 MGs (n = 13 mice, 4 litters) were analysed, among which 22 were positive for the presence of nGFP cells (26%). At P5, clones were scored according to their keratin expression into basal (K14+), luminal (K8+) and bipotent clones (K14+/K8+). Whole mounts of P21 K5CreER/Rosa-Confetti MG were analysed by confocal microscopy. At P21, 36 unipotent basal clones were analysed (from 12 MGs, n = 3 mice). See Supplementary Table 1 for further details.

Quantification of percent labelled cells. Whole-mounts stained for K14, K8 and GFP were analysed by confocal microscopy. Then, 2,417, 2,234 and 2,249 total cells from three different mice were quantified in K14rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP; and 2,428, 2,031, 2,315 and 2,601 cells from four different mice were analysed in K14rtTA/TetO-Cre/Rosa-tdTomato. The proportion of labelled LCs in K14rtTA/TetO-Cre/Rosa-tdTomato or K14rtTA/TetO-Cre/Rosa-ΔNp63-IRES-GFP was quantified as the ratio of K8+ Tomato+ or GFP+ LCs over total K8+ LCs, whereas the proportion of labelled BCs was quantified as the ratio of K14+ Tomato+ or GFP+ BCs over K14+ total BCs.

Mammary cell preparation. E14 Lgr5-GFP embryos were collected and the whole skin containing the MGs was dissected. Tissues were digested in 300 U ml^{-1} collagenase (Sigma cat. no. C0130) + 300 µg ml^{-1} hyaluronidase (Sigma cat. no. 4272) diluted in HBSS for 1 h 30 min at 37 °C under shaking. EDTA at a final concentration of 5 mM was added for 3 min, followed by two washes in 10% FBS/PBS and 2% FBS/PBS before filtration through a 40 µm mesh. Adult MGs were dissected and the lymph nodes removed. Tissues were briefly washed in HBSS, and chopped into 1 mm^3 pieces. Chopped tissues were digested in HBSS + 300 U ml^{-1} collagenase (Sigma cat. no. C0130) + 300 µg ml^{-1} hyaluronidase (Sigma cat. no. 4272) for 2 h at 37 °C under agitation. Physical dissociation using a P1000 pipette was done every 15 min throughout the enzymatic digestion. EDTA 5 mM was added for 5 min, followed by 0.25% trypsin-EGTA for 1 min, before filtration through a 70 µm mesh, and two successive washes in 2% FBS/PBS.

Cell labelling, flow cytometry and sorting. Samples were incubated in 250 µl of 2% FBS/PBS with fluorochrome-conjugated primary antibodies for 30 min, with shaking every 10 min. Primary antibodies were washed with 2% FBS/PBS, and cells were resuspended in 2.5 mg ml^{-1} DAPI (Invitrogen D1306) before analysis. The following primary antibodies were used: APC-conjugated anti-CD45 (1:100, clone 30-F11, eBiosciences), APC-conjugated anti-CD31 (1:100, clone 390, eBiosciences), APC-conjugated anti-CD140a (1:100, clone APA5, eBiosciences) and PE-conjugated anti-CD49f (1:200, clone GoH3, eBiosciences) for embryos; PECy7-conjugated anti-CD24 (1:100, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1:100, clone eBioHMb1-1, eBiosciences), PE-conjugated anti-CD45 (1:100, clone 30-F11, eBiosciences), PE-conjugated anti-CD31 (1:100, clone MEC 13.33, BD Biosciences), PE-conjugated anti-CD140a (1:100, clone APA5, eBiosciences) for adult MGs. Data analysis and cell sorting were performed on a FACSAria sorter using FACS DiVa software (BD Biosciences). Dead cells were excluded with DAPI; $CD45^+$, $CD31^+$ and $CD140a^+$ cells were excluded (Lin^+) before analysis of the GFP^+ cells.

Microarray processing and analysis. Sorted CD49f Hi Lgr5-GFP Hi cells (2,000 cells per sample, n = 3 biological replicates) were collected directly in 45 µl of lysis buffer (20 mM dithiothreitol (DTT), 10 mM Tris–HCl pH 7.4, 0.5% SDS, 0.5 mg ml^{-1} proteinase K). Samples were then lysed at 65 °C for 15 min and frozen. RNA isolation, amplification and microarray were performed in the Functional Genomics Core, Barcelona. cDNA synthesis, library preparation and amplification were performed as described in ref. [59]. Microarrays were then performed on the Mouse Genome 430 PM strip Affymetrix array at IRB Functional Genomics Core (Barcelona). All results were normalized using the RMA normalization algorithm using the R-bioconductor affy package with standard parameters[60,61]. Cross experiment normalization was further performed to eliminate the batch effect using non-parametric empirical Bayes frameworks for adjusting data implemented in the ComBat function of the Surrogate Variable Analysis package (SVA) in R-bioconductor[62]. The transcriptional profiles of Lgr5 cells were normalized with the transcriptional profiles of adult BCs arising from K5CreER/Rosa-YFP mice and LCs arising from adult K8CreER/Rosa-YFP (n = 2 for each sample, previously described in ref. [17]). Only probes upregulated or downregulated by at

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

least 1.5-fold were considered in the analysis. Genes up- or downregulated were defined as having at least one probe displaying a 1.5-fold change. Venn diagrams were generated using Venny 2.0. The hypergeometric *P* value associated with each comparison between two signatures (calculated using the R statistical tool) corresponds to the probability of observing an intersection of a given size by chance only, knowing the number of genes tested on a microarray.

RNA-seq and analysis of bulk samples. A total of 40,000 LCs and 5,000 BCs were isolated by FACS as described above and collected into kit lysis buffer. RNA was extracted using an Absolutely RNA Nanoprep kit (Stratagene). RNA quality was checked using a Bioanalyzer 2100 (Agilent Technologies). Indexed cDNA libraries were obtained using the Ovation Solo RNA-Seq System (NuGen) following the manufacturer's recommendations. Multiplexed libraries (18 pM) were loaded on flow cells and sequences were produced using a NovaSeq 6000 S2 Reagent Kit (200 cycles) from a NovaSeq 6000 System (Illumina). Approximately 19 million paired-end reads per sample were mapped against the mouse reference genome (GRCm38/mm10) using STAR software to generate read alignments for each sample. Annotations Mus_musculus.GRCm38.87.gtf were obtained from ftp.Ensembl.org. After transcripts assembly, gene level counts were obtained using HTSeq. The fold changes of mean gene expression for the duplicates were used to calculate the level of differential gene expression. Heatmaps of the 500 most variable genes across the eight samples and corresponding clustering dendrograms were drawn with the heatmap.2 function of the R package gplots. Euclidian distance, with the complete linkage agglomeration method, was used for clustering.

GSEA analysis. For Fig. 2, GSEA analysis was performed using the ranked fold change of probe expression values between Lgr5 and BCs or LCs and genes upregulated in LCs or BCs for the displayed data set[63]. For Fig. 6, GSEA analysis was performed using the ranked fold change of gene expression values between BCs and LCs or between LCΔNp63 and LCs for genes expressed with at least 10 reads per 20 million aligned reads in RNA-seq counts and genes upregulated in LCΔNp63 compared to LCs or EMPs signatures from a single cell.

GO analysis of the multipotent signatures of the Venn diagrams. Genes upregulated in each subset of Venn diagrams were tested for enrichment in each GO class using the DAVID web server[64,65]. Statistically significant enrichments correspond to those presenting a corrected *P* value (for example, Bonferroni or Benjamini) smaller or equal to 0.05, although some genes involved in non-statistically significant GO classes were plotted for their biological relevance.

scRNA-seq. scRNA-seq were generated using a modified Smartseq-2 protocol[66]. A 1 µl volume of lysis buffer in 384-well PCR plates for cell sorting was prepared with 0.4% vol/vol Triton-X lysis buffer, 2.5 mM dNTPs, 2.5 µM oligo-dT30-VN and ERCC (external RNA controls consortium) controls at a final dilution of 1:100 million. Reverse transcription and PCR was performed according to the Smartseq-2 protocol with reduced volumes: 1 µl of reverse transcription mix instead of 5.7 µl and 3 µl PCR Master mix instead of 12.5 µl. cDNA was amplified for 24 cycles and cleaned using HighPrep PCR beads (MagBio Genomics) at a 0.8× ratio on a Hamilton STAR liquid handler, eluted in 30 µl elution buffer (EB) (Qiagen) and transferred to 384_PP acoustic plates (LabCyte). DNA quantification was performed with a Picogreen assay (Thermofisher), and a subset of samples were selected for quality control using an Agilent 2100 BioAnalyser using a High Sensitivity DNA kit. After initial quality control, 24 samples were discarded with cDNA yields of less than 21 ng. These samples were replaced with 24 cells from another 384-well plate from the same cell sort by plate reformatting with an acoustic dispenser (LabCyte Echo 525).

Library preparation continued from Smartseq-2. Next-generation sequencing library preparation was performed using a Nextera XT DNA library preparation kit with volumes reduced by one-tenth (Illumina) using an acoustic dispenser. In brief, 100 pg of cDNA in a volume of 500 nl was tagmented by adding 1.5 µl Tn5-buffer mix and incubating for 10 min at 55 °C. Tagmented samples were barcoded with Nextera index sets A–D and amplified with 11 cycles of PCR. After PCR, all samples were pooled and cleaned using HighPrep PCR beads at a 0.6× ratio. Library pools were eluted in buffer EB, and quality control was performed using an Agilent 2100 BioAnalyser and High Sensitiviy DNA chip before adjusting to a concentration of 4 nM. The diluted pools were quantified using a KAPA qPCR library quantification kit on a LightCycler 480 (Roche) before a final dilution to 2 nM. The pool of 384 samples was sequenced on two lanes of a HiSeq2500 in high output mode v4 chemistry with 1× 100 bp read length.

Single-cell bioinformatics analysis. Sequencing reads were trimmed for adapter sequences using cutadapt (version 1.13) and reads were aligned to the GRCm38 reference genome including ERCC sequences using STAR with default parameters (version 2.5.2b)[67]. The expression count matrix was generated using HTSeq (version 0.6.0)[68] on GENCODE M12 transcript annotations, and counts for each protein coding gene were collapsed. Quality control was performed using the Scater R package (version 1.2.0)[69]. Cells that complied with one of the following conditions were excluded: had fewer than 10^5 counts, showed expression of fewer than 2,500 unique genes, had more than 20% counts belonging to ERCC sequences, or had more than 8% counts belonging to mitochondrial sequences. BCs and EMPs that showed no expression of either K5 or K14, and LCs that did not express K8 were further excluded. Cells coming from a row F of the 384-well plate that showed systematic mixing of LC and BC markers were excluded from further analysis due to a likely pipetting error. Of the 377 samples that passed sequencing, 221 passed quality control. Genes for which fewer than 20 counts were observed across the complete data set were excluded from further analysis. Read counts were normalized using *scran* with default parameters (version 1.2.2). Clustering using the SC3 R package (version 1.3.18)[40] and PCA was performed using the prcomp function in R, plots were generated using the ggplot2 R package (version 2.2.1). We chose $k=4$ (all cells), $k=2$ (for EMP-only and LC vs BC clustering) for SC3 as this best represented the heterogeneity in our data set and recapitulated the studied cell lineages. For cluster marker gene discovery we set the thresholds to all genes with an AUC higher than 0.8 and *P* value lower than 0.01. BC/LC-specific markers were determined by filtering marker genes identified by SC3 on non-EMP cells and retaining only marker genes that were expressed in 75% of the respective population and less than 50% of the opposite population. The adjusted proportion of specific markers for each cell was computed by counting the number of specific LC/BC markers over the total number of LC/BC-specific markers and correcting for differences in sensitivity by modelling the linear relationship between the adjusted proportion of markers detected and the total number of genes detected for each cell. Heatmaps were generated using a modified version of the *gplots* R package (3.0.1). Cell-cycle phase was automatically assigned using the *scran* package and cells not in G1 phase were excluded from further analysis. Gene regulatory network analysis was performed using SCENIC[48] with default parameters. Regulon AUC thresholds for binary activity determination were manually assessed and adjusted for 425 regulons expressed. Pearson correlations between regulon AUCs and the adjusted proportion of specific LC/BC markers was computed, and the corresponding *P* value was FDR corrected using the Benjamini–Hochberg method. To determine the enrichment of p63-inferred target genes in p63-overexpressing LC cells, we computed the probability under a binomial model with $P=0.12$ (fraction of genes with 1.5-fold enrichment in ΔNp63-expressing LC cells compared to WT LC cells) for which 18 (or more) of 48 trials were a success.

Statistics and reproducibility. All experiments were repeated at least three times with similar results (unless a different number of repeats is stated in a figure legend). The methods used, *P* values and *N* are indicated in the figure legends. No statistical method was used to predetermine sample size. All experimental mice used in this study were females of mixed genetic backgrounds. No animals were excluded from the study. No method of randomization was used. The investigators were not blinded to allocation during experiments or outcome assessment.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Microarray, RNA-seq and single cell RNA sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE109711. Previously published microarray data that were re-analysed here are available under accession code GSE69290.

Source data for Figs. 1d, 1l, 1q, 6d and 7c are provided in Supplementary Table 1.

References

56. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
57. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
58. Perl, A. K., Wert, S. E., Nagy, A., Lobe, C. G. & Whitsett, J. A. Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl Acad. Sci. USA* **99**, 10482–10487 (2002).
59. Gonzalez-Roca, E. et al. Accurate expression profiling of very small cell populations. *PLoS One* **5**, e14418 (2010).
60. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
61. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
62. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
63. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

64. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
65. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
66. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
69. McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Cédric BLANPAIN

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The confocal imaging data-sets generated were merged and displayed with the ZEN 2 software. Flow cytometry data analysis and cell sorting was performed on a FACSAria sorter using the FACS DiVa software (BD Biosciences).
Data analysis	The confocal imaging data-sets generated were analysed with the ZEN 2 software. Flow cytometry data analysis was performed on a FACSAria sorter using the FACS DiVa software (BD Biosciences). All microarray results were normalized using the RMA normalization algorithm using R-bioconductor affy package with standard parameters. Cross experiment normalization was further performed to eliminate the batch effect using non-parametric empirical Bayes frameworks for adjusting data implemented in ComBat function of the Surrogate Variable Analysis package (SVA) in R-bioconductor. Venn diagrams were generated using Venny 2.0. GSEA analysis was performed using ranked fold change of probe expression values between Lgr5 and BCs or LCs and genes upregulated in LCs or BCs for the displayed dataset. Genes up-regulated in each subset of the Venn Diagrams were tested for enrichment in each Gene Ontology class using the DAVID web server. Sequencing reads were trimmed for adapter sequences using cutadapt (version 1.13) and reads were aligned to the GRCm38 reference genome including ERCC sequences using STAR with default parameters (version 2.5.2b)79. The expression count matrix was generated using HTSeq (version 0.6.0)80 on GENCODE M12 transcript annotations and counts for each protein coding gene were collapsed. Quality control was performed using the scater R package (version 1.2.0)81. Cells that complied with one of the following conditions were excluded: had fewer than 105 counts, showed expression of fewer than 2000 unique genes, had more than 20% counts belonging to

ERCC sequences, had more than 8% counts belonging to mitochondrial sequences. BCs and EMPs that showed no expression of neither K5 nor K14, and LCs that did not express K8 were further excluded. Out of the 377 samples which passed sequencing, 281 passed quality control. Genes for which less than 20 counts were observed across the complete dataset were excluded from further analysis. Read counts were normalized using scran with default parameters (version 1.2.2) 82. Clustering using the SC3 R package (version 1.3.18)49 and PCA was performed using the prcomp function in R, plots were generated using the ggplot2 R package (version 2.2.1). We chose k=4 (all cells), k=2 (for EMP-only and LC vs BC clustering) for SC3 as this best represented the heterogeneity in our dataset and recapitulated the studied cell lineages. For cluster marker gene discovery we set the thresholds to all genes with an AUC higher than 0.8 and p-value lower than 0.01. Heatmaps were generated using a modified version of the gplots R package (3.0.1). Cell-cycle phase was automatically assigned using the scran package.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Microarray, RNAseq and single cell RNA sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE109711. Previously published microarray data that were re-analysed here are available under accession code GSE69290.
Source data for Figure 1d, 1l, 1q, 6d and 7c have been provided as Supplementary Table 1.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. All experiments were repeated at least three times with similar results, except for microarray and RNAseq for which experiments were repeated twice.
Data exclusions	For the single cell RNA sequencing, after initial quality control 24 samples were discarded with cDNA yields of less than 21ng. These samples were replaced with 24 cells from another 384 well plate from the same cell sort by plate reformatting with an acoustic dispenser (LabCyte Echo 525). Cells coming from a row F of the 384-well plate which showed systematic mixing of LC and BC markers were excluded from further analysis due to a likely pipetting error. Out of the 377 samples which passed sequencing, 221 passed quality control. Genes for which less than 20 counts were observed across the complete dataset were excluded from further analysis.
Replication	All attempts at replication were successfull and are shown, n is described in legends.
Randomization	The experiments were not randomized. Transgenic mice of mixed background were used and compared to control mice of mixed background.
Blinding	The investigators were not blinded to allocation during experiments and outcome assessment. Blinding was not possible as the same investigator processed the animals and analysed the data.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Research animals
☒	☐ Human research participants

Antibodies

Antibodies used

The following primary antibodies were used: anti-GFP (chicken, 1:1000, ab13970, Abcam), anti-K8 (rat, 1:1000, Troma-I, Developmental Studies Hybridoma Bank, University of Iowa), anti-K14 (rabbit or chicken, 1:1000, Thermo), anti-K5 (rabbit, 1:1000, PRB-160P-0100, Covance), anti-CD49f-PE (rat, 1:100, clone GoH3, eBiosciences), anti-p63 (rabbit, 1:500, clone EPR5701, Abcam), anti-SMA-Cy3 (mouse, 1:500, clone 1A4, Sigma), anti-smMHC (rabbit, 1:100, BT562, Biomedical Technologies), anti-Sox9 (rabbit, 1:5000, AB5535, Millipore), anti-FoxA1 (rabbit, 1:100, clone EPR10881, Abcam), anti-ER (rabbit, 1:300, sc542, Santa Cruz). The following secondary antibodies, diluted 1:400, were used: anti-rabbit (A21206), anti-rat (A21208), anti-chicken (A11039) conjugated to AlexaFluor488 (Molecular Probes), anti-rabbit (711-295-152), anti-rat (712-295-153), anti-chicken (703-295-155) Rhodamine Red-X or anti-rabbit (711-605-152), anti-rat (712-605-153), anti-chicken (703-605-155) Cy5 (JacksonImmunoResearch).

Validation

Antibodies specificity was controlled by using sections of tissues known to express high levels of the antigen, as well as sections of tissues known to lack expression of the antigen.
Antibodies used in this study were described previously in Van Keymeulen et al, Nature 2011, Van Keymeulen et al, Nature 2015, Wuidart et al, Genes and dev 2016, Van Keymeulen et al, Cell Reports 2017.

Research animals

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Animals/animal-derived materials

Lgr5-EGFP-IRES-CreER15 and Rosa-tdTomato55 mice were obtained from the Jackson Laboratory. Rosa-Confetti13 mice were provided by H. Clevers; K14rtTA transgenic mice56 were provided by Elaine Fuchs; TetO-Cre mice57 were provided by Andreas Nagy; Rosa26-ΔNp63-IRES-GFP mice51 were provided by Wim Declercq. The generation of K5CreER and of K8rtTA were previously described8, 10. Mice colonies were maintained in a certified animal facility in accordance with European guidelines. The experiments were approved by the local ethical committee (CEBEA) under protocols #477 and #527.The study is compliant with all relevant ethical regulations regarding animal research.
All experimental mice used in this study were females of mixed genetic backgrounds. Mice were analysed at embryonic stages E14, E15 and E17, during postnatal development at P1, P5, P21, P60, 7w and in adult mice (over 8w), as indicated in figure legends.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ Magnetic resonance imaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Mammary cell preparation – E14 Lgr5-GFP embryos were collected and the whole skin containing the mammary buds was dissected. Female tissues were placed in HBSS + 300U/ml collagenase (Sigma) + 300µg/ml hyaluronidase (Sigma) and digested for 1h30 at 37°C under shaking. EDTA at a final concentration of 5mM was added for 3min to the resultant organoid suspension, followed by a wash in 10% FBS/PBS and labelling in 2% FBS/PBS before filtration through a 40µm mesh. Adult mammary glands

were dissected and the lymph nodes removed. Tissues were briefly washed in HBSS, and chopped in 1mm3 pieces. Chopped tissues were placed in HBSS + 300 U/ml collagenase (Sigma) + 300µg/ml hyaluronidase (Sigma) and digested 2h at 37°C under agitation. Physical dissociation using a P1000 pipette was done every 15mins throughout the enzymatic digestion time. EDTA at a final concentration of 5mM was added for 5 min to the resultant organoid suspension, followed by 0,25% Trypsin-EGTA for 1 min before filtration through a 70-µm mesh, 2 successive washes in 2% FBS/PBS and labelling.

Cell labelling, flow cytometry and sorting – Samples were incubated in 250µl of 2% FBS/PBS with fluorochrome-conjugated primary antibodies for 30min, with shaking every 10min. Primary antibodies were washed with 2% FBS/PBS, and cells were resuspended in 2.5mg/ml DAPI (Invitrogen) before analysis. The following primary antibodies were used: APC-conjugated anti-CD45 (1:100, clone 30-F11, eBiosciences), APC-conjugated anti-CD31 (1:100, clone 390, eBiosciences), APC-conjugated anti-CD140a (1:100, clone APA5, eBiosciences) and PE-conjugated anti-CD49f (1:200, clone GoH3, eBiosciences) for embryos; PECy7-conjugated anti-CD24 (1:100, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1:100, clone eBioHMb1-1, eBiosciences), PE-conjugated anti-CD45 (1:100, clone 30-F11, eBiosciences), PE-conjugated anti-CD31 (1:100, clone MEC 13.33, BD Biosciences), PE-conjugated anti-CD140a (1:100, clone APA5, eBiosciences) for adult MGs. Data analysis and cell sorting was performed on a FACSAria sorter using the FACS DiVa software (BD Biosciences). Dead cells were excluded with DAPI; CD45+, CD31+ and CD140a+ cells were excluded (Lin+) before analysis of the GFP+ cells.

Instrument

FACSAria 3 sorter (BD Biosciences)

Software

FACS DiVa software (BD Biosciences)

Cell population abundance

To control cell sorting quality, 5000 cells of the different populations of interest were sorted in PBS and their FACS profile was analyzed using exactly the same gating strategy (post-sort control). Minimum 90% of the sorted cells were located within the sorting gate. Cells usually showed a reduced fluorescent labelling, due to bleaching following the first sort, explaining why a fraction of the cells are not located within the sorting gate.

Gating strategy

Supplementary Figure 1. Gating strategy for flow cytometry analysis. A-G, Unicellular suspension of skin and mammary bud cells from Lgr5-IRES-GFP E14 embryos stained for Lin (CD31, CD45, CD140a) in APC and CD49f in PE were gated as shown in A to eliminate debris, doublets were discarded with gate shown in B followed by gate showed in C, the living cells were gated by DAPI dye exclusion as shown in D, the non-epithelial Lin positive cells were discarded in E. The CD49f Hi cells were gated as shown in F and the GFP+ cells were gated as shown in G. H-O, Unicellular suspension of mammary cells from adult K8rtTA/TetOCre/DNp63-IRES-GFP, induced at P30 and analysed at P45, stained for Lin (CD31, CD45, CD140a) in PE, CD24 in PECy7 and CD29 in APC, were gated as shown in H to eliminate debris, doublets were discarded with gates shown in I followed by gate shown in J, the living cells were gated by DAPI dye exclusion as shown in K, the non-epithelial Lin positive cells were discarded in L and the GFP+ cells were gated as shown in M. CD24 and CD29 expression was studied in Lin- cells (N) or in YFP+ cells (O). The CD24+CD29Lo gate corresponds to luminal cells (LC), while CD24+CD29Hi gate corresponds to basal cells (BC). The stromal population corresponds to the cells labelled due to the leakiness of the Tet-O-Cre, as described previously in reference 14.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.