Repression of endogenous retroviruses prevents antiviral innate immunity and is required for mammary gland development

Alexandra Avgustinova[1,8#], Carmelo Laudanna[1*], Mónica Pascual-García[1*], Quirze Rovira[2], Magdolna Djurec[1], Andres Castellanos[1], Uxue Urdiroz-Urricelqui[1], Domenica Marchese[3], Neus Prats[1], Alexandra Van Keymeulen[4], Holger Heyn[3,5], Juan M. Vaquerizas[2,6], and Salvador Aznar Benitah[1,7#§]

[1] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[2] Max Planck Institute for Molecular Biomedicine, Muenster, Germany

[3] CNAG-CRG, Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[4] Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles (ULB), Brussels, Belgium

[5] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[6] MRC London Institute of Medical Sciences, Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK

[7] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

[8] current address: Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain

§- lead author

- corresponding authors (salvador.aznar-benitah@irbbarcelona.org; aavgustinova@fsjd.org)

* - these authors contributed equally to this work

SUMMARY

The role of heterochromatin in cell fate specification during development is unclear. We demonstrate that loss of the lysine 9 of histone H3 (H3K9) methyltransferase G9a in the mammary epithelium results in *de novo* chromatin opening, leading to aberrant formation of the mammary ductal tree, impaired stem cell potential, disrupted intraductal polarity and loss of tissue function. G9a loss derepresses long terminal repeat (LTR) retroviral sequences, most prominently from the ERVK family. Transcriptionally activated endogenous retroviruses generate double-stranded DNA (dsDNA) that triggers an antiviral innate immune response, and knockdown of the cytosolic dsDNA sensor *Aim2* in G9a-knock out (G9acKO) mammary epithelium rescues mammary ductal invasion. Mammary stem cell transplantation into immunocompromised or G9acKO-conditioned hosts shows partial dependence of the G9acKO mammary morphological defects on the inflammatory milieu of the host mammary fat pad. Thus, altering chromatin accessibility of retroviral elements disrupts mammary gland development and stem cell activity through both cell autonomous and non-autonomous mechanisms.

INTRODUCTION

Modulation of chromatin accessibility through epigenetic regulation in non-coding regions might be important for correct fate specification during embryonic development, but it is unclear if this is critical for proper tissue development and maintenance. We recently showed that keratin 14 (K14)–specific deletion of the histone methyltransferase G9a leads to severe reduction of dimethylated H3K9 (H3K9me2) levels, concomitant with significant *de novo* chromatin opening in the epidermis (Avgustinova et al., 2018). H3K9me2 generates low chromatin accessibility, contributing to gene repression at euchromatin and to formation of heterochromatin (Tachibana et al., 2005). Despite the increased chromatin opening following epidermal deletion of G9a, very few protein-coding genes show changes in their expression, and epidermal development, morphology, and cell fate–specification remain largely unperturbed. However, upon cellular stress (in the form of DNA damage), epidermal stem cells lacking G9a are more susceptible to malignant transformation (Avgustinova et al., 2018). Epigenetic mechanisms can be largely dispensable for the maintenance of adult stem cell identity yet mediate the stem cell response to tissue-specific stresses (Avgustinova and Benitah, 2016). We thus wondered whether increased chromatin accessibility following G9a deletion might have different outcomes in other tissues.

Here we investigated how chromatin opening following G9a ablation affects mammary gland development and cell fate specification. The mouse mammary gland develops largely postnatally, making it an excellent model system for studying tissue development and differentiation. At birth, the mammary epithelium consists of a rudimentary ductal tree, which undergoes rapid expansion in response to hormonal signals during puberty, until the homeostatic adult state is reached (Visvader and Stingl, 2014). During adulthood, the mammary epithelium remains plastic and is sensitive to hormonal cues, facilitating lobulo-alveolar differentiation during pregnancy and involution post-pregnancy. We demonstrate that G9a loss results in severely disrupted development of the mammary epithelium, impaired stem cell potential, and profound alterations in epithelial polarity and extracellular matrix

deposition. Intriguingly, despite the striking disruption of development in G9a-ablated mammary epithelia, lineage specification and fidelity remained largely unperturbed. However, mammary epithelial cells lacking G9a aberrantly expressed LTR endogenous retroviruses, which can activate an antiviral innate immune response. Blocking the cellular response to LTR retroviral derepression ameliorates morphological defects in mammary epithelia lacking G9a. Finally, we provide evidence that suggests that both cell autonomous and non-autonomous mechanisms contributed to the observed morphological abnormalities in G9acKO mammary glands.

RESULTS

G9a Loss Delays Mammary Gland Development

Mice carrying a K14-specific knockout of G9a (henceforth, G9acKO; Figure 1A) exhibited severely reduced levels of G9a and H3K9me2 in both luminal and basal cells of the mammary epithelium (Figures 1B, 1C, and S1A–1C). Although luminal cells do not express cytokeratin 14, loss of G9a and H3K9me2 in this lineage is expected, as both lineages originate from a common K14-positive progenitor cell during embryonic development (Jonkers et al., 2001; Sun et al., 2010).

Ductal invasion into the mammary fat pad was severely delayed in early-pubertal (4-week-old) and pubertal (6-week-old) G9acKO mice (Figure 1D), although most G9acKO mammary glands filled the entire mammary fat pad in 12-week-old (henceforth, adult) mice. Morphologically, the 6-week-old (henceforth, pubertal) G9acKO mammary tree was characterized by severely distended and abnormal terminal end buds (TEBs) and a slight increase in the proportion of tip versus branch segments (consistent with an overall smaller mammary tree) (Figures 1E–1G). Both pubertal and adult G9acKO mammary epithelia displayed distended ducts (Figure 1H), while adult G9acKO mammary epithelia were more branched, with a large number of abortive ductal tips and overall shorter tip- and ductal branch-segments, compared to wild-type (WT) littermate controls (Figure 1I).

Aberrant Cell Polarity and Extracellular Matrix in G9acKO Mammary Epithelium

The bilayered mammary epithelium comprises two principal lineages: 1) luminal cells, which line the ductal lumen and are responsible for hormone sensing and milk production during pregnancy; and 2) basal cells, which envelop the luminal layer, directly contact both the stroma and the basement membrane, and provide the contractile force for milk ejection during lactation. WT mammary ducts had correct expression of both layers of luminal cells (expressing cytokeratin 8 [CK8]) and basal cells (expressing αSMA); in contrast, the

G9acKO mammary epithelium had intermingled luminal and basal cells (as shown by αSMA and CK8) (in both ducts and TEBs in pubertal mice, and mainly at the tip segments in adult mice), and the basal cells had a columnar (rather than spindle) shape, reminiscent of luminal cells (Figures 1J–1N and S1D-S1G).

In G9acKO mammary glands, neither the tight-junction protein zona occludens 1 (ZO-1) nor the apical membrane domain marker p-ezrin (Thr567)/p-radixin (Thr564)/p-moesin (Thr558) (pERM) (Godde et al., 2014) were limited to the apical surface of the luminal epithelial cells, as would be expected (Ewald et al., 2008), indicating a loss of apico-basal polarity (Figures 2A, S2A, and S2B). Strikingly, ZO-1 also marked the baso-lateral surface of a subset of luminal cells, identified additional rudimentary lumina within the multilayered G9acKO mammary epithelium and showed nuclear localization in adult G9acKO epithelia, consistent with immature cell–cell contacts (Gottardi et al., 1996) (Figures 2A and S2A). Furthermore, E-cadherin and pERM were aberrantly localized on the apical surface of the G9acKO basal epithelial cells (Figure S2B).

Finally, collagen IV, one of the main constituents of the mammary basement membrane, was more abundant and no longer confined to the epithelial/stromal interface of the basal cells in G9acKO mammary glands (Figures 2B and S2C). Indeed, collagen deposition (as measured by Masson´s trichome staining) was significantly increased around G9acKO mammary ducts (Figures 2C, 2D, S2D, and S2E). Of note, both estrogen receptor alpha (ERα)-positive and ERα-negative luminal lineages were correctly specified in G9acKO mammary epithelia, albeit with a slight decrease in ERα-positive mature luminal cells (Figures S2F and S2G).

Postpartum, tissue expansion and alveologenesis were incomplete in G9acKO mothers on day 1 of lactation (Figures 2E and S2H), concomitant with a strong decrease in H3K9me2 levels (Figure 2F) and increased collagen deposition (Figures 2G and 2H). G9acKO mothers failed

to produce milk, and their pups starved on postnatal day 1 if not cross-fostered to WT mothers (data not shown). The delay in alveolar differentiation and aberrant collagen deposition were present in mid-pregnancy (Figures S2I–S2L), foretelling the failure of G9acKO mammary tissue function.

Severely Reduced Stem Cell Potential of G9acKO Mammary Epithelium

We next used the Rosa26YFP-reporter cassette in our transgenic mouse line (see Figure 1A) to distinguish the mammary epithelium from stroma by flow cytometry, by following YFP-expressing cells (Figure S3A). As compared to WT mice, pubertal G9acKO had a significantly lower proportion of YFP-positive mammary epithelial cells in mammary fat pads; however, the YFP-positive proportion was recovered in adult G9acKO fat pads, underscoring a developmental delay rather than failure (Figures 3A, 3B, and S3B). Surprisingly, there were no differences in the relative percentage of luminal and basal cells at either developmental time point (Figure 3C).

Assessing the stem cell potential of FACS-sorted cells, we observed that colony-forming potential was severely compromised (with both reduced colony number and size) in both luminal and basal G9acKO cells plated in 2D or 3D (Figures 3D–3G and S3C–S3H). However, the gold standard for mammary stem cell potential is epithelial cell transplantation into a cleared mammary fat pad (Shackleton et al., 2006). As mammary basal cells in G9acKO mice had an altered morphology (and thus perhaps distinct expression of cell-specific markers, biasing FACS isolation), we chose to transplant sorted total mammary epithelium from WT or G9acKO mice (rather than basal cells only). All mice transplanted with 5000 FACS-isolated WT mammary epithelial cells successfully developed mammary trees; in contrast, almost none of the mice transplanted with G9acKO cells did, and the few outgrowths that developed were small and displayed an abnormal morphology reminiscent of invasive organoids (rather than ductal trees) (Figures 3H–3J). Collectively, these findings

indicated that deletion of G9a resulted in a clear reduction of mammary stem cell potential, which was at least in part cell autonomous.

Lineage specification in G9acKO mammary epithelial cells is unaltered

To test for lineage fidelity and specification, we analyzed individual WT or G9acKO mammary epithelial cells (FACS-sorted from virgin adults, YFP+/Lin–; Figure S3A) using single-cell RNA sequencing (scRNA-seq). We identified the previously-described four cell clusters (Pal et al., 2017): basal, luminal progenitor (LP), luminal intermediate, and mature luminal (ML) cells (Figures 4A, S4A, and S4B). Notably, we also identified a fifth cluster adjacent to LP cells and defined by the expression of cell cycle–related genes (Figure S4A; Table S1); we termed these cells "proliferative LPs". Identities of the mammary cell populations were confirmed based on expression of lineage markers (Figure S4C). No major differences in co-expression of the basal cell markers αSMA, keratin 14, and p63 were observed, although the proportion of G9acKO basal cells expressing all thee markers was slightly reduced (Figures S4D and S4E). Mammary epithelial cell clusters and lineage score expression (based on Pal et al., 2017) between WT and G9acKO mammary epithelial cells were largely unchanged (Figures 4B and S4F), except for a substantial reduction in proliferative LPs in G9acKO mammary epithelial cells (Figure 4A). Further, focusing on the basal cell cluster only , we identified a proliferative subcluster, again reduced in G9acKO mammary epithelia (Figure S4G; Table S2). Staining for cell proliferation marker Ki67 confirmed a reduction in proliferation of both luminal and basal G9acKO mammary lineages (Figures 4C, 2D, and S4H). We concluded that lineage fidelity and specification were essentially retained in G9acKO mammary epithelial cells.

Strikingly, few genes (25, 29, or 32) were differentially expressed for the Basal, LP, or ML cell clusters, respectively, comparing G9acKO to WT (FDR < 0.05, FC > 1.5; Table S3). The genes upregulated in G9acKO basal cells versus WT included the transcriptional repressor of mesenchymal genes *Klf4* (Tiwari et al., 2013), keratin 18 (which is typically luminal), and

several cytoskeleton-associated genes (*Tuba1c*, *Tuba4a*, *Tubb2a*, and *Tubb4b*) (Figure 4E), consistent with the columnar (epithelial) appearance of G9acKO basal cells. Our single-cell level expression data suggested that upregulation of these genes was the result of increased median expression levels per cell and a slight expansion of a basal population already present in basal WT cells (data not shown). Notably, the cytokine genes *Cxcl1*, *Cxcl2*, and *Csf3* were strongly upregulated in both basal and LP G9acKO cells (Figures 4E–4G), suggesting a potential paracrine effect of G9a deletion.

Skewed immune environment in G9acKO mammary gland

To gain higher resolution of differentially expressed genes (DEGs) after G9a deletion, we performed bulk RNA-seq of FACS-isolated mammary luminal and basal epithelial cells from virgin adult WT or G9acKO mice. We found a total of 1235 DEGs (498 down-, 737 upregulated) in basal cells, and of 683 DEGs (92 down-, 590 upregulated) in luminal cells (FDR < 0.05, FC > 1.5; Figures 5A and S5A; Table S4). Gene ontology (GO) analysis of upregulated DEGs in G9acKO cells indicated a strong activation of immune response pathways (Figures 5B and S5B; Table S4), with numerous cytokines and chemokines mis-expressed in both luminal and basal cells (Table S4), consistent with the scRNA-seq results (Figure 4E–G). For instance, the GO categories "cellular response to interleukin-1" and "cellular response to interferon γ/β" were over-represented in basal cells, while "cytokine-cytokine receptor interaction" and "graft-versus-host disease" were the most significantly over-represented KEGG pathways (Table S4). Luminal cells showed upregulated DEGs in the GO category "pyroptosis", which is an immunogenic type of programmed cell death (Jorgensen et al., 2017; Shi et al., 2017), and key pyroptotic genes (e.g., *Aim2, Casp1, Gsdmc2, Naip5, Naip6,* and *Nlrp1b*) were upregulated in luminal and/or basal G9acKO cells (Figures 5C and S5C). Finally, array-based whole-genome expression profiling on pubertal WT or G9acKO luminal and basal cells (Figure S5D; Table S5) showed a substantial overlap in DEGs between puberty and adulthood, notably including the pyroptosis-related *Aim2*, *Casp1*, and *Naip5* genes (Figure S5E). We concluded that these DEGs characterized the

mammary epithelium of G9acKO mice in general (rather than a specific developmental time point). Note that the gene for G9a, *Ehmt2*, was not detected as a DEG by the Affymetrix gene expression arrays due to probe targeting to *Ehmt2* regions retained in G9acKO animals. Loss of *Ehmt2* expression was confirmed by real-time quantitative PCR specific for the deleted exon 27 of *Ehmt2* (Figure S5F).

Caspase-1 (the pyroptosis effector caspase) had elevated RNA, protein, and activity levels in G9acKO mammary epithelium versus WT (Figures 5D-F and S5G-S5I), indicating activation of a functional inflammasome upon G9a loss. Caspase-1 is activated by diverse triggers, including viruses, bacteria, and toxins (Jorgensen et al., 2017); however, co-upregulation of the pyroptotic sensor protein Aim2, which specifically senses cytosolic double-stranded DNA (dsDNA) (Fernandes-Alnemri et al., 2009; Hornung et al., 2009), suggested cytosolic dsDNA as a trigger. Interestingly, the GO categories "positive regulation of MDA-5 signaling pathway" and "positive regulation of RIG-I signaling pathway" were significantly under-represented in G9acKO versus WT luminal cells (Figure S5B; Table S4). The MDA-5 and RIG-I signaling pathways form part of the cytosolic DNA sensing machinery of the cGAS-STING pathway (Zevini et al., 2017), which has been reported to be negatively regulated by Aim2 inflammasome-based pyroptotic cytosolic DNA sensing (Corrales et al., 2016; Nakaya et al., 2017).

Histological staining revealed that cytosolic dsDNA levels were strongly elevated in both luminal and basal cells in G9acKO versus WT mammary epithelium (Figures 5G and S5J). We ruled out that this cytosolic dsDNA came from bacterial or viral infection (as mice were kept in specific pathogen-free conditions, and same-cage littermates were used as controls), suggesting self-DNA as a potential source. Notably, cell death levels were unchanged (Figures S5K and S5L), suggesting that G9a loss only triggered altered cytokine expression but not pyroptotic cell death. Similarly, Aim2-inflammasome activation in the epidermis leads to cytokine production but not pyroptotic cell death (Naik et al., 2017).

Finally, consistent with a pyroptosis-like inflammatory signaling, we observed a slight but significant increase of infiltrating inflammatory monocytes, a (non-significant) increase of infiltrating neutrophils, and an increase of infiltrating T lymphocytes, in G9acKO mice versus WT littermates (Figure 5H-5J). There were no differences in the infiltration of B-cells, macrophages, or NK cells (data not shown).

Lack of G9a results in differentially accessible chromatin in LTR retrotransposon regions

To assess whether reduced H3K9me2 levels due to G9a loss led to increased chromatin opening, we performed an assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (Buenrostro et al., 2013) on FACS-isolated luminal or basal cells from WT or G9acKO mice (Figure S3A). We found increased accessibility for about 65% of differentially open chromatin regions (8445 of 12857) in G9acKO basal cells, and for about 81% (2138 of 2638) in G9acKO luminal cells (Figures 6A and S6A); these were on average longer than those in WT cells (Figures 6B and S6B) and were enriched in intergenic regions and regions containing the LTR retrotransposon class of transposable elements (TEs) (Figures 6C and S6C). In contrast, in WT luminal and basal cells, differentially open regions were predominantly in genic (promoter/exon/intron/TTS) regions (Figures 6C and S6C). This suggests that deletion of G9a preferentially opened intergenic chromatin regions containing LTR retrotransposons, which, if derepressed, could be reverse-transcribed through a dsDNA intermediate in the cytosol and hence trigger the Aim2 inflammasome signaling cascade (Johnson, 2019).

To test this hypothesis further, we analyzed the enrichment of different classes of TEs in differential ATAC peaks from WT or G9acKO mammary epithelial cells. Strikingly, as compared to WT cells, differentially open chromatin regions in G9acKO basal cells were enriched for LTR retrotransposon and LINE classes of TEs, specifically for the ERVK and

ERVL-MaLR TE families (Figures 6D). LTR retrotransposon and LINE classes of TEs were also over-represented in chromatin regions differentially open in G9acKO luminal cells, albeit no single TE family was significantly enriched (Figure S6D). We saw no examples in which G9a loss resulted in increased chromatin accessibility at loci of all copies of any individual TE; for instance, <10% of copies of the TE RLTR10D2 overlap with differentially accessible chromatin (Figures 6E and S6F). This suggests that H3K9me2 on its own is not a general mechanism of repression of any TE in mammary epithelial cells, but rather works in concert with other transcriptional and epigenetic mechanisms to establish stable repression of TEs.

LTR retrotransposons are specifically derepressed in G9acKO mice via chromatin opening

Re-visiting the total RNA-seq data (see Figure 5), we observed differential expression of 20 and 38 TEs (predominantly LTR retrotransposons) in luminal and basal G9acKO cells, respectively (FDR < 0.01) (Figures 6F and S6G). Four LTR retrotransposons from the ERVK family (namely, RLTR44A, -53B, -10D2, and -44B) were upregulated and highly enriched for differential chromatin opening, in basal and/or luminal G9acKO mammary epithelial cells (Figure S6H). Thus, chromatin opening in TE regions led to their transcriptional derepression in G9acKO mammary epithelial cells.

To determine if the derepressed LTR retrotransposons in G9acKO cells were full-length (e.g., functional and able to generate cytosolic dsDNA), we mapped their genomic locations and compared these to annotations of full-length LTR retrotransposon subfamilies (as available). We used *DERfinder* (see Star Methods) to identify differentially expressed clusters (DECs) along the genome. Focusing on intergenic DECs (to reduce noise from gene expression), we found a total of 777 intergenic DECs in basal G9acKO cells (725 up-, 52 downregulated; median length 1867 bp [95% CI 1490–2158 bp]), and 174 intergenic DECs in luminal G9acKO cells (160 up-, 14 downregulated; median length 2121 bp [95% CI 1548–2904 bp]) (FDR < 0.05; Table S6). By mining the TEs annotated to upregulated DECs for internal LTR

retrotransposon regions flanked by LTRs (to identify full-length LTR retroviral elements), we found that the MMVL30-INT and IAP-D-INT internal regions flanked by LTRs were repeatedly derepressed in G9acKO basal and luminal cells versus WT (Figures 6G and S6I). MMVL30-INT is the internal region of the VL30 retrovirus subfamily, which accumulates as cytosolic dsDNA when derepressed and triggers an anti-viral immune response (Herquel et al., 2013). Notably, 18 of 86 known full-length VL30 genomic copies (Markopoulos et al., 2016) were derepressed in G9acKO mammary epithelium versus WT, highlighting the VL30 subfamily of retroviruses as a potential source of cytosolic dsDNA in G9acKO mammary epithelia. Full-length retroviral activation can also lead to genome re-integration, which triggers DNA damage in the host cell (Johnson, 2019). Indeed, we observed that the DNA damage marker γH2Ax was increased in G9acKO mammary epithelia versus WT (Figures 6H-6J, S6J, and S6K).

Notably, upregulated DECs were closer to differentially opened chromatin regions in G9acKO than in WT epithelial cells (Figures 6K and S6L). These results suggest that DEC (and thus TE) derepression in G9acKO cells is regulated by *de novo* chromatin opening, which could cause the altered cytokine and inflammatory milieu in G9acKO fat pads.

Aim2 upregulation depends on TE derepression and is responsible for developmental delay of G9a-null mammary epithelium

LINE, RLTR44A, and MERVK10D3 TE upregulation was maintained *in vitro* in G9acKO mammary epithelial cells as compared to WT controls, but was abrogated following treatment with lamivudine (a nucleoside reverse-transcriptase inhibitor)(Figure S7A), concomitant with reduced cytosolic dsDNA levels (Figure S7B). Lamivudine treatment also abolished the upregulation of *Aim2* and *Casp1* (Figure S7C), in G9acKO as compared to WT mammary epithelial cells. Thus, TE derepression and cytosolic dsDNA accumulation are functionally linked to the activation of the Aim2-dependent inflammasome in G9acKO mammary epithelial cells. We therefore next asked whether disrupting the pyroptotic pathway *in vivo*

could ameliorate the G9acKO mammary developmental defects. We injected shCTR or shAim2 lentiviral particles intraductally (through the nipple) in early-pubertal (4-week-old) WT or G9acKO mice and assessed mammary outgrowth after 2 weeks (in pubertal mice). shAim2 knockdown (Figure S7D) rescued mammary fat pad invasion defects (but not the distended TEB phenotype) of G9acKO mice and had no effect on their WT littermates (Figures 7A and 7B). These data indicated that the dsDNA sensor Aim2 (and its related activation of the pyroptotic pathway) was responsible (at least in part) for the G9acKO ductal invasion defects.

Adaptive immune response underlying G9acKO mammary stem cell defects

To test whether an autoimmune-like environment in G9acKO mammary fat pads affected the cleared mammary fat pad repopulating potential of G9acKO mammary epithelial cells, we transplanted 5000 FACS-isolated, YFP-positive WT or G9acKO mammary epithelial cells into immunocompromised athymic nude mice. In contrast to transplants into immunocompetent WT recipients (Figures 3H-3J), more than half of immunocompromised fat pads supported the outgrowth of organized G9acKO mammary tree structures, with two fat pads even being fully populated (Figures 7C and 7D). Nonetheless, G9acKO luminal and basal stem cell potential was severely impaired *in vitro* (see Figures 3D-3G and S3C-S3H). Therefore, we concluded that reduced G9acKO stem cell potential is the result of both cell autonomous and cell non-autonomous affects in G9acKO mammary fat pads.

Normal mammary morphogenesis is disrupted by changes in the stromal mammary milieu in the absence of G9a

To determine the influence of the altered cytokine milieu and collagen content in G9acKO fat pads on morphology, we transplanted WT mammary epithelial cells into non-cleared fat pads of WT or G9acKO recipients (note that non-cleared recipient fat pads allowed continuous niche conditioning by the endogenous mammary epithelium). Indeed, while WT cells transplanted into WT recipients appeared normal, WT cells transplanted into G9acKO fat

pads displayed distended TEBs, thickened ducts, and increased lateral branching, thereby partially recapitulating the G9acKO phenotype (Figures 7E-7H). Therefore, we conclude that the defects observed in the development of the mammary epithelium with G9a loss consisted of cell autonomous components (i.e. decreased stem cell potential; see Figure 3) as well as cell non-autonomous components (i.e., altered paracrine signaling).

DISCUSSION

Whether chromatin accessibility in non-coding regions plays a role in determining cell fate and tissue development is not known. We now show that the histone methyltransferase G9a is indispensable for mammary gland function and organization of the stem cell pool, yet is not required for mammary epithelial cell fate determination. We provide evidence for both cell autonomous (i.e. decreased stem cell potential) and cell non-autonomous (i.e. paracrine signaling, extracellular matrix deposition, and immune cell recruitment) aspects of the observed defects. We suggest that these phenotypes are due to the derepression of specific transposable elements that, once activated, trigger an antiviral immune response. This altered immune environment even precludes the outgrowth of WT mammary cells transplanted into the mammary fat pads of G9acKO mice.

We previously reported a substantial decrease in stem cell potential following G9a loss in epidermal stem cells (Avgustinova et al., 2018). Intriguingly, epidermal stem cells devoid of G9a also upregulated the expression of genes involved in an antiviral innate immune response, such as *Aim2*. However, in stark contrast to the severely aberrant G9acKO mammary glands reported here, epidermal development, morphology, and maintenance were largely unscathed, and gene expression was not majorly altered, in mice lacking G9a in epidermal cells. This phenotypic difference between the epidermis and mammary gland in the absence of G9a might be explained by the ability of the epidermis (as a consequence of its natural barrier function) to withstand a response to pathogens (e.g., viruses) without altering its development or homeostasis. For instance, hair follicle stem cells exposed to an

inflammatory stress retain an epigenetic memory of pyroptotic genes, such as *Aim2* and *Casp1*, that allows them to respond more efficiently to exposure to subsequent rounds of inflammatory stress (Naik et al., 2017).

G9acKO mammary epithelia are disorganized and multilayered (Figures 1J-1N and S1D-S1G), reminiscent of hyperproliferative lesions. The formation of such hyperplastic lesions is often associated with an imbalance between cell proliferation and cell death in favor of increased proliferation. However, G9acKO luminal and basal cells are less proliferative than their WT counterparts (Figures 4C-4D and S4G-S4H), arguing against neoplastic transformation causing the G9acKO mammary duct polarity defects. Consistently, G9acKO mice do not develop spontaneous mammary tumours even during aging (up to 95 weeks) (data not shown). Although we cannot exclude that the altered inflammatory environment in G9acKO mice precludes development of *bona fide* malignant lesions from a pre-neoplastic state, we propose rather that the aberrant G9acKO mammary tissue polarity and multilayering stem from increased tissue stiffness, which hinders TEB outgrowth and is evidenced by increased collagen deposition in G9acKO as compared to WT mammary fat pads (Figures 2B-2D and S2C-S2E). Consistently, WT mammary ducts that develop following transplantation into G9acKO hosts display larger TEBs than those transplanted into WT recipients (Figures 7F and 7H). We conclude that the G9acKO stromal milieu is (in part) responsible for the morphological defects observed in G9acKO mammary glands.

Surprisingly, in light of the severe morphological defects, the cell fate specification of the G9acKO mammary epithelia remained largely intact; the only changes were in the decreased levels of proliferative basal and LP populations (Figures 4A and S4; Tables S1 and S2). The slight reduction of ERα-positive luminal cells in G9acKO mammary ducts (Figures S2F and S2G) could be partially responsible (through a decrease in pro-proliferative signals) for the reduced proliferative LP cells; however, this is not likely to be the only mechanism, as ERα-

negative G9acKO basal and epidermal cells also displayed reduced proliferative capacity (Figures 4D, S4G; see (Avgustinova et al., 2018).

Although DNA methylation is considered the principal mechanism of TE repression in higher eukaryotes, DNMT-independent mechanisms have also been reported (Hutnick et al., 2010), including H3K9 methylation and heterochromatin formation (Karimi et al., 2011; Liu et al., 2014; Matsui et al., 2010). H3K9me2 has been implicated in L1 germline repression (Di Giacomo et al., 2014) as well as in ERV repression in embryonic stem cells (Maksakova et al., 2013) and in ovarian cancer epithelial cells (Liu et al., 2018). Mechanistically, TEs (and LTRs in particular) can act as *cis*-regulatory "promotor" elements to drive gene expression (Burns, 2017; Chuong et al., 2017); however, this effect was negligible for LTR derepression in G9acKO mammary epithelial cells (see Table S6). Further, full-length TEs can be detected by the cell as non-self-DNA, triggering an inflammatory response (Chuong et al., 2017; De Cecco et al., 2019). Indeed, in G9acKO mammary epithelial cells, we showed altered cytokine and chemokine gene expression, ECM deposition, and immune cell infiltration, in addition to LTR retrotransposon derepression. Immune cells are active participants in mammary gland development, through secretion of matrix metalloproteases and tissue inhibitors of matrix metalloproteases (Reed and Schwertfeger, 2010; Tan et al., 2014), which affect TEB morphology and branching morphogenesis (Page-McCaw et al., 2007; Wiseman et al., 2003). This is consistent with both the mammary defects in G9acKO mice and the aberrant mammary structures formed by WT mammary epithelial cells in G9acKO mice.

Our results suggest that aberrant activation of TEs prevents the proper development of the mammary gland. This phenotype might be related to derepressed TEs associated with age-related inflammation (Benayoun et al., 2019; De Cecco et al., 2019), which may contribute to cancer initiation (Burns, 2017; Saini et al., 2020). Future work is necessary to determine the precise mechanisms that regulate G9a-mediated TE repression during development, and if and how these are deregulated during aging and cancer.

Limitations of the study

G9a represses ERVL elements in embryonic stem cells (ESCs) (Maksakova et al., 2013), ERVL and ERVL-MaLR elements in mouse epiblasts (Zylicz et al., 2015), and L1 elements in spermatogonia (Di Giacomo et al., 2014). While we observed that the ERVL, ERVL-MaLR and L1 TE families were significantly enriched in differentially open chromatin regions in G9acKO basal cells (Figure 6D), we observed no differential RNA expression of these TE families (Figures 6F and S6G). Thus, we suggest that G9a can repress the ERVL/ERVL-MaLR/L1 TEs, but that continued transcriptional repression in homeostatic mammary epithelial cells is maintained by additional repressive mechanisms (e.g. DNA and/or histone methylation). Elucidating the underlying molecular intricacies of these TE repressive mechanisms in addition to G9a activity (and cooperation with chromatin remodeling complexes and transcriptional repressors) requires future studies.

ACKNOWLEDGEMENTS

S.A.B. was supported by the European Research Council (ERC), the Government of Cataluña (SGR grant). C.L., by the European Union's H2020 R&I programme – Marie Skłodowska-Curie GA No. 754510; M.D., by the Spanish Association for Cancer Research (AECC); U.U., by a BIST PhD fellowship; H.H., by the SAF2017-89109-P; AEI/FEDER, UE; and J.M.V, by the Max Planck Society, the Deutsche Forschungsgemeinschaft (DFG) Priority Programme SPP2202 Spatial Genome Architecture in Development and Disease (Project Ref VA 1456/1), the European Union's H2020–Marie Sklodowska-Curie GA No 643062 - ZENCODE-ITN, and the Medical Research Council, UK. The IRB Barcelona is a Severo Ochoa Center of Excellence (MINECO award SEV-2015-0505). We thank V.A. Raker for manuscript editing.

AUTHOR CONTRIBUTIONS

A.A. and S.A.B. designed the study; A.A. carried out most experiments; M.P.G, M.D., A.C. and U.U-U assisted with experiments; N.P. performed the immunohistochemical analyses; A.V.K. assisted with mammary transplants; C.L. and Q.R. performed computational analyses; D.M. and H.H. generated scRNA-seq data and assisted with analyses; J.M.V directed TE computational analysis; A.A. and S.A.B. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. G9a Loss Leads to Delayed Mammary Gland Development and Mammary Lineage Intermingling

(A) Schematic of G9acKO mouse line generation. *Ehmt2*, gene encoding G9a. FP, fluorescent protein, of either eYFP or tdTomato, as indicated. (B, C) Immunohistochemistry (B) and quantification (C) of H3K9me2 in mammary fat pads from 6-week-old (pubertal) WT or G9acKO mice. n = 4 mice per group. In (B), nuclei were counterstained with haematoxylin. The dashed line delineates epithelium/stroma boundary. Scale bar, 25 μm. In (C), data shown are mean ± SEM of independent animals. Two-sided Student's *t*-test. (D) Carmine-alum staining of mammary wholemounts from WT or G9acKO mice (4-, 6-, or 12-weeks old) (stitched images, see Star Methods). Dashed line indicates mammary tree invading front. LN, lymph node. Images are representative. 4-week-old mice: n = 3; 6-week-old mice: n = 8 (WT), n = 6 (G9acKO); adult mice: n = 3 (WT), n = 6 (G9acKO). Scale bar, 2 mm. (E) Magnification of areas highlighted in (D). Asterisks indicate distended terminal end buds of G9acKO mammary ducts. Scale bar, 0.5 mm. (F–I) Quantification of mammary epithelium in pubertal or adult mice (WT or G9acKO) for area in mammary fat pad filled by epithelium (F), fraction of branch versus tip segments (G), median lumen area (H), and mean branch and tip segment length (I). Data are mean ± SEM. In (F), each data point represents an independent animal. Pubertal mice: n = 8 (WT), n = 6 (G9acKO); adult mice: n = 3 (WT), n = 6 (G9acKO)

mice. Two-sided Student's *t*-test. In (G), pubertal mice: n = 7 (WT), n = 6 (G9acKO); adult mice: n = 3. Two-way ANOVA with Dunnett multiple comparisons test. In (H), each data point represents an independent animal. n = 4 mice for each group. Two-sided Student's *t*-test. In (I), each data point represents an ROI (with 3 ROIs per animal). Pubertal mice: n = 7 (WT), n = 6 (G9acKO); adult mice: n = 3. Two-way ANOVA with Dunnett multiple comparisons test. (J) Immunofluorescence for αSMA (red) and cytokeratin 8 (green) on mammary fat pads from pubertal WT or G9acKO mice. n = 4 mice for each group. Nuclei were counterstained with DAPI (blue). Two examples are shown for G9acKO to capture phenotypic heterogeneity. Dashed line delineates epithelium/stroma boundary. Scale bar, 25 μm. (K–N) Endogenous Rosa26tdTomato signal (red; K, M) or immunofluorescence staining (L, N) for CK8 (green) and αSMA (gray) on mammary wholemounts from pubertal WT or G9acKO mice. Nuclei were counterstained with DAPI (blue). G9a loss leads to cellular intermingling and basal cell multilayering in both ductal segments (K, L) and terminal end buds (M, N). Scale bar, 50 μm (K, M) or 25 μm (L, N). See also Figure S1.

Figure 2. G9a loss disrupts tissue polarity and function. (A–C) Immunofluorescence studies of mammary fat pads from pubertal WT or G9acKO mice, staining: (A) ZO-1 (green); nuclei are counterstained with DAPI (blue). Arrowheads indicate aberrant lumina. Two examples are shown for G9acKO to capture phenotypic heterogeneity; (B) collagen IV; or (C) Masson´s trichome staining. For (B, C), nuclei were counterstained with haematoxylin. n = 4 mice for each group. Scale bar, 25 μm. (D) Quantification of thickness of Masson's trichome layer around mammary ducts in mammary fat pads from pubertal WT or G9acKO mice. Each data point represents an individual duct. WT, n = 56 ducts from 4 mice; G9acKO, n = 47 ducts from 5 mice. Two-sided Student's *t*-test. (E–H) Immunofluorescence analyses of mammary fat pads at lactation day 1 from WT or G9acKO mice, staining for: (E) αSMA (red) and cytokeratin 8 (green), with nuclei counterstained with DAPI (blue). A lack of alveolar differentiation in G9acKO mammary epithelia is evident; (F) H3K9me2; the dashed

line delineates the epithelium/stroma boundary; (G) collagen IV; and (H) Masson´s trichome staining. For (F–H), nuclei were counterstained with haematoxylin. n = 3 mice per group. Scale bar, 25 μm. See also Figure S2.

Figure 3. Mammary stem cell potential is impaired following G9a loss. (A–C) Representative FACS plots showing gating of luminal and basal cells (A), and quantification of YFP-positive mammary epithelial cells (B) or luminal/basal cells (C), of pubertal (6-week-old) or adult (12-week-old) WT or G9acKO mice, as indicated. Data are mean ± SEM. Pubertal mice, n = 5 (WT), n = 3 (G9acKO); adult mice, n = 3 (WT), n = 8 (G9acKO). Two-sided Student's *t*-test (B) or two-way ANOVA with Dunnett multiple comparisons test (C). (D–G) 2D (D, E) or 3D (F, G) colony formation assay of adult WT or G9acKO basal cells, showing representative images of three independent experiments (D, F) and quantification of colony area (E, G). Scale bar, 5 mm (D) or 50 μm (F). In (E, G), data are mean ± SEM of individual colonies pooled from triplicate wells. Two-sided Student's *t*-test. (H) Pie charts representing percentage of mammary fat pad area filled by each cleared fat pad mammary transplant (5000 mammary epithelial cells). Transplant recipients were WT mice. Transplant donor genotypes are indicated. (I) Representative images of cleared fat pad transplants of the indicated genotypes (stitched images, see Star Methods). Scale bar, 2 mm. (J) Immunofluorescence for endogenous fluorescent protein (YFP or tdTomato, depending on donor) to visualise cleared fat pad mammary transplant morphology. Scale bar, 50 μm. Color swap experiments were performed with equivalent results. See also Figure S3.

Figure 4. Correct lineage specification in G9acKO mammary epithelia. (A) UMAP plots of the transcriptomes of total epithelial cells of mammary glands from adult WT or G9acKO mice, generated using 10× genomics, separated by genotype. n = 2 mice per group. Major clusters (with the contribution of WT:G9acKO cells indicated) were: basal (cluster 2; 3226:1760 cells); luminal progenitors (LP, cluster 0; 3338:3648 cells), mature luminal (ML, cluster 1; 1660: 3338 cells), luminal intermediate (cluster 3; 464:651 cells), and proliferative

LPs (cluster 4; 251:23 cells). (B) Expression of known lineage-specific signatures was tested in the epithelial cell clusters separated by genotype (see Star Methods), to confirm their identities. Color indicates expression score of each indicated signature. (C) Immunofluorescence for cytokeratin 8 (green), Ki67 (red), and cytokeratin 8 (gray) on mammary fat pads from adult WT or G9acKO mice. Nuclei were counterstained with DAPI (blue). n = 3 (WT), n = 5 (G9acKO) mice. Arrowheads indicate proliferative luminal cells; arrows indicate proliferative basal cells. Scale bar, 25 μm. For visualisation, the brightness of the red channel was modified uniformly across all images. Quantification was performed on raw data. (D) Quantification of Ki67-positive (proliferative) cells in (C). Cells in which Ki67 colocalizes with cytokeratin 14 (proliferative basal cells) or cytokeratin 8 (proliferative luminal cells). Data are mean ± SEM of individual animals. For pubertal mice, n = 3 (WT), n = 4 (G9acKO); for adult mice, n = 3 (WT), n = 5 (G9acKO). Two-way ANOVA with Dunnett multiple comparisons test. (E) MA plots highlighting DEGs between WT and G9acKO cells by cluster. Genes with FDR < 0.05 and FC >1.5 are indicated for basal, LP, and ML cells (see Table S2). Highly differentially expressed cytokines are highlighted. (F) Expression of differentially expressed cytokines to determine their patterning amongst the cell clusters. Colour indicates expression level. (G) Quantification of (E). For statistics, see DEG analysis in (E) and Table S2. See also Figure S4.

Figure 5. Expression profiling reveals activation of pyroptosis in G9acKO mammary epithelium (A) Heatmap showing unsupervised hierarchical clustering of DEGs of FACS-isolated, WT or G9acKO basal cells from adult mice. (B) GO pathways deregulated in G9acKO basal cells from adult mice. (C) MA plot showing DEGs in G9acKO versus WT basal cells from adult mice. Pyroptosis-associated genes are highlighted; significant DEGs of the pyroptosis pathway are indicated in bold. (D) Immunohistochemistry for caspase-1 on mammary fat pads from adult WT or G9acKO mice. Nuclei were counterstained with haematoxylin. Scale bar, 25 μm. n = 5 (WT), n = 4 (G9acKO) mice. Representative images are shown. (E) Representative FACS plot of fluorochrome-labelled inhibitors of caspases

(FLICA) caspase 1 activity assay in WT or G9acKO basal cells. n = 4 (WT) and n = 3 (G9acKO) mice. (F) Quantification of (E). Data are mean ± SEM for individual animals. Two-sided Student's *t*-test. (G–J) Analyses of WT or G9acKO mammary fat pads from adult mice, showing (G) immunofluorescence for dsDNA (red), with nuclei counterstained with DAPI (blue). Representative images are shown. n = 5 (WT), n = 4 (G9acKO) mice; (H) quantification of infiltrating inflammatory monocytes (n = 7) or neutrophils (n = 6 [WT], n = 7 [G9acKO]; (I) immunohistochemistry for CD3, with nuclei counterstained with haematoxylin. Arrowheads indicate CD3-positive lymphocytes within the mammary epithelium. Representative images are shown. n = 4 (WT), n = 7 (G9acKO) mice; and (J) quantification of (I). For (G, I), scale bar, 25 μm. For (H, J), two-sided Student's *t*-test; data are mean ± SEM of individual animals. See also Figure S5.

Figure 6. Increased chromatin opening in G9acKO mammary epithelium correlates with derepression of LTRs (A–G) Analysis of FACS-isolated WT or G9acKO basal cells, showing: (A) correlation heatmap of regions of differential chromatin opening; (B) quantification of length of differentially open chromatin regions; (C) annotation of location of differentially open chromatin regions; (D) dot plot of enrichment of TEs in differential ATAC peaks; (E) heatmap showing ATAC-seq read coverage of individual copies of RLTR10D2 subfamily of LTR retrotransposons. Each row represents an individual RLTR10D2 locus (+/– 10 kb). RLTR10D2 copies that are in open chromatin (cluster 1) or closed chromatin (cluster 2) are indicated. The mappability scores of the individual loci demonstrate that differences in ATAC-seq read coverage are not due to regional mappability; (F) MA plot showing differentially expressed TEs; significant ones (ajd. p-value < 0.01) are highlighted in green and subclassified as LTR or non-LTR; and (G) example of increased chromatin accessibility leading to full-length IAP derepression that results in pervasive transcription. (H) Immunohistochemistry for γH2AX on mammary fat pads from adult (WT or G9acKO) mice. Nuclei were counterstained with haematoxylin. Scale bar, 25 μm. n = 4 mice per group. Representative images are shown. (I) Quantification of (E). Each data point indicates an ROI;

3 ROIs per mouse were used. (J) Quantification of γH2AX staining of WT or G9acKO mammary epithelial cells by flow cytometry. Each data point represents an individual animal. n = 4 per group. (K) Violin plot of the distance between differentially expressed clusters (DECs) and the closest differentially open chromatin regions in basal cells, highlighting an association between G9acKO DEC upregulation and chromatin opening. In (B, I, and J), two-sided Student's *t*-test; in (C), Fisher´s exact test with FDR correction for multiple testing. See also Figure S6.

Figure 7. Paracrine deregulation contributes to G9acKO morphological defects. (A) Carmine-alum staining of mammary wholemounts from pubertal WT or G9acKO mice, injected intraductally with shCTR or shAim2 lentiviral particles (stitched images, see Star Methods). Dashed line indicates mammary tree invading front. LN, lymph node. Scale bar, 2 mm. (B) Quantification of (A). Data are mean ± SEM of individual 4th mammary fat pads. n = 4 (WT), n = 3 (G9aKO shCTR), and n = 5 (G9acKO shAim2). Two-way ANOVA with Dunnett multiple comparisons test. (C, D) Representative images (stitched images, see Star Methods) (D) and quantification (of percentages of filled area) (C) of cleared mammary fat pad transplants of WT or G9acKO mammary epithelial cells into nude mice. Transplant donor genotypes are indicated. Fisher´s exact test. Scale bar, 2 mm. (E, F) Representative immunofluorescence images of endogenous tdTomato fluorescent protein, to visualise fat pad transplant morphology from WT donors. Transplant recipient genotype is indicated. Scale bar, 100 μm. (G) Pie charts representing percentage of area filled by donor mammary epithelium (tdTomato-positive) in each individual, non-cleared fat pad mammary transplant. Fisher´s exact test. (H) Quantification of median terminal end bud (TEB) area of (G). Data are mean ± SEM of individual transplants. n = 3 for WT recipients, n = 6 for G9acKO recipients. Outgrowths (rudimentary organoid-like mammary structures with no TEBs) in two G9acKO recipients were not considered. Two-sided Student's *t*-test. See also Figure S7.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Salvador Aznar Benitah (salvador.aznar-benitah@irbbarcelona.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

RNA-seq (total and single-cell) and ATAC-seq data that support the findings of this study are deposited in the Gene Expression Omnibus (GEO) with accession number GEO: GSE149978. All other data supporting the findings of this study are available from the lead author on reasonable request. The software and algorithms for data analyses used in this study are all well-established from previous work and are referenced throughout the manuscript.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse strains. Animal (*Mus musculus*) experiments were approved by the Ethical Committee for Animal Experimentation (CEEA) of the Scientific Park of Barcelona (PCB) and the Government of Catalunya and fully complied with their ethical regulations. All immunocompetent mice were maintained in a C57/Bl6 genetic background. Rosa26YFP, Rosa26tdTomato, K14-Cre and athymic nude mice were obtained from Jackson Laboratories. $G9a^{fl/fl}$ mice were kindly supplied by Dr. Yoichi Shinkai (Saitama University). Only female mice were used for studies. The developmental stage of the animals used is indicated in the figures and figure legends. Animals were housed in specific pathogen free (SPF) conditions.

Allocation to WT or G9acKO groups was determined by the animal´s genotype. Littermate controls (kept in the same cage) were used for all experiments.

Cell lines. NIH-3T3 feeder cells and HEK293T cells (both from ATCC) were maintained in high-glucose DMEM (41965039, Life Technologies) containing 10% FBS (10270106, Life Technologies), 2 mM L-glutamine (25030024, Life Technologies), and 100 μg/ml penicillin and streptomycin (15140122, Life Technologies) at 37°C and 5% CO_2. Prior to addition of mammary epithelial cells (also see "Clonogenic assays"), NIH-3T3 feeder cells were treated with 4 μg/mL mitomycin-C (M4287, Sigma-Aldrich) at 37°C for 2 h to arrest proliferation. Luminal cells were cultured using the EpiCult-B Mouse Medium kit (05610, StemCell Technologies) according to the manufacturer's instructions. Basal cells were maintained in FAD medium (three-parts DMEM [41965039], one-part Ham´s F12 [21765029]; both from Life Technologies) containing 10% FBS (10270106, Life Technologies), 2 mM L-glutamine (25030024, Life Technologies), 1.8×10^{-4} M adenine (A3159, Sigma-Aldrich), 0.5 g/ml hydrocortisone (386698, Sigma-Aldrich), 5 μg/ml insulin, 1×10^{-10} M cholera enterotoxin (100, Gentaur Molecular Products BVBA), 10 ng/ml EGF (315-09, Peprotech), and 10 μM ROCK-inhibitor Y-27632 (Y0503, Sigma-Aldrich). Primary cell cultures were maintained in normoxia conditions (37°C, 10% CO_2, and 3% O_2).

METHOD DETAILS

Clonogenic assays. For 2D colony formation assays, 500 luminal or basal cells were sorted directly into 6-well plates containing cell-type appropriate medium (see "Cell culture") and 1×10^5 mitomycin-C–treated NIH-3T3 feeder cells. After 7–10 days, colonies were fixed in ice-cold acetone:methanol (1:1 mixture; 34859 and 322415 respectively, both from Sigma-Aldrich) for 30 s and then air-dried. Colonies were stained in Wright´s Giemsa (WG16, Sigma-Aldrich) for 3 min, washed, and air-dried. For 3D colony formation assays, FACS-sorted cells were resuspended in 100% Matrigel (354234, BD Biosciences), and gels were

allowed to set for 30 min at 37°C before cell-appropriate culture medium was added. Spheroids were cultured for 10–14 days before fixation in 4% paraformaldehyde. Quantification was carried out using Fiji software (version 2.0.0-rc-68) (Schindelin et al., 2012).

Single-cell preparation. Mammary epithelial cells were isolated using collagenase/hyaluronidase (07912, StemCell Technologies) digestion as detailed in the manufacturer's protocol. Briefly, mammary fat pads of 6- or 12-week-old female mice (as indicated in the figure legends) were digested at 37°C for 6 h under moderate agitation (750 rpm); red blood cells were then lysed using RBC Lysis Buffer (00-4333-57, eBioscience), and mammary epithelial fragments were further digested in 0.25% pre-warmed trypsin/EDTA (25200056, Life Technologies) for 2 min. Trypsin activity was neutralized with 2% FBS in HBSS (14175053, Life Technologies), followed by a final digest in pre-warmed 5 mg/ml Dispase II (D4693, Sigma-Aldrich) containing 1 mg/ml DNase I (DN25, Sigma-Aldrich) for 2 min. Cell suspensions were filtered through 40-μm cell strainers (SPL Life Sciences) and stained for flow cytometry.

For mammary-infiltrating immune-cell analysis, a shorter dissociation protocol was followed to maximize immune cell viability, as described previously (Unsworth et al., 2016). Briefly, 4^{th} inguinal mammary fat pads were collected, and lymph nodes were carefully removed; 4^{th} inguinal mammary fat pads were then mechanically dissociated using a McIlwain Tissue Chopper (The Mickle Laboratory Engineering Co. LTD), and minced tissue was digested under gentle agitation (300 rpm) using 1.5 mg/ml collagenase I (C0130, Sigma-Aldrich) and 40 U/ml DNase I (DN25, Sigma-Aldrich) in calcium-free EMEM (BE06-174G, Lonza) for 60 min at 37°C. Cells were pelleted, and red blood cells were lysed using RBC Lysis Buffer, washed twice in PBS (H3BE17-516F, Lonza), filtered through 40 μm cell strainers (SPL Life Sciences) and stained for flow cytometry.

Flow cytometry. For flow cytometry analysis and sorting of mammary epithelial populations, single-cell suspensions from individual animals were re-suspended to 1×10^7 cells/ml in PBS and labelled with EpCAM-PE, CD29-APC, and lineage-BV605 (CD31; CD45; TER119). Cells were re-suspended in 2 μg/ml DAPI (32670, Sigma-Aldrich) to exclude dead cells. Due to the presence of the ROSA26-YFP cassette in mice, mammary epithelial cells were positive for YFP. Luminal cells (EpCAMhigh/CD29low/lineage–) and basal cells (EpCAMlow/CD29high/lineage–) were gated from YFP+ live cells and FACS-sorted using a BD FACSAria Fusion flow cytometer (BD Biosciences). For mammary transplants, mice containing ROSA26-tdTomato instead of ROSA26-YFP cassette were used, and total tdTomato+ mammary epithelium was sorted.

For immune analysis, an additional Fc-blocking step was performed prior to antibody incubation. The innate immune compartment was assayed and gated as previously described (Yu et al., 2016). The adaptive immune compartment was characterized following well-established protocols (Unsworth et al., 2016). Detailed information on all antibodies used is available in the Key Resources Table.

Mammary transplants. Mammary transplants were performed as previously described (Shackleton et al., 2006). FACS-sorted WT or G9acKO mammary epithelial cells (5000 Lin– cells per transplant) were resuspended in PBS containing 50% growth factor–reduced Matrigel (356230, BD Biosciences) and 0.04% Trypan blue (T8154, Sigma-Aldrich) and injected in a 10 μl volume into the 4th inguinal fat pad of 21-day-old female WT, G9acKO, or athymic nude mice. Prior clearing of endogenous mammary epithelium is indicated in the relevant figure legends. Recipient glands were removed for evaluation at 8 weeks post-transplant. Wholemounts of mammary outgrowths were stained as indicated below. An outgrowth was defined as an epithelial structure comprising ducts arising from a central point, with lobules and/or terminal end buds.

Lentiviral production. Lentiviral particles were produced in HEK293T cells transfected with shCTR (pLKO.1, as a TRC control; a gift from David Root (Addgene plasmid # 10879)), or shAim2 (TRCN0000096107, Sigma-Aldrich). Viral supernatants were concentrated using Lenti-X Concentrator (631231, Takara Bio) and viral particles were quantified using the Lenti-X p24 Rapid Titer Kit (632200, Takara Bio) according to the manufacturer´s instructions.

Intraductal lentiviral injections. Intraductal injections of lentiviral particles were performed as previously described (Krause et al., 2013). Briefly, 4-week-old WT or G9acKO animals were anesthetized using isofluorane, and the hair in the nipple area was removed using over-the-counter hair removal cream. shCTR or shAim2 lentiviral particles (3×10^6) in 3 μl PBS were injected into the nipple of the 4th inguinal fat pad using a 34 G blunt end metal hub needle (7803-05, Hamilton Company) affixed to a 50 μl Hamilton syringe (7637-01, Hamilton Company). After 2 weeks, the 4th inguinal fat pads were dissected and processed for downstream analyses.

Caspase activity assay. Caspase 1 activity was assessed using the fluorochrome-labeled inhibitors of caspases (FLICA) 660 Caspase-1 Assay Kit (9122, ImmunoChemistry) according to the manufacturer´s instructions. Briefly, 1×10^6 mammary cells were incubated with 1× FLICA 660 reagent at 37°C for 1 h. Cells were washed twice, incubated with primary antibodies (EpCAM-PE, CD31-BV605, CD45-BV605, or TER-119-BV605), washed again, and re-suspended in 2 μg/mL DAPI (32670, Sigma-Aldrich) to be able to exclude dead cells. FLICA 660 levels within basal cells (EpCAMlow/Lin–) and luminal cells (EpCAMhigh/Lin–) were analysed on a BD FACSAria Fusion flow cytometer (BD Biosciences).

Mammary wholemounts. For Carmine Alum-stained mammary wholemounts, 4th inguinal mammary fat pads were dissected and fixed in Carnoy's fixative (60% absolute ethanol [1.00983.2500, VWR], 30% chloroform [C2432, Sigma-Aldrich], and 10% glacial acetic acid [1018302500, VWR]) at room temperature overnight, rehydrated in sequential washes with decreasing concentrations of ethanol, and stained overnight in Carmine Alum Stain (0.2% carmine [C1022, Sigma-Aldrich] and 0.5% aluminum potassium sulfate [A7167, Sigma-Aldrich]). Wholemounts were once again dehydrated in ethanol and cleared in xylene (534056, Sigma-Aldrich) before mounting in CV Mount (14046430011, Leica Microsystems). Images were acquired on an Olympus MVX10 Macroscope and stitched together to create images covering the entire wholemount using MosaicJ plugin in Fiji. Quantification of branch and tip lengths was done using the ruler tool in Fiji. A branch was defined as the length of duct between two branch points. A tip was the length of branch from the most distal branch point to the tip of the mammary epithelium.

For fluorescent wholemounts, 4th inguinal fat pads were dissected, fixed in 4% formaldehyde (15710, Aname) at room temperature for 2 h, and blocked in PBS (H3BE17-516F, Lonza) containing 10% BSA (A7906, Sigma-Aldrich), 5% FBS (10270106, Life Technologies), and 1% Triton X (T8787, Sigma-Aldrich) at room temperature for 3 h. Wholemounts were incubated with primary antibodies diluted in blocking solution at room temperature overnight, washed three times in PBS, and incubated with secondary antibodies and DAPI diluted in blocking solution at room temperature for 3 h. Mammary fat pads were washed and mounted using Vectashield (H-1000, Vector laboratories) and imaged using a Leica TCS SP5 confocal microscope.

Immunostaining. Tissues were fixed in 10% formalin (HT501128, Sigma-Aldrich) for 3 h at room temperature and embedded in paraffin; 4-µm sections were stained. For immunofluorescence, sections were blocked in 10% donkey serum (D9663, Life Technologies) for 1 h at room temperature and stained with primary antibodies overnight at 4°C. Sections were washed three times in PBS before incubation with secondary antibodies

for 1 h at room temperature. Slides were washed three times in PBS containing 2 µg/mL DAPI (D9542, Sigma-Aldrich) and mounted in Vectashield (H-1000, Vector laboratories). Immunofluorescence images were acquired using a Leica TCS SP5 or a Zeiss ELYRA PS.1 confocal microscope. On occasion, the brightness of individual channels was adjusted in Fiji to aid visualization. Such modifications are indicated in the figure legends and were performed uniformly across all images presented for comparison. Raw data were used for quantification.

For immunohistochemistry, sections were stained with primary antibodies at room temperature for 2 h, washed twice, and then incubated with secondary antibodies, followed by a 5-min incubation with DAB (K346711-2, Dako). Sections were counterstained with hematoxylin (CS70030-2, Dako) and mounted using Toluene-Free Mounting Medium (CS70530-2, Dako).

Stained immunohistochemistry sections were scanned using a high-resolution NanoZoomer 2.0 HT (Hamamatsu); images were quantified using the positive cell detection tool or the ruler tool of QuPath software (version 0.1.2) (Bankhead et al., 2017).

ATAC sequencing. Library preparation for ATAC (assay for transposase-accessible chromatin) sequencing used 5×10^4 luminal or basal cells (FACS-sorted) from four WT or three G9caKO mice, as previously described (Buenrostro et al., 2013). Samples were sequenced on a HiSeq2500 sequencer (Illumina) using V4 chemistry, generating 50-bp paired-end reads. After adapter-cleaning and quality correction using *Trimmomatic* (version 0.36) (Bolger et al., 2014), paired-end reads were aligned to the *mm10* genome (UCSC) using BowTie 2 Aligner (version 2.3.3.1) (Langmead and Salzberg, 2012). Duplicate reads were removed using SAMtools (version 1.3.1) (Li et al., 2009). Read alignment was offset as previously described (Buenrostro et al., 2013). To normalize for sequencing depth, the number of reads of all samples was downsampled to match the number of reads in the sample with the lowest coverage. Peaks were called using *MACS2* (version 2.0.10) (Feng et al., 2012) and an FDR < 0.05. Differential peaks were determined using the DiffBind package (version

2.4.8) (Stark and Brown, 2011) in R (version 3.5.2) and an FDR < 0.05. The distribution of the differential regions within genomic features was conducted with HOMER (version 2015-03-22) (Heinz et al., 2010).

Whole-genome expression analysis. For whole-genome expression profiling of mammary epithelial cells from 12 week old animals using total RNA sequencing, $2–5 \times 10^5$ luminal or basal cells were sorted and lysed in TRIzol (15596018, Thermo Scientific). RNA was extracted following the manufacturer´s recommendations and treated with TURBO DNA-free Kit (AM1907, Thermo Fisher) prior to library preparation using TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat Kit (RS-122-2201, Illumina). Samples were sequenced on a HiSeq2500 sequencer (Illumina) using V4 chemistry, generating 125-bp single reads. After adapter-cleaning and quality correction using Trimmomatic (version 0.36), paired-end reads were aligned to the mm10 genome (UCSC) using TopHat Aligner (version 2.1.1) (Kim et al., 2013). Only uniquely aligning reads were used. For differentially expressed genes (DEGs), read count was performed using featureCounts (version 1.6.0) (Liao et al., 2014) and DEGs were called using DESeq2 (version 1.26.0) (Love et al., 2014), with adjusted p-value < 0.05 and fold-change > 1.5.

Gene ontology was performed using DAVID (version 6.7) (Huang da et al., 2009) Genes from pyroptosis pathway were identified using the GO term GO:0070269 (Biological process, pyroptosis).

For whole-genome expression profiling of mammary epithelial cells from 6-week old animals, RNA was isolated and libraries prepared from 1000 cells as previously described (Gonzalez-Roca et al., 2010). cDNA was hybridized to the Clariom S Mouse Array (901570, Thermo Scientific) and scanned on a GeneAtlas Imaging station (Affymetrix). After RMA normalisation and batch correction using the *ComBat* function of the *SVA* package in R, differentially expressed genes were determined using the *limma* package (version 3.32.10) (Ritchie et al., 2015) in R and a FDR < 0.05.

For differentially expressed clusters (DECs), expressed regions along the whole genome were identified using the *DERfinder* package (version 1.20.0) in R (Collado-Torres et al., 2017), using a read cutoff of 5, maxClusterGap = 2000 L, and maxRegionsGap = 0. DECs identified using DESeq2 (FDR < 0.05) were filtered for intergenic regions. DECs were annotated using the annotateRegions function in R. TEs within DECs were considered differentially expressed.

Real-time quantitative PCR (RT-qPCR). RNA was isolated as described above, followed by cDNA synthesis using the RevertAid First Strand cDNA Synthesis Kit (K1621, Thermo Scientific) according to the manufacturer´s instructions. RT-qPCR analysis was performed using the SYBR Select Master Mix (4472918, Life Technologies) according the manufacturer´s instructions. The reference gene *Polr2a* was used for normalisation. Primer sequences are indicated in Table S7.

Single-cell RNA-sequencing (scRNA-seq). Total mammary epithelium (YFP+ cells; see Supplemental Figure 2A) were FACS-isolated from two WT and two G9acKO littermates. Cell concentration and viability were verified by counting with a TC20™ Automated Cell Counter (Bio-Rad Laboratories, S.A). Cells were partitioned into Gel Bead-In-Emulsions in a Chromium Controller (10x Genomics) with a target cell recovery of 5000 cells. cDNA sequencing libraries were prepared using the Single-cell 3′ mRNA kit (V3; 10× Genomics) following the manufacturer's instructions. Briefly, after GEM-RT clean-up, cDNA was amplified for 11 cycles, and cDNA quality control and quantification were performed on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies). cDNA libraries were indexed by PCR using the PN-220103 Chromiumi7 Sample Index Plate. Size distribution and concentration of 3′ cDNA libraries were verified on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies). Finally, cDNA libraries were sequenced on a HiSeq 4000

sequencer (Illumina) to obtain approximately 25,000–50,000 paired-end reads per cell, respectively.

scRNA-seq analysis. Alignment of scRNA.seq analyses was completed utilizing the 10× Genomics Cell Ranger pipeline (version 2.2.0). Each library was aligned to an indexed mm10 genome (USCS) using Cell Ranger Count and processed using the Seurat pipeline (version 3.0.1)(Stuart et al., 2019) in R (version 3.5.2). Only genes expressed in five or more cells were considered for the analysis. Cells were trimmed to those with >200 but <4000 unique genes, <25% mitochondrially aligned reads, and <50% of ribosomal reads. Doublets were removed using DoubletFinder and a doublet formation rate of 3%. Gene expression levels were normalized and scaled using the NormalizeData and ScaleData functions in Seurat.
For each sample, the first 10 principal components (PCs) were selected for clustering. To cluster cells, a K-nearest neighbor (KNN) graph constructed on a Euclidean distance matrix in PCA space was calculated and then converted to a shared nearest neighbor (SNN) graph, in order to find highly interconnected communities of cells (Xu and Su, 2015). Cells were then clustered using the Louvain method to maximize modularity (Waltman and van Eck, 2013). Datasets were anchored and integrated using the Seurat v3 integrative analysis (FindIntegrationAnchors and IntegrateData functions) with a chosen dimensionality of 20. Integrated data were then normalized and scaled as described above. Post anchoring, PCA was performed, and the first 30 PCs were used for UMAP dimensionality reduction, and subsequent clustering using the default Louvain implementation.
The first 20 PCs were then used to construct an SNN matrix using the FindNeighbors function in Seurat with k.param set to 20. We then identified clusters using the FindClusters command with the resolution parameter set to 0.2 (or 0.1 in the case of basal subcluster identification). Marker genes per cluster were calculated using Seurat's FindAllMarkers() function and the "wilcox" test option. The full list of cluster markers is provided in Table S1. To find cluster-specific DEGs between conditions, a GLM-based method for single-cell differential expression analysis (MAST) within Seurat was used with default parameters.

Cluster annotation was done using the AddModuleScore function using signatures genes for basal, luminal progenitor (LP), and mature luminal (ML) cell types obtained from previously published RNA-seq data (Pal et al., 2017).

Transposable element differential expression analysis. Before aligning, bulk RNA-seq sequencing reads were trimmed using BBduk software (Bushnell et al., 2017) with the following parameters: ktrim=r, k=23, mink=11, hdist=1, tbo, tpe. Trimmed reads were mapped using STAR aligner (version STAR_2.5.2b) (Dobin et al., 2013) with the parameters --sjdbOverhang 100, --winAnchorMultimapNmax 200 and --outFilterMultimapNmax 100 to the GRCm38 *Mus musculus* reference genome. Alignment files were sorted and indexed using samtools (version 1.5) (Li et al., 2009) . To count reads aligning to genes and TEs considering uniquely-mapping reads, TEtranscript software (version 2.0.3) (Jin et al., 2015) was used with the parameters --stranded reverse and --mode uniq. Gene annotations were downloaded from Gencode (Release M12, GRCm38.p5) and repeat annotations were downloaded from RepeatMasker (http://www.repeatmasker.org/). Repeats with the same id were merged. Differential expression analysis was performed in R (version 3.4) using the DESeq2 package (version 1.18.1) (Love et al., 2014) . TEs with adjusted p-value ≤ 0.01 were considered significantly differentially expressed. The lfcShrink function from DESeq2 was used for visualization purposes.

Differential ATAC-seq peak enrichment analysis. Enrichment analysis of ATAC-seq peaks around differentially expressed genes and TEs was performed in R (version 3.4). For enrichment around differentially expressed genes, ATAC-seq peaks in WT and G9acKO were overlapped with differentially expressed genes (gene body + 1Kb upstream of TSS) to construct a contingency table, and Fisher's exact test was used to determine significant association. For enrichment on TEs, presence or absence of ATAC-seq peak in WT or G9acKO was overlapped with all TE copies of each TE sub-family and tested for association

with Fisher's exact test. After testing every TE sub-family, p-values were corrected for multiple testing using the p.adjust function with fdr method.

Heatmaps (Figures 4E and S4F) were generated using deepTools (version 3.3.1) (Ramirez et al., 2016) with option scale-regions to align the 5′ end and 3′ end of all the TE copies of the specific TE. Flanking region of 10 Kb upstream and downstream was included binned into bins. ATAC-seq open chromatin signal in the TE body was used to differentiate two clusters with --kmeans 2 option of plotheatmap function from deepTools.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample sizes, statistical tests, definitions of error bars, and p-values are indicated in figure legends. All statistical tests were two-sided, and $p > 0.05$ was considered as not significant (ns). Each experiment was repeated successfully at least twice with similar results. Statistical analyses were performed using Prism 8 (GraphPad) software. Flow cytometry data were analysed using FlowJo 10 (Treestar). Adobe Photoshop CS6 and Adobe Illustrator CS6 were used for figure presentation.

KEY RESOURCES TABLE

SUPPLEMENTAL TABLES

Table S1. Cluster-defining genes identified by single-cell RNA-seq analysis of FACS-isolated mammary epithelial cells from 12 week old mice. Related to Figure 4.

Table S2. Basal subclusters identified by single-cell RNA-seq analysis. Related to Figure 4.

Table S3. Differentially expressed genes of adult WT versus G9acKO FACS-isolated mammary epithelial cells identified in single-cell RNA-seq, by cluster. FDR < 0.05; FC > 1.5. Related to Figure 4.

Table S4. Differentially expressed genes and GO analysis of adult WT versus G9acKO FACS-isolated luminal or basal cells identified in total RNA-seq. Immune response pathways (such as the GO categories "cellular response to interleukin-1" and "cellular response to interferon gamma/beta") and KEGG pathways ("cytokine-cytokine receptor interaction" and "graft-versus-host disease") were over-represented in basal cells. FDR < 0.05; FC > 1.5. Related to Figure 5.

Table S5. Differentially expressed genes between pubertal WT and G9acKO FACS-isolated luminal or basal cells identified in whole genome expression arrays. FDR < 0.05. Related to Figure 5.

Table S6. Differentially expressed intergenic clusters between adult WT and G9acKO luminal or basal cells, identified in total RNA-seq. FDR < 0.05. The closest downstream coding gene and the TEs contained within the DEC are indicated. Related to Figure 6.

Table S7. List of probes and primers. Related to STAR Methods.

REFERENCES

Avgustinova, A., and Benitah, S.A. (2016). Epigenetic control of adult stem cell function. Nat Rev Mol Cell Biol *17*, 643-658.

Avgustinova, A., Symeonidi, A., Castellanos, A., Urdiroz-Urricelqui, U., Sole-Boldo, L., Martin, M., Perez-Rodriguez, I., Prats, N., Lehner, B., Supek, F., *et al.* (2018). Loss of G9a preserves mutation patterns but increases chromatin accessibility, genomic instability and aggressiveness in skin tumours. Nat Cell Biol *20*, 1400-1409.

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., *et al.* (2017). QuPath: Open source software for digital pathology image analysis. Scientific Reports *7*, 16878.

Benayoun, B.A., Pollina, E.A., Singh, P.P., Mahmoudi, S., Harel, I., Casey, K.M., Dulken, B.W., Kundaje, A., and Brunet, A. (2019). Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. Genome Res *29*, 697-709.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Burns, K.H. (2017). Transposable elements in cancer. Nat Rev Cancer *17*, 415-424.

Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. PLOS ONE *12*, e0185056.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet *18*, 71-86.

Collado-Torres, L., Nellore, A., Frazee, A.C., Wilks, C., Love, M.I., Langmead, B., Irizarry, R.A., Leek, J.T., and Jaffe, A.E. (2017). Flexible expressed region analysis for RNA-seq with derfinder. Nucleic Acids Res *45*, e9.

Corrales, L., Woo, S.R., Williams, J.B., McWhirter, S.M., Dubensky, T.W., Jr., and Gajewski, T.F. (2016). Antagonism of the STING Pathway via Activation of the AIM2 Inflammasome by Intracellular DNA. J Immunol *196*, 3191-3198.

De Cecco, M., Ito, T., Petrashen, A.P., Elias, A.E., Skvir, N.J., Criscione, S.W., Caligiana, A., Brocculi, G., Adney, E.M., Boeke, J.D., *et al.* (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature *566*, 73-78.

Di Giacomo, M., Comazzetto, S., Sampath, S.C., Sampath, S.C., and O'Carroll, D. (2014). G9a co-suppresses LINE1 elements in spermatogonia. Epigenetics Chromatin *7*, 24.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., and Werb, Z. (2008). Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. Dev Cell *14*, 570-581.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728-1740.

Fernandes-Alnemri, T., Yu, J.W., Datta, P., Wu, J., and Alnemri, E.S. (2009). AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. Nature *458*, 509-513.

Godde, N.J., Sheridan, J.M., Smith, L.K., Pearson, H.B., Britt, K.L., Galea, R.C., Yates, L.L., Visvader, J.E., and Humbert, P.O. (2014). Scribble modulates the MAPK/Fra1 pathway to disrupt luminal and ductal integrity and suppress tumour formation in the mammary gland. PLoS Genet *10*, e1004323.

Gonzalez-Roca, E., Garcia-Albeniz, X., Rodriguez-Mulero, S., Gomis, R.R., Kornacker, K., and Auer, H. (2010). Accurate expression profiling of very small cell populations. PLoS One *5*, e14418.

Gottardi, C.J., Arpin, M., Fanning, A.S., and Louvard, D. (1996). The junction-associated protein, zonula occludens-1, localizes to the nucleus before the maturation and during the remodeling of cell-cell contacts. Proc Natl Acad Sci USA *93*, 10779-10784.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Herquel, B., Ouararhni, K., Martianov, I., Le Gras, S., Ye, T., Keime, C., Lerouge, T., Jost, B., Cammas, F., Losson, R., *et al.* (2013). Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. Nat Struct Mol Biol *20*, 339-346.

Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E., and Fitzgerald, K.A. (2009). AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. Nature *458*, 514-518.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Hutnick, L.K., Huang, X., Loo, T.C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. J Biol Chem *285*, 21082-21091.

Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics *31*, 3593-3599.

Johnson, W.E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses. Nat Rev Microbiol *17*, 355-370.

Jonkers, J., Meuwissen, R., van der Gulden, H., Peterse, H., van der Valk, M., and Berns, A. (2001). Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. Nat Genet *29*, 418-425.

Jorgensen, I., Rayamajhi, M., and Miao, E.A. (2017). Programmed cell death as a defence against infection. Nat Rev Immunol *17*, 151-164.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., *et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676-687.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Krause, S., Brock, A., and Ingber, D.E. (2013). Intraductal injection for localized drug delivery to the mouse mammary gland. Journal of visualized experiments : JoVE.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

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

Liu, M., Thomas, S.L., DeWitt, A.K., Zhou, W., Madaj, Z.B., Ohtani, H., Baylin, S.B., Liang, G., and Jones, P.A. (2018). Dual Inhibition of DNA and Histone Methyltransferases Increases Viral Mimicry in Ovarian Cancer Cells. Cancer Res *78*, 5754-5766.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev *28*, 2041-2055.

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

Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of

the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin *6*, 15.

Markopoulos, G., Noutsopoulos, D., Mantziou, S., Gerogiannis, D., Thrasyvoulou, S., Vartholomatos, G., Kolettas, E., and Tzavaras, T. (2016). Genomic analysis of mouse VL30 retrotransposons. Mob DNA *7*, 10.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927-931.

Naik, S., Larsen, S.B., Gomez, N.C., Alaverdyan, K., Sendoel, A., Yuan, S., Polak, L., Kulukian, A., Chai, S., and Fuchs, E. (2017). Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. Nature *550*, 475-480.

Nakaya, Y., Lilue, J., Stavrou, S., Moran, E.A., and Ross, S.R. (2017). AIM2-Like Receptors Positively and Negatively Regulate the Interferon Response Induced by Cytosolic DNA. MBio *8*.

Page-McCaw, A., Ewald, A.J., and Werb, Z. (2007). Matrix metalloproteinases and the regulation of tissue remodelling. Nat Rev Mol Cell Biol *8*, 221-233.

Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A.C., Wilcox, S., Fu, N., Liu, K.H., Jackling, F.C., *et al.* (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat Commun *8*, 1627.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

Reed, J.R., and Schwertfeger, K.L. (2010). Immune cell location and function during post-natal mammary gland development. J Mammary Gland Biol Neoplasia *15*, 329-339.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47-e47.

Saini, S.K., Ørskov, A.D., Bjerregaard, A.M., Unnikrishnan, A., Holmberg-Thydén, S., Borch, A., Jensen, K.V., Anande, G., Bentzen, A.K., Marquard, A.M., *et al.* (2020). Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers. Nat Commun *11*, 5660.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.

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

Shi, J., Gao, W., and Shao, F. (2017). Pyroptosis: Gasdermin-Mediated Programmed Necrotic Cell Death. Trends Biochem Sci *42*, 245-254.

Stark, R., and Brown, G. (2011). DiffBind: differential binding analysis of ChIP-Seq peak data. R package version *100*.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

Sun, P., Yuan, Y., Li, A., Li, B., and Dai, X. (2010). Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. Histochem Cell Biol *133*, 213-221.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev *19*, 815-826.

Tan, J., Buache, E., Alpy, F., Daguenet, E., Tomasetto, C.L., Ren, G.S., and Rio, M.C. (2014). Stromal matrix metalloproteinase-11 is involved in the mammary gland postnatal development. Oncogene *33*, 4050-4059.

Tiwari, N., Meyer-Schaller, N., Arnold, P., Antoniadis, H., Pachkov, M., van Nimwegen, E., and Christofori, G. (2013). Klf4 is a transcriptional regulator of genes critical for EMT, including Jnk1 (Mapk8). PLoS One *8*, e57329.

Unsworth, A., Anderson, R., Haynes, N., and Britt, K. (2016). OMIP-032: Two multi-color immunophenotyping panels for assessing the innate and adaptive immune cells in the mouse mammary gland. Cytometry A *89*, 527-530.

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

Waltman, L., and van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. The European Physical Journal B *86*, 471.

Wiseman, B.S., Sternlicht, M.D., Lund, L.R., Alexander, C.M., Mott, J., Bissell, M.J., Soloway, P., Itohara, S., and Werb, Z. (2003). Site-specific inductive and inhibitory activities of MMP-2 and MMP-3 orchestrate mammary gland branching morphogenesis. J Cell Biol *162*, 1123-1133.

Xu, C., and Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics *31*, 1974-1980.

Yu, Y.R., O'Koren, E.G., Hotten, D.F., Kan, M.J., Kopin, D., Nelson, E.R., Que, L., and Gunn, M.D. (2016). A Protocol for the Comprehensive Flow Cytometric Analysis of Immune Cells in Normal and Inflamed Murine Non-Lymphoid Tissues. PLoS One *11*, e0150606.

Zevini, A., Olagnier, D., and Hiscott, J. (2017). Crosstalk between Cytoplasmic RIG-I and STING Sensing Pathways. Trends Immunol *38*, 194-205.

Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C., and Surani, M.A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. Elife *4*.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-collagen IV	abcam	Cat#ab6586; RRID:AB_305584
Chicken polyclonal anti-cytokeratin 14	Covance	Cat#SIG-3476; RRID:AB_10718041
Rat monoclonal anti-cytokeratin 8 (TROMA-I)	Developmental Studies Hybridoma Bank	Cat#TROMA-I; RRID:AB_531826
Mouse monoclonal anti-dsDNA (35I9 DNA)	abcam	Cat#ab27156; RRID:AB_470907
Mouse monoclonal anti-dsDNA (AE-2)	Merck-Millipore	Cat#MAB1293; RRID:AB_94097
Mouse monoclonal anti-E-cadherin (36)	BD Biosciences	Cat#610182; RRID:AB_397581
Mouse monoclonal anti-ERα (1D5)	Dako	Cat#M7047; RRID:AB_2101946
Rabbit polyclonal anti-G9a	Cell Signaling	Cat#3306; RRID:AB_2097647
Mouse monoclonal anti-gH2ax (JBW301)	Merck-Millipore	Cat#05-636; RRID:AB_309864
Mouse monocolonal anti-H3K9me2 (mAbcam 1220)	abcam	Cat#ab1220; RRID:AB_449854
Mouse monoclonal anti-Ki67 (MM1)	Leica	Cat#NCL-Ki67-MM1; RRID:AB_442101
Rat monoclonal anti-mouse CD16/32 (93)	BioLegend	Cat#101320; RRID:AB_1574975
Rabbit polyclonal anti-phospho-Ezrin (Thr567)/Radixin (Thr564)/Moesin (Thr558) (pERM)	Cell Signaling	Cat#3141; RRID:AB_330232
Rabbit polyclonal anti-zonula occludens-1 (ZO-1)	Merck-Millipore	Cat#AB2272; RRID:AB_10807434
Rat monoclonal anti-CD19 (6D5); PerCP	BioLegend	Cat#115532; RRID:AB_2072926
Rat monoclonal anti-Ly6C (HK1.4); PerCP-Cy5.5	BioLegend	Cat#128011; RRID:AB_1659242
Rat monoclonal anti-CD24 (M1/69); PE-Cy7	BioLegend	Cat#101822; RRID:AB_756048
Rat monoclonal anti-CD8a (53-6.7); PE-Cy7	BioLegend	Cat#100722; RRID:AB_312761
Armenian hamster monoclonal anti-CD3e (1145-2C11); PE	BioLegend	Cat#100308; RRID:AB_312673
Rat monoclonal anti-EpCAM (G8.8); PE	BioLegend	Cat#118205; RRID:AB_1134176
Rat monoclonal anti-CD62L (MEL-14); BV711	BioLegend	Cat#104445; RRID:AB_2564215

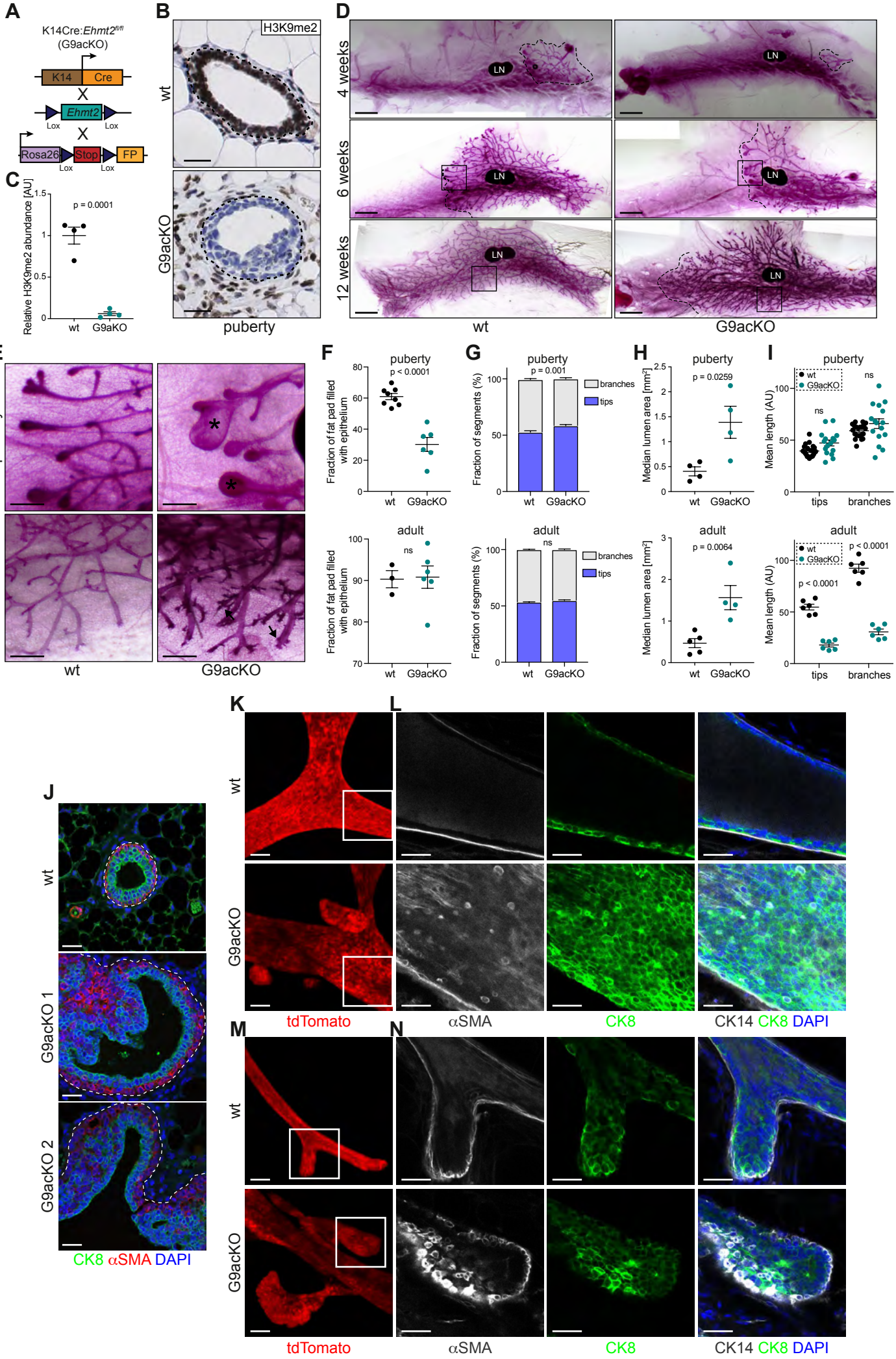
Rat monoclonal anti-IA/IE (M5/114.15.2); BV650	BioLegend	Cat#107641; RRID:AB_2565975
Rat monoclonal anti-CD31 (390); BV605	BioLegened	Cat#102427; RRID:AB_2563982
Rat monoclonal anti-CD45 (30-F11); BV605	BioLegend	Cat#103140; RRID:AB_2562342
Rat monoclonal anti-TER-119 (TER-119); BV605	BioLegend	Cat#116239; RRID:AB_2562447
Armenian hamster monoclonal anti-CD11c (N418); BV421	BioLegend	Cat#117329; RRID:AB_10897814
Rat monoclonal anti-CD11b (M1/70); APC-Cy7	BioLegend	Cat#101225; RRID:AB_830641
Rat monoclonal anti-EpCAM (G8.8); APC-Cy7	BioLegend	Cat#118218; RRID:AB_2098648
Armenian hamster monoclonal anti-CD29 (HMb1-1); APC	BioLegend	Cat#102216; RRID:AB_492833
Rat monoclonal anti-CD44 (IM7); APC	BioLegend	Cat#103012; RRID:AB_312963
Mouse monoclonal anti-CD64 (X54-5/7.1); APC	BioLegend	Cat#139305; RRID:AB_11219205
Rat monoclonal anti-CD4 (RM4-5); AF700	BioLegend	Cat#100536; RRID:AB_493701
Rat monoclonal anti-Ly6G (1A8)	BioLegend	Cat#127621; RRID:AB_10640452
anti-FVS510	BD Biosciences	Cat#564406; RRID:AB_2869572
Mouse monoclonal anti-α-smooth muscle actin (clone 1A4)	Sigma-Aldrich	Cat#A5228; RRID:AB_262054
Goat anti-chicken IgY secondary Ab, AF647	Life Technologies	Cat#A21449; RRID:AB_1500594
Goat anti-mouse IgG secondary Ab, AF488	Life Technologies	Cat#A32723; RRID:AB_2633275
Donkey anti-mouse IgG secondary Ab, AF568	Life Technologies	Cat#A10037; RRID:AB_2534013
Donkey anti-rabbit IgG secondary Ab, AF568	Life Technologies	Cat#A21206; RRID:AB_2535792
Donkey anti-rabbit IgG secondary Ab, AF594	Life Technologies	Cat#A21207; RRID:AB_141637
Donkey anti-rat IgG secondary Ab, AF488	Life Technologies	A21208; RRID:AB_141709
Rabbit monoclonal anti-p63 (EPR5701)	abcam	Cat#ab124762; RRID:AB_10971840
Bacterial and Virus Strains		
TRC1.5-pLKO.1-shRNA(*Aim2*)	Sigma-Aldrich	TRCN0000096107
pLKO.1-TRC control	Addgene	Cat#10879; RRID:Addgene_10879
Chemicals, Peptides, and Recombinant Proteins		

Mitomycin-C	Sigma-Aldrich	Cat#M4287
EpiCult-B Mouse Medium kit	StemCell Technologies	Cat#05610
Ham´s F12	Life Technologies	Cat#21765029
Adenine	Sigma-Aldrich	Cat#A3159
Hydrocortisone	Sigma-Aldrich	Cat#386698
Cholera enterotoxin	Gentaur Molecular Products BVBA	Cat#100
Recombinant murine EGF	Peprotech	Cat#315-09
Y-27632	Sigma-Aldrich	Cat#Y0503
Acetone	Sigma-Aldrich	Cat#34859
Methanol	Sigma-Aldrich	Cat#322415
Wright´s Giemsa	Sigma-Aldrich	Cat#WG16
BD Matrigel Matrix	BD Biosciences	Cat#354234
Collagenase/hyaluronidase	StemCell Technologies	Cat#07912
RBC Lysis Buffer	eBioscience	Cat#00-4333-57
Trypsin/EDTA	Life Technologies	Cat#25200056
HBSS	Life Technologies	Cat#14175053
Dispase II	Sigma-Aldrich	Cat#D4693
DNase I	Sigma-Aldrich	Cat#DN25
Collagenase I	Sigma-Aldrich	Cat#C0130
EMEM (Ca^{2+}-free)	Lonza	Cat#BE06-174G
PBS	Lonza	Cat#H3BE17-516F
BD Matrigel Matrix (growth factor-reduced)	BD Biosciences	Cat#356230
Trypan blue	Sigma-Aldrich	Cat#T8154
DAPI	Sigma-Aldrich	Cat#32670
Absolute ethanol	VWR	Cat#1.00983.2500
Chloroform	Sigma-Aldrich	Cat#C2432
Glacial acetic acid	VWR	Cat#1.01830.2500
Carmine	Sigma-Aldrich	Cat#C1022
Aluminum potassium sulfate	Sigma-Aldrich	Cat#A7167
Xylene	Sigma-Aldrich	Cat#534056
CV Mount	Leica	Cat#14046430011
Formaldehyde	Aname	Cat#15710
BSA	Sigma-Aldrich	Cat#A7906
Triton X	Sigma-Aldrich	Cat#T8787
Vectashield	Vector laboratories	Cat#H-1000
Formalin	Sigma-Aldrich	Cat#HT501128
Donkey serum	Life technologies	Cat#D9663
DAB	Dako	Cat#K346711-2
Hematoxylin	Dako	Cat#CS70030-2

Toluene-Free Mounting Medium	Dako	Cat#CS70530-2
TRIzol	Thermo Scientific	Cat#15596018
High-glucose DMEM	Life Technologies	Cat#41965039
FBS	Life Technologies	Cat#10270106
SYBR Select Master Mix	Life Technologies	Cat#4472918
L-glutamine	Life Technologies	Cat#25030024
Penicillin/Streptomycin	Life Technologies	Cat#15140122
Critical Commercial Assays		
TURBO DNA-free Kit	Thermo Fisher	Cat#AM1907
Clariom S Mouse Array	Thermo Scientific	Cat#901570
RevertAid First Strand cDNA Synthesis Kit	Thermo Scientific	Cat#K1621
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat	Illumina	Cat#RS-122-2201
FLICA 660 Caspase-1 Assay Kit	ImmunoChemistry	Cat#9122
Lenti-X Concentrator	Takara Bio	Cat#631231
Lenti-X p24 Rapid Titer Kit	Takara Bio	Cat#632200
Deposited Data		
Raw and analysed data	This paper	GEO: GSE149978
Experimental Models: Cell Lines		
NIH/3T3	ATCC	ATCC® CRL-1658™
HEK293T/17	ATCC	ATCC® CRL-11268™
Experimental Models: Organisms/Strains		
Mouse: B6.129X1-*Gt(ROSA)26Sor*$^{tm1(EYFP)Cos}$/J (Rosa26YFP)	The Jackson Laboratory	Cat#006148
Mouse: Crl:NU(Ico)-*Foxn1*^{nu} (Swiss nude)	Charles River Laboratories	Cat#620
Mouse: B6N.Cg-Tg(KRT14-cre)1Amc/J (K14-Cre)	The Jackson Laboratory	Cat#018964
Mouse: B6.Cg-Gt(ROSA)26Sor$^{tm9(CAG-tdTomato)Hze}$/J (Rosa26tdTomato)	The Jackson Laboratory	Cat#007909
Software and Algorithms		
BowTie2 (version 2.3.3.1)	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
SAMtools (version 1.3.1)	Li et al., 2009	http://samtools.sourceforge.net
MACS2 (version 2.0.10)	Feng et al., 2012	https://github.com/macs3-project/MACS/wiki/Install-macs2
DiffBind (version 2.4.8)	Stark and Brown, 2011	https://www.bioconductor.org/packages/release/bioc/html/DiffBind.ht
HOMER (version 2015-03-22)	Heinz et al., 2010	http://homer.ucsd.edu/homer/
Fiji (version 2.0.0-rc-68)	Schindelin et al.,	https://imagej.net/Fiji

	2012	
QuPath (version 0.1.2)	Bankhead et al., 2017	https://qupath.github.io
TopHat (version 2.1.1)	Kim et al., 2013	https://ccb.jhu.edu/software/tophat/index.shtml
DESeq2 (version 1.26.0)	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DAVID (version 6.7)	Huang da et al., 2009	https://david.ncifcrf.gov
limma (version 3.32.10)	Ritchie et al., 2015	https://www.bioconductor.org/packages/release/bioc/html/limma.htm
DERfinder (version 1.20.0)	Collado-Torres et al., 2017	https://www.bioconductor.org/packages/derfinder
Seurat (version 3.0.1)	Stuart et al., 2019	https://satijalab.org/seurat/
BBduk	Bushnell et al., 2017	http://jgi.doe.gov/data-and-tools/bbtools
STAR aligner (version STAR_2.5.2b)	Dobin et al., 2013	https://github.com/alexdobin/STAR
TEtranscript (version 2.0.3)	Jin et al., 2015	https://github.com/mhammell-laboratory/TEtranscripts
deeptools (version 3.3.1)	Ramirez et al., 2016	https://deeptools.readthedocs.io/en/develop/
Prism (version 8.4.2)	GraphPad	https://www.graphpad.com
FlowJo (version 10.4.2)	Treestar	https://www.flowjo.com
Adobe Photoshop CS6	Adobe	https://www.adobe.com/products/photoshop.html
Adobe Illustrator CS6	Adobe	https://www.adobe.com/products/illustrator.html
Trimmomatic (version 0.36)	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
Other		
34 G blunt end metal hub needle	Hamilton Company	Cat#7803-05
Hamilton syringe (50 µl)	Hamilton Company	Cat#7637-01

Figure 1


A
K14Cre:Ehmt2fl/fl
(G9acKO)
K14
Cre
Ehmt2
Lox
Lox
Rosa26
Stop
FP
Lox
Lox
B
H3K9me2
wt
G9acKO
puberty
C
Relative H3K9me2 abundance [AU]
p = 0.0001
wt
G9aKO
D
4 weeks
6 weeks
12 weeks
LN
wt
G9acKO
E
puberty
adult
wt
G9acKO
F
puberty
p < 0.0001
Fraction of fat pad filled with epithelium
adult
ns
G
puberty
p = 0.001
Fraction of segments (%)
branches
tips
adult
ns
H
puberty
p = 0.0259
Median lumen area [mm²]
adult
p = 0.0064
I
puberty
wt
G9acKO
ns
Mean length (AU)
tips
branches
adult
p < 0.0001
p < 0.0001
J
wt
G9acKO 1
G9acKO 2
CK8 αSMA DAPI
K
L
wt
G9acKO
tdTomato
αSMA
CK8
CK14 CK8 DAPI
M
N
wt
G9acKO
tdTomato
αSMA
CK8
CK14 CK8 DAPI

Figure

Figure 2

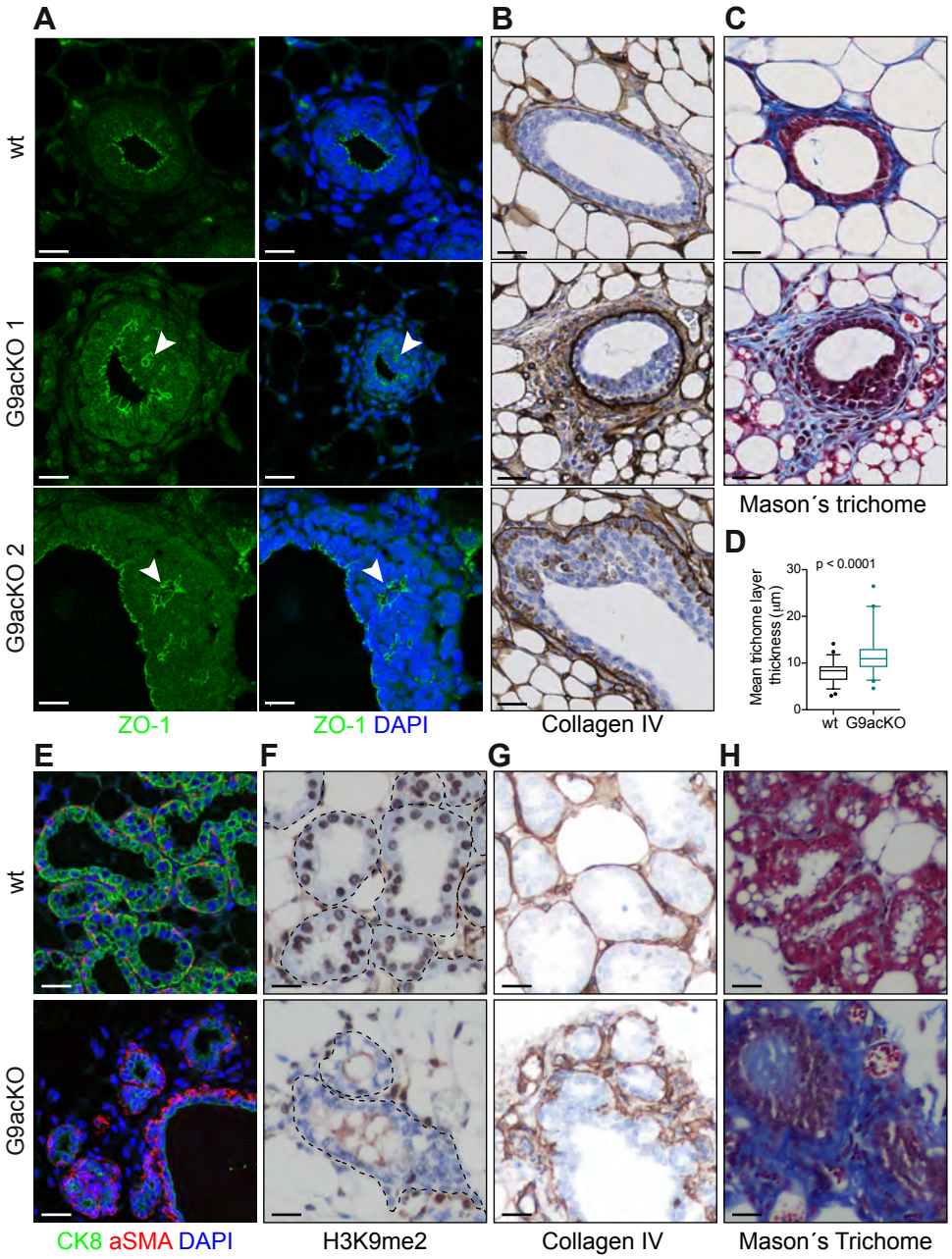


Figure
Figure 3

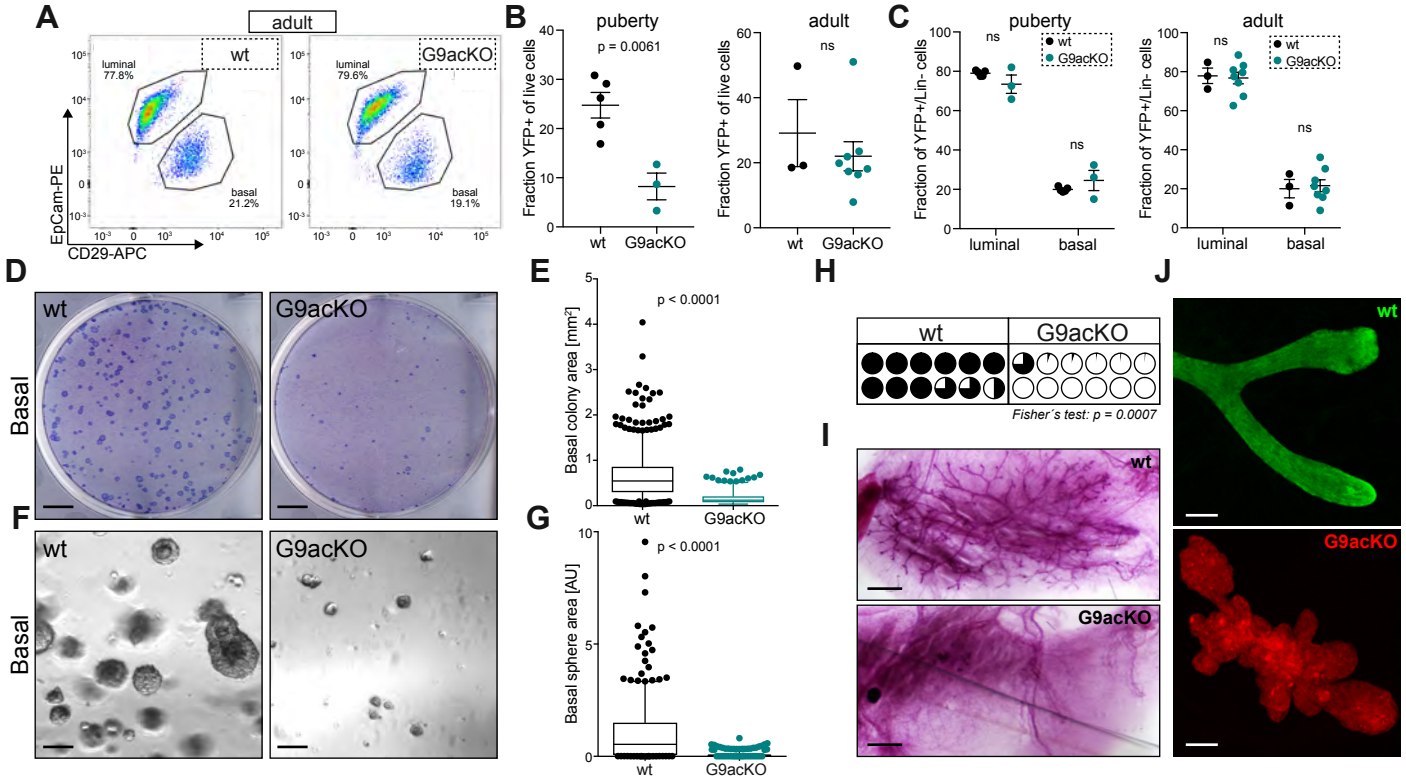
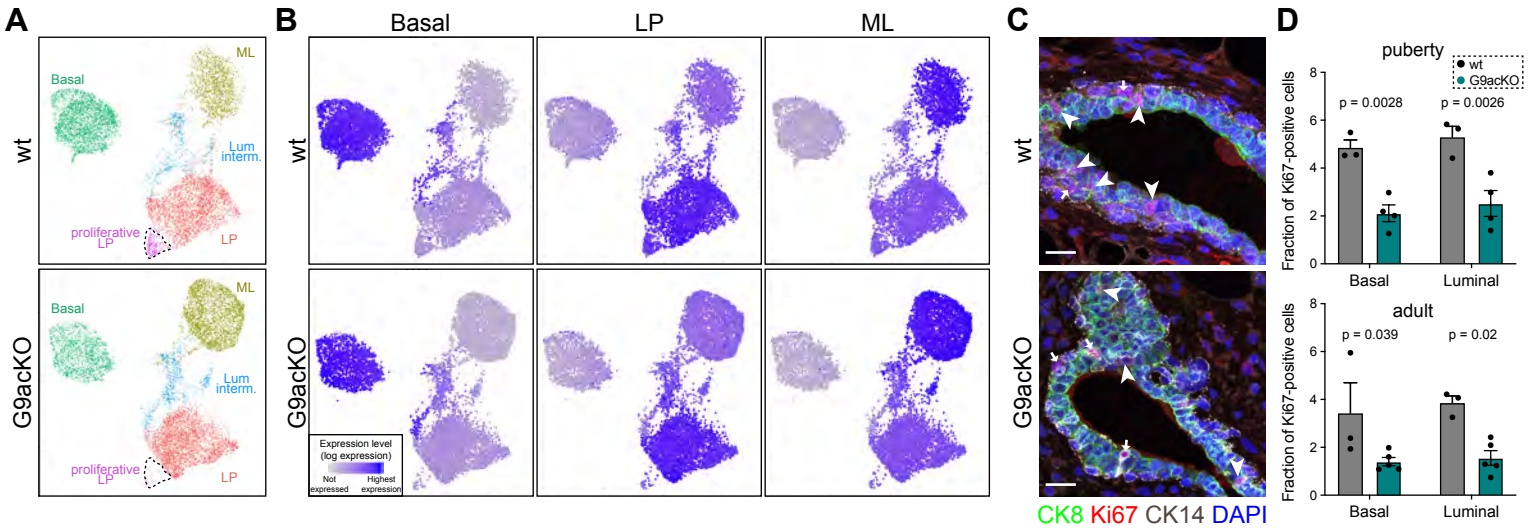
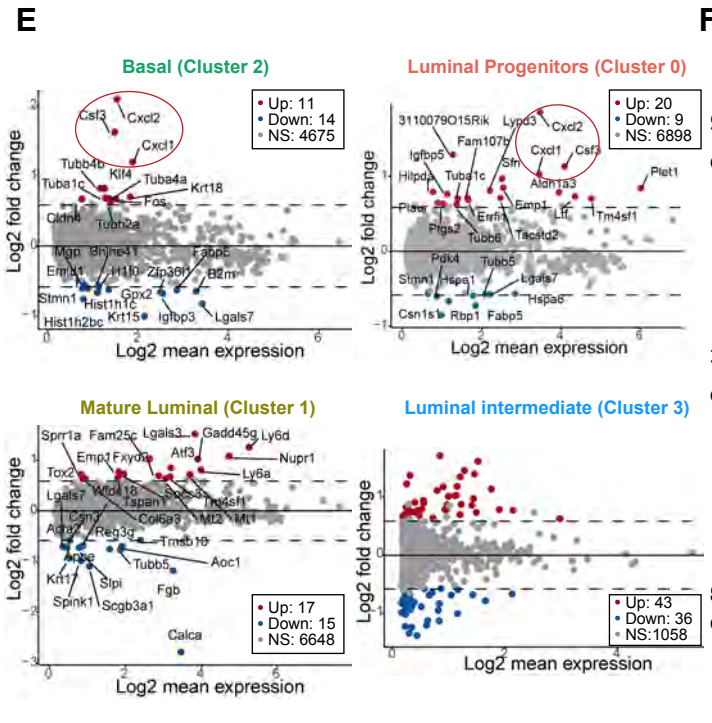
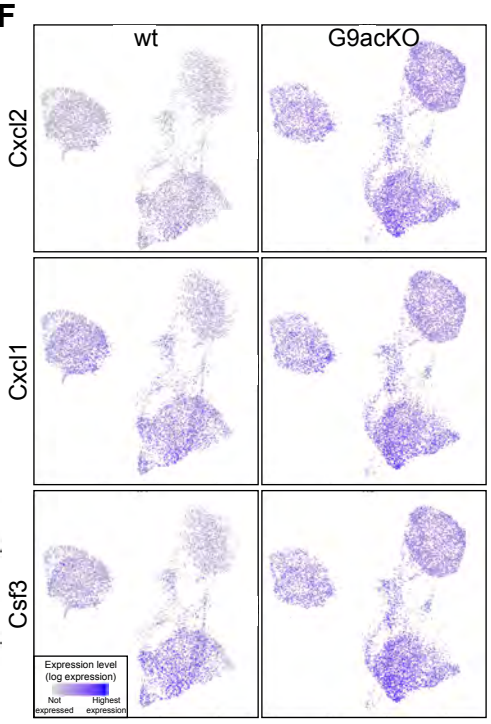
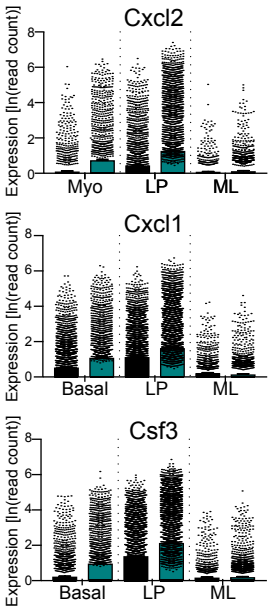


Figure
Figure 4









Figure

Figure 6

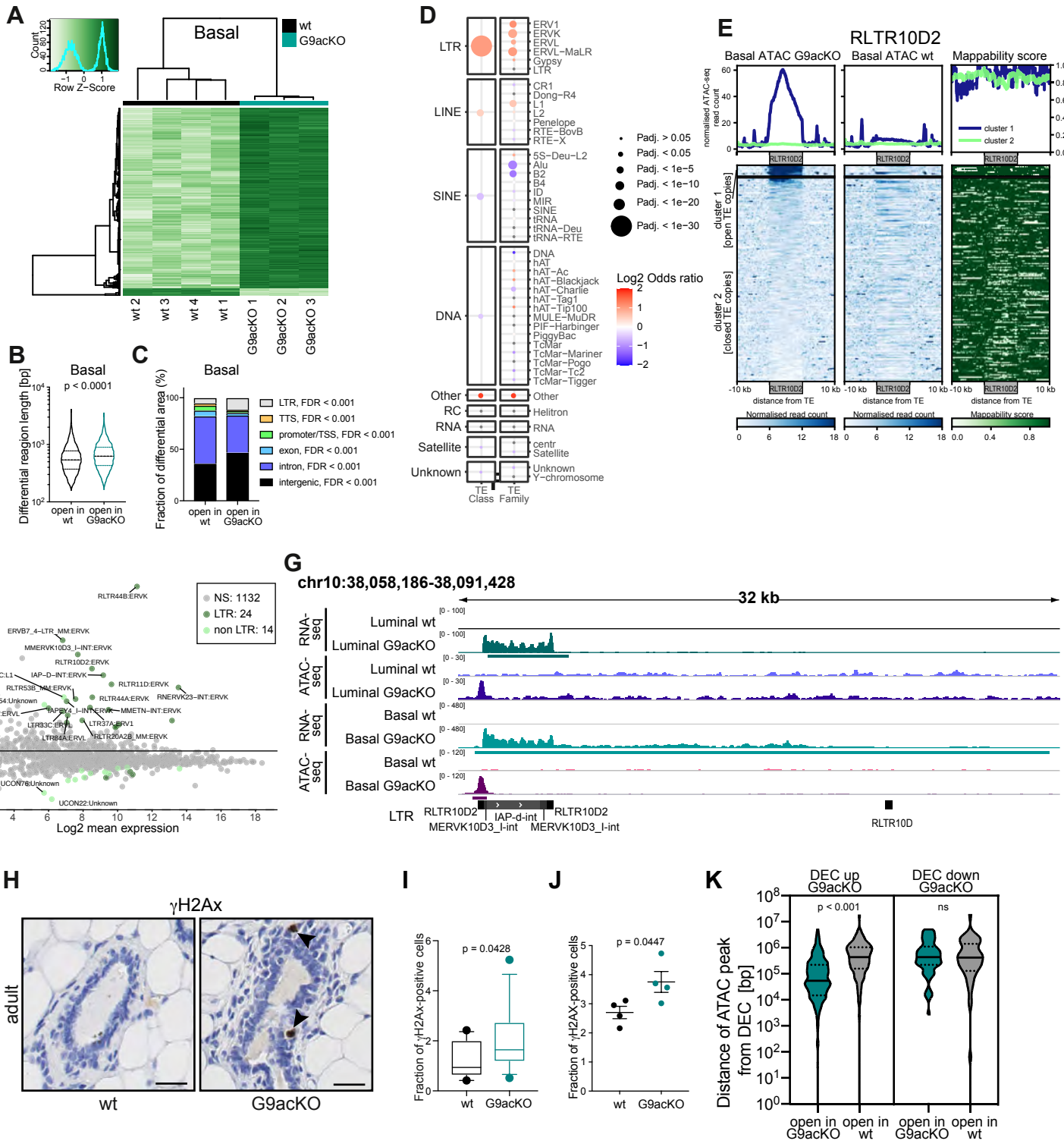


Figure 7

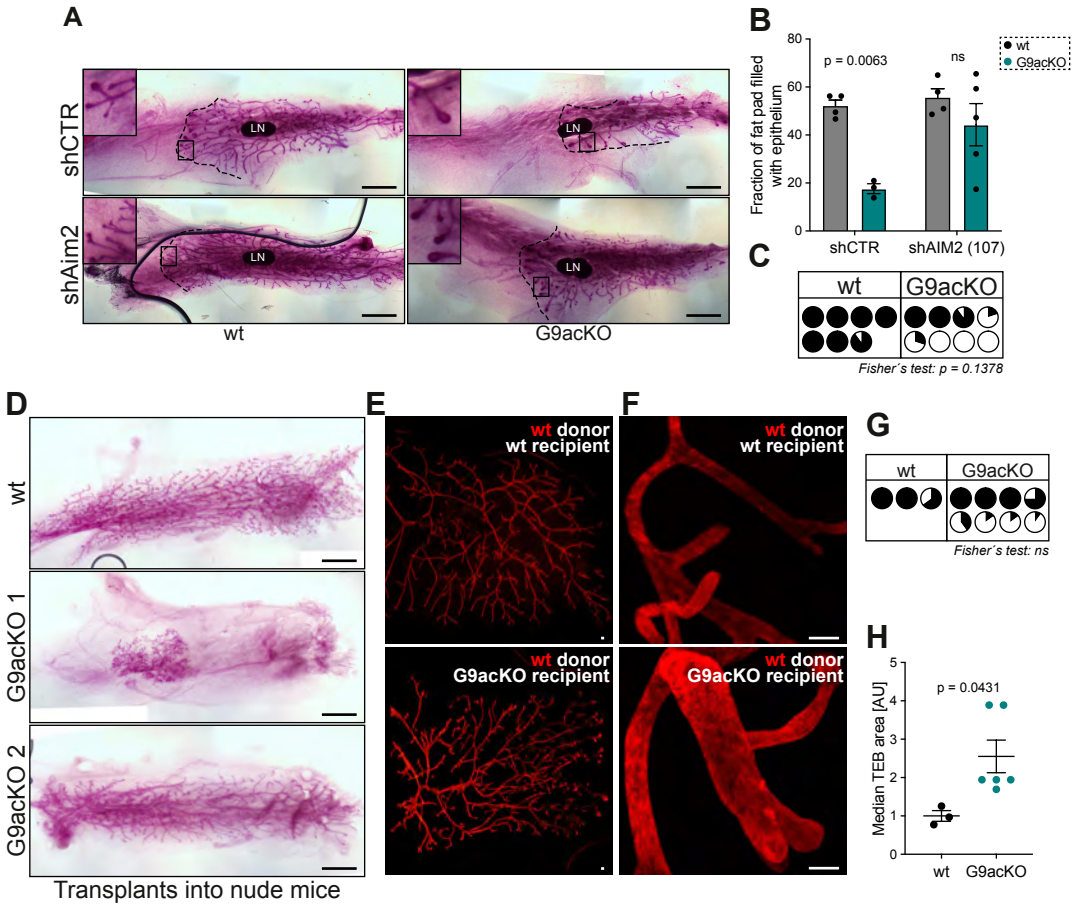
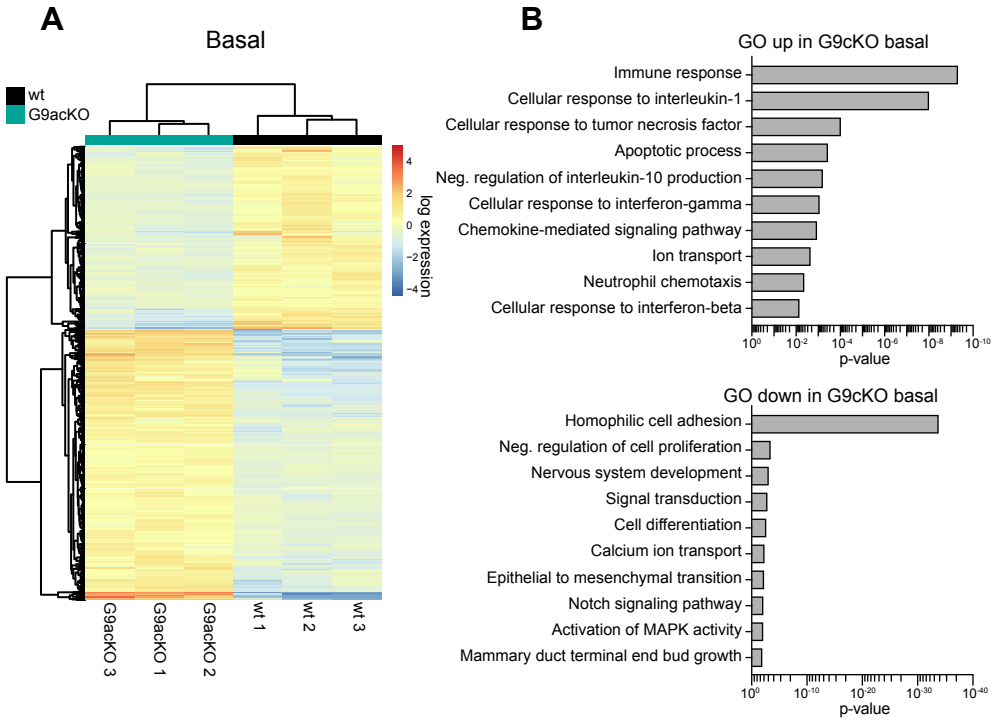
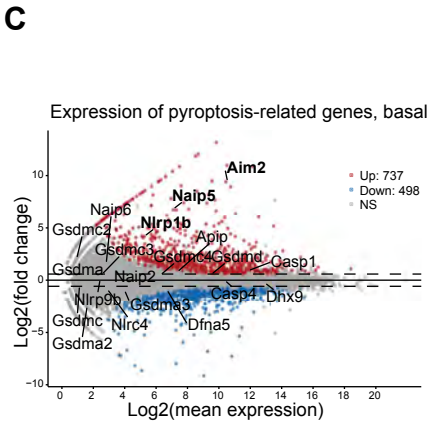
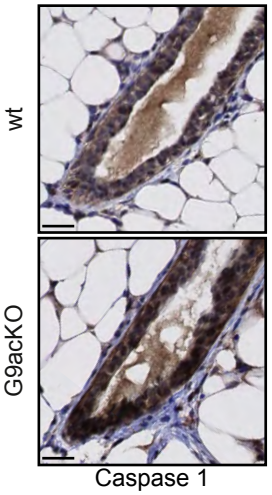
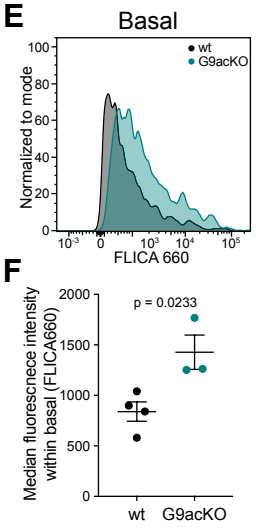

A
shCTR
shAim2
LN
wt
G9acKO
B
Fraction of fat pad filled with epithelium
p = 0.0063
ns
wt
G9acKO
shCTR
shAIM2 (107)
C
wt
G9acKO
Fisher´s test: p = 0.1378
D
wt
G9acKO 1
G9acKO 2
Transplants into nude mice
E
wt donor
wt recipient
wt donor
G9acKO recipient
F
wt donor
wt recipient
wt donor
G9acKO recipient
G
wt
G9acKO
Fisher´s test: ns
H
Median TEB area [AU]
p = 0.0431
wt
G9acKO

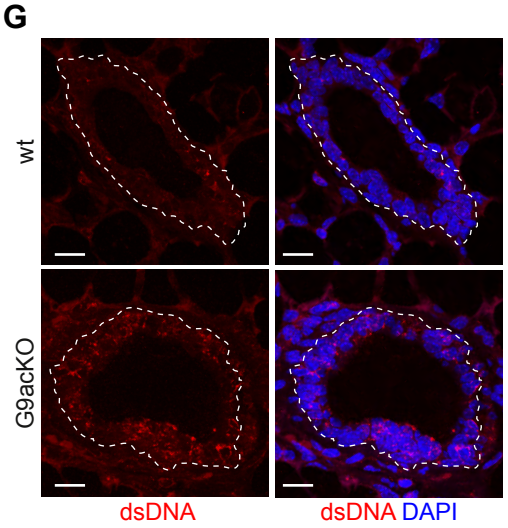
Figure 5

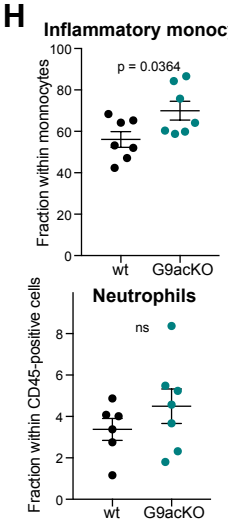

A
Basal
wt
G9acKO
log expression
G9acKO 3
G9acKO 1
G9acKO 2
wt 1
wt 2
wt 3
B
GO up in G9cKO basal
Immune response
Cellular response to interleukin-1
Cellular response to tumor necrosis factor
Apoptotic process
Neg. regulation of interleukin-10 production
Cellular response to interferon-gamma
Chemokine-mediated signaling pathway
Ion transport
Neutrophil chemotaxis
Cellular response to interferon-beta
p-value
GO down in G9cKO basal
Homophilic cell adhesion
Neg. regulation of cell proliferation
Nervous system development
Signal transduction
Cell differentiation
Calcium ion transport
Epithelial to mesenchymal transition
Notch signaling pathway
Activation of MAPK activity
Mammary duct terminal end bud growth
p-value

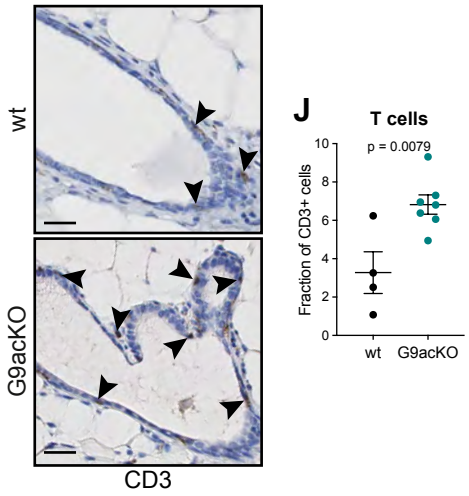

C
Expression of pyroptosis-related genes, basal
Up: 737
Down: 498
NS
Aim2
Naip5
Naip6
Nlrp1b
Gsdmc2
Apip
Gsdmd3
Gsdmc4
Gsdmd
Casp1
Gsdma
Naip2
Nlrp9b
Gsdma3
Casp4
Dhx9
Gsdmc
Nlrc4
Dfna5
Gsdma2
Log2(fold change)
Log2(mean expression)


D
wt
G9acKO
Caspase 1


E
Basal
wt
G9acKO
Normalized to mode
FLICA 660
F
p = 0.0233
Median fluorescnece intensity within basal (FLICA660)
wt
G9acKO


G
wt
G9acKO
dsDNA
dsDNA DAPI


H
Inflammatory monocytes
p = 0.0364
Fraction within monnocytes
wt
G9acKO
Neutrophils
ns
Fraction within CD45-positive cells
wt
G9acKO


I
wt
G9acKO
CD3
J
T cells
p = 0.0079
Fraction of CD3+ cells
wt
G9acKO

SUPPLEMENTARY INFORMATION

Figure S1. G9a-ablation leads to loss of mammary gland organisation; related to Figure 1. (A) Immunohistochemistry for H3K9me2 on 12-week-old (adult) WT and G9acKO mammary fat pads. Nuclei were counterstained with haematoxylin. Images shown are representative of n = 5 (WT), n = 4 (G9acKO) independent animals. Scale bar, 25 μm. (B) Quantification of H3K9me2 abundance within mammary epithelium in adult WT and G9acKO mammary fat pads. Data are mean ± SEM independent animals; n = 5 (WT), n = 4 (G9acKO). Two-sided Student's *t*-test. (C) Immunofluorescence for G9a (red) cytokeratin 8 (green) and cytokeratin 14 (gray) on adult WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). White dashed line delineates epithelium/stroma boundary. Arrowheads highlight G9a-positive stromal cells in G9acKO mammary fat pads. Scale bar, 25 μm**.** Images shown are representative of n = 5 (WT) , n = 4 (G9acKO) independent animals. (D) Immunofluorescence for αSMA (red) and cytokeratin 8 (green) on adult WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). Two examples are shown for G9acKO to capture phenotypic heterogeneity. Dashed line indicates epithelium/stroma boundary. Scale bar, 25 μm**.** Images shown are representative of n = 5 (WT), n = 4 (G9acKO) independent animals. For visualisation purposes, the brightness of the blue and green channels has been modified uniformly across all images presented for comparison. (E) Quantification of number of mammary epithelial ducts with that display lineage intermingling. Data are mean ± SEM of independent animals; n = 4 (WT), n = 4 (G9acKO) mice at 6 weeks (puberty),, n = 5 (WT), n = 4 (G9acKO) mice at 12 weeks (adult). Two-sided Student's *t*-test. (F-G) Endogenous Rosa26tdTomato signal (red, F) and immunofluorescence staining (G) for CK8 (green) and αSMA (gray) on adult WT and G9acKO mammary wholemounts. Nuclei were counterstained with DAPI (blue). G9a loss leads to cellular intermingling and basal cell multi-layering preferentially in terminal end bud regions (F, G). Scale bar, 50 μm (F) or 25 μm (G).

Figure S2. G9a Loss Disrupts Tissue Polarity and Function; related to Figure 2. (A) Immunofluorescence for ZO-1 (green) on 12-week-old (adult) WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). Two examples are shown for G9acKO to capture phenotypic heterogeneity. Arrowheads indicate aberrant lumina. Note nuclear localisation in G9acKO luminal cell indicative of immature cell-cell junction formation. Images shown are representative of n = 5 (WT) , n = 4 (G9acKO) mice. Scale bar, 25 μm**.** (B) Immunofluorescence for pERM (green), E-cadherin (red), and cytokeratin 14 (K14) (gray) on 6-week-old (pubertal) WT and G9acKO mammary fat pads (n = 4 mice for each group). Nuclei were counterstained with DAPI (blue). Two examples are shown for G9acKO to capture phenotypic heterogeneity. Arrowheads indicate pERM mislocalisation. White dashed line delineates luminal/basal boundary. Scale bar, 50 μm. For visualisation purposes, the brightness of the blue and red channels were modified uniformly across all images presented for comparison. (C) Immunohistochemistry for collagen IV on adult WT and G9acKO mammary fat pads. Nuclei were counterstained with haematoxylin. n = 5 (WT) , n = 4 (G9acKO) mice. Scale bar, 25 μm. (D) Masson's trichome staining on adult WT and G9acKO mammary fat pads. Nuclei were counterstained with haematoxylin. n = 5 (WT), n = 4 (G9acKO) mice. Scale bar, 25 μm. (E) Quantification of thickness of Masson's trichome layer around mammary ducts in adult WT and G9acKO mammary fat pads. Each data point represents an individual duct. WT, n = 181 ducts from 5 mice; G9acKO, n = 125 ducts from 4 mice. Two-sided Student's *t*-test. (F) Immunofluorescence for Estrogen receptor-alpha (ERα) (green), cytokeratin 8 (red), and cytokeratin 14 (gray) on adult WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). Asterisks highlight ERα-positive luminal cells. Scale bar, 25 μm**.** Images shown are representative of n = 5 (WT), n = 4 (G9acKO) independent animals. For visualisation purposes, the brightness of the green and gray channels has been modified uniformly across all images presented for comparison. Quantification was performed on raw data. (G) Quantification of (F). Data are mean ± SEM. Two-sided Student's *t*-test. (H, I) Carmine-alum staining (stitched images, see Star Methods) of WT or G9acKO mammary wholemounts at day 1 of lactation (I) of gestation day 14.5 (I). LN, lymph node. n = 3 mice for each group. Scale bar, 2 mm. (J) Immunofluorescence for αSMA (red) and cytokeratin 8 (green) on WT and G9acKO mammary fat pads at gestation day 14.5 showing lack of alveolar differentiation in G9acKO mammary epithelia (n = 3 mice for each group). Nuclei were counterstained with DAPI (blue). Scale bar, 25 μm. (K, L) Immunohistochemistry for H3K9me2 (K) or collagen IV (L) in mammary fat pads from WT or G9acKO mice at gestation day 14.5 (n = 3 mice per group). Nuclei were counterstained with haematoxylin. Scale bar, 25 μm.

Figure S3. Impaired Mammary Stem Cell Potential upon G9a Loss; related to Figure 3. (A) Gating strategy for FACS isolation. Total mammary epithelium, or luminal and basal cells, were isolated as indicated in each experiment. (B) Representative FACS plots showing gating of luminal and basal cells from 6-week-old (pubertal) WT and G9acKO mice. (C) 2D colony formation assay of adult WT and G9acKO luminal cells. Representative images of three independent experiments are shown. Scale bar, 5 mm. (D, E) Quantification of number of colonies (D), or colony area (E), shown in (C). Data are mean ± SEM of individual colonies pooled from triplicate wells. Two-sided Student's *t*-test. (F) 3D colony formation assay of adult WT and G9acKO luminal cells. Representative images of three independent experiments are shown. Scale bar, 50 μm. (G) Quantification of colony area in (F). Data are mean ± SEM of individual colonies pooled from triplicate wells. Two-sided Student's *t*-test. (H) Quantification of number of colonies in Figure 2D (2D clonogenic with basal cells). Data are mean ± SEM of individual colonies pooled from triplicate wells. Two-sided Student's *t*-test.

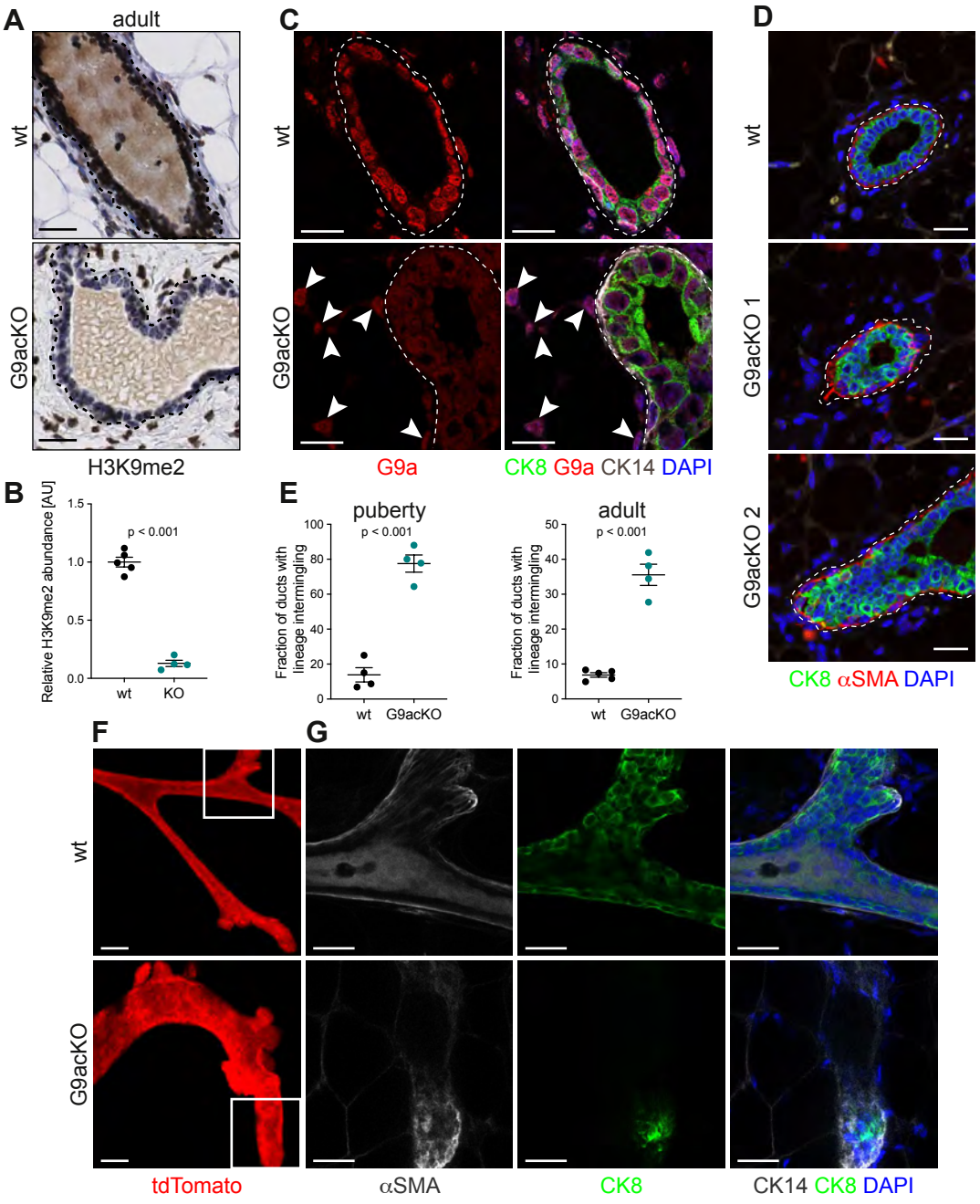
Figure S4. Mammary Lineages are Correctly Specified in G9acKO Mice; related to Figure 4. (A) UMAP plot of the transcriptomes of 18,359 total epithelial cells from virgin 12-week-old (adult) mammary glands (WT, n = 8939 cells from 2 mice; G9acKO, n = 9420 cells from 2 mice, shown together) generated using 10× genomics. The major clusters were identified as basal (cluster # 2; 4986 cells), luminal progenitors (LP, cluster 0; 6986 cells), mature luminal (ML, cluster 1; 4998 cells), luminal intermediate (cluster 3; 1115) and proliferative luminal progenitors (prolif. LPs, cluster 4; 274 cells), as indicated. (B) Heatmap showing expression of the top 15 DEGs for each cluster identified in (A). (C) The mammary epithelial cell clusters separated by genotype were interrogated for the expression of six known lineage marker expression: *Krt14*, *Acta2*, and *Sparc* for basal cells, *Krt19* for luminal cells, *Elf5* for luminal progenitors, and *Prlr* for mature luminal cells. Colour indicates expression levels. (D) Immunofluorescence for αSMA (green), p63 (red) and cytokeratin 14 (gray) on adult WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). Scale bar, 25 μm**.** Images shown are representative of n = 3 independent animals (each group). For visualisation purposes, the brightness of the green channel has been modified uniformly across all images presented for comparison. (E) Subdivision of basal cells as identified from (A) by expression of the basal cell markers αSMA, CK14 and p63. WT, n = 3226 cells from 2 mice; G9acKO, n = 1760 cells from 2 mice, separated by genotype. Fisher´s exact test with FDR correction for multiple testing. (F) Violin plots showing quantification of cluster expression of known lineage-specific signatures for these cell types (see also Figure 4B). Median and quartiles are indicated. (G) UMAP plot of subclustering of basal cells, as identified from (A). WT, n = 3226 cells from 2 mice; G9acKO, n = 1760 cells from 2 mice, separated by genotype. Basal subclusters identified were proliferative basal (subcluster 2-1; WT, 122 cells; and G9acKO, 10 cells) and bulk basal cells (subcluster 2-0; WT, 3104 cells; and G9acKO, 1750 cells), as indicated; also see Table S2. (H) Immunofluorescence for cytokeratin 8 (green), Ki67 (red) and cytokeratin 8 (gray) on 6-week-old (pubertal) WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). Arrowheads indicate proliferative luminal cells; arrows indicate proliferative basal cells. n = 3 (WT), n = 4 (G9acKO) mice. Scale bar, 25 μm. For visualisation purposes, the brightness of the red channel has been modified uniformly across all images presented for comparison. Quantification was performed on raw data.

Figure S5. Pyroptosis in G9acKO Mammary Epithelium; related to Figure 5. (A) Heatmap showing unsupervised hierarchical clustering of DEGs between FACS-isolated WT and G9acKO luminal cells at 6 weeks of age. (B) GO pathways deregulated in G9acKO luminal cells at 12 weeks of age. (E) MA plot showing DEGs in G9acKO versus WT luminal cells at 12 weeks of age. Pyroptosis-associated genes are highlighted; significant DEGs of the pyroptosis pathway are indicated in bold. (D) Heatmap showing unsupervised hierarchical clustering of DEGs between WT and G9acKO basal or luminal cells at 6 weeks of age. (E) Venn diagrams highlighting overlap between differentially expressed genes between WT and G9acKO basal and luminal cells at 6 weeks-old (pubertal) versus 12 weeks-old (adult) mice. (F) Real-time quantitative PCR of *Ehmt2* (gene name of G9a) expression in WT and G9acKO luminal or basal mammary epithelial cells from 6-week-old mice, confirming G9a loss in pubertal G9acKO mammary cells. Data are mean ± SEM of individual animals. n = 3 (WT), n = 5 (G9acKO) mice. Two-way ANOVA with Dunnett multiple comparisons test. (G) Immunohistochemistry for caspase 1 on 6-week-old (pubertal) WT and G9acKO mammary fat pads. Nuclei were counterstained with haematoxylin. Scale bar, 25 μm. n = 4 mice for each group. (H) Representative FACS plot of FLICA caspase 1 activity assay in WT and G9acKO luminal cells. n = 4 (WT) , n = 3 (G9acKO). (I) Quantification of (H). (J) Immunofluorescence for dsDNA (red) on pubertal WT and G9acKO mammary fat pads. Nuclei were counterstained with DAPI (blue). Scale bar, 25 μm. n = 4 mice for each group. (K) TUNEL staining on pubertal or adult WT and G9acKO mammary fat pads. Scale bar, 25 μm. n = 4 puberty mice for each group, , n = 5 (WT) , n = 4 (G9acKO) adult mice. (L) Quantification of (K). Data points represent ROIs; three ROIs were used per mouse. Two-sided Student's *t*-test.

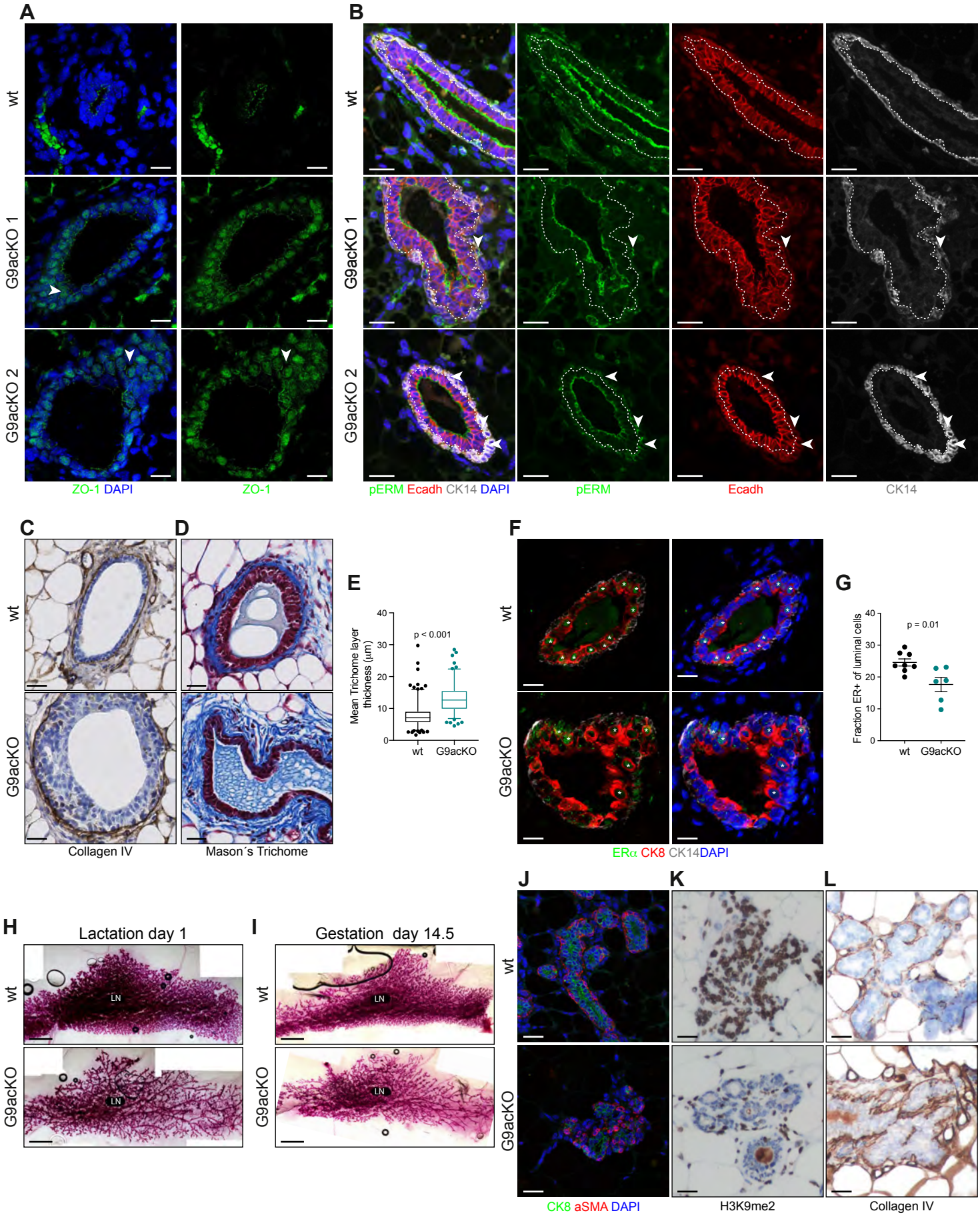
Figure S6. Increased chromatin opening in G9acKO mammary epithelium concomitant with derepression of LTRs; related to Figure 6. (A) Correlation heatmap showing regions of differential chromatin opening between FACS-isolated WT and G9acKO luminal cells. (B) Quantification of length of differentially open chromatin regions in luminal cells. Two-sided Student's *t*-test. (C) Annotation of location of differentially open chromatin regions luminal cells. Fisher´s exact test with FDR correction for multiple testing. (D) Dot plot of enrichment of TEs in differential ATAC peaks, luminal cells. (E) Percentage of TE copies covered by ATAC peaks for each element. (F) Heatmap showing ATAC-seq read coverage of individual copies RLTR10D2 subfamily of LTR retrotransposons in WT or G9acKO luminal cells. Each row represents an individual RLTR10D2 locus. RLTR10D2 copies that are in accessible chromatin (cluster 1) or closed chromatin (cluster 2) are indicated. The mappability score of the individual loci is shown, demonstrating that differences in ATAC-seq read coverage are not due to differences in regional mappability. (G) MA plot showing differentially expressed TEs in G9acKO versus WT luminal cells. Differentially expressed TEs (ajd. p-value < 0.01) are highlighted in green and subclassified into LTR or non-LTR. (H) Enrichment of G9acKO chromatin opening in differentially expressed LTRs from the ERVK family. (I) Example of upregulation of the VL30 family of retroviruses. The full-length VL30 element (see Markopoulos et al., 2016) is indicated. (J) Immunohistochemistry for γH2AX on 6-week-old (pubertal) WT and G9acKO mammary fat pads. Nuclei were counterstained with haematoxylin. Scale bar, 25 μm. n = 4 mice for each group. (K) Quantification of (J). Each data point indicates an ROI; 3 ROIs per mouse were used. Two-sided Student's *t*-test. (L) Violin plot of the distance between DECs and the closest differentially open chromatin regions in luminal cells, highlighting an association between G9acKO DEC upregulation and chromatin opening.

Figure S7. Cytosolic dsDNA is Dependent on TE expression; related to Figure 7. (A) Real-time quantitative PCR of expression of different TE subfamilies in WT and G9acKO mammary epithelial cells cultured in the presence of vehicle or 1 mM lamivudine (a nucleoside reverse-transcriptase inhibitor). Data are mean ± SEM of technical replicates. Equivalent results were obtained from 3 independent experiments. P-values between G9acKO (vehicle) and G9acKO (lamivudine) are shown. Two-way ANOVA with Dunnett multiple comparisons test. N/D –not determined (expression below qPCR detection limit). (B) Immunofluorescence for dsDNA (green) on cultured WT and G9acKO mammary epithelial cells. Nuclei were counterstained with DAPI (gray). Scale bar, 10 μm. (C) Real-time quantitative PCR of *Aim2* and *Casp1* expression in WT and G9acKO mammary epithelial cells cultured in the presence of vehicle or 1 mM lamivudine. Data are mean ± SEM of technical replicates. Equivalent results were obtained from 3 independent experiments. P-values between G9acKO (vehicle) and G9ackO (lamivudine) are shown. Two-way ANOVA with Dunnett multiple comparisons test. N/D –not determined (expression below qPCR detection limit). (D) Real-time quantitative PCR of *Aim2* in NIH3T3 mouse fibroblasts in response to shCTR or shAim2 lentivirus. Data are mean ± SEM of technical replicates. Equivalent results were obtained from 2 experiments. N/D, not determined (expression below qPCR detection limit).

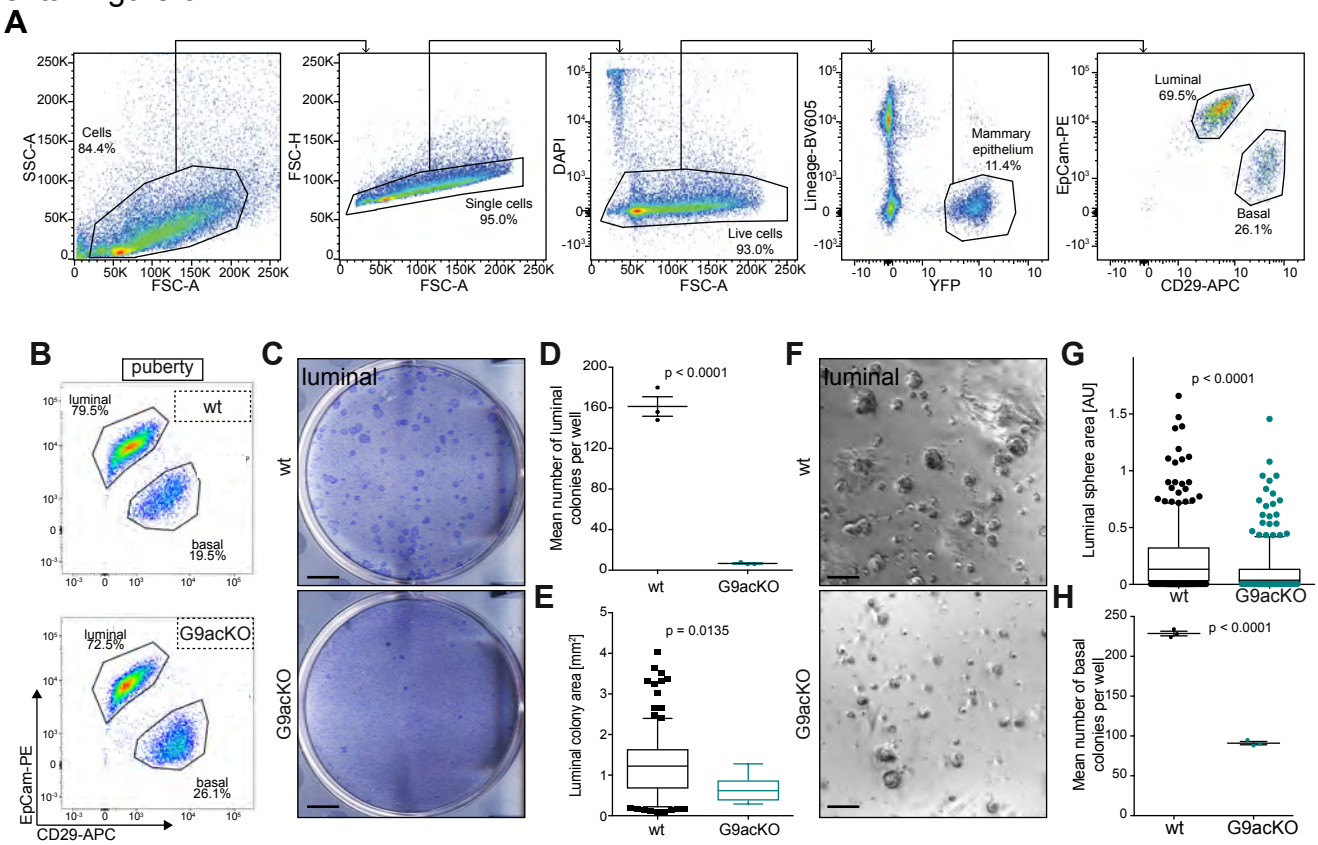
Supplemental Figure 1



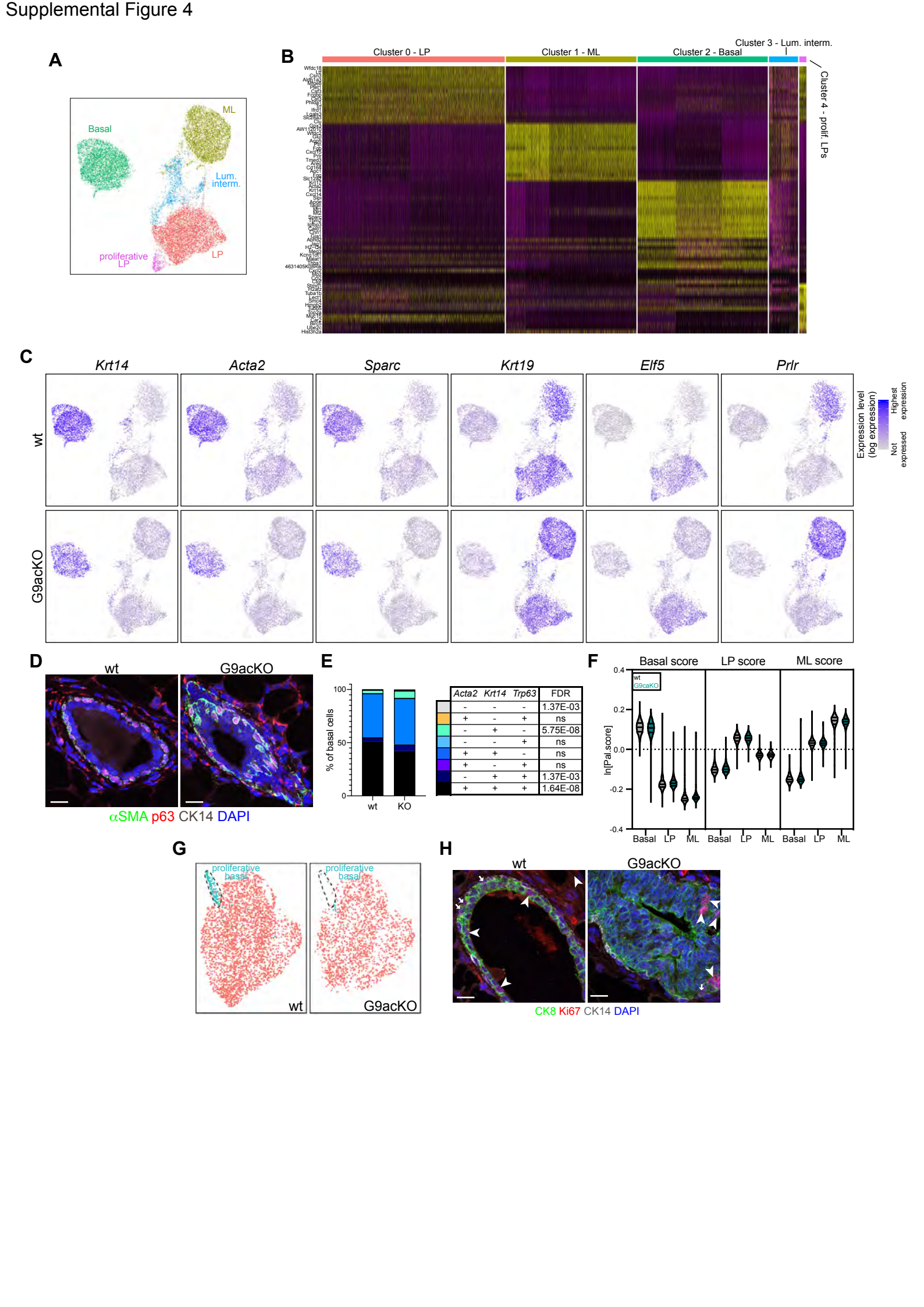
Supplemental Figure 2



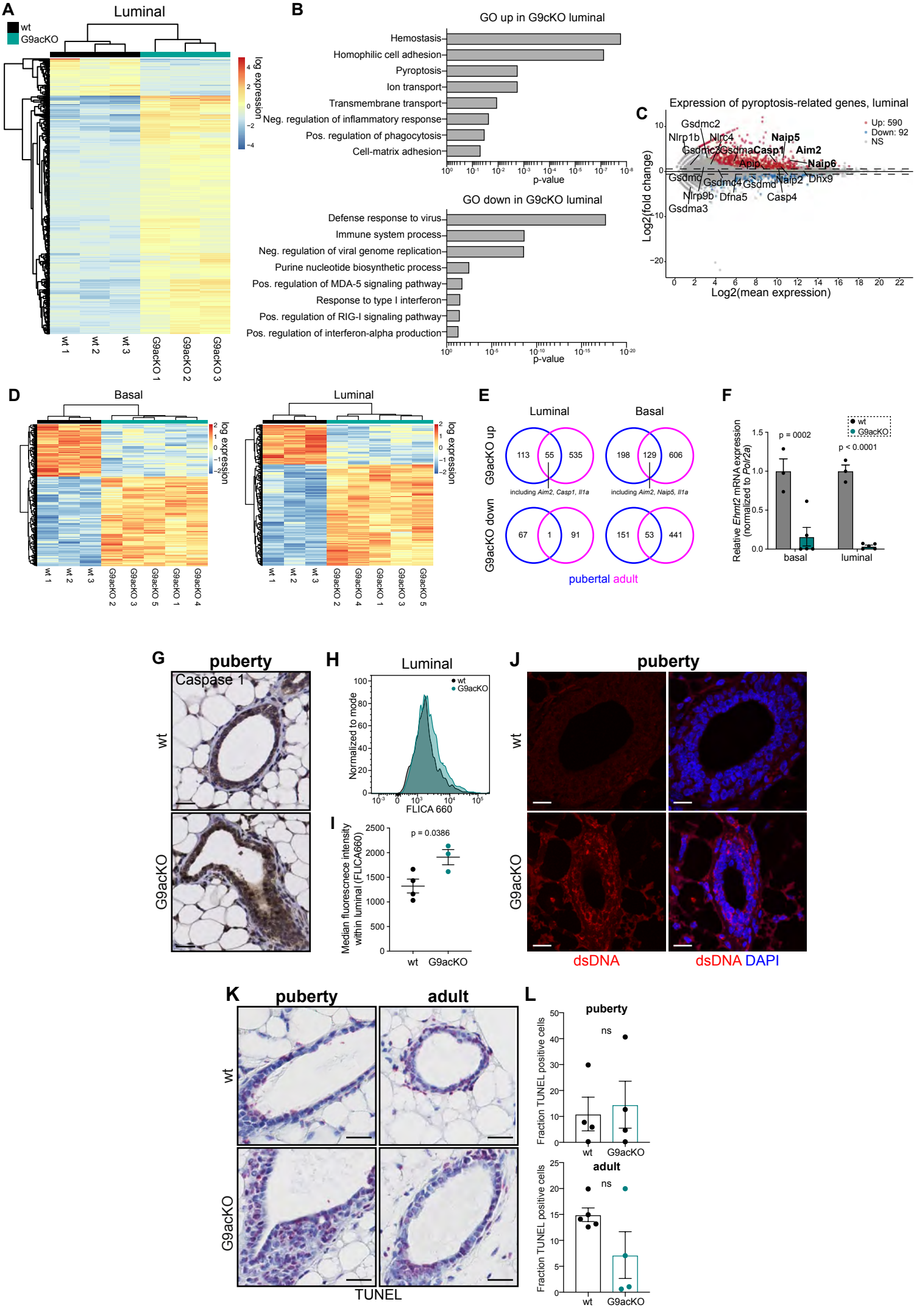
Supplemental Figure 3



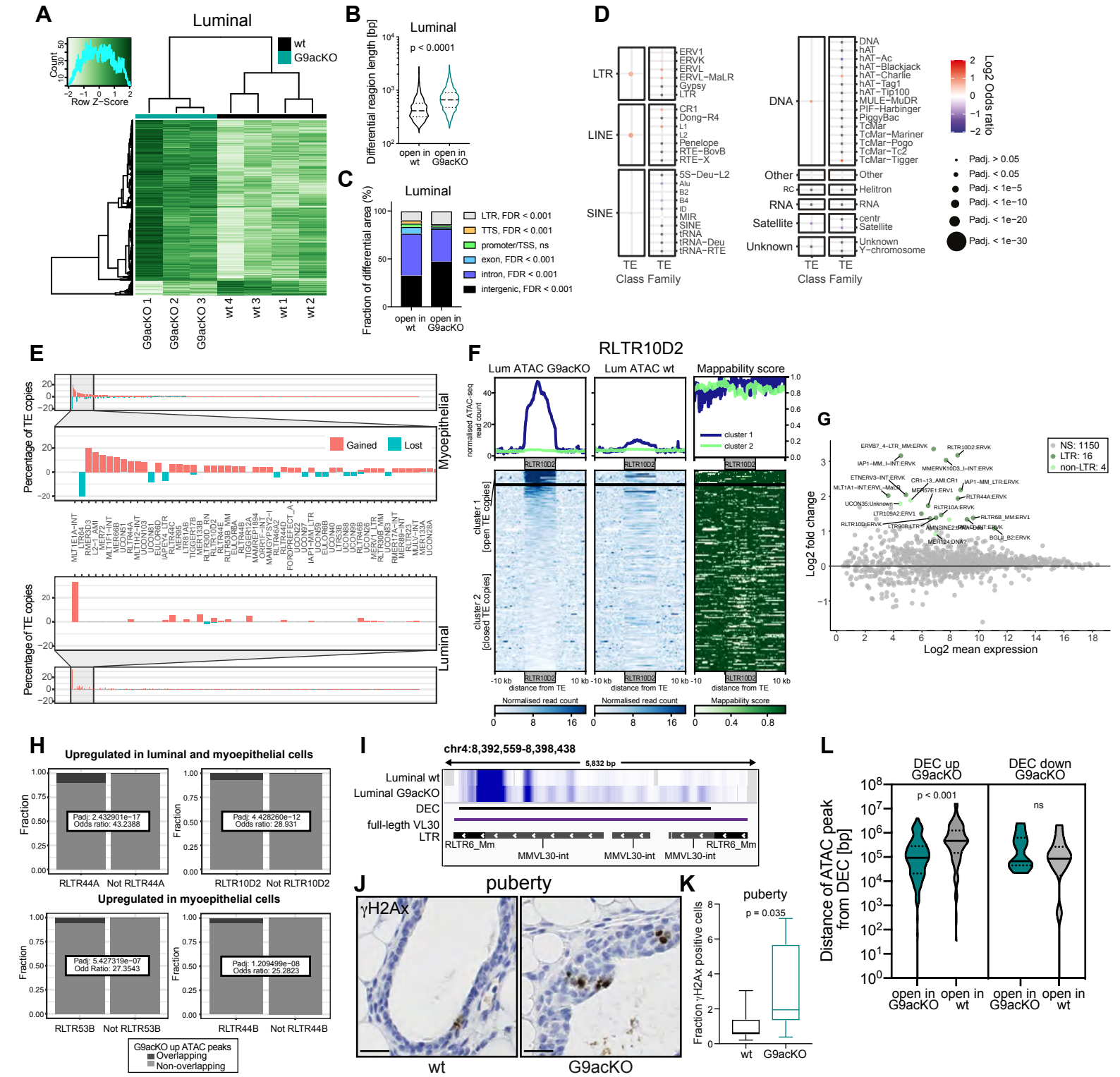
Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7