1 **Pregnancy Reduces Il33+ Hybrid Progenitor Accumulation in the Aged Mammary Gland**

2

3 Andrew Olander[1*], Cynthia M Ramirez[2*], Veronica Haro Acosta[1], Paloma Medina[1,3,5], Sara

4 Kaushik[1], Vanessa D Jonsson[3,4], Shaheen S Sikandar[1,4,5#]

5

6 **Affiliations**

7 [1]Department of Molecular, Cell and Developmental Biology, University of California - Santa Cruz

8 [2]Department of Applied Mathematics, University of California - Santa Cruz

9 [3]Department of Biomolecular Engineering, University of California - Santa Cruz

10 [4]Genomics Institute, University of California - Santa Cruz

11 [5]Institute for the Biology of Stem Cells, University of California - Santa Cruz

12

13 *These authors contributed equally to the work

14 [#]Corresponding author: Shaheen Sikandar, ssikanda@ucsc.edu

15

16 **ABSTRACT:**

17 Aging increases breast cancer risk while an early first pregnancy reduces a woman's life-long risk.

18 Several studies have explored the effect of either aging or pregnancy on mammary epithelial cells

19 (MECs), but the combined effect of both remains unclear. Here, we interrogate the functional and

20 transcriptomic changes at single cell resolution in the mammary gland of aged nulliparous and

21 parous mice to discover that pregnancy normalizes age-related imbalances in lineage

22 composition, while also inducing a differentiated cell state. Importantly, we uncover a minority

23 population of *Il33*-expressing hybrid MECs with high cellular potency that accumulate in aged

24 nulliparous mice but is significantly reduced in aged parous mice. Functionally, IL33 treatment of

25 basal, but not luminal, epithelial cells from young mice phenocopies aged nulliparous MECs and

26 promotes formation of organoids with *Trp53* knockdown. Collectively, our study demonstrates that

27   pregnancy blocks the age-associated loss of lineage integrity in the basal layer through a

28   decrease in *Il33+* hybrid MECs, potentially contributing to pregnancy-induced breast cancer

29   protection.

30

31   **INTRODUCTION**

32   A woman's risk of breast cancer increases with age (median age at diagnosis = 62 years[1]),

33   but an early first pregnancy (below the age of 30) significantly reduces lifetime risk[2–8]. Previous

34   studies that have examined changes shortly after pregnancy suggest that post-pregnancy

35   mammary epithelial cells (MECs) have a reduced stem cell capacity[9,10] and increased expression

36   of differentiation markers[11]. On the other hand, the studies that examined changes with aging in

37   nulliparous (no pregnancy) MECs have found an accumulation of dysfunctional luminal

38   progenitors[12], lineage infidelity[13], and altered differentiation programs[14,15]. Importantly, there is not

39   a clear understanding of how these processes combine and evolve with aging, i.e., how does

40   pregnancy alter the process of aging in mammary stem/progenitor cells?

41   Most of the previous studies focused on time points immediately post-pregnancy or 40

42   days post-involution (~3.3 human years), yet pregnancy-induced protection does not take effect

43   until almost 10 years post-birth[4,16,17]. To address this gap in knowledge, we simulated conditions

44   that mimic an early first pregnancy (20-30 years in humans, 3-8 months in mice) and the post-

45   menopausal stage (>50 years in humans, ~18 months in mice). The long-term effect of pregnancy

46   on aging of MECs is important to determine because 75% breast cancer diagnoses occur over

47   the age of 50[18], while most women in the United States have their first pregnancy between 20

48   and 33 years of age[19,20]. Using this 18-month time point in conjunction with a 3-month control, we

49   interrogate the aging process of MECs with and without pregnancy while removing confounding

50   factors affecting human samples[21]. Our study not only delineates the combined effects of aging

51   and pregnancy, but also uncovers a previously unknown *Il33+* hybrid MEC population that

52   accumulates with age and is reduced in aged mice that have undergone pregnancy.

53

RESULTS

Pregnancy induces long lasting changes in cell fate decisions and reduces organoid formation

To understand whether pregnancy alters the aging of the mammary gland, we compared 18-month-old parous mice (18m P) that have undergone multiple pregnancies between 3-8 months of age with 3-month-old nulliparous (3m NP) and 18-month-old nulliparous (18m NP) (**Fig. 1a**). H&E staining and whole mount imaging showed no major differences in ductal density, morphology, or complexity of branching between 18m NP and 18m P (**Extended Data Fig. 1a, b**).

The mammary gland is a bilayered tree-like structure composed of two main epithelial cell lineages, luminal and basal cells[22]. Luminal cells can be further subdivided into two functionally distinct cell types: hormone receptor (HR)-high and -low cells[23]. To understand how these populations evolve with aging and pregnancy, we performed flow cytometry analysis using previously established markers of MEC lineages, basal (CD49F^{high}/EPCAM$^{med-low}$) and luminal (CD49F^{med}/EPCAM^{high}) (**Extended Data Fig. 2**). We found that in 18m NP mice, there was a significant increase in the basal population (increased basal:luminal ratio) as compared to 3m NP mice (**Fig. 1b, d**), consistent with previous studies[24]. However, 18m P mice showed a basal:luminal ratio lower than 18m NP but still higher than 3m NP mice, suggesting that pregnancy partially normalizes the aged-induced expansion of basal cells (**Fig. 1b, d**).

Within the luminal compartment, our analysis showed a minor but non-significant increase in HR-low luminal cells (CD14+/CD33-) and a significant decrease in HR-high luminal cells (CD14-/CD133+) in the 18m NP mice (**Fig. 1c, e**), consistent with previous studies in 13-14 month mice[23]. However, in contrast to the normalization of basal:luminal ratio in 18m P mice, we found a further increase in HR-low and a decrease in HR-high luminal cells in these mice (**Fig 1c, e**). Thus, our data suggests that while the age-induced expansion of basal cells is normalized by pregnancy,

79 the luminal cells retain a residual involution program, with an increased proportion of CD14+ HR-

80 low luminal cells.

81   To understand whether there are functional differences in luminal and basal cells between

82 the 3m NP, 18m NP and 18m P mice, we performed *in vitro* organoid assays. In contrast to

83 previous studies that have shown an age-induced increase in organoid formation[24], we found that

84 3m and 18m NP MECs had similar capacity to form organoids (**Fig. 1, f-i**). However, both luminal

85 and basal cells from 18m P mice had a significantly decreased capacity of organoid formation as

86 compared to 18m NP mice (**Fig. 1f-i**). These findings suggest that MECs from the aged parous

87 glands are diminished in clonogenicity, supporting the hypothesis that pregnancy reduces

88 regenerative potential.

89

90 **Pregnancy promotes a differentiated cell state but also reverses aged-associated**

91 **transcriptional programs**

92   To determine transcriptomic changes in response to aging and pregnancy, we sorted

93 luminal and basal cells and performed bulk RNA sequencing from 3m NP, 18m NP, and 18m P

94 mice. Principal component analysis (PCA) of gene expression values showed that luminal and

95 basal samples formed two distinct clusters (**Fig. 2a**), as expected. Moreover, gene expression for

96 established lineage markers showed that sorting was specific for basal and luminal cells, with

97 enrichment of *Krt14*, *Krt5*, and *Acta2* in the basal cells and *Krt8*, and *Krt18* in luminal cells (**Fig.**

98 **2b**).

99   To understand the long-term effects of pregnancy on gene expression in MECs, we

100 performed differential gene expression analysis between 18m NP and 18m P for basal and

101 luminal cells. Although luminal cells undergo significant functional and transcriptomic changes

102 throughout pregnancy[25], we found only 45 differentially expressed genes (DEGs) at p-adj of =<

103 0.1 that are persistently altered in luminal cells (**Data Table 1**). Interestingly, we found 74 DEGs

104 in basal cells at the same p-adj threshold (**Data Table 2**). Previous studies suggest that parity has

105 the most pronounced impact on the transcriptome of luminal cells[25], but our data suggest that

106 basal cells also retain a robust transcriptional memory of pregnancy that persists with age. Basal

107 cells from 18m P mice displayed consistent upregulation of genes associated with differentiation

108 towards a mesenchymal lineage and contractile functions, such as *Myod1, Fmod,* and *Vim* (**Fig.**

109 **2c**). Moreover, 18m P basal cells had increased expression of milk-protein genes, such as

110 *Csn1s1* and *Csn2*, suggesting that basal cells upregulate pathways of milk-production during

111 pregnancy or give rise to milk producing cells[26], or vice versa[27]. We also detected a modest down

112 regulation of genes associated with stem/progenitor cells (*Tspan8* and *Bcl11b*)[28,29] and growth

113 factor receptors (*Erbb3* and *Egr2*). Likewise, luminal cells from 18m P mice had higher expression

114 of genes involved in alveolar differentiation (*Csn1s1, Csn2, Car2,* and *Apobr*) and down regulation

115 of genes involved in stem/progenitor function (*Runx3* and *Bst1*, **Fig. 2d**). Taken together, these

116 results suggest that pregnancy establishes long-term transcriptomic changes that reflect

117 increased differentiation not only in luminal cells but also in basal cells. Interestingly, by comparing

118 the gene expression patterns between 3m NP, 18m NP and 18m P mice we found that expression

119 of *Tspan8*, that marks hormone-responsive stem cells[28] and growth factor receptors (*Erbb3* and

120 *Egr2*) increases with aging in basal cells but is reduced in 18m P mice (**Fig. 2c**). Similar patterns

121 were observed for luminal cells in stem/progenitor genes (*Runx3* and *Bst1*), the RANKL decoy

122 receptor, *Tnfrsf11b,* and the *Wnt* inhibitor, *Nkd2*, among others (**Fig. 2d,** see Data tables 1-4).

123   To identify specific pathways that are altered by pregnancy and persist with age, we

124 performed gene ontology analysis on basal or luminal cells in 18m P and 18m NP mice. We found

125 significant enrichment of genes involved in cytoskeletal remodeling, extracellular matrix, and

126 contractile functions (**Fig. 2e**). Similarly, 18m P luminal cells showed enrichment of pathways

127 involved in regulation of peptide metabolism, antimicrobial immune responses, and collagen

128 structure (**Fig. 2f**), which have been proposed to play key roles during alveologenesis,

129 involution[30], and immune development of offspring[31,32]. These data suggest that pregnancy

130 normalizes age-induced transcriptional changes in luminal and basal epithelial cells, while also

131 inducing a differentiated state.

132

Single-cell RNA-sequencing identifies a minority population of hybrid MECs in aged

nulliparous mice

135 Recent studies have demonstrated that minority populations of stem/progenitor cells

136 evolve with age and are likely precursors of tumor initiation[24,33]. To investigate the cellular

137 heterogeneity and aging-associated transcriptional changes that are altered with pregnancy at

138 single-cell resolution, we performed single-cell RNA-sequencing (scRNA-seq) on mammary

139 glands from 18m NP and 18m P mice. To increase the power of our analysis, we integrated single

140 cell transcriptomes from the Tabula Muris[34] 18m mammary gland dataset (see methods). After

141 filtering cells with low-expressing genes, high mitochondrial counts, and doublets (see methods),

142 we identified epithelial cells (basal, HR-high luminal, HR-low luminal), immune cells (B cells, T

143 cells, macrophages), endothelial cells, and fibroblasts using the expression of well-established

144 marker genes (**Fig. 3a, Extended Data Fig. 3a**). Intriguingly, we identified a minority population

145 of cells co-expressing basal and luminal markers (hybrid) (**Fig. 3c**), that are present in 18m NP

146 mice (~90% of cell cluster) but are significantly reduced in 18m P mice (~10% of cell cluster) (**Fig.**

147 **3b**). Moreover, we found no differences in the proportion of B cells, T cells, macrophages,

148 endothelial cells, and HR-low luminal cells between 18m P and 18m NP (**Fig. 3b**). However, 18m

149 P had a lower proportion of basal cells and HR-high luminal cells (**Fig. 3b**), consistent with our

150 flow cytometry analyses (**Fig. 1b-e**). Moreover, we did not find differences in cell cycle between

151 18m NP and 18m P mice (**Extended Data Fig. 3b**).

152 To investigate the unknown hybrid population further, we calculated basal cell and luminal

153 cell gene signature enrichment scores for all cells based on previously published transcriptomic

154 analyses of adult MECs[35]. We found that the hybrid MEC population lies between the luminal and

155 basal populations, implying that hybrid MECs display a mixed gene signature of both basal and

156   luminal lineages (**Fig. 3d**). Pseudotime inference using PAGA[36] and Slingshot[37] also predicted

157   the hybrid MECs lineage trajectory between basal and luminal populations (**Fig. 3e, Extended**

158   **Data Fig. 4a, b**).

159   To understand whether these hybrid MECs represent an immature cell type, we applied

160   CytoTRACE2[38], to agnostically predict cellular potency. Intriguingly, we found that 33% hybrid

161   MECs are predicted to be multipotent, while majority of basal and luminal cells were predicted to

162   be oligopotent and differentiated, respectively. Moreover, the average cellular potency of hybrid

163   MECs was between basal and luminal cells (**Fig. 3f, Extended Data Fig. 4c**). Intriguingly, the

164   hybrid MECs are also enriched in expression of *Krt6a* (**Extended Data Fig. 5a**), which has been

165   previously identified as a marker of bipotent luminal progenitor cells in the mammary gland[39].

166   KRT6a+ cells are normally found only in the luminal layer and do not express the basal marker

167   *Krt14*. However, a previous report showed that a population of KRT14+/KRT6a+ cells expands

168   during early stages of pregnancy[40]. Analysis of the Pal et al. (2017) scRNA-seq dataset containing

169   all mammary gland developmental stages showed that *Krt6a* expression peaks during early

170   pubertal development, particularly within terminal end buds (TEBs)[41] (**Extended Data Fig. 5b, c**).

171   Taken together, these data suggest that aging coincides with aberrant activation of *Krt6a*

172   expression likely conferring a cell state similar to TEBs, but this is diminished by pregnancy.

173

174   **Hybrid KRT6a+ cells are present in the basal epithelium of aged nulliparous mice but not**

175   **aged parous mice.**

176   To expand our analysis from mRNA to protein, we performed immunofluorescence

177   staining on 3m NP, 18m NP, and 18m P mammary glands for KRT6a, together with KRT8 (luminal

178   marker) and KRT5 (basal marker). We found a greater than 6-fold increase in the percentage of

179   KRT6a+ cells in the 18m NP (6.51%) mice relative to 3m NP (~1.05%) (**Fig. 3g, h**), suggesting

180   that these hybrid MECs accumulate over the course of aging. However, 18m P mice had a small

181   percentage of KRT6a+ cells (<1%), less than the 3m NP (**Fig. 3g, h**), consistent with our single-

182   cell analysis. Moreover, we found an increased percentage of KRT5+ basal cells and decreased

183   percentage of KRT8+ luminal cells in 18m NP mice (**Extended Data Fig. 5d**) consistent with our

184   flow cytometry data (**Fig. 1b, d**), but 18m P cells were more similar to 3m NP cells.

185   To further understand the localization of the KRT6a+ cells, we quantified the percentage

186   of KRT6a+ cells in the luminal and basal layer (i.e., KRT6a+/KRT8+ or KRT6a+/KRT5+). We

187   found that 95% of KRT6a+ cells in 3m NP glands were positive for KRT8 and resided in the

188   luminal layer (**Fig. 3i**), consistent with previous studies[39]. In contrast, over 50% of KRT6a+ cells

189   in the 18m NP gland were positive for KRT5 and localized to the basal layer. Interestingly, in the

190   18m P mice, the localization of the KRT6a+ cells was closer to 3m NP mice (~10%, **Fig. 3i**). This

191   suggests that the KRT6a+ hybrid MECs that accumulate with age are likely due to the loss of

192   lineage integrity of basal cells, but this integrity is maintained in mice that have undergone

193   pregnancy.

194

195   **IL33 treatment of young basal cells phenocopies aged nulliparous mammary glands**

196   In addition to *Krt6a*, the hybrid MEC population identified in 18m NP mice was uniquely

197   enriched in the expression of *Interleukin 33 (Il33)* (**Fig. 3c, Extended Data Fig. 6a**), also known

198   as alarmin. IL33 has been previously implicated in activating pro-tumorigenic signaling

199   pathways[42–44], cancer stem cell maintenance[45], and establishing an immunosuppressive

200   microenvironment[46], but its role in the normal mammary gland is unknown. Analysis of the Pal et

201   al. (2017) scRNA-seq dataset showed that *Il33* is also highly expressed in the TEBs during

202   puberty, (**Extended Data Fig. 6b, c**), similar to Krt6a (**Extended Data Fig. 5b, c**) suggesting that

203   aging likely upregulates an early developmental program in basal cells.

204   To determine whether IL33 treatment affects primary MECs *in vitro*, we cultured basal

205   (CD49f^hi/EPCAM^lo/med) or luminal (CD49f^lo/EPCAM^hi) cells from 3m NP mice with recombinant

206   IL33. Treatment of basal and luminal cells with IL33 resulted in an increased number of basal-

207   derived organoids (**Fig. 4a**) but no increase in organoids derived from luminal cells (**Extended**

208 **Data Fig. 6d**). Interestingly, IL33 treatment of basal organoids for 2 weeks inhibited spontaneous

209 *in vitro* luminal cell differentiation, resulting in a ~2-fold increase in the proportion of basal cells

210 (**Fig. 4b**). Moreover, IF staining and 3D imaging demonstrated an increase in the percentage of

211 KRT6a and KRT5 upon treatment with IL33 (**Fig. 4c, Extended Data Fig. 6e**), but no significant

212 change in KRT8+ luminal cells (**Extended Data Fig. 6f, g**). These results suggest IL33 treatment

213 of 3m NP basal cells phenocopies the increased proportion of basal cells and accumulation of

214 KRT6a+ hybrid MECs observed in aged nulliparous mice (**Fig. 3i**).

215 To test whether IL33 treatment increases clonogenicity in cells with tumor-promoting

216 mutations, we infected 3m NP basal and luminal cells with lentiviral shRNA targeting *Trp53*[47] and

217 exposed them to IL33. Consistent with our results in normal epithelial cells (**Fig. 4a, Extended

218 Data Fig. 6d**), we found that *Trp53* knockdown resulted in increased clonogenicity of basal cells

219 when treated with IL33 but not luminal cells (**Fig. 4d**). Thus, our data suggest that the age-induced

220 increase in *Il33*+ hybrid MECs likely confers a survival advantage as cells acquire oncogenic

221 mutations.

222

223 **DISCUSSION**

224 Our study resolves the complex relationship between aging and pregnancy in the

225 mammary gland, such as basal-cell bias and differentiation. We report that pregnancy not only

226 normalizes the age-induced expansion of basal cells but also concurrently reduces the capacity

227 of basal cells to form organoids. In contrast, luminal cells in aged parous mice retain an involution

228 specific signature, which potentially makes them more susceptible to immune surveillance[48–51].

229 Hybrid MECs expressing luminal and basal genes have been identified previously and are

230 associated with a less-differentiated phenotype, higher plasticity, and are capable of tumor

231 initiation[52–55] but precise molecular signatures of these cells are not yet defined. Our work

232 identifies a previously unknown *Il33*+ hybrid MEC population that accumulates with age in the

233 basal layer. Majority of studies on breast tumor initiation implicate luminal cells as the cell of

234   origin[56], that lose lineage integrity to express basal markers during tumor initiation and also with

235   aging[12–15]. However, our study suggests that basal cells can acquire hybrid features with aging

236   by turning on pathways active in luminal cells during early pubertal development. In support of

237   this model, studies have shown that basal cells can acquire hybrid phenotypes upon

238   transformation[57] and form luminal-like tumors[52,54], which are the predominant subtypes found in

239   postmenopausal breast cancers[58]. Future work will determine the precise contribution of basal

240   cells to tumor initiation with aging.

241   The postnatal mammary gland development and maintenance is largely driven by lineage

242   committed progenitor cells that expand during pregnancy to form the milk-producing cells[59–61].

243   Studies have also shown that there are rare bipotent mammary stem cells in the basal layer that

244   can expand with pregnancy and contribute to the alveolar lineage[61,62]. We speculate that during

245   pregnancy some basal cells either turn on an alveolar gene expression program or directly

246   contribute to the alveolar lineage, leading to the normalization of the basal:luminal ratio and a

247   decrease in the hybrid MEC population in aged parous mice.

248   IL33 has been extensively studied in inflammation and immune cell responses[63] but its

249   role in epithelial cell plasticity is unknown. Our results show that IL33 increases organoid

250   formation capacity and induces a KRT6a+ hybrid state in basal cells. Recent work has uncovered

251   a role for IL33 in pancreatic cancer where *Il33* expression is induced post-injury and cooperates

252   with mutant *Kras* to promote neoplastic transformation[64]. Notably, when cells acquire hybrid

253   signatures due to expression of *Pik3ca^{H1047R}* in luminal cells, they dramatically increase the

254   expression of *Il33*[54]. Moreover, studies have demonstrated a cell-intrinsic role of IL33 in B-cell

255   development[65], repair during skin injury[66], and promoting inflammation in chronic pancreatitis[67].

256   These studies, combined with our data, suggest that *Il33* expression promotes a cell-state that is

257   proliferative and plastic, allowing for injury repair or reprogramming during transformation. Future

258   studies will determine the cell intrinsic role of *Il33* during mammary gland development, in

259   establishing a hybrid cell state, and tumorigenesis.

260 **METHODS**

261 **Mice**

262 Wild-type C57BL/6 mice (young), retired breeders and aged matched mice when appropriate

263 were purchased from Charles River Laboratories and The Jackson Laboratory. Simultaneously

264 an aging colony was maintained in house. The mice were recorded for estrous and 18m mice

265 were found to be non-cycling. All mice used for this study were maintained at the UCSC Animal

266 Facility/Vivarium in accordance with the guidelines of the Institutional Animal Care and Use

267 Committee (Protocol #Sikas2311dn).

268

269 **H&E staining and quantification**

270 H&E staining was performed as previously described[68]. Briefly, mammary glands were dissected

271 from parous and nulliparous mice, fixed in 4% PFA and paraffin embedded for histology. Sections

272 underwent a standard staining protocol, with five-minute incubation in 100% xylene and ethanol

273 (100%, 70%) followed by DI water for deparaffinization. Sections were stained with Hematoxylin

274 for 30 seconds, then bluing solution (NH_3OH + MilliQ H2O, 2 mins) and Eosin for 1min followed

275 by dehydration and mounting (Permount). Visualization of mammary gland ducts was performed

276 using the Leica Widefield microscope, and ductal quantification was conducted using Fiji.

277

278 **Mammary gland whole mounts and branching analysis**

279 Inguinal mammary glands were removed, adhered to a Superfrost Plus Microscope slide

280 (Fisherbrand Cat. No. 1255015) and incubated in Carnoy's Fixative overnight at room

281 temperature. Mammary glands were then incubated in 70% ethanol, 50% ethanol, then DI water

282 for 15 minutes each. Fixed mammary glands were stained overnight with Carmine Alum at room

283 temperature, then dehydrated in 15-minute incubations in increasing ethanol grades, followed by

284 dehydration in xylene and mounted in Permount.

285   Images of the entire mammary gland wholemount for both parous and nulliparous samples were

286   imported as a TIFF file onto the Fiji/ImageJ software, where a three-by-three tiled ROI was chosen

287   from a consistent distance from the lymph node. The ROI was processed to remove background

288   noise. The image was skeletonized, and the branches dilated to improve clarity. A Sholl analysis

289   (neuroanatomy) plugin was used on the skeletonized image, and a vertical radius from the side

290   of the image was set as a consistent measurement throughout all samples to account for

291   decreased branching normally seen at the edges of an image. After setting the starting radius as

292   zero microns and selecting the best-fitting Sholl methods, the Sholl analysis program was run.

293   This generated a "Sholl Log Plot," which indicated branching intersections over an area, as well

294   as a "Sholl Regression Coefficient (k)", which described branching complexity. A k value closer

295   to zero indicated a higher branching complexity. A "Sholl Mask (heat map)" was also generated

296   to visually depict regions of higher branching complexity, where the red areas indicated maximal

297   branching.

298

299   Immunofluorescent staining of mammary gland paraffin sections

300   Mammary glands were fixed in 4% paraformaldehyde (PFA) in PBS and embedded in paraffin for

301   immunostaining. 5μm sections were deparaffinized, dehydrated, and autoclaved for 15 min in

302   Tris-EDTA buffer [10mM Tris 1mM EDTA (pH 9.0)] for antigen retrieval. Tissue sections were

303   incubated overnight at 4°C with primary antibodies diluted in tris-buffered saline (TBS) + 5%

304   bovine serum albumin (BSA) (antibodies are listed in table 5). Samples were subsequently

305   washed (2X) with TBS + 0.05% Tween for 10 min and were incubated with donkey anti-rat Alexa

306   Flour 647 (1:400), donkey anti-chicken Alexa Flour 488 (1:400), donkey anti-rabbit Alexa Flour

307   594 (1:400) conjugated secondary antibodies (Jackson ImmunoResearch Laboratories) in TBS +

308   5% BSA 0.1% Tween for 1 hour at room temperature (RT). Samples were subsequently washed

309   (3X) with TBS + 0.1% Tween and were incubated with DAPI in TBS + 5% BSA + 0.1% Tween

310   (1:10,000) for 10 min. All the immunofluorescence sections and cells were mounted in

311   Fluoromount-G (Genesee). Images were acquired by a Solamere Spinning Disk Confocal

312   microscope. Images were processed using Fiji.

313

Immunofluorescent staining of primary MEC organoids

315   Organoids were collected at day 12-14 in culture and recovered from the growth factor-reduced

316   Matrigel by incubating with Cell Recovery Solution (Corning) for 1 hour on ice with gentle shaking.

317   5mL collection tubes and pipette tips were precoated in PBS + 5% BSA. Organoids were washed

318   with cold PBS (2X) before being fixed with 4% PFA in PBS for 45 min on ice. Fixed organoids

319   were then washed in PBS (1X) and incubated with 0.2% glycine in PBS for 20 min at room

320   temperature. After a final wash, fixed organoids were resuspended in 0.05% sodium azide in PBS

321   and stored at 4°C.

322   Organoids were permeabilized in cold 100% methanol for 10 minutes on ice, washed PBS (1X),

323   and incubated in blocking buffer (0.1% BSA, 0.3% Triton-X, 5% normal donkey serum) for 3h at

324   room temperature with light shaking followed by overnight incubation with primary antibodies in

325   blocking buffer at 4°C with light shaking. The next day they were washed (3X) in PBS + 0.3%

326   Triton-X and incubated overnight with secondary antibodies in a blocking buffer at 4°C with light

327   shaking in the dark. Organoids were again washed (3X) in PBS + 0.3% Triton-X and incubated

328   with DAPI for 10 min at room temperature. After a final round of washes (3X), organoids were

329   transferred to μ-Slide 8 Well Glass Bottom slide (Ibidi) pre-coated in poly-L-lysine (Sigma-Aldrich)

330   and imaged on a Solamere Spinning Disk Confocal microscope. Images were processed using

331   Fiji.

332

Organoid assays

334   Primary MECs were sorted into complete organoid media as previous described[68,69] (Advanced

335   DMEM F/12, 10% FBS, 1% PSA, 50ng/mL EGF, 100ng/mL Noggin, 250ng/mL R-Spondin-1, 1X

336   N2, 1X B27, 1X GlutaMAX, 10mM HEPES) supplemented with Y-27632 (10uM). 1,000 cells were

337 plated in 96-well ultra-low attachment plates seeded with a 50uL mixture of growth factor-reduced

338 Matrigel and irradiated L-Wnt3a-secreting-3T3 feeder cells (11,000 feeder cells per 50uL growth

339 factor-reduced Matrigel). Organoids were cultured at 37C, 5% CO_2 with added humidity. On day

340 4, Y-27632 supplemented media was removed and replaced with complete organoid media.

341 Fresh media was added every other day thereafter. Basal and luminal organoids were imaged at

342 day 6 and 10 in culture, respectively, on a Zeiss Live Cell microscope and analyzed on Biodoc.AI

343 to collect data on organoid count and average size. For IL33 treatment experiments, organoids

344 were cultured as described above with the addition of IL33 (0ng/mL, 5ng/mL, 10ng/mL).

345

Mammary gland digestion and processing

347 L2-5 and R2-5 mammary glands were harvested, minced, and chemically digested overnight in

348 Advanced DMEM F/12 with 1% PSA, gentle collagenase/hyaluronidase, and DNAse I at 37C, 5%

349 CO_2, and added humidity as previously described[70]. Briefly, partially digested glands were then

350 mechanically digested by pipetting with a serological pipette until no tissue pieces were visible.

351 Digested glands were washed with staining buffer (Hank's Balanced Salt Solution, 2% Bovine

352 Calf Serum, 1% PSA) and centrifuged (1500RPM) at 4C for 5 minutes. Red blood cells were lysed

353 with 5mL of ACK Lysis buffer for 5 minutes, and cells were washed with 15mL of staining buffer.

354 Cells were treated with 0.25% Trypsin with EDTA and gently pipetted continuously for 2-3 minutes

355 to digest the basement membrane. Cells were then treated with DNAse I and Dispase and

356 pipetted continuously for 2-3 minutes to prevent clumping. The single-cell suspension was then

357 filtered through a 40um mesh strainer and pelleted via centrifugation (1500RPM, 4C, 5 minutes).

358 Cells were then resuspended in staining buffer and transferred to FACS tubes for staining.

359

Fluorescence-activated cell sorting (FACS)

361 Cells were stained with antibodies listed in Supplemental Table 5 for 15 minutes at room

362 temperature, as previously described[68,69]. Stained cells were then washed with a staining buffer,

363   resuspended with DAPI (1:10,000), filtered, and analyzed on a BD Biosciences FACSAria cell

364   sorter. See Extended Data Fig. 2 for FACS gating strategies. Data were analyzed using FlowJo

365   software (10.10.0).

366

367   **Lentiviral Infection**

368   For *Trp53* knockdown experiments, pSicoR-GFP-sh*Trp53* was a gift from Tyler Jacks (Addgene

369   plasmid #12090; http://n2t.net/addgene:12090 ; RRID:Addgene_12090). Viruses were produced

370   in 293T cells using the second- generation lentiviral system and transfection using Lipofectamine

371   2000 (Life Technologies) as previously described[70]. Supernatants were collected at 48 hours,

372   filtered with a 0.45µm filter, and precipitated with lentivirus precipitation solution (Alstem LLC) per

373   the manufacturer's instructions. Viral titers were determined by flow cytometry analyses of 293T

374   cells infected with serial dilutions of concentrated virus.

375

376   **Single-cell RNA-seq cell preparation and sequencing**

377   Libraries were generated using the 10x Genomics Chromium Next GEM Single Cell 5' Reagent

378   Kits v2 (Dual Index) (cat#1000264) following the manufacturer's protocol. Single-cell suspensions

379   were prepared from mammary glands as described above and encapsulated using the Chromium

380   Controller to generate Gel Beads-in-Emulsion (GEMs), allowing for barcoding of individual cells.

381   Post-GEM generation, reverse transcription (GEM-RT) was performed within each GEM, followed

382   by cleanup and cDNA amplification. The cDNA underwent fragmentation, end repair, and A-

383   tailing. Dual index adapters were ligated to the cDNA fragments, which were then amplified using

384   PCR. The libraries were quantified using a Qubit Flex Fluorometer (Invitrogen Q33327) and their

385   size distribution was assessed using an Agilent TapeStation 4200. Sequencing libraries were

386   constructed to target 200 million paired-end reads (400 million total) to achieve a coverage of

387   20,000 reads per cell from a targeted capture of 10,000 cells. Sequencing was performed on an

388   Illumina NovaSeq 6000 following the manufacturer's recommendations.

Single-cell data pre-processing and quality control

Sequencing data was processed using the 10x Genomics Cell Ranger pipeline. Reads were aligned to the mouse reference genome (mm10). We performed quality control for downstream analysis, removing (1) genes that were detected in less than 3 cells, (2) cells with less than 200 genes, (3) cells with gene counts < 600 or > 8,000, (4) cells with total counts of UMIs per cell < 2,000 or > 12,000, and (5) cells with mitochondrial gene ratio > 1.5%. The mitochondrial gene ratio is defined as the percentage of UMIs mapped to mitochondrial genes compared to non-mitochondrial genes within each cell. Doublets were identified using Scrublet[73], resulting in the removal of 588 cells. Data preprocessing and analysis steps below were implemented using the Scanpy framework version 1.19[71].

Single cell data integration, dimension reduction, and cell type annotation

We combined our single cell data set with an 18 month nulliparous pre-processed single cell data set from Tabula Muris Senis[34]. The UMI counts for each cell were normalized using a target sum of 1e4, and log transformed, with an added pseudocount of 1. This resulted in a combined dataset of 10,001 cells and 13,892 genes. Principal component analysis (PCA) was conducted to produce a reduced dataset that was used as input to correct for technical variation due to samples using Harmonypy, version 0.0.4[72]. Neighborhood graphs were calculated on batch corrected reduced data and used to conduct Leiden[73] unsupervised clustering, with a resolution of 0.09, resulting in the identification of nine distinct clusters. These clusters were manually annotated using well-established marker genes. Data was visualized using Uniform Manifold Approximation and Projection (UMAP, version 0.5.2)[74].

Cell-cycle analysis, basal and luminal gene scoring

Cell cycle enrichment analysis was conducted using the Regev lab cell cycle gene list[75] to classify each cell into its corresponding cell cycle phase. To classify epithelial cells, we used previously

415 published marker genes identified for Basal, HR-high Luminal, and HR-low Luminal cells from

416 transcriptomic analyses of adult mammary epithelial cells (MECs)[35]. Enrichment scores for these

417 marker genes were calculated for each of the four epithelial cell types (Basal epithelial, HR-high

418 Luminal, HR-low Luminal, and Hybrid MECs). Cell cycle and epithelial cell type classifications

419 were conducted using the 'sc.tl.score_genes' function from the Scanpy library[71].

420

Pseudotime inference and cellular potency prediction

421

422 Pseudotime analysis was performed on the epithelial cell types (basal, HR-low, HR-high, and

423 hybrid). Partition-based graph abstraction (PAGA, version 1.2)[36] analysis was performed using

424 Scanpy (version 1.9.6). The diffusion map was computed using the basal cell type as the root cell.

425 To ensure robustness, pseudotime trajectory was also inferred using Python's implementation of

426 Slingshot (pyslingshot, version 0.1.3)[37], with basal cells as the start node. Slingshot pseudotime

427 ordering scores were scaled between 0 and 1 for comparison with PAGA results. To characterize

428 cellular potency, we used Python's implementation of CytoTRACE 2 (version 1.0.0)[38]. Figures

429 were generated using a combination of Scanpy[71], Seaborn[76], and matplotlib python packages[77].

430 To assess the statistical significance of the differences in pseudotime between epithelial cell

431 types, we performed pairwise comparison using the Wilcoxon Rank Sum Test using Benjamini-

432 Hochberg correction method to adjust the p-values and control the false discovery rate. For this

433 analysis we used 'scipy.stats.mannwhitneyu' function[78] for the Wilcoxon Rank Sum Test and

434 'multipletests' from the 'statsmodels.stats.multitest' library for Benjamini-Hochberg testing

435 correction[79].

436

Library Construction, Quality Control and Bulk RNA Sequencing

437

438 Bulk RNA sequencing was performed by Novogene on sorted basal (CD49f[hi]/EPCAM[low-med]) and

439 luminal (CD49f[low]/EPCAM[hi]) populations from 3m NP, 18m NP and 18m P mice (n=3 mice/group).

440 RNA was isolated according to manufacturer's instructions (Qiagen RNEasy Plus Micro Kit, Cat.

441 No. 74034). Messenger RNA was purified from total RNA using poly-T oligo-attached magnetic

442 beads. After fragmentation, the first strand cDNA was synthesized using random hexamer

443 primers, followed by the second strand cDNA synthesis using either dUTP for directional library

444 or dTTP for non-directional library[80]. For the non-directional library, it was ready after end repair,

445 A-tailing, adapter ligation, size selection, amplification, and purification. For the directional library,

446 it was ready after end repair, A-tailing, adapter ligation, size selection, USER enzyme digestion,

447 amplification, and purification. The library was checked with Qubit and real-time PCR for

448 quantification and bioanalyzer for size distribution detection. Quantified libraries will be pooled

449 and sequenced on Illumina platforms, according to effective library concentration and data

450 amount.

451

452 **Data Quality Control**

453 Raw data (raw reads) of fastq format were processed through fastp software. In this step, clean

454 data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N

455 and low quality reads from raw data. Q20, Q30 and GC content were calculated. All the

456 downstream analyses were based on the clean data with high quality.

457

458 **Reads mapping to the reference genome**

459 Reference genome (GRCm39/mm39) and gene model annotation files were downloaded from

460 genome website directly. Index of the reference genome was built usingHisat2 v2.0.5 and paired-

461 end clean 1 reads were aligned to the reference genome using Hisat2 v2.0.5. We selected

462 Hisat2[81] as the mapping tool for that Hisat2 can generate a database of splice junctions based on

463 the gene model annotation file and thus a better mapping result than other non-splice mapping

464 tools.

465

466 **Quantification of gene expression level**

467   featureCounts[82] v1.5.0-p3 was used to count the reads numbers mapped to each gene. Gene

468   expression was then converted to Fragments Per Kilobase of transcript sequence per Millions

469   base pairs sequenced (FPKM), which takes the effects into consideration of both sequencing

470   depth and gene length on counting of fragments. Analysis was performed by Novogene.

471

472   **Differential expression analysis**

473   Differential expression[83] analysis of two conditions (three biological replicates per condition) was

474   performed using the DESeq2Rpackage[84] (1.20.0) on raw gene expression values by Novogene.

475   DESeq2 provides statistical routines for determining differential expression in digital gene

476   expression data using a model based on the negative binomial distribution. Differentially

477   expressed genes were identified by comparing 3m NP vs. 18m NP and 18m P vs 18m NP for

478   both luminal and basal cell samples. The resulting P-values were adjusted using the Benjamini-

479   Hochberg's approach for controlling the false discovery rate. Genes with an adj-p<=0.1 found by

480   DESeq2 were assigned as differentially expressed. Principal Component Analysis (PCA) plots

481   and heatmaps of differentially expressed genes were generated using the NovoMagic analysis

482   platform provided by Novogene. PCA plots are based on FPKM normalized gene expression

483   values. Heatmaps of differentially expressed genes display $log_2(FC+1)$ values. Representative

484   genes from the top 30 DEGs are shown. See Data tables 1-4 for complete gene list.

485

486   **Enrichment analysis of differentially expressed genes**

487   Gene Ontology[85] (GO) enrichment analysis of differentially expressed genes was implemented by

488   the cluster Profiler R package, in which gene length bias was corrected. GO terms with corrected

489   p-value less than 0.1 were considered significantly enriched by differential expressed genes.

490

491   **Statistical analysis**

492   All graphs display the average as central values, and error bars indicate ± SD unless otherwise

493   indicated. P values are calculated using paired or unpaired t test, ANOVA, Wilcoxon rank-sum

494   test, and Mann-Whitney U test, as indicated in the figure legends. All P and Q values were

495   calculated using Prism (10.2.2) or Python, unless otherwise stated. For animal studies, sample

496   size was not predetermined to ensure adequate power to detect a prespecified effect size, no

497   animals were excluded from analyses, experiments were not randomized, and investigators were

498   not blinded to group allocation during experiments.

499

500   **DATA AVAILABILITY**

501   Data generated or analyzed during this study are included in this published article (and its

502   supplemental information files). Data needed to evaluate the conclusions in the paper are present

503   in the paper and/or the Supplemental Materials. Source data files will be made available upon

504   request. Single-cell and bulk RNA-sequencing data generated in this study have been deposited

505   in the Gene Expression Omnibus with the primary accession code GSE272933 (bulk) and

506   GSE272932 (single-cell). All bioinformatics tools used in this study are published and publicly

507   available.

508

509

510

511

512

513

514

515

516

References:

1. SEER*Explorer Application [Internet]. [cited 2023 Apr 15]. Available from: https://seer.cancer.gov/statistics-network/explorer/application.html?site=1&data_type=1&graph_type=2&compareBy=sex&chk_sex_3=3&chk_sex_2=2&rate_type=2&race=1&age_range=1&hdn_stage=101&advopt_precision=1&advopt_show_ci=on&hdn_view=0&advopt_show_apc=on&advopt_display=2#resultsRegion0

2. MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, Salber EJ, Valaoras VG, Yuasa S. Age at first birth and breast cancer risk. Bull World Health Organ. 1970;43(2):209–221. PMCID: PMC2427645

3. Albrektsen G, Heuch I, Tretli S, Kvåle G. Breast cancer incidence before age 55 in relation to parity and age at first and last births: a prospective study of one million Norwegian women. Epidemiology. 1994 Nov;5(6):604–611. PMID: 7841242

4. Albrektsen G, Heuch I, Hansen S, Kvåle G. Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects. Br J Cancer. Nature Publishing Group; 2005 Jan;92(1):167–175.

5. Russo J, Tay LK, Ciocca DR, Russo IH. Molecular and cellular basis of the mammary gland susceptibility to carcinogenesis. Environ Health Perspect. 1983 Mar;49:185–199. PMCID: PMC1569117

6. Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in breast cancer. Breast Cancer Res. 2005;7(3):131–142. PMCID: PMC1143568

7. Medina D. Breast Cancer: The Protective Effect of Pregnancy. Clinical Cancer Research. 2004 Jan 1;10(1):380s–384s.

8. Russo J, Balogh GA, Heulings R, Mailo DA, Moral R, Russo PA, Sheriff F, Vanegas J, Russo IH. Molecular basis of pregnancy-induced breast cancer protection. European Journal of Cancer Prevention. 2006 Aug;15(4):306.

9. Siwko SK, Dong J, Lewis MT, Liu H, Hilsenbeck SG, Li Y. Evidence That an Early Pregnancy Causes a Persistent Decrease in the Number of Functional Mammary Epithelial Stem Cells—Implications for Pregnancy-Induced Protection Against Breast Cancer. Stem Cells. 2008 Dec 1;26(12):3205–3209.

10. Choudhury S, Almendro V, Merino VF, Wu Z, Maruyama R, Su Y, Martins FC, Fackler MJ, Bessarabova M, Kowalczyk A, Conway T, Beresford-Smith B, Macintyre G, Cheng YK, Lopez-Bujanda Z, Kaspi A, Hu R, Robens J, Nikolskaya T, Haakensen VD, Schnitt SJ, Argani P, Ethington G, Panos L, Grant M, Clark J, Herlihy W, Lin SJ, Chew G, Thompson EW, Greene-Colozzi A, Richardson AL, Rosson GD, Pike M, Garber JE, Nikolsky Y, Blum JL, Au A, Hwang ES, Tamimi RM, Michor F, Haviv I, Liu XS, Sukumar S, Polyak K. Molecular Profiling of Human Mammary Gland Links Breast Cancer Risk to a p27+ Cell Population with Progenitor Characteristics. Cell Stem Cell. 2013 Jul 3;13(1):117–130.

11. Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, Zhong H, Slifker M, Peri S, Ross EA, López de Cicco R, Zhai Y, Nguyen T, Sheriff F, Russo IH, Su Y, Arslan AA, Bordas P,

557   Lenner P, Åhman J, Landström Eriksson AS, Johansson R, Hallmans G, Toniolo P, Russo
558   J. Genomic signature of parity in the breast of premenopausal women. Breast Cancer
559   Research. 2019 Mar 28;21(1):46.

560   12. Shalabi SF, Miyano M, Sayaman RW, Lopez JC, Jokela TA, Todhunter ME, Hinz S, Garbe
561   JC, Stampfer MR, Kessenbrock K, Seewaldt VE, LaBarge MA. Evidence for accelerated
562   aging in mammary epithelia of women carrying germline BRCA1 or BRCA2 mutations. Nat
563   Aging. Nature Publishing Group; 2021 Sep;1(9):838–849.

564   13. Pelissier Vatter FA, Schapiro D, Chang H, Borowsky AD, Lee JK, Parvin B, Stampfer MR,
565   LaBarge MA, Bodenmiller B, Lorens JB. High-Dimensional Phenotyping Identifies Age-
566   Emergent Cells in Human Mammary Epithelia. Cell Reports. 2018 Apr 24;23(4):1205–1219.

567   14. Garbe JC, Pepin F, Pelissier FA, Sputova K, Fridriksdottir AJ, Guo DE, Villadsen R, Park M,
568   Petersen OW, Borowsky AD, Stampfer MR, LaBarge MA. Accumulation of Multipotent
569   Progenitors with a Basal Differentiation Bias during Aging of Human Mammary Epithelia.
570   Cancer Research. 2012 Jul 15;72(14):3687–3701.

571   15. Pelissier FA, Garbe JC, Ananthanarayanan B, Miyano M, Lin C, Jokela T, Kumar S,
572   Stampfer MR, Lorens JB, LaBarge MA. Age-Related Dysfunction in Mechanotransduction
573   Impairs Differentiation of Human Mammary Epithelial Progenitors. Cell Reports. Elsevier;
574   2014 Jun 26;7(6):1926–1939. PMID: 24910432

575   16. Lambe M, Hsieh C cheng, Trichopoulos D, Ekbom A, Pavia M, Adami HO. Transient
576   Increase in the Risk of Breast Cancer after Giving Birth. New England Journal of Medicine.
577   Massachusetts Medical Society; 1994 Jul 7;331(1):5–9.

578   17. Husby A, Wohlfahrt J, Øyen N, Melbye M. Pregnancy duration and breast cancer risk. Nat
579   Commun. Nature Publishing Group; 2018 Oct 23;9(1):4255.

580   18. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA: A Cancer Journal for
581   Clinicians. 2022;72(1):7–33.

582   19. Osterman MJK, Hamilton BE, Martin JA, Driscoll AK, Valenzuela CP. Births: Final Data for
583   2021. Natl Vital Stat Rep. 2023 Jan;72(1):1–53. PMID: 36723449

584   20. Mathews TJ. Mean Age of Mothers is on the Rise: United States, 2000–2014. 2016;(232).

585   21. Reed AD, Pensa S, Steif A, Stenning J, Kunz DJ, Porter LJ, Hua K, He P, Twigger AJ, Siu
586   AJQ, Kania K, Barrow-McGee R, Goulding I, Gomm JJ, Speirs V, Jones JL, Marioni JC,
587   Khaled WT. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult
588   human breast. Nat Genet. Nature Publishing Group; 2024 Apr;56(4):652–662.

589   22. Fu NY, Nolan E, Lindeman GJ, Visvader JE. Stem Cells and the Differentiation Hierarchy in
590   Mammary Gland Development. Physiological Reviews. American Physiological Society;
591   2020 Apr;100(2):489–523.

592   23. Li CMC, Shapiro H, Tsiobikas C, Selfors LM, Chen H, Rosenbluth J, Moore K, Gupta KP,
593   Gray GK, Oren Y, Steinbaugh MJ, Guerriero JL, Pinello L, Regev A, Brugge JS. Aging-
594   Associated Alterations in Mammary Epithelia and Stroma Revealed by Single-Cell RNA
595   Sequencing. Cell Reports. 2020 Dec 29;33(13):108566.

24. Dong Q, Gao H, Shi Y, Zhang F, Gu X, Wu A, Wang D, Chen Y, Bandyopadhyay A, Yeh IT, Daniel BJ, Chen Y, Zou Y, Rebel VL, Walter CA, Lu J, Huang C, Sun LZ. Aging is associated with an expansion of CD49f^{hi} mammary stem cells that show a decline in function and increased transformation potential. Aging. 2016 Nov 15;8(11):2754–2776. PMID: 27852980

25. Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, Khaled WT. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat Commun. Nature Publishing Group; 2017 Dec 11;8(1):2128.

26. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, Wilcox S, Fu N, Liu KH, Jackling FC, Davis MJ, Lindeman GJ, Smyth GK, Visvader JE. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat Commun. Nature Publishing Group; 2017 Nov 20;8(1):1627.

27. Walker MR, Amante JJ, Li J, Liu H, Zhu LJ, Feltri ML, Goel HL, Mercurio AM. Alveolar progenitor cells in the mammary gland are dependent on the β4 integrin. Developmental Biology. 2020 Jan 1;457(1):13–19.

28. Fu NY, Rios AC, Pal B, Law CW, Jamieson P, Liu R, Vaillant F, Jackling F, Liu KH, Smyth GK, Lindeman GJ, Ritchie ME, Visvader JE. Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. Nat Cell Biol. Nature Publishing Group; 2017 Mar;19(3):164–176.

29. Cai S, Kalisky T, Sahoo D, Dalerba P, Feng W, Lin Y, Qian D, Kong A, Yu J, Wang F, Chen EY, Scheeren FA, Kuo AH, Sikandar SS, Hisamori S, Weele LJ van, Heiser D, Sim S, Lam J, Quake S, Clarke MF. A Quiescent Bcl11b High Stem Cell Population Is Required for Maintenance of the Mammary Gland. Cell Stem Cell. Elsevier; 2017 Feb 2;20(2):247-260.e5. PMID: 28041896

30. Dallas DC, Murray NM, Gan J. Proteolytic Systems in Milk: Perspectives on the Evolutionary Function within the Mammary Gland and the Infant. J Mammary Gland Biol Neoplasia. 2015 Dec;20(0):133–147. PMCID: PMC4637187

31. Hassiotou F, Geddes DT. Immune Cell–Mediated Protection of the Mammary Gland and the Infant during Breastfeeding. Advances in Nutrition. 2015 May 1;6(3):267–275.

32. Goldman AS, Chheda S, Garofalo R. Evolution of Immunologic Functions of the Mammary Gland and the Postnatal Development of Immunity. Pediatr Res. Nature Publishing Group; 1998 Feb;43(2):155–162.

33. Bai H, Liu X, Lin M, Meng Y, Tang R, Guo Y, Li N, Clarke MF, Cai S. Progressive senescence programs induce intrinsic vulnerability to aging-related female breast cancer. Nat Commun. Nature Publishing Group; 2024 Jun 17;15(1):5154.

34. Almanzar N, Antony J, Baghel AS, Bakerman I, Bansal I, Barres BA, Beachy PA, Berdnik D, Bilen B, Brownfield D, Cain C, Chan CKF, Chen MB, Clarke MF, Conley SD, Darmanis S, Demers A, Demir K, de Morree A, Divita T, du Bois H, Ebadi H, Espinoza FH, Fish M, Gan Q, George BM, Gillich A, Gòmez-Sjöberg R, Green F, Genetiano G, Gu X, Gulati GS, Hahn O, Haney MS, Hang Y, Harris L, He M, Hosseinzadeh S, Huang A, Huang KC, Iram T, Isobe T, Ives F, Jones RC, Kao KS, Karkanias J, Karnam G, Keller A, Kershner AM, Khoury N,

Kim SK, Kiss BM, Kong W, Krasnow MA, Kumar ME, Kuo CS, Lam J, Lee DP, Lee SE, Lehallier B, Leventhal O, Li G, Li Q, Liu L, Lo A, Lu WJ, Lugo-Fagundo MF, Manjunath A, May AP, Maynard A, McGeever A, McKay M, McNerney MW, Merrill B, Metzger RJ, Mignardi M, Min D, Nabhan AN, Neff NF, Ng KM, Nguyen PK, Noh J, Nusse R, Pálovics R, Patkar R, Peng WC, Penland L, Pisco AO, Pollard K, Puccinelli R, Qi Z, Quake SR, Rando TA, Rulifson EJ, Schaum N, Segal JM, Sikandar SS, Sinha R, Sit RV, Sonnenburg J, Staehli D, Szade K, Tan M, Tan W, Tato C, Tellez K, Dulgeroff LBT, Travaglini KJ, Tropini C, Tsui M, Waldburger L, Wang BM, van Weele LJ, Weinberg K, Weissman IL, Woszczyna MN, Wu SM, Wyss-Coray T, Xiang J, Xue S, Yamauchi KA, Yang AC, Yerra LP, Youngyunpipatkul J, Yu B, Zanini F, Zardeneta ME, Zee A, Zhao C, Zhang F, Zhang H, Zhang MJ, Zhou L, Zou J, The Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature. Nature Publishing Group; 2020 Jul;583(7817):590–595.

35. Kendrick H, Regan JL, Magnay FA, Grigoriadis A, Mitsopoulos C, Zvelebil M, Smalley MJ. Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. BMC Genomics. 2008 Dec 8;9(1):591.

36. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biology. 2019 Mar 19;20(1):59.

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

38. Kang M, Armenteros JJA, Gulati GS, Gleyzer R, Avagyan S, Brown EL, Zhang W, Usmani A, Earland N, Wu Z, Zou J, Fields RC, Chen DY, Chaudhuri AA, Newman AM. Mapping single-cell developmental potential in health and disease with interpretable deep learning [Internet]. bioRxiv; 2024 [cited 2024 Jul 12]. p. 2024.03.19.585637. Available from: https://www.biorxiv.org/content/10.1101/2024.03.19.585637v1

39. Bu W, Chen J, Morrison GD, Huang S, Creighton CJ, Huang J, Chamness GC, Hilsenbeck SG, Roop DR, Leavitt AD, Li Y. Keratin 6a marks mammary bipotential progenitor cells that can give rise to a unique tumor model resembling human normal-like breast cancer. Oncogene. Nature Publishing Group; 2011 Oct;30(43):4399–4409.

40. Smith GH, Mehrel T, Roop DR. Differential keratin gene expression in developing, differentiating, preneoplastic, and neoplastic mouse mammary epithelium. Cell Growth Differ. 1990 Apr;1(4):161–170. PMID: 1707299

41. Pal B, Chen Y, Milevskiy MJG, Vaillant F, Prokopuk L, Dawson CA, Capaldo BD, Song X, Jackling F, Timpson P, Lindeman GJ, Smyth GK, Visvader JE. Single cell transcriptome atlas of mouse mammary epithelial cells across development. Breast Cancer Res. 2021 Jun 29;23(1):69. PMCID: PMC8243869

42. Kim JY, Kim G, Lim SC, Choi HS. IL-33-Induced Transcriptional Activation of LPIN1 Accelerates Breast Tumorigenesis. Cancers. Multidisciplinary Digital Publishing Institute; 2021 Jan;13(9):2174.

43. Kim JY, Lim SC, Kim G, Yun HJ, Ahn SG, Choi HS. Interleukin-33/ST2 axis promotes epithelial cell transformation and breast tumorigenesis via upregulation of COT activity. Oncogene. Nature Publishing Group; 2015 Sep;34(38):4928–4938.

44. Liu J, Shen JX, Hu JL, Huang WH, Zhang GJ. Significance of Interleukin-33 and Its Related Cytokines in Patients with Breast Cancers. Frontiers in Immunology [Internet]. 2014 [cited 2023 Apr 16];5. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2014.00141

45. Hu H, Sun J, Wang C, Bu X, Liu X, Mao Y, Wang H. IL-33 facilitates endocrine resistance of breast cancer by inducing cancer stem cell properties. Biochem Biophys Res Commun. 2017 Apr 8;485(3):643–650. PMID: 28216163

46. Jovanovic IP, Pejnovic NN, Radosavljevic GD, Pantic JM, Milovanovic MZ, Arsenijevic NN, Lukic ML. Interleukin-33/ST2 axis promotes breast cancer growth and metastases by facilitating intratumoral accumulation of immunosuppressive and innate lymphoid cells. Int J Cancer. 2014 Apr 1;134(7):1669–1682. PMID: 24105680

47. Ventura A, Meissner A, Dillon CP, McManus M, Sharp PA, Van Parijs L, Jaenisch R, Jacks T. Cre-lox-regulated conditional RNA interference from transgenes. Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences; 2004 Jul 13;101(28):10380–10385.

48. Jena MK, Jaswal S, Kumar S, Mohanty AK. Molecular mechanism of mammary gland involution: An update. Developmental Biology. 2019 Jan 15;445(2):145–155.

49. Watson CJ, Kreuzaler PA. Remodeling mechanisms of the mammary gland during involution. The International Journal of Developmental Biology. UPV/EHU Press; 2011 Nov 29;55(7-8–9):757–762.

50. Stein T, Morris JS, Davies CR, Weber-Hall SJ, Duffy MA, Heath VJ, Bell AK, Ferrier RK, Sandilands GP, Gusterson BA. Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. Breast Cancer Research. 2003 Dec 18;6(2):R75.

51. Stein T, Salomonis N, Gusterson BA. Mammary Gland Involution as a Multi-step Process. J Mammary Gland Biol Neoplasia. 2007 Mar 1;12(1):25–35.

52. Langille E, Al-Zahrani KN, Ma Z, Liang M, Uuskula-Reimand L, Espin R, Teng K, Malik A, Bergholtz H, Ghamrasni SE, Afiuni-Zadeh S, Tsai R, Alvi S, Elia A, Lü Y, Oh RH, Kozma KJ, Trcka D, Narimatsu M, Liu JC, Nguyen T, Barutcu S, Loganathan SK, Bremner R, Bader GD, Egan SE, Cescon DW, Sørlie T, Wrana JL, Jackson HW, Wilson MD, Witkiewicz AK, Knudsen ES, Pujana MA, Wahl GM, Schramek D. Loss of Epigenetic Regulation Disrupts Lineage Integrity, Induces Aberrant Alveogenesis, and Promotes Breast Cancer. Cancer Discov. 2022 Dec 2;12(12):2930–2953. PMCID: PMC9812400

53. Christin JR, Wang C, Chung CY, Liu Y, Dravis C, Tang W, Oktay MH, Wahl GM, Guo W. Stem Cell Determinant SOX9 Promotes Lineage Plasticity and Progression in Basal-like Breast Cancer. Cell Reports. 2020 Jun 9;31(10):107742.

717   54. Van Keymeulen A, Lee MY, Ousset M, Brohée S, Rorive S, Giraddi RR, Wuidart A,
718   Bouvencourt G, Dubois C, Salmon I, Sotiriou C, Phillips WA, Blanpain C. Reactivation of
719   multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature. 2015 Sep
720   3;525(7567):119–123. PMID: 26266985

721   55. Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, Britschgi A, Eichlisberger T,
722   Kohler H, Aina O, Cardiff RD, Bentires-Alj M. PIK3CAH1047R induces multipotency and
723   multi-lineage mammary tumours. Nature. Nature Publishing Group; 2015
724   Sep;525(7567):114–118.

725   56. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, MacKay A,
726   Grigoriadis A, Tutt A, Ashworth A, Reis-Filho JS, Smalley MJ. *BRCA1* Basal-like Breast
727   Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. Cell
728   Stem Cell. 2010 Sep 3;7(3):403–417.

729   57. Lloyd-Lewis B, Gobbo F, Perkins M, Jacquemin G, Huyghe M, Faraldo MM, Fre S. In vivo
730   imaging of mammary epithelial cell dynamics in response to lineage-biased Wnt/β-catenin
731   activation. Cell Rep. 2022 Mar 8;38(10):110461. PMCID: PMC7615182

732   58. Oprean CM, Negru SM, Popovici DI, Saftescu S, Han RA, Dragomir GM, Hoinoiu T, Dema
733   A. Postmenopausal Breast Cancer in Women, Clinical and Epidemiological Factors Related
734   to the Molecular Subtype: A Retrospective Cohort Study in a Single Institution for 13 Years.
735   Follow-Up Data. Int J Environ Res Public Health. 2020 Dec;17(23):8722. PMCID:
736   PMC7727650

737   59. Davis FM, Lloyd-Lewis B, Harris OB, Kozar S, Winton DJ, Muresan L, Watson CJ. Single-
738   cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of
739   stem/progenitor cell progeny. Nat Commun. Nature Publishing Group; 2016 Oct
740   25;7(1):13053.

741   60. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N,
742   Dekoninck S, Blanpain C. Distinct stem cells contribute to mammary gland development and
743   maintenance. Nature. 2011 Nov;479(7372):189–193.

744   61. Rios AC, Fu NY, Lindeman GJ, Visvader JE. In situ identification of bipotent stem cells in
745   the mammary gland. Nature. Nature Publishing Group; 2014 Feb;506(7488):322–327.

746   62. van Amerongen R, Bowman AN, Nusse R. Developmental Stage and Time Dictate the Fate
747   of Wnt/β-Catenin-Responsive Stem Cells in the Mammary Gland. Cell Stem Cell. Elsevier;
748   2012 Sep 7;11(3):387–400. PMID: 22863533

749   63. Cayrol C, Girard JP. Interleukin-33 (IL-33): A critical review of its biology and the
750   mechanisms involved in its release as a potent extracellular cytokine. Cytokine. 2022 Aug
751   1;156:155891.

752   64. Alonso-Curbelo D, Ho YJ, Burdziak C, Maag JLV, Morris JP, Chandwani R, Chen HA,
753   Tsanov KM, Barriga FM, Luan W, Tasdemir N, Livshits G, Azizi E, Chun J, Wilkinson JE,
754   Mazutis L, Leach SD, Koche R, Pe'er D, Lowe SW. A gene–environment-induced epigenetic
755   program initiates tumorigenesis. Nature. Nature Publishing Group; 2021
756   Feb;590(7847):642–648.

65. Stier MT, Mitra R, Nyhoff LE, Goleniewska K, Zhang J, Puccetti MV, Casanova HC, Seegmiller AC, Newcomb DC, Kendall PL, Eischen CM, Peebles RS Jr. IL-33 Is a Cell-Intrinsic Regulator of Fitness during Early B Cell Development. The Journal of Immunology. 2019 Sep 15;203(6):1457–1467.

66. Yin H, Li X, Hu S, Liu T, Yuan B, Gu H, Ni Q, Zhang X, Zheng F. IL-33 accelerates cutaneous wound healing involved in upregulation of alternatively activated macrophages. Mol Immunol. 2013 Dec;56(4):347–353. PMID: 23911389

67. Kurimoto M, Watanabe T, Kamata K, Minaga K, Kudo M. IL-33 as a Critical Cytokine for Inflammation and Fibrosis in Inflammatory Bowel Diseases and Pancreatitis. Front Physiol [Internet]. Frontiers; 2021 Oct 25 [cited 2024 Jul 17];12. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.781012/full

68. Sikandar SS, Kuo AH, Kalisky T, Cai S, Zabala M, Hsieh RW, Lobo NA, Scheeren FA, Sim S, Qian D, Dirbas FM, Somlo G, Quake SR, Clarke MF. Role of epithelial to mesenchymal transition associated genes in mammary gland regeneration and breast tumorigenesis. Nat Commun. Nature Publishing Group; 2017 Nov 21;8(1):1669.

69. Zabala M, Lobo NA, Antony J, Heitink LS, Gulati GS, Lam J, Parashurama N, Sanchez K, Adorno M, Sikandar SS, Kuo AH, Qian D, Kalisky T, Sim S, Li L, Dirbas FM, Somlo G, Newman A, Quake SR, Clarke MF. LEFTY1 Is a Dual-SMAD Inhibitor that Promotes Mammary Progenitor Growth and Tumorigenesis. Cell Stem Cell. 2020 Aug 6;27(2):284-299.e8.

70. Sikandar SS, Gulati GS, Antony J, Fetter I, Kuo AH, Ho WHD, Haro-Acosta V, Das S, Steen CB, Pereira TA, Qian D, Beachy PA, Dirbas FM, Red-Horse K, Rabbitts TH, Thiery JP, Newman AM, Clarke MF. Identification of a minority population of LMO2+ breast cancer cells that integrate into the vasculature and initiate metastasis. Science Advances. American Association for the Advancement of Science; 2022 Nov 9;8(45):eabm3548.

71. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biology. 2018 Feb 6;19(1):15.

72. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P ru, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. Nature Publishing Group; 2019 Dec;16(12):1289–1296.

73. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. Nature Publishing Group; 2019 Mar 26;9(1):5233.

74. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Internet]. arXiv; 2020 [cited 2024 Jul 17]. Available from: http://arxiv.org/abs/1802.03426

75. scanpy_usage/180209_cell_cycle/data/regev_lab_cell_cycle_genes.txt at master · scverse/scanpy_usage [Internet]. GitHub. [cited 2024 Jul 17]. Available from: https://github.com/scverse/scanpy_usage/blob/master/180209_cell_cycle/data/regev_lab_cell_cycle_genes.txt

796   76. Waskom ML. seaborn: statistical data visualization. Journal of Open Source Software. 2021
797   Apr 6;6(60):3021.

798   77. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering.
799   2007 May;9(3):90–95.

800   78. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E,
801   Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ,
802   Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW,
803   VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR,
804   Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P. SciPy 1.0: fundamental algorithms
805   for scientific computing in Python. Nat Methods. Nature Publishing Group; 2020
806   Mar;17(3):261–272.

807   79. Josef Perktold, Skipper Seabold, Kevin Sheppard, ChadFulton, Kerby Shedden,
808   jbrockmendel, j-grana6, Peter Quackenbush, Vincent Arel-Bundock, Wes McKinney, Ian
809   Langmore, Bart Baker, Ralf Gommers, yogabonito, s-scherrer, Yauhen Zhurko, Matthew
810   Brett, Enrico Giampieri, yl565, Jarrod Millman, Paul Hobson, Vincent, Pamphile Roy, Tom
811   Augspurger, tvanzyl, alexbrc, Tyler Hartley, Fernando Perez, Yuji Tamiya, Yaroslav
812   Halchenko. statsmodels/statsmodels: Release 0.14.2 [Internet]. Zenodo; 2024 [cited 2024
813   Jul 23]. Available from: https://zenodo.org/doi/10.5281/zenodo.593847

814   80. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H,
815   Soldatov A. Transcriptome analysis by strand-specific sequencing of complementary DNA.
816   Nucleic Acids Research. 2009 Oct 1;37(18):e123.

817   81. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying
818   mammalian transcriptomes by RNA-Seq. Nat Methods. Nature Publishing Group; 2008
819   Jul;5(7):621–628.

820   82. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
821   assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923–930.

822   83. Anders S, Huber W. Differential expression analysis for sequence count data. Genome
823   Biology. 2010 Oct 27;11(10):R106.

824   84. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
825   seq data with DESeq2. Genome Biology. 2014 Dec 5;15(12):550.

826   85. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq:
827   accounting for selection bias. Genome Biology. 2010 Feb 4;11(2):R14.

828

829

830

831

842

843 **AUTHOR CONTRIBUTIONS**

844 S.S.S. and A.O conceived and designed the study. A.O performed most of the experiments and

845 analyzed the data with assistance from V.H.A and S.K and under the supervision of S.S.S. V.H.A

846 collected samples for single-cell RNA sequencing and processed samples for bulk-RNA

847 sequencing. P.M verified the differential gene expression analysis from bulk RNA sequencing

848 data. C.M.R analyzed single-cell RNA sequencing data under supervision of V.D.J. A.O and

849 S.S.S. wrote the manuscript with contributions from C.M.R and V.D.J. All authors commented on

850 the manuscript.

851

852 **COMPETING INTERESTING**

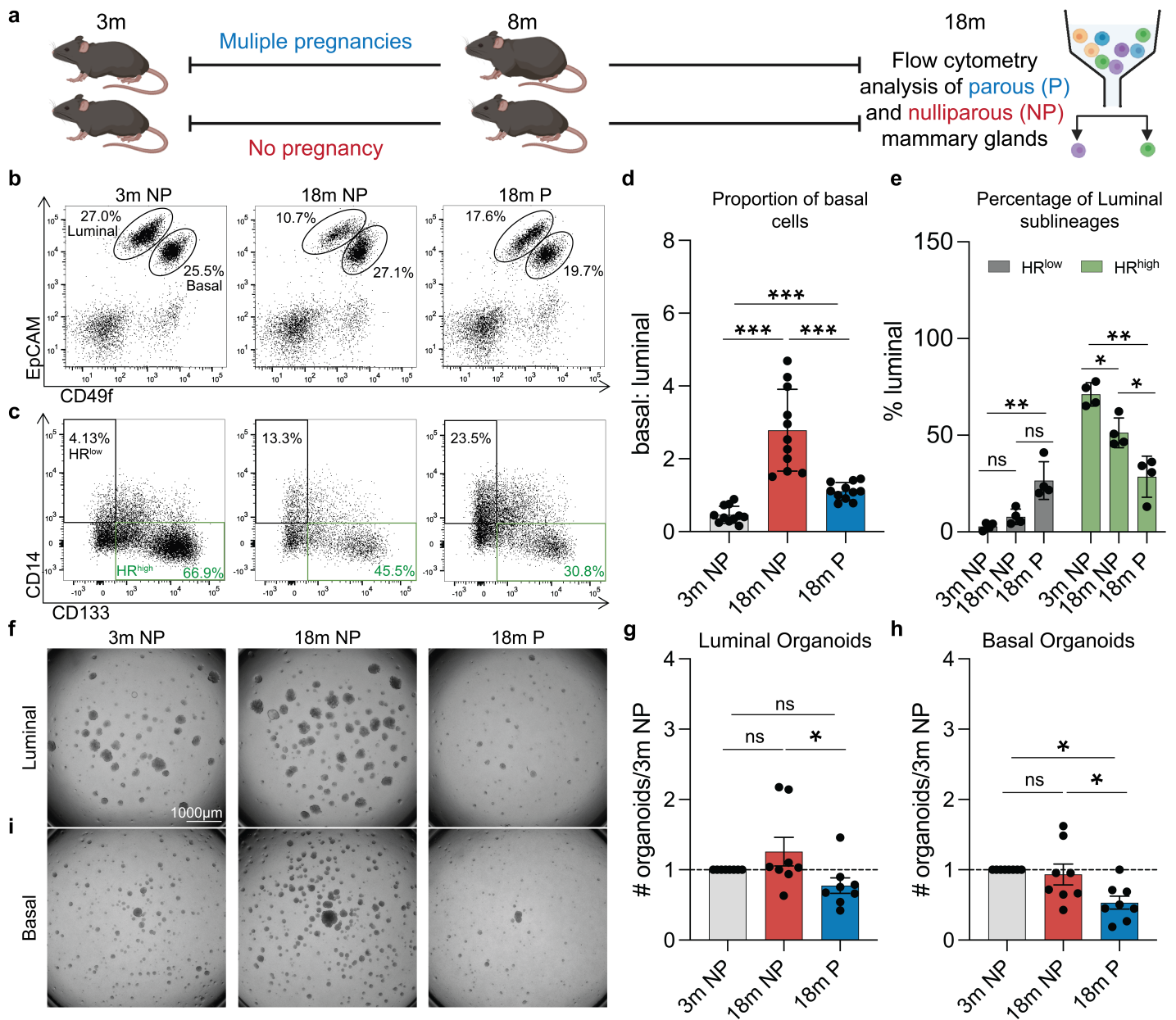853 The authors declare no competing interests.

854

855 **EXTENDED DATA:**

856 Extended data figures 1-6

857 Tables 1-5

858   **MAIN FIGURES AND LEGENDS**

859

860   **Figure 1. Pregnancy induces persistent changes in cell fate decisions and reduces**

861   **organoid formation.** (a) Schematic of flow cytometry analysis for 3-month nulliparous (3m NP),

862   18-month nulliparous (18m NP), and 18-month parous (18m P). Created with Biorender.com. (b)

863   Representative flow cytometry plots of DAPI-/CD45-/CD31-/Ter119- single cells from mouse

864   mammary tissues. Basal (EPCAM$^{med/low}$CD49f^{high}) and luminal (EPCAM^{high}CD49f$^{med/low}$) cells are

865   denoted by representative gates. (c) Quantification of flow cytometry data shown in (b) (n = 11

866   mice). (d) Representative flow cytometry plots of DAPI-/Lineage-/EPCAM^{high}CD49f$^{med/low}$ single

867   cells from mouse mammary tissues. (e) Quantification of flow cytometry data shown in (d) (n = 4

868   mice). (f) Representative images of primary organoids derived from 3m NP, 18m NP, and 18m

869   P luminal cells. (g) Quantification of luminal organoid number. (h) Quantification of basal

870   organoid number. (i) Representative images of primary organoids derived from 3m NP, 18m NP,

871   and 18m P basal cells. n = 8 mice for experiments in f-i. Statistical significance was determined

872   by performing ANOVA (d, e) or paired t tests (g, i). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

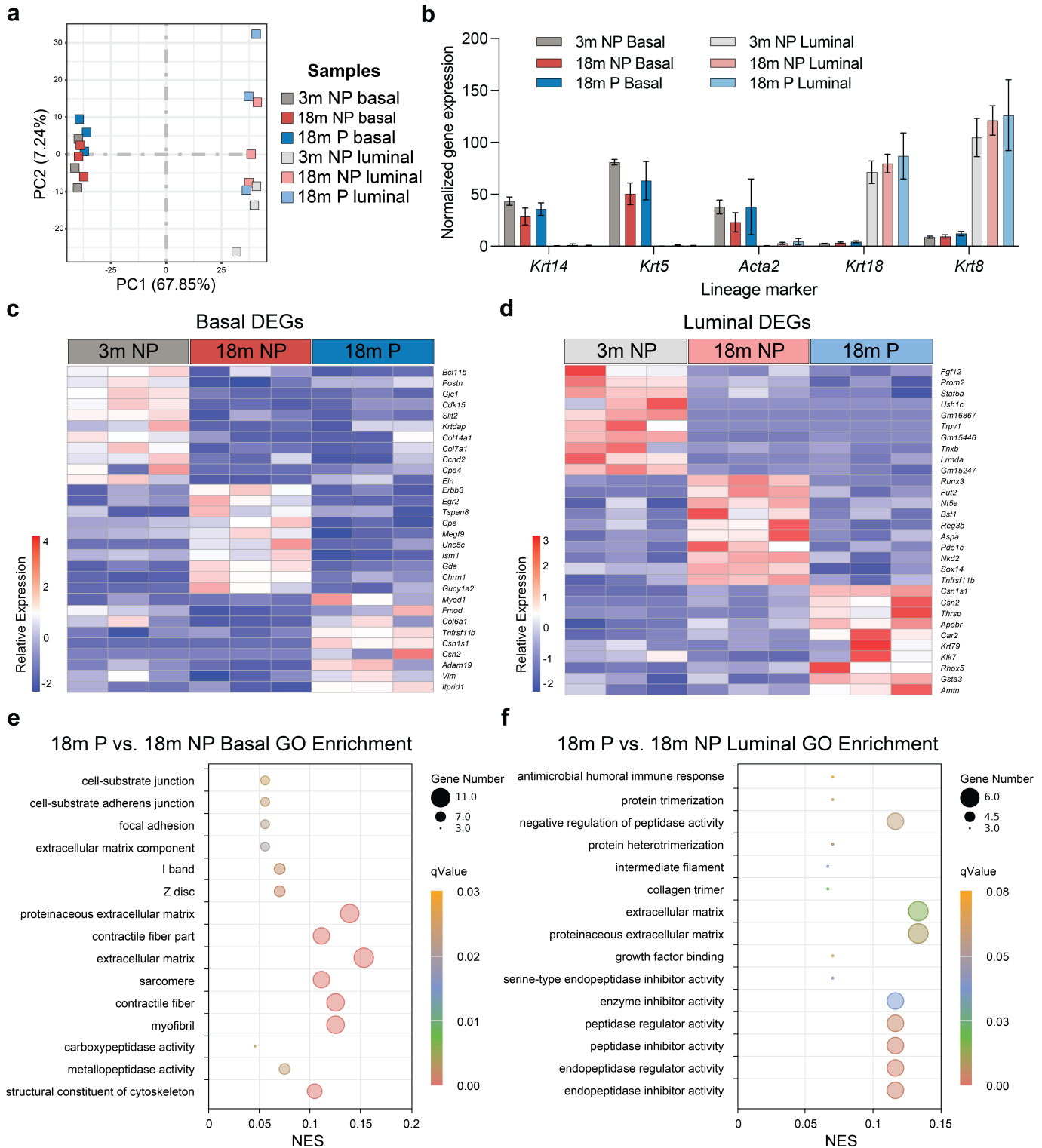873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

Figure 2. Bulk RNA sequencing in aged nulliparous and parous mammary glands reveals long-lasting transcriptomic changes. (a) Principal Component Analysis (PCA) of basal and luminal RNA-seq samples from 3m NP, 18m NP, and 18m P mice. (b) Bar plot depicting the

895   expression (FPKM x10^3) of basal (*Krt14, Krt15, Acta2*) and luminal (*Krt18, Krt8*) marker genes

896   across samples. (c) Heatmap of differentially expressed genes (DEGs) across 3m NP, 18m NP,

897   and 18m P basal cells and (d) luminal cells. Relative expression reflects log2(FC+1) values.

898   Significance was determined using an adj-p value of <= 0.1. (e, f) Dotplot of Gene Ontology

899   (GO) results for 18m P basal (e) and luminal (f) cells, relative to 18m NP. GO results were

900   selected from the top 15 hits with the lowest adj-p value denoted as q-value (< 0.03 for basal, <

901   0.08 for luminal), NES = Normalized Enrichment Score. Gene number refers to the number of

902   genes from the input gene list that are present in the GO gene list.

903
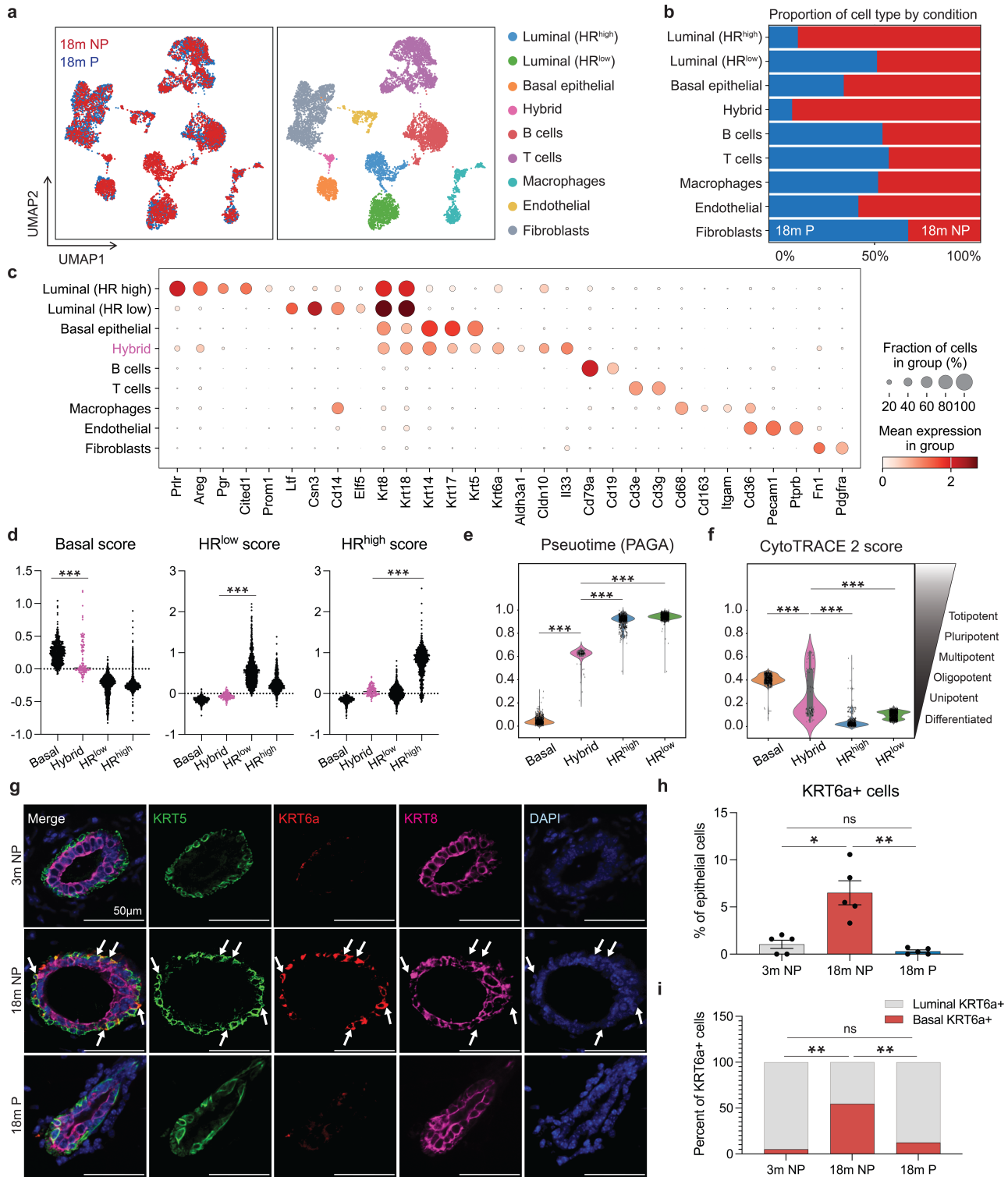
904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

Figure 3. Single cell RNA sequencing identifies a hybrid population of MECs in aged nulliparous mice but is reduced in aged parous mice. (a) UMAP plot of scRNA-seq data generated from the mammary glands of 18m nulliparous (NP, red) and parous (P, blue) mice (n = 3 mice) and integrated with Tabula Muris (n=10,001). UMAPs are colored by condition (left)

926   and annotated cell type (right). (b) Bar plot showing the proportion of cell types by condition. (c)

927   Dotplot of gene expression for selected marker genes of mammary cell lineages. (d) Basal, HR-

928   low, and HR-high gene expression scores across epithelial clusters (black), including a minority

929   population of hybrid MECs (pink). (e) Violin plot showing the distribution of inferred partition-

930   based graph abstraction (PAGA) pseudotime scores across epithelial cell clusters. (f) Violin plot

931   comparing predicted developmental potential (obtained through CytoTRACE 2) across epithelial

932   cell clusters. (g) Representative immunofluorescence (IF) stains against KRT5 (basal cells,

933   green), KRT6a (hybrid MECs, red), and KRT8 (luminal cells, magenta). Nuclei are visualized

934   using DAPI (blue). Hybrid MECs localized to the basal layer are indicated by white arrows. (h)

935   Quantification of all KRT6a+ hybrid MECs in the epithelium. (i) Quantification of the localization

936   of KRT6a+ hybrid MECs (basal or luminal). Statistical significance was determined by

937   performing one-way ANOVA with Tukey test (d) or unpaired t tests with Welch's correction to

938   account for unequal standard deviations (h). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical tests

939   for scRNA-sequencing are described in the methods section. n = 5 mice for IF experiments.
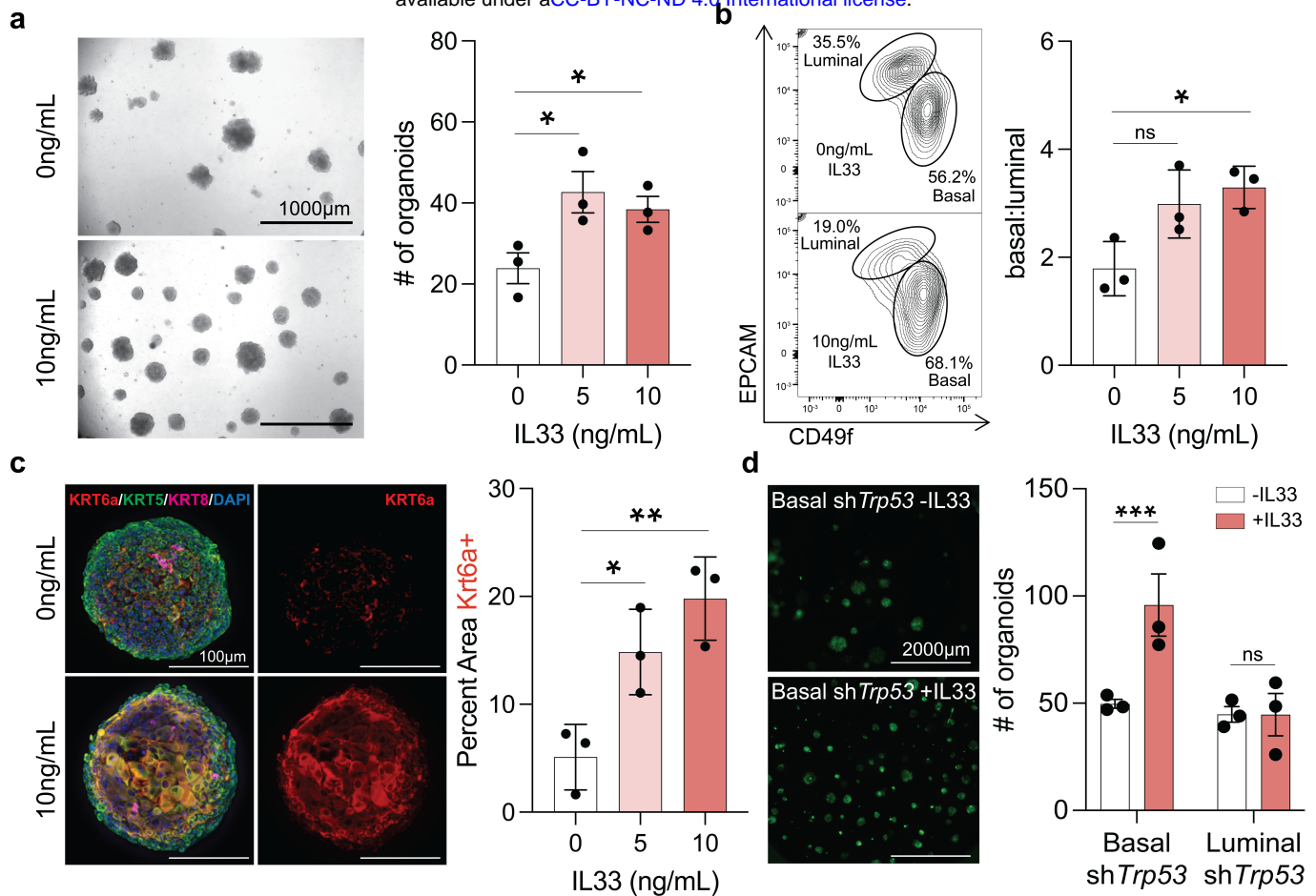
940   Scale bar = 50 μm.

941

942

943

944

945

946

947

948

949

950

Figure 4. **IL33 treatment of young basal cells phenocopies aged nulliparous MECs.** (a)

Representative images of primary basal organoids from 3m NP mice treated with no Il33

treatment (top left) and with IL33 treatment (bottom left). Quantification of the number of

organoids formed with 0, 5, or 10ng/mL of IL33. (b) Representative flow cytometry plots of basal

and luminal cells from IL33-treated organoids (left) and the quantified proportion of basal cells

relative to the luminal population (right). (c) Representative IF stains (left) against KRT5 (basal

cells), KRT6a (hybrid MECs), and KRT8 (luminal cells) on IL33-treated organoids. Nuclei are

visualized using DAPI. Quantification of KRT6a+ area in basal organoids cultured with 0, 5, or

10ng/mL of IL33 (right). (d) Representative images of preneoplastic (sh*Tp53*) basal organoids

treated with (top left, 10ng/mL) or without IL33 (bottom left). Quantification of the number of

sh*Trp53*- basal organoids formed with or without IL33 treatment after 12 days in culture (right).

Statistical significance was determined by performing unpaired t tests (a-c) or 2-way ANOVA

963   with Šídák's multiple comparisons test. * p <0.05, ** p < 0.01, *** p < 0.001. n = 3 mice. Scale

964   bar = 1000 μm (a, d) and 100 μm (c).