



**HAL**  
open science

## Luminal progenitors undergo partial epithelial-to-mesenchymal transition at the onset of basal-like breast tumorigenesis

Camille Landragin, Melissa Saichi, Pacôme Prompsy, Adeline Durand, Jérémy Mesple, Amandine Trouchet, Marisa Faraldo, Hélène Salmon, Céline Vallot

### ► To cite this version:

Camille Landragin, Melissa Saichi, Pacôme Prompsy, Adeline Durand, Jérémy Mesple, et al.. Luminal progenitors undergo partial epithelial-to-mesenchymal transition at the onset of basal-like breast tumorigenesis. 2022. hal-03860732

**HAL Id: hal-03860732**

**<https://hal.science/hal-03860732>**

Preprint submitted on 22 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Luminal progenitors undergo partial epithelial-to-mesenchymal transition during basal-like**  
2 **breast tumorigenesis**

3

4 Camille Landragin<sup>1,2,\*</sup>, Melissa Saichi<sup>1,2,\*</sup>, Pacôme Prompsy<sup>1,2</sup>, Adeline Durand<sup>1,2</sup>, Jérémy Mesple<sup>3</sup>,  
5 Amandine Trouchet<sup>1,4</sup>, Marisa Faraldo<sup>5,6</sup>, Anne Vincent-Salomon<sup>7,8</sup>, Hélène Salmon<sup>3</sup>, Céline Vallot<sup>1,2,4</sup>

6

7 Affiliations:

8 1 CNRS UMR3244, Institut Curie, PSL University, Paris, France

9 2 Translational Research Department, Institut Curie, PSL University, Paris, France

10 3 INSERM U932, Institut Curie, PSL University, Paris, France

11 4 Single Cell Initiative, Institut Curie, PSL University, Paris, France

12 5 CNRS UMR3215, Institut Curie, PSL University, Paris, France

13 6 INSERM U934, Institut Curie, PSL University, Paris, France

14 7 Department of Pathology-Genetics and Immunology, Institut Curie, PSL University, Paris, France.

15 8 INSERM U934, Institut Curie, PSL University, Paris, France

16

17 \* co-first authors

18

19 Correspondence should be addressed to: [celine.vallot@curie.fr](mailto:celine.vallot@curie.fr)

20

21

22

23

24 **ABSTRACT**

25

26 Basal-like breast tumors have been proposed to originate from luminal progenitor cells. Yet how  
27 luminal cells transform into invasive cancer cells remains poorly understood. Here, we reconstitute the  
28 sequence of events leading luminal progenitor cells to tumor formation in a *BRCA1*-deficient context,  
29 studying mammary glands at single-cell resolution in mouse and human. Luminal progenitors can  
30 actually tolerate multiple genomic alterations, including *TP53* loss, without transforming. Switch to  
31 tumoral state is rather initiated with a partial epithelial-to-mesenchymal (EMT) transition, following a  
32 senescent-like process. In humans, signature of such pre-tumoral states is detected in early stage  
33 basal-like tumors that rarely recur. Studying biopsies of *BRCA1* carriers, we further show the existence  
34 of cells in abnormal luminal states with activated pre-tumoral signature. Our results support partial  
35 EMT as a player of initial epithelial transformation and open perspectives for detection of pre-tumoral  
36 events.

37

38

39

40

41

## 42 INTRODUCTION

43 Triple-negative breast cancer (TNBC) refers to a subgroup of aggressive breast cancers defined by the  
44 lack of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor  
45 receptor 2 (HER2) accounting for 15–20% of all breast tumors (Onitilo et al. 2009). Along with  
46 transcriptional heterogeneity, TNBC is characterized by complex genomes, dictated by high genetic  
47 instability and complex patterns of copy number alterations and chromosomal rearrangements (R. Gao  
48 et al. 2016; Engebraaten, Vollan, and Børresen-Dale 2013). Defects in double-stranded DNA repair  
49 mechanisms are indeed characteristic of TNBC, as a result of either germline or somatic mutations in  
50 *BRCA1/2* and other genes involved in DNA repair (Timms et al. 2014; Stefansson et al. 2011). In this  
51 genetically unstable context, there is a chaotic de-structuration of the mammary gland, with recurrent  
52 loss of proper cell identity. Part of these cancers harbor basal-like phenotypes, expressing an  
53 incomplete set of basal markers but with high intra-tumor heterogeneity (Marra et al. 2020; Bianchini  
54 et al. 2016). Interestingly, *BRCA1*-deficient tumors are suspected to originate from luminal progenitor  
55 cells of the gland, implicating a recurrent switch or loss in cell identity during tumorigenesis (Molyneux  
56 et al. 2010; Lim et al. 2009). Recent data indicate that *Brca1*-deficiency in the mammary gland induces  
57 aberrant alveolar differentiation of luminal progenitors, suggesting early phenotypic defects in the  
58 mammary gland of a *Brca1*-deficient individual (Bach et al. 2021). However, the tumor-initiating events  
59 leading to the emergence of tumor cells *per se* remain unknown.

60 Studying early steps of tumorigenesis is not feasible solely based on human tumor samples which are  
61 complex stacks of molecular alterations acquired over time. Animal models enable the isolation of a  
62 continuum of states from normal to pathologic gland to precisely map the evolution of the  
63 physiological mammary gland towards tumorigenesis. In the case of basal-like breast cancers, models  
64 with *Brca1/Trp53* deficiency in luminal progenitors have been shown to mimic formation of human  
65 basal-like breast cancers (Selbert et al. 1998; Molyneux et al. 2010). *TP53* mutations remain the most  
66 common genetic alteration in basal-like cancers (85%, (Cancer Genome Atlas Network 2012)). In  
67 *BRCA1*-germline carriers, *TP53* mutation was actually shown to be among the earliest events in tumor

68 formation (Martins et al. 2012). In this context, a mouse model with conditional deletion of *Trp53* and  
69 *Brca1* in the luminal compartment of the mammary gland appears as an apropos model to catch the  
70 rare transforming events leading an HDR-deficient luminal progenitor to tumorigenesis. In contrast to  
71 humans, where these events are extremely rare, the deletion in the mouse of these genes in multiple  
72 cells of the mammary glands greatly enhances our ability to detect the transitioning states from  
73 aberrant luminal progenitor to basal-like breast cancer phenotype.

74 Here, we first map steps of *Brca1*-tumorigenesis *in vivo*, with a focus on epithelial cells to catch rare  
75 pre-tumoral epithelial states. We combine single-cell transcriptomics and multiplex imaging in a Blg-  
76 Cre *Trp53<sup>F1/F1</sup>* and *Brca1<sup>F1/F1</sup>* mouse model, where *Trp53* and *Brca1* are deleted in luminal progenitor  
77 cells (Molyneux et al. 2010). We identify cells transitioning from luminal progenitor to tumor  
78 phenotype, expressing p16 and with highly remodeled genomes. These cells are partially switching to  
79 a mesenchymal phenotype while retaining their epithelial characteristics and activating angiogenesis.

80 We furthermore discover that these cells have undergone a major epigenomic crisis with a disruption  
81 of their heterochromatin through the accumulation of multiple heterochromatin foci. In human tumor  
82 cohorts, signature of such pre-tumoral states is found in basal-like tumors of low stage that rarely  
83 recur. Finally, studying juxta-tumoral tissue of *BRCA1* carriers with single-nuclei transcriptomics, we  
84 show the existence of aberrant luminal states, attesting to the enhanced plasticity of the luminal  
85 compartment in *BRCA1* carriers. We detect a population of luminal cells with abnormal activation of  
86 *FOXC1* transcription factor, pivotal mesenchymal regulator in physiological development. We propose  
87 that partial EMT occurs at the initiation of *BRCA1* tumorigenesis, and is not only restricted to its well-  
88 known pivotal role in cell dissemination and metastasis of cells from tumors.

## 89 RESULTS

### 90 Identification of p16-high luminal cells with mesenchymal traits

91 Virgin Blg-Cre *Trp53<sup>fl/fl</sup>* *Brca1<sup>fl/fl</sup>* females develop mammary tumors at a median age of 5.2 months. In  
92 order to map early state transitions in the mammary gland, we profiled mammary epithelium from  
93 animals at various time points (2.7, 3.2 and 5.2 months, n=12), as well as from three tumors (Fig. 1A,  
94 n=15 mice in total). To maximize our chances of identifying tumor-initiating cells, among the n=15  
95 mice, we profiled the mammary epithelium of 2 mice at 5.2 months of age, with no apparent tumor,  
96 but belonging to a litter of an animal with a tumor. Among these individuals at 5.2 months of age, we  
97 observed multiple lesions - less than 0,5 mm - within the mammary epithelium (Fig. 1A, black arrows).  
98 Part of the collected samples were enriched for epithelial fraction to further increase our chances of  
99 identifying rare phenotypic states within the *Brca1/Trp53* deficient mammary epithelium (see  
100 Methods).

101 With single-cell transcriptomics, we identify n=11,113 epithelial cells (Fig. 1B) from all samples. Using  
102 reference lists of physiological markers (Bach et al. 2021, 2017; Watson and Khaled 2008), we identify,  
103 in 2.7-, 3-, and 5.2-month samples, physiological cell populations of the mammary gland: basal cells  
104 (*Krt5*) and clusters of luminal cells (*Krt8*) - luminal hormone-sensing (Luminal H-S, *Prlr*), luminal  
105 progenitor (LP, *Aldh1a3*) and secretory alveolar cells (*Avd*, *Csn2*) (Fig. 1C, Fig. S1A). The abnormal  
106 presence of secretory alveolar cells in the mammary gland of virgin mice at all timepoints, confirms  
107 the abnormal differentiation of luminal progenitors into alveolar cells in *Brca1/Trp53* deficient  
108 mammary glands, which had recently been observed during *Brca1* tumorigenesis (Bach et al. 2021).

109 With the objective to disentangle the earliest steps of *Brca1*-tumorigenesis, we next focused on  
110 clusters of cells belonging to earliest time points prior to tumor detection and including the smallest  
111 tumor (Fig. 1D, n= 3,525 cells). With unsupervised clustering, we identify a cluster of cells in between  
112 normal luminal compartments and tumor cells (green cluster, Fig. 1D) characterized by an unequivocal  
113 activation of *Cdkn2a/p16* compared to both LP and Avd (Fig. S1B-C, Table S1). This partition originates

114 mainly from pre-tumoral glands with lesions (at 5.2 months of age) (Fig. S1D, adj.p value < 5.0 10<sup>-2</sup>,  
115 Fisher's test), but few cells also belong to mammary glands of 2.7 and 3 month-old animals.  
116 In terms of identity, these cells show a significant down-regulation of genes characteristic of luminal  
117 compartment, compared to LP and Avd cells - e.g *Krt8*, *Krt18*, *Csn2* (Fig. 1E, Fig. S1B, Table S1). In  
118 addition to a partial repression of epithelial cytokeratins, a series of transcriptional changes testify of  
119 dampened epithelial characteristics and acquisition of mesenchymal features: (i) upregulation of *Vim*,  
120  *and  *(Fig. 1E, Table S1), indicative of changes in cytoskeleton and extracellular matrix, and (ii)  
121 down-regulation of *Cdh1* and several Claudin genes (*Cldn4*, *3* and *1*, Table S1), indicative of the  
122 dissolution of adherens and tight junctions. Such intermediary epithelial/mesenchymal state recalls  
123 partial EMT phenomena (Yang et al. 2021). In addition, cells from the p16-high cluster display a specific  
124 downregulation of *Lmna* (Table S1), indicative of diminished nuclear stiffness, potentially increasing  
125 their migration potential (Harada et al. 2014). This cell population also presents a higher cycling rate  
126 than normal epithelial cells, similar to that of tumor cells (Fig. S1E).**

127 We next sought to validate the detection of p16-high cycling cells *in situ*. We performed multiplex  
128 immunohistochemistry on paraffin-embedded formalin-fixed (FFPE) sections from mice with lesions  
129 or tumors, and control Cre- mammary gland, studying for over 70,090 cells: i) cell identity (*Krt8*, *Krt5*),  
130 ii) cell cycle status (p16, Ki67), and iii) epithelial to mesenchymal plasticity (EMP) (*E-cadherin*, *N-*  
131 *cadherin*, *Vim*) (Fig. 1F, Fig. S1). In lesions, we could detect patches of p16-high cells and vimentin-  
132 based filaments (Fig. 1F, Fig. S1I). In full grown tumors, in addition to p16 and vimentin, we also  
133 detected N-Cadherin and N-Cadherin/E-Cadherin-positive cells (Fig. S1F). While we did not observe  
134 any p16 positive cells in the Cre- animals (Fig. 1G), they were found within juxta-tumoral tissues of  
135 tumor-bearing animals, as isolated single cells located in the inner part of the duct, within the luminal  
136 compartment. Such initial localization, in addition to the co-occurrence of p16 and *Krt8* stainings (Fig.  
137 1F, Fig. S1G) confirms their luminal origin (89% of the total p16 positive cells). Within juxta-tumoral  
138 tissue and lesions, more than 32% of p16 positive cells are Ki67+, supporting their capacity to escape  
139 p16-mediated cell cycle arrest (Fig. S1H-I). The proliferative index of p16 positive cells is significantly

140 higher in juxta-tumoral duct and lesions, compared to p16 positive cells in tumors (15% Ki67+) and  
141 control mammary epithelial cells (6% Ki67+), suggesting that p16 positive cells at the onset of *Brc1*  
142 tumorigenesis are particularly proliferative (Fig. 1F, Fig. S1H).

143

#### 144 **Partial EMT in luminal cells precedes tumor formation**

145 We then sought to reconstitute the potential timeline of state transitions leading luminal progenitors  
146 to tumor formation. With the algorithm PHATE (Moon et al. 2019) – which works without any specified  
147 root state – we first show the existence of a progression between the luminal towards the tumor cell  
148 state, identifying the p16-high EMT cell state as the intermediary cell state (Fig. S2A). With the PAGA  
149 algorithm, we further map and quantify connections between cell states (Wolf et al. 2019) and show  
150 that the p16-high EMT state is the most connected to all other cell states found in the dataset (Fig.  
151 2A). In addition, the tumor cell states are only reachable through this central hub. These results  
152 position the p16-high EMT state as a prerequisite for luminal progenitors to transform and become  
153 tumor cells. We will refer to this state as ‘pre-tumoral state’ for the rest of the study.

154 We next further studied which biological pathways characterized the transition from luminal  
155 progenitor to the pre-tumoral state. We first used the Slingshot algorithm to infer a pseudotime  
156 ordering of the cells, taking the LP cluster as the root (Fig. 2B). Next, we identified the most correlated  
157 genes to Slingshot pseudo-time values and performed pathway enrichment and quantification using  
158 MsigDB Hallmark as a reference (Liberzon et al. 2015). We show that while transitioning from LP to  
159 tumor state, pre-tumoral cells activate angiogenesis and EMT (Fig. 2C), while inhibiting pathways of  
160 apoptosis and estrogen response (Fig. S2B). Such transcriptional signatures endorse the pre-tumoral  
161 nature of this intermediate state: angiogenesis and inhibition of apoptosis are canonical hallmarks of  
162 cancer cells (Hanahan and Weinberg 2016) - meant to enable fast growth of cells.

163 As loss of *BRCA1* impairs homologous repair (HR) mechanisms and leads to a major genomic crisis  
164 (Scully and Livingston 2000; Polak et al. 2017) - it appeared critical to understand when such a genomic  
165 crisis was occurring, and how it related to the state transitions we were observing. We used our

166 scRNAseq epithelial dataset to quantify Copy Number Variation (CNV) (Fig. S2C, Patel et al. 2014),  
167 taking the basal cells as reference, as the Cre is not expressed in these cells (Molyneux et al. 2010). For  
168 each cell, we calculated the percentage of their genome displaying CNVs (Fig. 2D). LP cells of tumor-  
169 free & lesion-free animals already display a high percentage of CNVs (median 23 %), even at 2.7 months  
170 (median 22.4%), compared to basal cells (Fig. 2B). Such observations imply that the LP compartment  
171 can tolerate numerous CNVs following *Brca1/Trp53* deletion, without any rapid phenotypic  
172 consequence. Rates of genome rearrangement in pre-tumoral cells are among the highest of the LP  
173 compartment (median 25%), yet their maximum rate does not exceed what is observed in the LP  
174 population (Fig. 2D). Our data show that the major outburst of CNVs occurs in the LP compartment  
175 prior to any tumor formation, and before the activation of p16 and partial EMT process. Such findings  
176 are in agreement with previous genomic studies which proposed that a large fraction of copy number  
177 alterations were acquired in short punctuated bursts at early stages of tumor formation (R. Gao et al.  
178 2016). In terms of clonality, if the LP compartment is multi-clonal and tumors rather poly-clonal, half  
179 of the pre-tumoral cell population is multi-clonal, (47%, Fig. S2C-D), showing that multiple cells can  
180 switch to the pre-tumoral state.

181 To decipher the molecular mechanisms driving initial state transitions *in vivo*, we next examined gene  
182 regulatory networks using two complementary approaches. With SCENIC (Aibar et al. 2017), we  
183 inferred the master regulons - defined as expressed Transcription Factors (TFs) with their putative  
184 expressed direct targets- active in our scRNAseq *in vivo* clusters (Fig. S2E). Out of 12 regulons, *Twist1*  
185 regulon (set of 27 target genes) was the only regulon positively correlated to pseudotime values (Fig.  
186 2E) and had a significantly higher expression (adj. p-value=  $3.0 \cdot 10^{-14}$ ) in the pre-tumoral cluster as  
187 compared to the LP and Avd ones (Fig. S2F). We also performed a TF enrichment test on the markers  
188 of the pre-tumoral state (Table1,  $\log_2FC > 0.8$ ) using Chea3 (Keenan et al. 2019) which, in contrast to  
189 SCENIC, does not rely on the expression of the TF itself, which can be hardly detected with scRNA-seq  
190 approaches. In addition to *Twist1*, previously identified, the top candidate TF predicted to drive the  
191 pre-tumoral expression programs (Fig. 2G) are key EMT-associated TF (*Twist2*, *Snail2*) as well as *Prrx2*,

192 a recently discovered EMT inducer in breast cancer (Lv et al. 2017), and several forkhead box (FOX)  
193 family of proteins.

194

#### 195 **Pre-tumoral state displays scars of past senescent-like process**

196 The top marker of the pre-tumoral state is *Cdkn2a/p16* (Fig. S1B; Table S1), a marker of cell-cycle arrest  
197 and senescence (Collado and Serrano 2010; Koppelstaetter et al. 2008; Di Micco et al. 2021; Campisi  
198 and d'Adda di Fagagna 2007). In line with this, the transcriptional profile of the pre-tumoral state is  
199 significantly enriched for senescence-related hallmark signatures (REACTOME\_Senescence Associated  
200 Secretory Phenotype, adj. p-value <  $2.0 \cdot 10^{-2}$ ). In addition, cells from the pre-tumoral cluster express a  
201 pro-senescence secreted factor, *Igfbp4* (Fig. 3), that can trigger senescence in neighboring cells  
202 (Severino et al. 2013). Yet, pre-tumoral cells are cycling (Fig. S1E, Fig. 1H), suggesting they have  
203 overcome cell cycle arrest. Combined over-expression in these cells of *Cdk4* and *Ccnd1* (Table S1; adj.  
204 p-value <  $2.0 \cdot 10^{-5}$ ), that together promote G1 to S transition, could for example help cells bypass cell-  
205 cycle arrest imposed by p16 overexpression (Roupakia, Markopoulos, and Kolettas 2021).

206 We looked for additional markers of senescence associated to p16 upregulation (Collado and Serrano  
207 2010) within tissues over different time points: presence of B-galactosidase (Bgal) and senescence-  
208 associated heterochromatin foci (SAHF). We could not quantify any Bgal within juxta-tumoral or tumor  
209 sections (Fig. S3A), however we identified SAHF-like structure in lesions by immunofluorescence (Fig.  
210 3A). SAHF were initially defined as main cores enriched in H3K9me3 mark, coated by enriched rings in  
211 H3K27me3 (Aird and Zhang 2013; Paluvai, Di Giorgio, and Brancolini 2020). As H3K9me3 mark is  
212 already organized into foci, corresponding to chromocenters, in non-senescent cells in mice (Probst  
213 and Almouzni 2008), we chose H3K27me3 staining to study changes in heterochromatin organization  
214 during tumorigenesis.

215 As expected, in control mammary glands from Cre- mice, H3K27me3 staining revealed one single foci  
216 per cell (Fig.3A-B), corresponding to the inactive X (Xi), whereas the remaining staining is  
217 homogeneously diffused in the nucleus (Fig. 3A-B). In lesions and tumors, we observed the

218 accumulation of multiple H3K27me3 foci (Fig. 3A-B), attesting a disruption of heterochromatin in these  
219 cells. In addition, H3K27me3-enriched foci tended to accumulate in ring-like structures, surrounding  
220 nuclear regions devoid of DNA as attested by negative DAPI staining - possibly corresponding to  
221 nucleoli (Cmarko et al. 2008).

222 We next investigated whether this major nuclear reorganization was associated with changes in  
223 genomic distribution of H3K27me3 during tumorigenesis. First, when comparing breadth of peaks  
224 across autosomes from tumor and normal cells, we could not identify any significant spreading of the  
225 H3K27me3 marks in the autosomal regions in tumor samples (Fig. S3B). At a more local scale, when  
226 comparing the genomic distribution of H3K27me3 in tumors versus cells of the physiological gland  
227 (Fig. S3E), we show that tumor cells present recurrent epigenomic differences with normal cells (Fig.  
228 S3C, PC1). In particular, several cell cycle genes (*Cdkn2a*, *Cdk12*, *Cdk6*) - both inhibitors and inducers -  
229 display a recurrent loss of repressive H3K27me3 enrichment in tumors (Fig. 3D), suggesting that local  
230 epigenomic remodeling could participate in the entry and exit of the cell cycle during tumorigenesis.  
231 Loss of H3K27me3 had already been shown to enable *Cdkn2a* transcriptional activation at the onset of  
232 senescence (Ito et al. 2018). Altogether, our data suggest that our pre-tumoral cells, now cycling, may  
233 have previously undergone a G1/S blockade and senescence-like phenomenon (Buj et al. 2021;  
234 Herranz and Gil 2018).

235

### 236 **Pre-tumoral signature of early luminal transformation is detected in low-stage basal-like human** 237 **tumors**

238 We next wanted to investigate whether we could find traces of our pre-tumoral state in human breast  
239 cancers. To do so, we defined a “mouse-derived pre-tumoral signature”, as the top over-expressed  
240 genes ( $\log_2FC > 0.8$  and adj. p-value  $< 5.0 \cdot 10^{-2}$ ) in the pre-tumoral mouse cluster compared to both LP  
241 and Avd compartments and studied its expression in publicly available large bulk RNAseq cohorts  
242 (Berger et al. 2018, Xu et al. 2015). First, studying the Pan Cancer cohort, we show that basal-like  
243 tumors have higher pre-tumoral scores than other breast cancer subtypes (Fig. 4A). Among basal-like

244 tumors, *BRCA1*-deficient tumors displayed slightly higher scores than *BRCA1* WT tumors (Fig. 4A). In  
245 addition, our mouse-derived pre-tumoral signature was significantly more expressed in early-stage (I)  
246 than late-stage tumors (II and III) (Fig. 4B).

247 We next derived a human pre-tumoral signature from our mouse pre-tumoral signature, selecting the  
248 genes most predictive of early stages in Pan Cancer basal-like tumors (n=13 genes out of 38, Fig. 4C).  
249 This refined signature included genes involved in EMT (*VIM*, *FN1* and *COL9A3*) and senescence-  
250 associated processes (*IGFBP4*, *SPP1*). *CDKN2A*/p16 was not included in the human pre-tumoral  
251 signature: its overexpression is specific to basal-like tumors (Fig. S4A) but not predictive of early stages.  
252 *CDKN2A* is indeed over-expressed in most basal-like tumors (n=128, 75%), irrespective of the stage,  
253 suggesting that its activation might be an early event that is kept throughout the life of the tumor. In  
254 contrast, genes from the human derived signature might correspond to genes overexpressed more  
255 specifically in early stages. We next tested the relevance of the human pre-tumoral signature in an  
256 independent dataset (CPTAC (Xu et al. 2015)). We confirmed that the human pre-tumoral signature is  
257 specific to low-stage basal-like tumors (Fig. 4D, E). Focusing on survival, we demonstrate that patients  
258 who display high expression scores of such pre-tumoral signature exhibit longer progression-free  
259 survival ( $p=2.2 \cdot 10^{-2}$ , Pan-Cancer), and a tendency to longer overall survival ( $p=8.3 \cdot 10^{-2}$ , CPTAC) (Fig. 4F-  
260 G). Overall, we demonstrate that the signature of initial luminal transformation - reflecting senescence  
261 and partial EMT processes - is detected specifically in early-stage basal-like tumors that more rarely  
262 recur.

263

#### 264 **Detection of abnormal luminal progenitors with pre-tumoral signature in *BRCA1* carriers**

265 We next tested whether we could detect signs of pre-tumoral states directly in luminal cells from  
266 *BRCA1* human carriers. In this context, in contrast to established human tumors and the mouse model  
267 presented above, only one copy of *BRCA1* is deficient and *TP53* is initially functional. We reasoned that  
268 the steps of early transformation would be detected in normal-like tissue of these patients, potentially  
269 looking (i) prior to tumor formation thanks to mastectomy samples, or (ii) in juxta-tumoral tissues. In

270 the latter case, we suspect that some luminal progenitors could be in a pre-tumoral state at that time,  
271 as at least one or a few of them had already transformed into a tumor.

272 We used a published dataset to investigate mammary epithelial cells of *BRCA1* carriers prior to tumor  
273 formation (Fig. S5, GSE161529, Pal et al. 2021). We performed the same semi-manual annotation  
274 procedure as in the first part of our study and focused solely on the mammary epithelial compartment  
275 from normal-like and *BRCA1*+/- pre-neoplastic samples (Fig. S5A). Studying all cells together, we could  
276 not detect clusters of epithelial cells specific to *BRCA1* carriers. Neither could we detect a pool of cells  
277 with higher CNV or pre-tumoral scores compared to normal LPs (Fig. S5B-C). To further search for  
278 unexpected cell states, we next studied each compartment with an independent principal component  
279 analysis (PCA). We show that LPs are the most affected cell type by *BRCA1* deficiency: the informative  
280 PCs with the highest explained variance were retrieved from the LP PCA projection (Fig. S5D, Methods).  
281 LPs in *BRCA1* carriers display transcriptional defects, as they aberrantly activate genes involved in  
282 mammary stem cell signatures and in senescence associated secretory phenotype (SASP), including  
283 the chemokine *IL6* and metalloproteinase *MMP3* (Fig. S5E-F, Table S2). Altogether, we did not detect  
284 aberrant luminal states in *BRCA1* deficient glands *per se*, prior to tumor formation, but identified  
285 activation of few senescence markers and pathways. In line with previous observations that *BRCA1*  
286 deficiency could lead to senescence-like states (Sedic et al. 2015). However, we could not detect signs  
287 of partial EMT or of our pre-tumoral state.

288 We next studied cells from juxta-tumoral tissues of *BRCA1* carriers, profiling 5 frozen biopsies using  
289 single nucleus RNA sequencing (Fig. 5A). All expected cell types and epithelial subtypes from a  
290 mammary gland were retrieved in the pooled datasets, averaging a total of 4,038 nuclei from the 5  
291 samples (Fig. 5B-C). Focusing on the epithelial compartment, we depicted the major subtypes using  
292 canonical markers; basal cells expressing *KRT5* and *KRT14*, mature luminal cells expressing *PRLR* and  
293 *ESR1*, luminal progenitors characterized by the expression of *ALDH1A3* and *KIT* (Fig. 5C). We also  
294 revealed the abnormal presence of alveolar differentiated cells (Avd, characterized by high *ELF5*  
295 expression) and a second cluster of LP (named LP\_2). The presence of alveolar cells attests an abnormal

296 differentiation process of luminal progenitors in *BRCA1* carriers, exactly as observed *in vivo* in mouse  
297 models above and by others (Bach et al. 2021). They were not detected in mastectomy samples,  
298 suggesting that aberrant differentiation either occurs only in some *BRCA1* carriers and not others,  
299 and/or only at the last stages prior to tumor formation.

300 The second cluster of luminal progenitors, LP\_2 cells, originated from all profiled samples (Fig. 5D),  
301 and harbored significant higher levels of genomic alterations, as compared to the remaining epithelial  
302 cell populations (average fraction of altered genome 8.4%, Fig. 5E). These cells exhibited the highest  
303 scores of the human pre-tumoral signature (Fig. 5F), and displayed among their top ten markers  
304 (compared to LP and Avd cells, Table S3): *FOXC1*, a mesenchymal master regulator during physiological  
305 development (Ray, Ryusaki, and Ray 2021; Haldipur et al. 2014), *INHBA/ActivinA* a ligand of the TGF-  
306 Beta pathway - a pathway known to trigger EMT (Zavadil and Böttinger 2005) - and *CCND1/cyclin D1*,  
307 just like mouse pre-tumoral cells (Fig. S5H). The TF *FOXC1* is suspected to play a role in EMT in several  
308 types of cancer cells (Zhu et al. 2017; Bloushtain-Qimron et al. 2008). In addition, LP\_2 cells partially  
309 lose their luminal identity, with a decreased expression of several luminal markers (*KRT7*, *KRT15*, *ELF5*  
310 & *ALDH1A3*, Fig. S5G), while others remain unchanged (*KRT8* e.g).

311 Altogether, LP\_2 cells show several similarities with mouse pre-tumoral cells: partial loss of luminal ID,  
312 expression program potentially orchestrated by TFs involved in EMT, an elevated level of genomic  
313 alterations, and cyclin D1 overexpression. We envisage these cells could (i) either be originating from  
314 the pool of *BRCA1*+/- luminal progenitors – impacted potentially by the presence of a tumor - or (ii)  
315 from the tumor itself; the former hypothesis being supported by the transcriptional similarity of LP\_2  
316 with LP\_1 cells.

317

318

## 319 **DISCUSSION**

320 Here we detect a continuum of rare state transitions occurring in epithelial cells prior to tumor  
321 formation, thanks to a mouse model launching tumorigenesis in multiple luminal progenitor cells.  
322 Luminal progenitor cells can tolerate multiple genomic alterations while not transforming and we  
323 actually show that partial epithelial to mesenchymal transition occurs at the initiation of tumor  
324 formation in these cells. In humans, despite the absence of such a continuum, we were still able to  
325 catch several abnormal luminal states in the mammary glands of *BRCA1* carriers, giving us a glimpse of  
326 potential human pre-tumoral states. In both species, we detect luminal progenitors partially losing  
327 their luminal identity and displaying signs of partial epithelial to mesenchymal transition. *In vivo*, we  
328 show that luminal progenitors, after encountering a senescence-like state, undergo partial EMT -  
329 driven by a myriad of potential transcription factors (*Twist1/2*, *Snail2*, Fox family members, *Prxx1/2*) -  
330 before forming tumors. In humans, we caught luminal progenitors activating the transcription factor  
331 *FOXC1*, a regulator of mesenchymal state during normal development ((Ray, Ryusaki, and Ray 2021))  
332 and suspected to play a role in cancer-associated EMT (Zhu et al. 2017; Yu et al. 2013; Li et al. 2018).  
333 Our data supports a role for EMT at the onset of tumorigenesis, suggesting it might not only be  
334 restricted to its well-known pivotal role in tumor cell dissemination and metastasis (Aggarwal et al.  
335 2021). Transient dampening of epithelial characteristics and acquisition of few mesenchymal-like  
336 features, could be essential for the rupture of the duct structure and formation of the initial tumor  
337 bud.

338 In both human and mouse *BRCA1* deficient contexts, we detect alveolar cells in mammary glands of  
339 non-gestating individuals, not only validating previous observations made in mice (Bach et al. 2021),  
340 but demonstrating the existence of such an aberrant differentiation process in humans. The  
341 unexpected presence of milk secreting cells could be a proxy of the extensive plasticity of the luminal  
342 progenitor compartment in such a context; as if in a *BRCA1* deficient context, there was a relaxation  
343 of the control of cell state in luminal progenitors. The occurrence of partial EMT in luminal progenitors  
344 could be another manifestation of the extensive plasticity of this compartment. What triggers exactly

345 the extensive plasticity of the LP population and subsequent state transitions to alveolar or pEMT state  
346 remains to be understood. *BRCA1* itself has recently been identified as a guardian of the epithelial  
347 states (Zhang et al. 2022) - inactivation of *BRCA1* by CRISPR leads to increased epithelial-to-  
348 mesenchymal plasticity in mammary cells. Alternatively, or in combination, state transitions could be  
349 driven by key TFs, as the ones detected in mouse and human pre-tumoral cells in our study (TWIST,  
350 SNAI or FOX family members). Another trigger of partial EMT could also be senescence - itself induced  
351 by extensive genomic rearrangements following *Trp53* and *Brca1* deletion. It has been proposed *in*  
352 *vitro* that EMT, driven by *Twist1* and 2, could help override Ras-induced senescence in mouse  
353 fibroblasts (Ansieau et al. 2008). In a therapy-induced senescence phenotype, it was also shown that  
354 senescence promotes reprogramming and cancer stemness (Milanovic et al. 2018), suggesting that  
355 non-genetic mechanisms could be tightly associated to the entry and exit of the senescent state in  
356 various contexts.

357 The time resolution of a mouse model was instrumental to partially bridge the gap between individual  
358 fixed observations made in humans from pre-tumoral and juxta-tumoral tissues and established basal-  
359 like tumors. There might be several differences between the timing of events in a mouse model and in  
360 humans, that will need to be further investigated with complementary mouse models. Here *in vivo*,  
361 both *Brca1* and *Trp53* are inactivated simultaneously in luminal progenitors. In humans, it was shown  
362 for *BRCA1* carriers that *TP53* mutation was actually among the earliest events in tumor formation  
363 (Martins et al. 2012), yet full *BRCA1* and *TP53* inactivation remains a stepwise process. Regarding  
364 activation of *CDKN2A/p16*, if ultimately the majority of basal-like breast cancers will express  
365 *CDKN2A/p16*, the timing of *CDKN2A* activation could differ between mice and humans. *CDKN2A*  
366 activation could depend on the *TP53* status of cells; cells undergoing major genomic rearrangement  
367 might rather over-express *CDKN1A/p21* in a *TP53* proficient context for example. In addition, our  
368 datasets suggest that pre-tumoral cells escape *CDKN2A*-associated cell cycle arrest through  
369 overexpression of *Ccnd1* and *Cdk4*; in contrast, in humans, overexpression of *CDKN2A* is frequently

370 associated with a loss of *RB1*, which we have not observed at the protein nor expression level in pre-  
371 tumoral cells (not shown). *RB1* could be lost later during tumorigenesis.

372 Finally, our work opens up several translational perspectives for the early interception of *BRCA1*  
373 tumorigenesis and potential patient stratification. Using mouse and human datasets, we were able to  
374 define a 'pre-tumoral signature', characteristic of the epithelial changes occurring at the onset of basal-  
375 like tumorigenesis. In human tumors, we show that this signature is specific to basal-like cancers. In  
376 addition, we show that it has prognostic potential: with basal-like tumors, patients with high pre-  
377 tumoral signature score have a significantly longer disease-free survival. Our results suggest that basal-  
378 like tumors with a high pre-tumoral signature score might have been detected at an earlier stage,  
379 hence with a better outcome. Our pre-tumoral gene signatures could constitute candidate biomarkers  
380 to detect early epithelial transformation and be favorable prognostic markers.

381 In terms of therapeutic targets, preventing the early state transitions occurring in the luminal  
382 progenitor compartment could be a relevant therapeutic avenue that we need to investigate. One  
383 strategy could be to target early epithelial-to-mesenchymal plasticity, targeting candidate TFs with  
384 proteolysis targeting chimeras, as shown for Brachyury or NF- $\kappa$ B (Samarasinghe and Crews 2021).  
385 Another Achilles heel of the pre-tumoral state could be the over-expression of *Ccnd1* that promotes  
386 the switch from G1 to S phase. In this line, cycling pre-tumoral cells might be particularly sensitive to  
387 CDK4/6 inhibitors.

388 **Methods**

389 **Animal models.** The generation of Brca1<sup>fl/fl</sup> and Trp53<sup>fl/fl</sup> mice has been previously described (Jonkers  
390 et al. 2001; Liu et al. 2007). Blg-Cre transgenic mice were purchased from The Jackson Laboratory. Mice  
391 strains were crossed to obtain Blg-Cre Trp53<sup>fl/fl</sup> Brca1<sup>fl/fl</sup> animals. Genotypes were determined by PCR  
392 (primers Cre: 3' CGAGTGATGAGGTTTCGCAAG 5' - 3' TGAGTGAACGAACCTGGTTCG 5'; primer Brca1 :  
393 3'TATCACCACTGAATCTCTACC 5' - 3' GACCTCAAACCTCTGAGATCCAC 5'; Trp53: 3'  
394 AAGGGGTATGAGGGACAAGG 5' - 3' GAAGACAGAAAAGGGGAGGG 5'). Mice were sacrificed by cervical  
395 dislocation. For each sample (gland or tumor), one piece was fixed in 4% paraformaldehyde (15710,  
396 Euromedex) for histological analysis, one piece was snap frozen in dry ice and stored at -80°C and one  
397 piece was kept fresh for the desired experimentation.

398

399 **Ethics statement.** All procedures used in the animal experimentations are in accordance with the  
400 European Community Directive (2010/63/EU) for the protection of vertebrate animals. The project has  
401 been approved by the ethics committee n°02265.02. We followed the international recommendations  
402 on containment, replacement and reduction proposed by the Guide for the Care and Use of Laboratory  
403 Animals (NRC 2011). We used as few animals as possible and minimized their suffering, no painful  
404 procedures were performed. The breeding, care and maintenance of the animals were performed by  
405 the Institut Curie animal facility (facility license #C75-05-18). Patients (n=5 juxta-tumoral tissue) gave  
406 informed consent for the use of their tissue in the study.

407

408 **Immunostaining.** Glands and tumors were fixed in 4%PFA/PBS at 4°C overnight, then washed with PBS  
409 (Gibco, 10010023) a first time for 1h and a second time at 4°C overnight. The samples were then passed  
410 through consecutive (50%, 60%, 70%) ethanol baths for 30 min each at room temperature. Paraffin  
411 embedding and sectioning (5µm) was performed by the experimental pathology department of Institut  
412 Curie. At the staining time, the slides are dewaxed by heating at 65°C for 1h and wash 2 times in Xylene  
413 10min, then rehydrated via consecutive bath: 2x Ethanol 100% (VWR 20821,31) 10min, 1x Ethanol 90%

414 5min, 1x Ethanol 80% 5min, 1x Ethanol 70% 5min, 1x Ethanol 50% 5min, 2x Water 5min. Retrieval  
415 treatment was performed by incubation in citrate buffer (C9999) for 20min at 95°C. After a 1h room  
416 temperature cooling, the slides are cleaned in PBS and permeabilized in permeabilization buffer (BSA  
417 2%, FBS 5%, Triton 0,3% in PBS) for 2h at room temperature. Primary Antibody incubation was done  
418 on blocking buffer (BSA 2%, FBS 5%, PBS) at 4°C overnight with Chicken Krt5 antibody 1:500 (905901),  
419 Rat Krt8 antibody 1:500 (MABT329), Rabbit H3K27me3 antibody 1:20 (C36B11), Rabbit p16 antibody  
420 1:100 (Abcam, ab211542). After 3 washes in PBS for 10 min each, incubation of the antibodies was  
421 performed for 2h at room temperature with: goat anti-rabbit Cy3 1:1000 (A10520), goat anti-rat Cy5  
422 1:1000 (A10525), goat anti-chicken Alexa Fluor 488 (A11039) 1:500, DAPI 0,5µg/ml. After 3 wash in  
423 PBS 10min, sections were mounted in Aquapoly mount media.

424

425 **LacZ staining.** Glands and tumors were directly fixed in PFA 4% for 2h and incubated in PBS, 30%  
426 Sucrose at least 24h. Samples were included in optimal cutting temperature OCT medium (23-730-751)  
427 in moulds and cooled on a metal support previously cooled on dry ice. The samples were stored at -  
428 80°C before being cut in a cryostat at -20°C in a 6µm section. Slides were stored at -80°C before use.  
429 For the staining, the slides were equilibrated at room temperature for 10-20 min and washed 3 times  
430 for 5 min at room temperature in the washing buffer: PBS, 2mM MgCl<sub>2</sub>, 1x Na-DOC, 0,02% NP40. After  
431 that, slides were incubated in the LacZ Stain: Washing solution, 10mM K<sub>3</sub>Fe, 10mM K<sub>4</sub>FE, 1,5 mg/ml  
432 X-Gal in a humidified chamber in the dark at 37°C for 4h to overnight. Slides were washed in a  
433 consecutive bath of: PBS for 1 min then for 15 min at room temperature, water for 15 min at room  
434 temperature and (optionally) Nuclear fast red for 5 min and 2 times in water for 5 min each. Sections  
435 were mounted in Aquapoly mount media.

436

437 **Microscopy, image acquisition and analysis.** Image acquisition of stained sections were done using a  
438 laser scanning confocal microscope (LSM780, Carl Zeiss) with a LD LCI PLAN-APO x40 or x65/08 NA oil  
439 objective. The acquisition parameters were: zoom 0.6; pixel size xy 554 nm; spectral emission filters

440 (bandwidth): 414-485 nm, 490-508 nm, 588-615 nm, 641-735 nm; laser wavelengths: 405, 488, 561  
441 and 633 nm. Images were captured using Metamorph. Image processing was performed using Fiji  
442 Software, version 1.0. The counting of  $\mu$ -HF was done in Fiji with a custom macro, for each nucleus, we  
443 selected the most representative Z, then the counting was done automatically with the AutoThreshold  
444 MaxEntropy.

445

446 **Multiplex histological staining.** Multiplexed IHC was performed according to the protocol developed  
447 by (Remark et al. 2016), with some adjustment. Tissues were baked at 60°C for 1h, deparaffinized in  
448 Xylene (Fisher Scientific, 10467270) and rehydrated. The heat-induced epitope retrieval was done with  
449 pH6.1 citrate buffer (Dako, S169984-2) or pH9 EDTA buffer (Dako, S236784-2) in a 95°C water bath for  
450 30 minutes for the first staining (otherwise 15min) followed by incubation in REAL peroxylase blocking  
451 solution (Agilent Dako, S202386-2) for 10 minutes. If the primary antibody was the same species as  
452 any antibody used in prior stains, another blocking step was added with Fab Fragment, only for anti-  
453 rabbit (Jackson ImmunoResearch Europe Ltd, 711-007-003) during 20 minutes. Protein block serum  
454 free (Agilent Dako, X090930-2) was added for 10 minutes. Primary antibody was incubated for 1 or 2  
455 hours at room temperature or overnight at 4°C. The primary antibody was detected using a secondary  
456 antibody directed against the first one, conjugated with horseradish peroxylase (Anti-rabbit: Agilent  
457 Dako, K400311-2) (Anti-rat: BioTechne, VC005-050) followed by chromogenic revelation with 3-amino-  
458 9-ethylcarabazole (AEC) (Agilent Dako, K3468). Slides were counterstained with hematoxylin (Thermo  
459 Scientific, 6765001) and mounted with Glycergel aqueous mounting medium (Dako, C056330-2). After  
460 scanning (Philips Ultra Fast Scanner 1.6 RA), tissues were bleached with ethanol baths and another  
461 cycle was performed starting with the heat induced epitope retrieval.

462

463 **Overlay of multiplex histological stainings.** Histological analysis was performed using the open-source  
464 image analysis QuPath software (QuPath-0.3.2, <http://qupath.github.io/>) (Bankhead et al., n.d.) and  
465 ImageJ/Fiji. We created a new QuPath project containing all scans of each slide which allow us to crop

466 and export (BioFormats plugin) and then overlay the images using Fiji script following these different  
467 steps: 1. Color deconvolution (separation of hematoxylin and AEC signal); 2. Alignment on hematoxylin  
468 images; 3. Creation of transformation matrix on AEC images; 4. For a part of the staining (Edac, Vim,  
469 Ki67) an automatic threshold using MaxEntropy was done to remove background, for the rest of the  
470 stainings (p16, Krt5, Krt8, Ncad) different threshold was determined using control cell signal (cf.  
471 Computational part). Each staining was colored as desired. To further analysis, the composite image  
472 was transferred back to QuPath. By hand, the different structures of the gland/tumors were annotated  
473 (duct, stroma, juxta-lesion or juxta-tumoral duct, lesion, tumor). To identify all the cells, we used the  
474 'cell detection' function based on hematoxylin nucleus staining. We then used the 'show detection  
475 measurement' function to export the annotation and the intensity signal for all staining for each cell  
476 and analyzed it in R.

477

478 **Multiplex histological data analysis strategy.** The resulting measurements were exported and  
479 analyzed in R (4.1.1). Briefly, high signal channels, corresponding to Ki67, Vim were thresholded by the  
480 Maximum Entropy algorithm, whereas the remaining channel markers were subjected to a custom  
481 thresholding approach. To identify true positive cells for each marker, mean "Cell" signal values were  
482 binarized as follows: - non-zero values of the Max Entropy thresholded markers were set to 1, whereas  
483 zero values were set to 0. To determine positive cells for p16, Ncad and Krt5, the local minimum after  
484 the highest peak was fitted on the density distribution of the merged cells from all the samples  
485 corresponding to each marker. Different thresholds were defined for each sample for the following  
486 markers: Krt8 and Ecad. Briefly, the "approxfunc" r interpolation function was applied on the density  
487 distribution of each marker on each sample, followed by an optimization step using the "optimize" r  
488 function to retrieve the local minimum within the interval of the density function. Higher values as  
489 compared to each threshold were set to 1, whereas smaller values were set to 0. basic r functions were  
490 used to calculate the percentages of positive cells for each or double positive for many markers, and

491 the ggplot package was used for graphical representations. Stromal cells were excluded in the  
492 analyses.

493

494 **Mammary gland / tumor dissociation and flow cytometry.** Samples were cut roughly with dissecting  
495 scissors and then with 2 scalpels for approximative 10 min. Then single cell dissociation was done by  
496 enzymatic digestion with 3mg/ml collagenase I (Roche, 11088793001) and 100U/ml hyaluronidase  
497 (Sigma-Aldrich, H3506) in complete media (HBSS (24020117), 5% SVF) during 1h30 under agitation at  
498 170 rpm at 37°C. Cells were then dissociated in PBS 0,25% Trypsin-Versen (Thermo Fisher Scientific,  
499 15040-033) prewarmed at 37°C for 1min30s with pipetting for 45s. The cell suspension was then  
500 treated with dispase 5 mg/ml (Sigma-Aldrich, D4693) and DNase 0,1 mg/ml (Roche, 11284932001) in  
501 complete media for 5 min at 37°C. A treatment with Red Blood cell lysis buffer (Thermo Fisher  
502 Scientific, 00-4333-57) was carried out then the suspension was filtered at 40µM before counting and  
503 FACS staining. Cell suspensions were stained 20 min in dark at 4°C with anti-CD45-APC 1:100  
504 (BioLegend, 103112), anti-CD31-APC 1:100 (BioLegend, 102510), anti-CD24-BV421 1:50 (BioLegend,  
505 101826), anti-CD49f-PE 1:50 (BioLegend, 313622). Cells were resuspended in cytometry media (PBS,  
506 BSA, EDTA). For the mammary gland samples, we either recovered the total epithelium or the luminal  
507 and basal cells populations separately.

508

509 **Single-cell / single-nuclei RNA-seq.** In accordance with the protocol of 10X Chromium manufacture,  
510 the cells were resuspended in PBS 0,04% BSA (Sigma, A8577). Depending on the samples,  
511 approximately 3000 or 4000 cells were loaded on the Chromium Single Cell Controller Instrument  
512 (Chromium single cell 3' v3 or 3' NextGem, 10X Genomics, PN-1000075) in accordance with the  
513 manufacturer's protocol. Libraries were prepared according to the same protocol. For the frozen  
514 samples, a nuclear extraction was performed before the microfluidic part.

515 Nuclei from frozen samples (around 20mg) were directly mechanically dissociated using dounce with  
516 pestle (15-20 strokes) in buffer containing 0,2U Superase IN (ThermoFisher, AM2696); 0,01M Tris HCL

517 pH 7,4; 0,01M NaCl; 0,003M MgCL2, 0,1M Igepal CA-630; 0,1M Tween-20; 0,001x Digitonin; 0,225%  
518 Sucrose in water. After vortexing briefly and pipette up and down 5-10 times with a micropipette,  
519 nuclei suspension is filtered with a 30um strainer. After a last centrifugation at 500g for 5 min at 4°C,  
520 nucleus are resuspended in PBS 0,2U SuperaseIN, 0,04% BSA, counted and processed on 10x  
521 Chromium.

522

523 **Bulk and single-cell ChIP-seq.** ChIP experiments were performed as previously described (Marsolier et  
524 al. 2022) using an anti-H3K27me3 antibody (Cell Signaling Technology, 9733 - C36B11). Bulk  
525 sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB, E7645S)  
526 according to the manufacturer's instructions. For single-cell experiments, cells were encapsulated on  
527 a custom microfluidic device as described before (Grosselin et al. 2019). Cells were stained with DAPI  
528 3µM or with 1µM CFSE during 15 min (CellTrace CFSE, ThermoFisher Scientific, Ref: C34554).

529

### 530 **COMPUTATIONAL ANALYSIS**

531 Code related to the following sections will be deposited on Github (<https://github.com/vallotlab>).

532

533 **Chromium 10X scRNAseq data pre-processing.** scRNAseq data acquisition was performed using the  
534 10X toolkit. Briefly, the CellRanger Software Suite (version 3.0.1) was used for demultiplexing, cell  
535 barcode assignment and further UMI (Unique molecular Identifier) quantification. The pre-built mm10  
536 reference genome proposed by 10X Genomics ([https://support.10xgenomics.com/single-cell-gene-](https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest)  
537 [expression/software/downloads/latest](https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest)) was used to align the reads. All the *in vivo* mouse datasets  
538 were analyzed together, without performing any batch correction. Doublet removal step was included  
539 in the 10X workflow, and was performed by the "emptyDrops" function from DropletUtils at an FDR of  
540 0.01.

541

542 **Quality Control (QC) for scRNAseq data analysis.** Low quality cells were defined as having aberrant  
543 values for the type and number of genes/UMIs detected. We evaluated the distribution of the total  
544 number of genes, molecules (UMIs) and the fraction of UMIs mapped to mitochondrial (MT) genes and  
545 set up thresholds to filter out those cells. Three upper cutoffs of 30% UMIs mapped to MT genes,  
546 10,000 genes and 100,000 nUMIs were fixed to get rid of outliers. Additionally, cells with less than  
547 1000 detected genes were excluded. This resulted in a total of 17,330 high quality cells, which were  
548 used for further analysis.

549  
550  
551 **scRNAseq data Normalization.** Normalization and variance stabilization were performed using the  
552 SCTransform method, implemented in the "SCTransform" function from the Seurat Suite. SCTransform  
553 omits the need for heuristic steps comprising log-transformation and pseudo-count addition, and  
554 results in improved downstream analytical steps. More recently, SCTransform also supports using the  
555 [glmGamPoi](#) package. Briefly, this method fits a "Gamma-Poisson Generalized Linear Model" to the  
556 overdispersed count matrices due to the high sparsity of the scRNAseq data, and results in a substantial  
557 improvement of the variance stabilization.

558  
559 **scRNAseq data dimension reduction and clustering.** Principal Component Analysis (PCA) was  
560 performed on the top 3000 Highly variable genes of the SCT assay from the SCTransform step, to  
561 reduce the data dimensionality. The top 60 PCs were further used to perform graph-based clustering  
562 and community (cell cluster) detection.

563 All the Uniform Manifold Approximation and Projection (UMAP) plots were computed using the  
564 "RunUMAP" Seurat function with default parameters ("uwot" as umap.method, n.neighbours=30,  
565 distance metric= "cosine", min.dist=0.3) and "random.state=42". The two-dimensional UMAP  
566 coordinates were calculated using the top 60 PCs previously computed on the SCT assay. For the sake  
567 of clarity, once the epithelial compartment is sub clustered, the same UMAP embeddings were used

568 to represent the “transitioning cell clusters”. Further “zoom ins” were performed using the  
569 corresponding umap coordinates of the cells of interest.

570

571 **Graph-based clustering and cell cluster identification.** Cell clustering was performed using a two-step  
572 wise approach, using the “FindNeighbours” and “FindClusters” respectively. Briefly, a k-Nearest  
573 Neighbours (kNN) graph is built on the dissimilarity matrix based on the pairwise euclidean distance  
574 between cells in the PCA space (using the previously computed 60 PCs). Edges are drawn between  
575 nodes (cells) with similar expression patterns (Jaccard Similarity). Edge weights are refined based on  
576 their shared overlap in their neighborhood.

577 “FindClusters” function was used to cluster the cells, using the Louvain algorithm as default, setting  
578 the resolution parameter to 1.2 to ensure an optimal granularity and stability of the cell clusters.

579

580 **scRNAseq cluster annotation.** Manual annotation of the cell clusters was performed on the merged  
581 samples on a two-steps basis. First, the cell clusters were annotated according to the major cell  
582 compartments, using well-established canonical markers. The latter included: Immune (*Ptprc+*, *Cd68+*,  
583 *Cd52+*), Epithelial (*Epcam*, *Krt5*, *Krt8*, *Elf5*), Endothelial (*Pecam1*, *Fabp4*, *Apold1*) and Fibroblasts (*Mgp*,  
584 *Dpep1*, *Col3a1*). Briefly, we computed the mean expression of each gene across the cells belonging to  
585 each cluster, to obtain a pseudo-bulked matrix containing only the genes of interest. A classical  
586 hierarchical clustering was performed on the clusters based on their correlation distance matrix to  
587 determine the cell cluster groups “Metaclusters” which displayed similar expression levels for each  
588 canonical gene signature. According to the dendrogram length, computed using the “ward.D” method,  
589 5 meta-clusters were identified. Each meta-cluster was assigned the cell type name for which the  
590 canonical genes were mostly expressed, as compared to the remaining genes. For instance, *COL3A1*  
591 displayed the highest expression level in the meta-cluster “1”. Therefore, all cell clusters previously  
592 defined (see **Graph-based clustering and cell cluster identification** section above) belonging to  
593 metacluster “1” are labeled as “Fibroblasts”.

594

595 **Refined Epithelial clusters annotation.** The epithelial compartment was further sub clustered to better  
596 explore the cell subtype/state heterogeneity between the control and the tumor samples.  
597 SCTransform, PCA dimension reduction and clustering steps were run on the subsetted clusters of  
598 interest. To achieve a high-resolutive cell subtype annotation, Differential expression (DE) was  
599 performed using the “FindAllMarkers” function. Briefly, a non-parametric Wilcoxon-Sum rank test was  
600 performed on a “1 cluster vs all” basis, setting a log2 Fold Change (FC) threshold at 0.5, and keeping  
601 only genes expressed in at least 30% of the cell clusters (to ensure expression homogeneity within the  
602 cluster). Associated p-values were corrected using the “Bonferroni” correction method, with a set  
603 threshold at 5%.

604 An automated function was designed to annotate the clusters. It takes as input the top 10 logFC ranked  
605 geneset for each cluster, and initially computes the contribution percentage of each tumor size feature  
606 of our dataset (control, control with primary lesions, small tumor (ST), medium tumor (MT) and large  
607 tumor (LT)) to each cell cluster. For a given cluster, if the major contributor is the control dataset, the  
608 function intersects the corresponding top genes with a knowledge-driven gene list of the known  
609 epithelial cell types (including basal, luminal progenitor, alveolar-differentiated, hormone-sensing ...)  
610 and labels the cluster with the corresponding cell type. If most of the cells (> 60%) were from tumor  
611 samples, the subtype name would be the concatenation of the top gene name with the tumor size  
612 symbol (ST, MT or LT).

613

614 **Differential expression.** Differential gene expression (DGE) analysis was performed using  
615 “FindMarkers” function. Non parametric Wilcoxon sum rank test was used to identify genes with an  
616  $\text{abs(FC)} > 0.5$  at an FDR of 0.05. To ensure cell cluster homogeneity, we set a lower cutoff of 30% of  
617 cells expressing a given gene.

618

619 **Pathway Enrichment Analysis (PEA).** Pathway Enrichment Analysis was performed on the significantly  
620 differentially expressed gene lists using the Hallmark collection from the Molecular Signature Database

621 (MSigDB). The latter was loaded into the R session using the “msigdb” package available on  
622 Bioconductor. Gene Set Enrichment Analysis was performed using the “enricher” function from the  
623 “msigdb” package. Only significantly enriched pathways (adjusted P-values < 0.05) were considered.

624

625 **Signature construction.** Transcriptional signatures were constructed from the gene lists contributing  
626 to each corresponding enriched pathway, using the “AUCell” package available on Github ([https://github.com/alexdaly/aucell](#)). Briefly,  
627 the genes of a given cell vs. gene data matrix are ranked based on their expression levels in each cell.  
628 AUCell computes then a Mann-Whitney U statistic (which is similar to AUC Area Under Curve), which is  
629 further used to evaluate gene signatures on the gene expression ranks of individual cells. We  
630 computed the gene signatures using the wrapper function “AddModuleScore\_AUCell”, giving as input a  
631 list of features, along with the Seurat object.

632

633 **Trajectory inference - Slingshot.** Pseudotime ordering of cells was conducted using Slingshot (Github  
634 [https://github.com/alexdaly/slingshot](#)), with default parameters, giving as input the UMAP coordinates and setting the starting cluster as  
635 the luminal progenitors “LP”, with stretch=2.

636 To ease the interpretation of the trajectory, we performed Slingshot only on the transitioning  
637 compartment, including (“LP”, Alveolar differentiated “Avd”, Luminal differentiated hormone-sensing  
638 “Luminal H-S”, and the annotated clusters of the small tumor. Downstream analytical steps were  
639 performed only on the longest branch starting from the “LP” and ending in the “Fgf8+ ST” cluster.

640

641 **Contribution of genes to a branch tree.** The aim of this section was to identify the most contributing  
642 genes to the transition observed from the Slingshot trajectory inference. To do so, a cell vs. gene  
643 expression matrix was created including the contributing cells to the longest branch, and the top 2000  
644 highly variable genes. We then applied a random forest regression model using 500 trees to predict  
645 the genes which contribute the most to predict pseudotime values (the response variable). The  
646 features (genes) were sorted according to their “variable.importance” parameter after the model was  
647 fit.

648

649 **Associated pathways to pseudotime values prediction.** We computed the mean expression values of  
650 the selected top 200 most important predictive genes to get pseudo-bulked matrices for the  
651 transitioning cells. To cluster the genes according to their profile correlation with pseudotime values,  
652 a pairwise-correlation matrix, followed by a hierarchical clustering were performed. 5 gene groups  
653 were obtained, each having a distinct profile along pseudotime. PEA (see below) was performed on  
654 each gene set, followed by a signature construction step and ultimately visualized on the UMAP  
655 embeddings.

656

657 **Partition-based graph abstraction (PAGA).** PAGA was performed using “scanpy” Python library loaded  
658 on RStudio using “reticulate” R package. Default parameters were used to construct the graph  
659 partition, and a threshold of 0.1 was set to preserve the highly connected nodes. Connectivity scores  
660 were extracted from the PAGA output, along with the nodes and edges connections. Centrality scores  
661 (number of edges) were computed by counting the number of edges that passed the cutoff (0.15) for  
662 each cell cluster.

663

664 **Potential of Heat-diffusion for Affinity-based Transition Embedding (PHATE).** PHATE was used as a  
665 visualization method to investigate continual progressions, branches and clusters in our data. Briefly,  
666 PHATE uses an information-geometric distance between cells (data points) to capture both local and  
667 global nonlinear structures, setting knn = 20, t (diffusion parameter) =40 as input parameters.

668

669 **Copy Number Variation (CNV) inference from scRNAseq data.** CNVs were inferred using inferCNV  
670 (<https://github.com/broadinstitute/infercnv>) with default parameters, taking as reference the basal  
671 cells. We extracted residual cell matrices, binarized the values using the 10th as lower and 90th  
672 percentile as higher thresholds, to get -1 (if the value < 10th percentile); +1 (if the value is higher than  
673 the 90th percentile) and 0 if the value is in between the two thresholds. To estimate the percentage  
674 of altered genome, we calculated the absolute value of binarized matrices, and counted the number

675 of 0s and 1s aggregated by chromosome. These values were added to the metadata of the scRNAseq  
676 Seurat object.

677

678 **TCGA\_Breast cancer dataset.** To compare the expression levels of CDKN2A, P16-signature, EMT and  
679 apoptosis pathway signatures, between non-diseased healthy tissues, tumor-adjacent normal tissue  
680 and tumor breast tissues, we harnessed breast tissues datasets from TCGA and GTEx consortia from  
681 normalized transcriptomic data available from Github  
682 (<https://github.com/mskcc/RNAseqDB/tree/master/data/normalized>). We constructed the gene  
683 signatures using the UCell package, and compared the tissue types using Wilcoxon T tests.

684

685 **scRNAseq data analysis of normal, preneoplastic and tumorigenic states in the human breast.** We  
686 downloaded the dataset from GEO, using the accession number: GSE161529. Briefly, we selected only  
687 the normal epithelium samples from pre-menopausal women (n=6), tumor samples (labeled as Triple  
688 Negative tumor, and Triple negative (Brca) tumor) (ntotal=8), and the nulliparous, pre-menopausal  
689 pre-neoplastic Brca1 samples (n=2). After sample merging, SCT normalization, dimension reduction  
690 and graph-based clustering, we selected the cell clusters expressing epithelial markers (Epcam, Krt8,  
691 Krt5) for further analysis. The same procedure was conducted on the epithelial compartment, followed  
692 by a finer annotation of the cell clusters using canonical markers of epithelial sub-populations. To point  
693 out the epithelial population which underwent major transcriptional modifications upon Brca1  
694 deficiency as compared to the normal population, we subset the luminal progenitor (LP), Basal and  
695 mature luminal (ML) clusters. For each subpopulation, principal component analysis (PCA) was  
696 performed, and the top 20 variable PCs were kept. To identify the main PC drivers of a  
697 normal/preneoplastic gradient, we tested whether the cell distributions along each PC coordinate  
698 were the same, using a Kosmogorov Smirnov nonparametric test. We selected the PCs with a  
699 significant p-value (<0.05) and a D-value > quantile(D-value,0.8). Alternatively, a linear regression  
700 method was tested to select the top predictive PCs to separate cells labeled as preneoplastic from

701 normal ones. Both methods indicated similar PCs. Next, to identify the epithelial sub-population for  
702 which the PCs were the most discriminant, we ranked the top “informative” PCs according to their  
703 percentage of variance explained. Pathway enrichment analysis was performed on the top genes  
704 (ranked by eigenvalues) which contributed most to the PC part corresponding to preneoplastic cells.

705

#### 706 **snRNAseq juxta-tumor data analysis**

707 Similar pre-processing steps were followed for the snRNAseq data analysis. Expression matrices from  
708 the 5 samples, obtained with the cellranger count suite were merged, filtered (setting 25% as the  
709 maximum rate of mitochondrial genes per cell), and integrated using Harmony algorithm. Similar semi-  
710 automatic approach was used to annotate the cell clusters according to major cell types (epithelial,  
711 adipocytes, fibroblasts, immune, fibroblasts and endothelial). Downstream analyses were only focused  
712 on the epithelial compartment, in which a refined annotation was performed to retrieve epithelial  
713 subtypes. Similarly, fraction of altered genome and scoring of the human-derived pre-tumoral state  
714 were estimated using inferCNV and UCell algorithms, respectively. Stack violin plots were plotted using  
715 a custom function.

716

717

#### 718 **DATA AVAILABILITY**

719 The datasets described in this study have been deposited in the public GEO repository GSE200444,  
720 accessible with the token gtoisgiftopnqr.

721

#### 722 **ACKNOWLEDGMENTS**

723

724 We thank Dr S. Fre for providing critical discussion. We also thank the animal facility, the sequencing  
725 and imaging platforms from Institut Curie. We thank Dr J. Jonkers for providing mouse strains.

726

#### 727 **FUNDING**

728

729 This work was supported by the ATIP Avenir program, by Plan Cancer, by the SiRIC-Curie program SiRIC  
730 Grants #INCa-DGOS-4654 and #INCa-DGOS-Inserm\_12554, support from Bettencourt-Schueller  
731 Foundation and by a starting ERC grant from the H2020 program #948528-ChromTrace (CV). High-  
732 throughput sequencing was performed by the ICGex NGS platform of the Institut Curie supported by  
733 the grants Equipex #ANR-10-EQPX-03, by the France Genomique Consortium from the Agence  
734 Nationale de la Recherche #ANR-10-INBS-09-08 ("Investissements d'Avenir" program), by the ITMO-  
735 Cancer Aviesan - Plan Cancer III and by the SiRIC-Curie program SiRIC Grant #INCa-DGOS- 4654.

736

737 **DECLARATION OF INTERESTS**

738

739 The authors declare no competing interests.

740

741

742

743

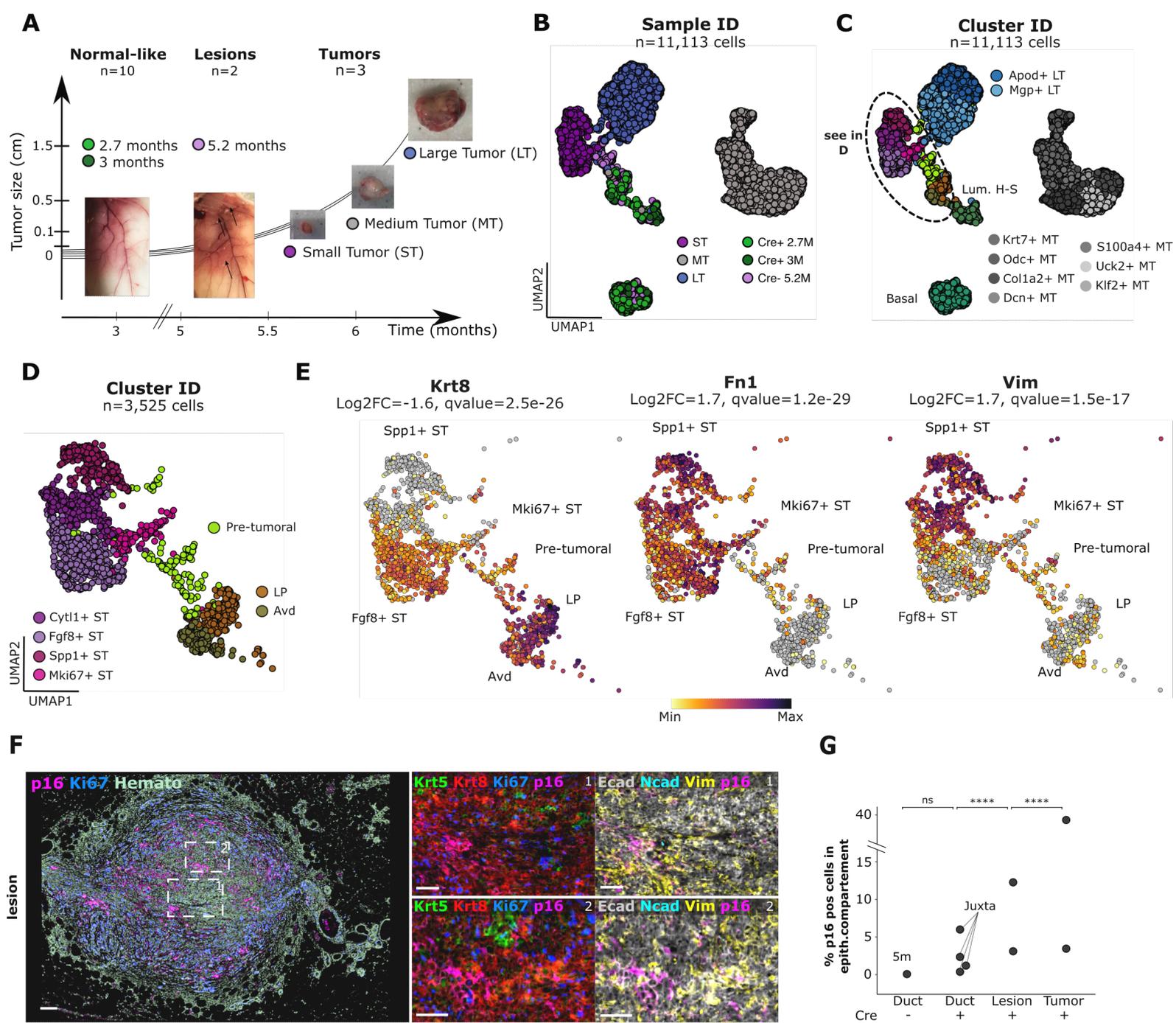
744

745

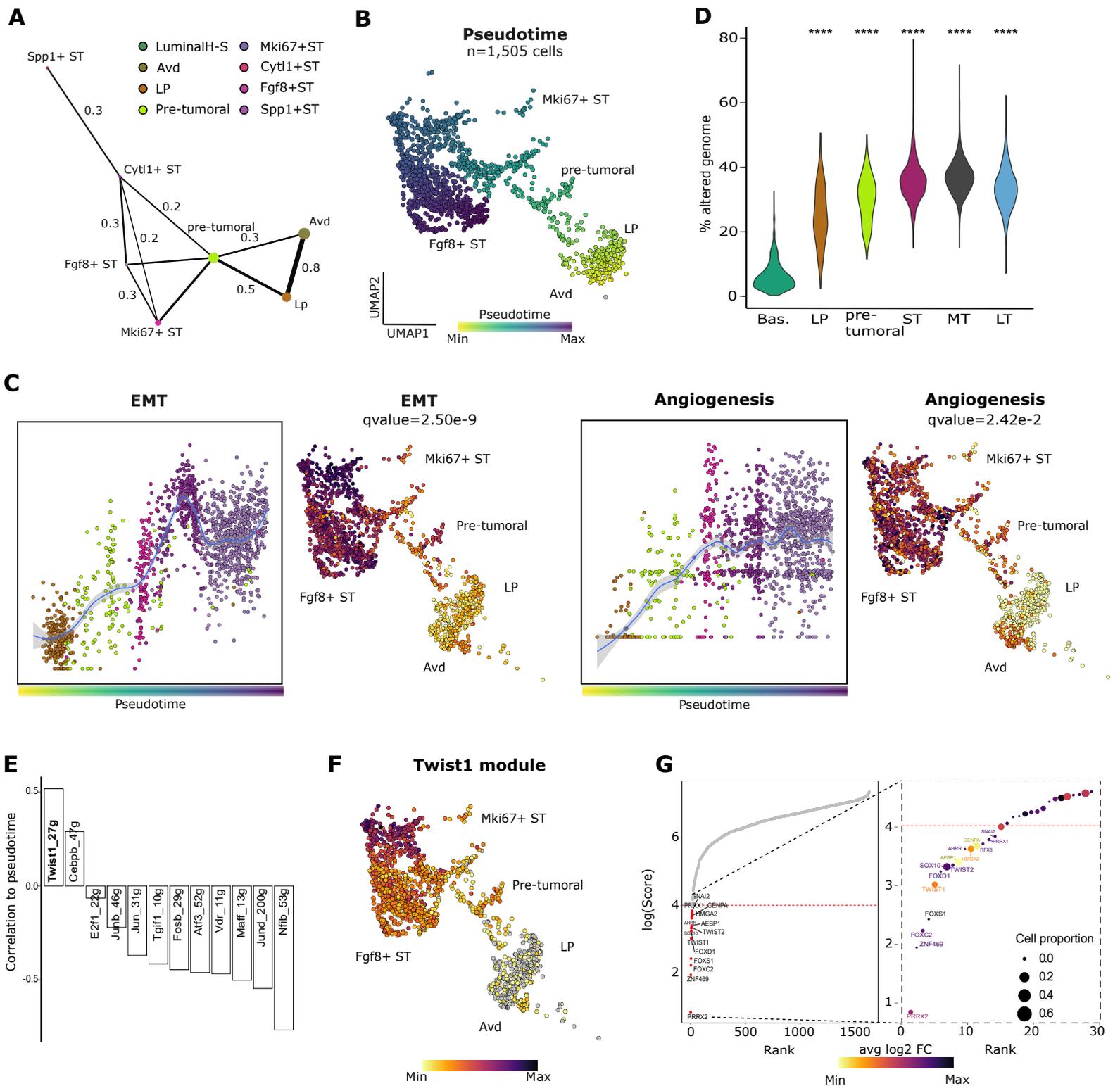
746

747

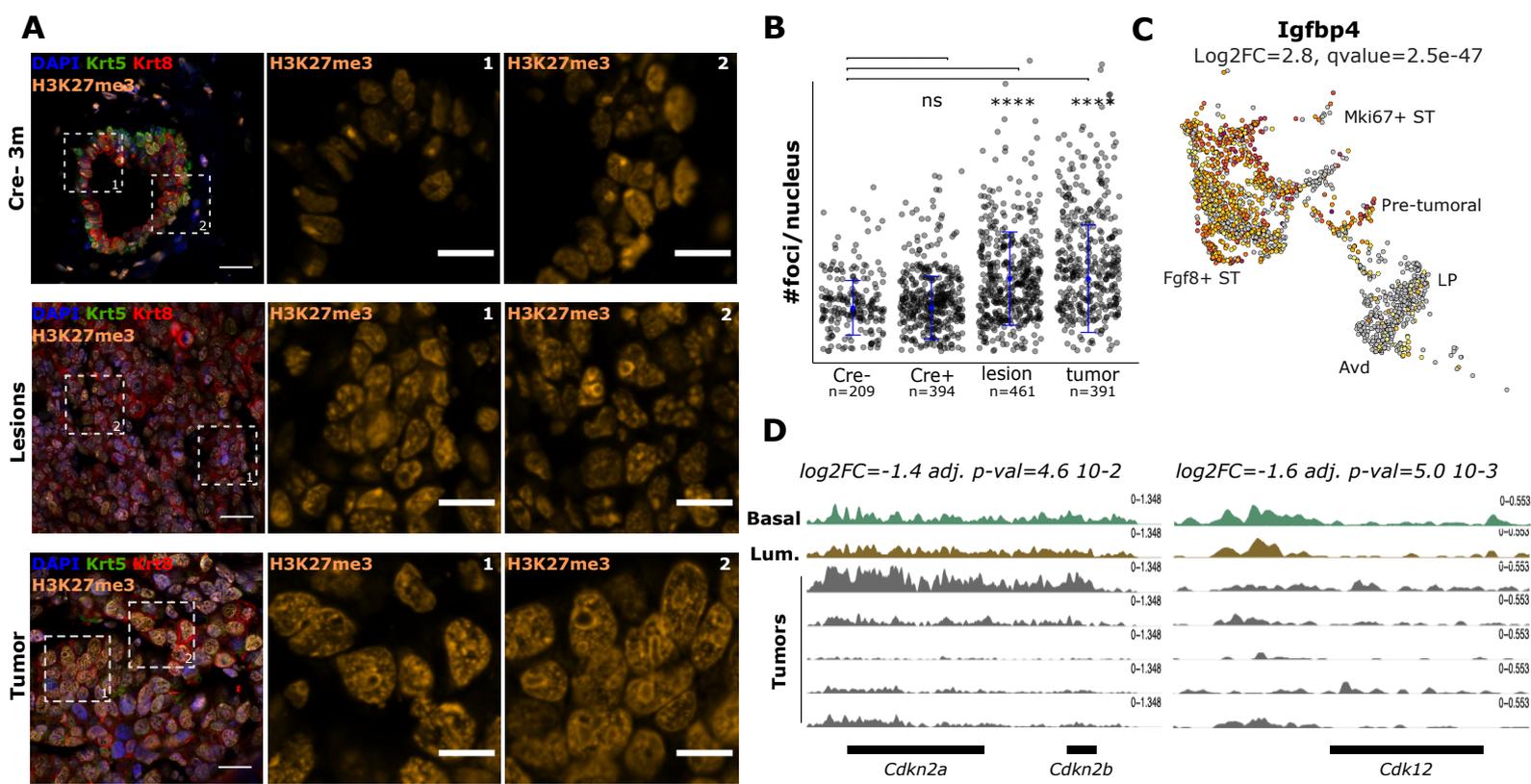
748

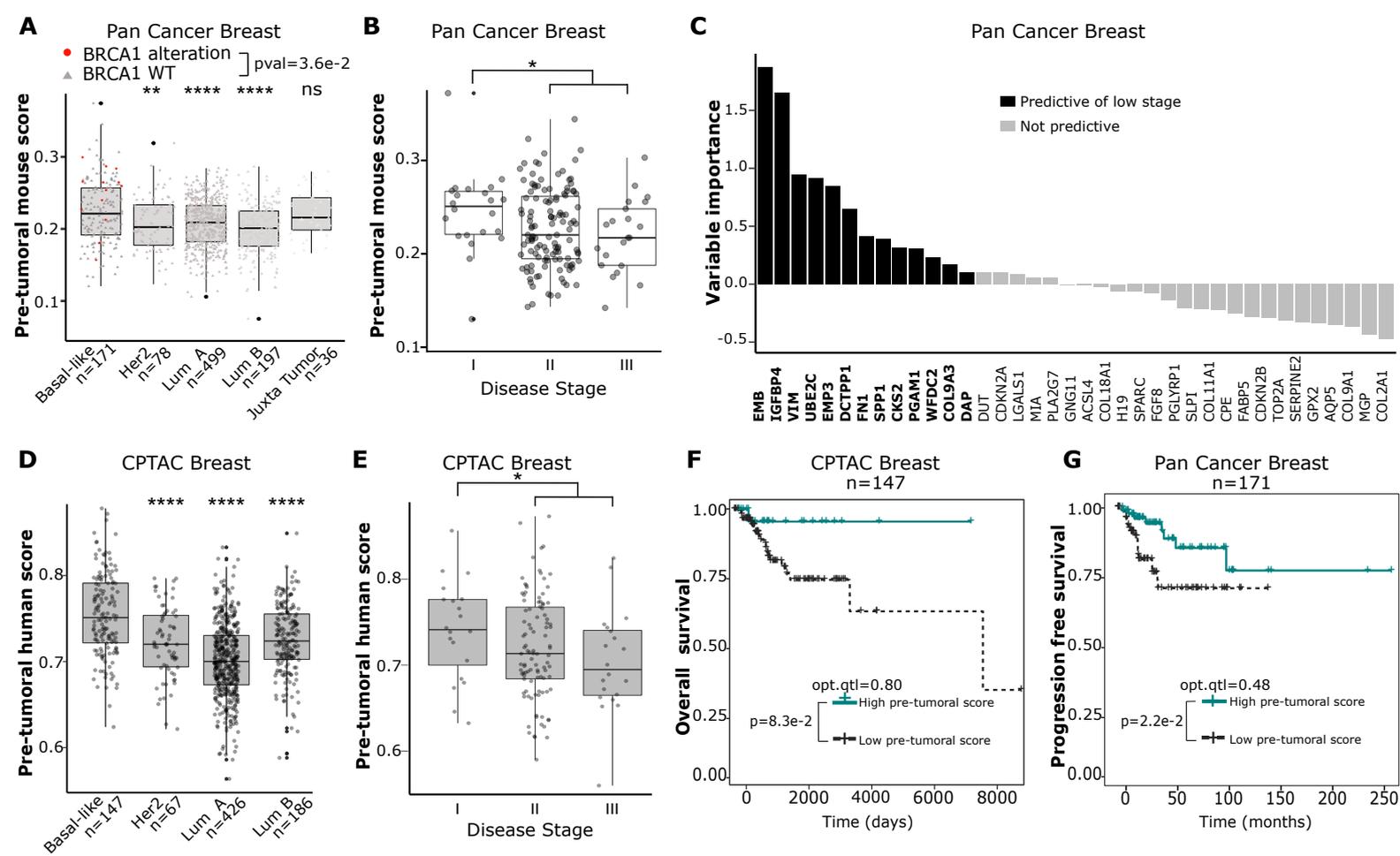


**Figure 1: Identification of p16-high luminal progenitors with mesenchymal features.** (A) Schematic representation of the timeframe and histological classification of the *in vivo* processed samples used for the scRNAseq. (B) UMAP representation of the epithelial cells; each dot represents a cell, and is colored according to the sample of origin. (C) UMAP colored according to cluster-based annotation. (D) Zoom in on the corresponding UMAP embeddings to the transitional cell clusters (see M&M). Cells were colored according to cluster-based annotation. (E) UMAPs representation with cells colored according to expression levels of top up-regulated and down-regulated genes in pre-tumoral cluster. (F) Left: Pseudo-colored multiplex IHC staining for p16 (pink), Ki67 (blue) in mammary gland with lesion, scale bar represents 100 $\mu$ m. For zooms, sections are colored according to identity markers Krt5 (green), Krt8 (red), cell cycle markers Ki67 (blue) and p16 (magenta) and EMT markers Ecad (white), Ncad (cyan) and Vimentin (yellow), scale bar represents 50 $\mu$ m. (G) Dot plot representation of the percentage of p16 positive (p16+) cells in normal-like ducts from juxta-tumoral regions, tumors and including a negative control "Duct 5m Cre-". Asterisks above each sample category represent the significance levels of pairwise comparisons between the 4 categories with Fisher's test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

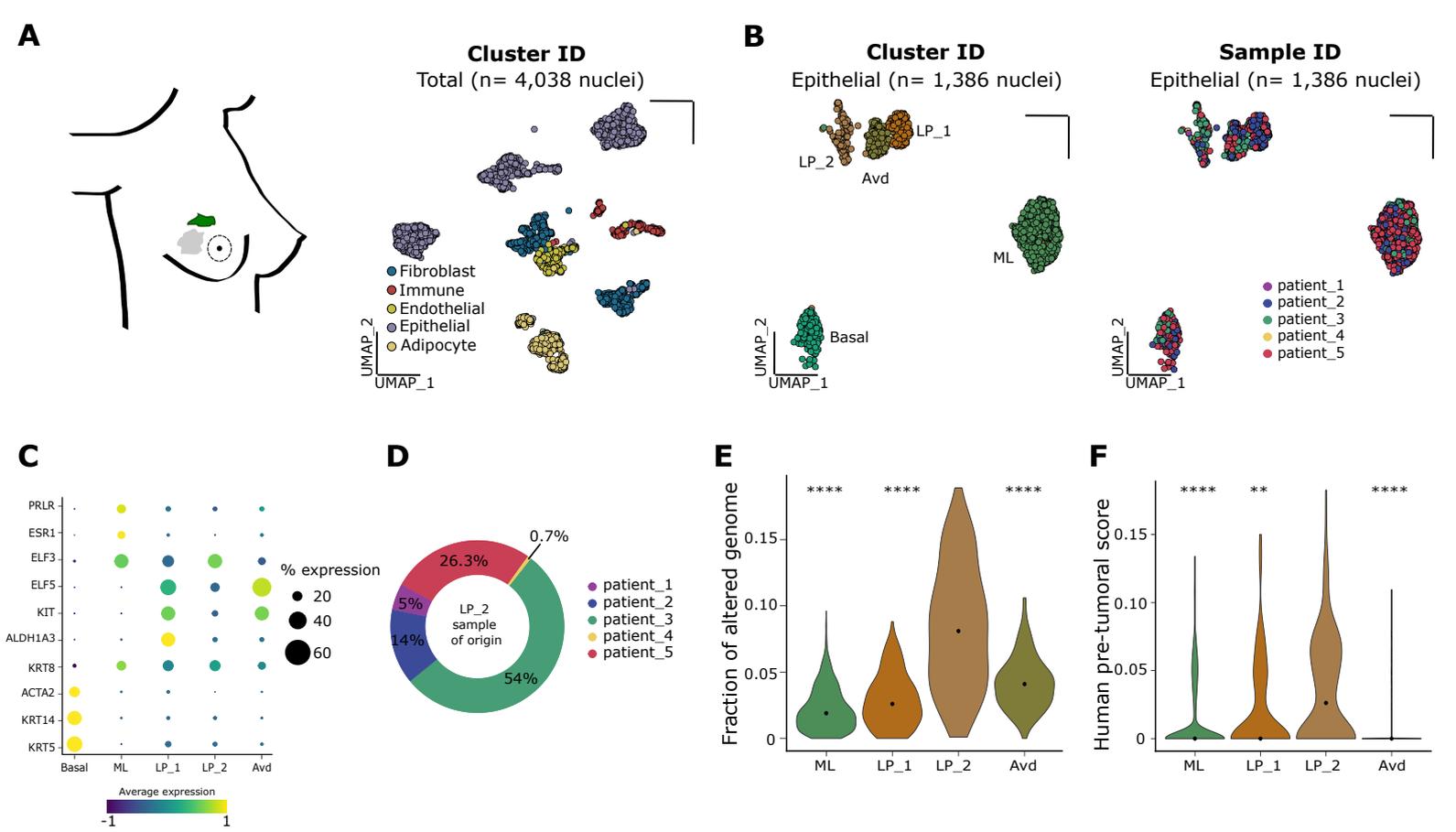


**Figure 2: Partial EMT occurs at the onset of tumorigenesis.** (A) Partition-based graph abstraction (PAGA) graphical representation of the transition cluster, previously represented on UMAP embedding (Fig. 1D); nodes are the clusters and the edge thickness quantifies the connectivity scores between the graph-partitions highlighted on the graph. (B) UMAP representation with cells colored according to inferred pseudotime values, using the Slingshot algorithm. (C) Left: Scatter plot representation of UCell scores for EMT and Angiogenesis signature, cells are ranked by increasing pseudotime values and colored according to their cluster ID. Right: UMAPs representation with cells colored according to Ucell scores for EMT and Angiogenesis signature. (D) Violin plot distribution of the percentage of altered genome per cells, grouped by cluster ID. Asterisks represent the significance levels of mean comparison with basal cells. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . (E) Barplot of correlation values between pseudotime & AUC enrichment scores of TF modules inferred by SCENIC. (F) UMAP representation colored according to AUC scores of the Twist1-module inferred by SCENIC. (G) Dot plot representation of the 30 top candidate TFs which could regulate the pre-tumoral expression program. TF enrichment was inferred using ChEA3.





**Figure 4: Pre-tumoral signatures in human breast cancers.** (A-B) Boxplot representation of scores for the pre-tumoral signature derived from our *in vivo* mouse scRNAseq data according to breast cancer type (A), basal-like tumor stage from the PAN CANCER breast dataset (B). (C) Barplot representation of the variable values output from Random forest-based classification of the tumor stage, top 13 genes were selected to be the most predictive genes. (D-E) Boxplot representation of scores for the human-derived (top 13 genes) pre-tumoral signature according to tumor subtype (D), and tumor stage (E) for CPTAC dataset. (F-G) Kaplan-Meier overall / disease-free survival curve for basal-like tumors, stratified according to expression score of the human-derived pre-tumoral signature in the CPTAC (F), and PAN CANCER (G) breast cancer datasets.



**Figure 5: Abnormal luminal progenitors in juxta-tumoral tissue of BRCA1-/+ carriers.** (A) Scheme representation of the experimental workflow set-up & UMAP representation of integrated snRNAseq data from 5 patients, colored by annotated cell type. (B) UMAP of the epithelial subset from (A), where cells were colored according to their corresponding epithelial subtype and sample of origin. (C) Dot plot of the log<sub>10</sub> mean expression of canonical cell markers, mean expression is color-coded, whereas the percentage of expressing cells within corresponding cluster is size-coded. (D) Donut plot representation of the sample of origin of the LP<sub>2</sub> cell cluster. (E) Violin plot showing the distribution of the fraction of altered genome, i.e with copy-number variation according to inferCNV. Asterisks represent the significance levels of median comparison with LP<sub>2</sub> cells. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, wilcoxon rank sum test. (F) Violin plot showing the distribution of human pre-tumoral signature scores, scores of each cluster were compared to scores of the LP<sub>2</sub> cluster using a wilcoxon rank sum test.

## 749 REFERENCES

- 750 Aggarwal, Vaishali, Catalina Ardila Montoya, Vera S. Donnenberg, and Shilpa Sant. 2021. "Interplay between  
751 Tumor Microenvironment and Partial EMT as the Driver of Tumor Progression." *iScience* 24 (2): 102113.
- 752 Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert  
753 Hulselmans, Florian Rambow, et al. 2017. "SCENIC: Single-Cell Regulatory Network Inference and  
754 Clustering." *Nature Methods* 14 (11): 1083–86.
- 755 Aird, Katherine M., and Rugang Zhang. 2013. "Detection of Senescence-Associated Heterochromatin Foci  
756 (SAHF)." *Methods in Molecular Biology* 965: 185–96.
- 757 Ansieau, Stéphane, Jeremy Bastid, Agnès Doreau, Anne-Pierre Morel, Benjamin P. Bouchet, Clémence Thomas,  
758 Frédérique Fauvet, et al. 2008. "Induction of EMT by Twist Proteins as a Collateral Effect of Tumor-  
759 Promoting Inactivation of Premature Senescence." *Cancer Cell* 14 (1): 79–89.
- 760 Bach, Karsten, Sara Pensa, Marta Grzelak, James Hadfield, David J. Adams, John C. Marioni, and Walid T.  
761 Khaled. 2017. "Differentiation Dynamics of Mammary Epithelial Cells Revealed by Single-Cell RNA  
762 Sequencing." *Nature Communications* 8 (1): 2128.
- 763 Bach, Karsten, Sara Pensa, Marija Zarocsinceva, Katarzyna Kania, Julie Stockis, Silvain Pinaud, Kyren A. Lazarus,  
764 et al. 2021. "Time-Resolved Single-Cell Analysis of Brca1 Associated Mammary Tumorigenesis Reveals  
765 Aberrant Differentiation of Luminal Progenitors." *Nature Communications* 12 (1): 1502.
- 766 Bankhead, Peter, Maurice B. Loughrey, José A. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D.  
767 Dunne, Stephen McQuaid, et al. n.d. "QuPath: Open Source Software for Digital Pathology Image  
768 Analysis." <https://doi.org/10.1101/099796>.
- 769 Berger, Ashton C., Anil Korkut, Rupa S. Kanchi, Apurva M. Hegde, Walter Lenoir, Wenbin Liu, Yuexin Liu, et al.  
770 2018. "A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers." *Cancer Cell* 33  
771 (4): 690–705.e9.
- 772 Bianchini, Giampaolo, Justin M. Balko, Ingrid A. Mayer, Melinda E. Sanders, and Luca Gianni. 2016. "Triple-  
773 Negative Breast Cancer: Challenges and Opportunities of a Heterogeneous Disease." *Nature Reviews.*  
774 *Clinical Oncology* 13 (11): 674–90.
- 775 Bloushtain-Qimron, Noga, Jun Yao, Eric L. Snyder, Michail Shipitsin, Lauren L. Campbell, Sendurai A. Mani, Min  
776 Hu, et al. 2008. "Cell Type-Specific DNA Methylation Patterns in the Human Breast." *Proceedings of the*  
777 *National Academy of Sciences of the United States of America* 105 (37): 14076–81.
- 778 Buj, Raquel, Kelly E. Leon, Marlyn A. Anguelov, and Katherine M. Aird. 2021. "Suppression of p16 Alleviates the  
779 Senescence-Associated Secretory Phenotype." *Aging* 13 (3): 3290–3312.
- 780 Campisi, Judith, and Fabrizio d'Adda di Fagagna. 2007. "Cellular Senescence: When Bad Things Happen to Good  
781 Cells." *Nature Reviews. Molecular Cell Biology* 8 (9): 729–40.
- 782 Cancer Genome Atlas Network. 2012. "Comprehensive Molecular Portraits of Human Breast Tumours." *Nature*  
783 490 (7418): 61–70.
- 784 Cmarko, D., J. Smigova, L. Minichova, and A. Popov. 2008. "Nucleolus: The Ribosome Factory." *Histology and*  
785 *Histopathology* 23 (10): 1291–98.
- 786 Collado, Manuel, and Manuel Serrano. 2010. "Senescence in Tumours: Evidence from Mice and Humans." *Nature Reviews*  
787 *Cancer*. <https://doi.org/10.1038/nrc2772>.
- 788 Di Micco, Raffaella, Valery Krizhanovsky, Darren Baker, and Fabrizio d'Adda di Fagagna. 2021. "Cellular  
789 Senescence in Ageing: From Mechanisms to Therapeutic Opportunities." *Nature Reviews. Molecular Cell*  
790 *Biology* 22 (2): 75–95.
- 791 Engebraaten, Olav, Hans Kristian Moen Vollan, and Anne-Lise Børresen-Dale. 2013. "Triple-Negative Breast  
792 Cancer and the Need for New Therapeutic Targets." *The American Journal of Pathology*.  
793 <https://doi.org/10.1016/j.ajpath.2013.05.033>.
- 794 Gao, Bowen, Ying Qu, Bingchen Han, Yoshiko Nagaoka, Makoto Katsumata, Nan Deng, Shikha Bose, Liting Jin,  
795 Armando E. Giuliano, and Xiaojiang Cui. 2017. "Inhibition of Lobuloalveolar Development by FOXC1  
796 Overexpression in the Mouse Mammary Gland." *Scientific Reports* 7 (1): 14017.
- 797 Gao, Ruli, Alexander Davis, Thomas O. McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei-Ching Tsai, et al. 2016.  
798 "Punctuated Copy Number Evolution and Clonal Stasis in Triple-Negative Breast Cancer." *Nature Genetics*  
799 48 (10): 1119–30.
- 800 Grosselin, Kevin, Adeline Durand, Justine Marsolier, Adeline Poitou, Elisabetta Marangoni, Fariba Nemati,  
801 Ahmed Dahmani, et al. 2019. "High-Throughput Single-Cell ChIP-Seq Identifies Heterogeneity of  
802 Chromatin States in Breast Cancer." *Nature Genetics* 51 (6): 1060–66.
- 803 Haldipur, Parthiv, Gwendolyn S. Gillies, Olivia K. Janson, Victor V. Chizhikov, Divakar S. Mithal, Richard J. Miller,  
804 and Kathleen J. Millen. 2014. "Foxc1 Dependent Mesenchymal Signalling Drives Embryonic Cerebellar

805 Growth." *eLife* 3 (December). <https://doi.org/10.7554/eLife.03962>.

806 Hanahan, Douglas, and Robert A. Weinberg. 2016. "The Hallmarks of Cancer." *Oxford Textbook of Oncology*.

807 <https://doi.org/10.1093/med/9780199656103.003.0001>.

808 Harada, Takamasa, Joe Swift, Jerome Irianto, Jae-Won Shin, Kyle R. Spinler, Avathamsa Athirasala, Rocky

809 Diegmiller, P. C. Dave P. Dingal, Irena L. Ivanovska, and Dennis E. Discher. 2014. "Nuclear Lamin Stiffness Is

810 a Barrier to 3D Migration, but Softness Can Limit Survival." *The Journal of Cell Biology* 204 (5): 669–82.

811 Herranz, Nicolás, and Jesús Gil. 2018. "Mechanisms and Functions of Cellular Senescence." *Journal of Clinical*

812 *Investigation*. <https://doi.org/10.1172/jci95148>.

813 Ito, Takahiro, Yee Voan Teo, Shane A. Evans, Nicola Neretti, and John M. Sedivy. 2018. "Regulation of Cellular

814 Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-

815 Dependent Pathways." *Cell Reports* 22 (13): 3480–92.

816 Jonkers, Jos, Ralph Meuwissen, Hanneke van der Gulden, Hans Peterse, Martin van der Valk, and Anton Berns.

817 2001. "Synergistic Tumor Suppressor Activity of BRCA2 and p53 in a Conditional Mouse Model for Breast

818 Cancer." *Nature Genetics*. <https://doi.org/10.1038/ng747>.

819 Keenan, Alexandra B., Denis Torre, Alexander Lachmann, Ariel K. Leong, Megan L. Wojciechowicz, Vivian Utti,

820 Kathleen M. Jagodnik, Eryk Kropiwnicki, Zichen Wang, and Avi Ma'ayan. 2019. "ChEA3: Transcription

821 Factor Enrichment Analysis by Orthogonal Omics Integration." *Nucleic Acids Research* 47 (W1): W212–24.

822 Koppelstaetter, Christian, Gabriele Schratzberger, Paul Perco, Johannes Hofer, Walter Mark, Robert Ollinger,

823 Rainer Oberbauer, et al. 2008. "Markers of Cellular Senescence in Zero Hour Biopsies Predict Outcome in

824 Renal Transplantation." *Aging Cell* 7 (4): 491–97.

825 Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo.

826 2015. "The Molecular Signatures Database Hallmark Gene Set Collection." *Cell Systems*.

827 <https://doi.org/10.1016/j.cels.2015.12.004>.

828 Li, Ji, Peter S. Choi, Christine L. Chaffer, Katherine Labella, Justin H. Hwang, Andrew O. Giacomelli, Jong Wook

829 Kim, et al. 2018. "An Alternative Splicing Switch in FLNB Promotes the Mesenchymal Cell State in Human

830 Breast Cancer." *eLife* 7 (July). <https://doi.org/10.7554/eLife.37184>.

831 Lim, Elgene, François Vaillant, Di Wu, Natasha C. Forrest, Bhupinder Pal, Adam H. Hart, Marie-Liesse Asselin-

832 Labat, et al. 2009. "Aberrant Luminal Progenitors as the Candidate Target Population for Basal Tumor

833 Development in BRCA1 Mutation Carriers." *Nature Medicine* 15 (8): 907–13.

834 Liu, Xiaoling, Henne Holstege, Hanneke van der Gulden, Marcelle Treur-Mulder, John Zevenhoven, Arno Velds,

835 Ron M. Kerkhoven, et al. 2007. "Somatic Loss of BRCA1 and p53 in Mice Induces Mammary Tumors with

836 Features of Human BRCA1-Mutated Basal-like Breast Cancer." *Proceedings of the National Academy of*

837 *Sciences of the United States of America* 104 (29): 12111–16.

838 Lv, Zhi-Dong, Hai-Bo Wang, Xiang-Ping Liu, Li-Ying Jin, Ruo-Wu Shen, Xin-Gang Wang, Bin Kong, Hui-Li Qu, Fu-

839 Nian Li, and Qi-Feng Yang. 2017. "Silencing of Prrx2 Inhibits the Invasion and Metastasis of Breast Cancer

840 Both In Vitro and In Vivo by Reversing Epithelial-Mesenchymal Transition." *Cellular Physiology and*

841 *Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*

842 42 (5): 1847–56.

843 Marra, Antonio, Dario Trapani, Giulia Viale, Carmen Criscitiello, and Giuseppe Curigliano. 2020. "Practical

844 Classification of Triple-Negative Breast Cancer: Intratumoral Heterogeneity, Mechanisms of Drug

845 Resistance, and Novel Therapies." *NPJ Breast Cancer* 6 (October): 54.

846 Marsolier, Justine, Pacôme Prompsy, Adeline Durand, Anne-Marie Lyne, Camille Landragin, Amandine

847 Trouchet, Sabrina Tenreira Bento, et al. 2022. "H3K27me3 Conditions Chemotolerance in Triple-Negative

848 Breast Cancer." *Nature Genetics* 54 (4): 459–68.

849 Martins, Filipe C., Subhajyoti De, Vanessa Almendro, Mithat Gönen, So Yeon Park, Joanne L. Blum, William

850 Herlihy, et al. 2012. "Evolutionary Pathways in BRCA1-Associated Breast Tumors." *Cancer Discovery* 2 (6):

851 503–11.

852 Milanovic, Maja, Dorothy N. Y. Fan, Dimitri Belenki, J. Henry M. Däbritz, Zhen Zhao, Yong Yu, Jan R. Dörr, et al.

853 2018. "Senescence-Associated Reprogramming Promotes Cancer Stemness." *Nature* 553 (7686): 96–100.

854 Molyneux, Gemma, Felipe C. Geyer, Fiona-Ann Magray, Afshan McCarthy, Howard Kendrick, Rachael Natrajan,

855 Alan MacKay, et al. 2010. "BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors

856 and Not from Basal Stem Cells." *Cell Stem Cell*. <https://doi.org/10.1016/j.stem.2010.07.010>.

857 Moon, Kevin R., David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim,

858 et al. 2019. "Visualizing Structure and Transitions in High-Dimensional Biological Data." *Nature*

859 *Biotechnology* 37 (12): 1482–92.

860 Onitilo, Adedayo A., Jessica M. Engel, Robert T. Greenlee, and Bickol N. Mukesh. 2009. "Breast Cancer Subtypes

861 Based on ER/PR and Her2 Expression: Comparison of Clinicopathologic Features and Survival." *Clinical*

862 *Medicine & Research* 7 (1-2): 4–13.

863 Pal, Bhupinder, Yunshun Chen, François Vaillant, Bianca D. Capaldo, Rachel Joyce, Xiaoyu Song, Vanessa L.

864 Bryant, et al. 2021. “A Single-Cell RNA Expression Atlas of Normal, Preneoplastic and Tumorigenic States in

865 the Human Breast.” *The EMBO Journal* 40 (11): e107333.

866 Paluvai, Harikrishnareddy, Eros Di Giorgio, and Claudio Brancolini. 2020. “The Histone Code of Senescence.”

867 *Cells*. <https://doi.org/10.3390/cells9020466>.

868 Patel, A. P., I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, et al. 2014. “Single-

869 Cell RNA-Seq Highlights Intratumoral Heterogeneity in Primary Glioblastoma.” *Science*.

870 <https://doi.org/10.1126/science.1254257>.

871 Pervolarakis, Nicholas, Quy H. Nguyen, Justice Williams, Yanwen Gong, Guadalupe Gutierrez, Peng Sun, Darisha

872 Jhutti, et al. 2020. “Integrated Single-Cell Transcriptomics and Chromatin Accessibility Analysis Reveals

873 Regulators of Mammary Epithelial Cell Identity.” *Cell Reports* 33 (3): 108273.

874 Polak, Paz, Jaegil Kim, Lior Z. Braunstein, Rosa Karlic, Nicholas J. Haradhavala, Grace Tiao, Daniel Rosebrock, et

875 al. 2017. “A Mutational Signature Reveals Alterations Underlying Deficient Homologous Recombination

876 Repair in Breast Cancer.” *Nature Genetics* 49 (10): 1476–86.

877 Probst, Aline V., and Geneviève Almouzni. 2008. “Pericentric Heterochromatin: Dynamic Organization during

878 Early Development in Mammals.” *Differentiation*. <https://doi.org/10.1111/j.1432-0436.2007.00220.x>.

879 Ray, Tania, Terry Ryusaki, and Partha S. Ray. 2021. “Therapeutically Targeting Cancers That Overexpress FOXC1:

880 A Transcriptional Driver of Cell Plasticity, Partial EMT, and Cancer Metastasis.” *Frontiers in Oncology*.

881 <https://doi.org/10.3389/fonc.2021.721959>.

882 Remark, Romain, Taha Merghoub, Niels Grabe, Geert Litjens, Diane Damotte, Jedd D. Wolchok, Miriam Merad,

883 and Sacha Gnjjatic. 2016. “In-Depth Tissue Profiling Using Multiplexed Immunohistochemical Consecutive

884 Staining on Single Slide.” *Science Immunology*. <https://doi.org/10.1126/sciimmunol.aaf6925>.

885 Roupakia, Eugenia, Georgios S. Markopoulos, and Evangelos Kolettas. 2021. “Genes and Pathways Involved in

886 Senescence Bypass Identified by Functional Genetic Screens.” *Mechanisms of Ageing and Development*

887 194 (March): 111432.

888 Samarasinghe, Kusal T. G., and Craig M. Crews. 2021. “Targeted Protein Degradation: A Promise for

889 Undruggable Proteins.” *Cell Chemical Biology* 28 (7): 934–51.

890 Scully, R., and D. M. Livingston. 2000. “In Search of the Tumour-Suppressor Functions of BRCA1 and BRCA2.”

891 *Nature* 408 (6811): 429–32.

892 Sedic, Maja, Adam Skibinski, Nelson Brown, Mercedes Gallardo, Peter Mulligan, Paula Martinez, Patricia J.

893 Keller, et al. 2015. “Haploinsufficiency for BRCA1 Leads to Cell-Type-Specific Genomic Instability and

894 Premature Senescence.” *Nature Communications* 6 (June): 7505.

895 Selbert, S., D. J. Bentley, D. W. Melton, D. Rannie, P. Lourenço, C. J. Watson, and A. R. Clarke. 1998. “Efficient

896 BLG-Cre Mediated Gene Deletion in the Mammary Gland.” *Transgenic Research* 7 (5): 387–96.

897 Severino, V., N. Alessio, A. Farina, A. Sandomenico, M. Cipollaro, G. Peluso, U. Galderisi, and A. Chambery.

898 2013. “Insulin-like Growth Factor Binding Proteins 4 and 7 Released by Senescent Cells Promote

899 Premature Senescence in Mesenchymal Stem Cells.” *Cell Death & Disease* 4 (November): e911.

900 Stefansson, Olafur Andri, Jon Gunnlaugur Jonasson, Kristrun Olafsdottir, Holmfridur Hilmarsdottir, Gudridur

901 Olafsdottir, Manel Esteller, Oskar Thor Johannsson, and Jorunn Erla Eyfjord. 2011. “CpG Island

902 Hypermethylation of BRCA1 and Loss of pRb as Co-Occurring Events in Basal/triple-Negative Breast

903 Cancer.” *Epigenetics: Official Journal of the DNA Methylation Society* 6 (5): 638–49.

904 Timms, Kirsten M., Victor Abkevich, Elisha Hughes, Chris Neff, Julia Reid, Brian Morris, Saritha Kalva, et al. 2014.

905 “Association of BRCA1/2 defects with Genomic Scores Predictive of DNA Damage Repair Deficiency among

906 Breast Cancer Subtypes.” *Breast Cancer Research*. <https://doi.org/10.1186/s13058-014-0475-x>.

907 Watson, Christine J., and Walid T. Khaled. 2008. “Mammary Development in the Embryo and Adult: A Journey

908 of Morphogenesis and Commitment.” *Development* 135 (6): 995–1003.

909 Wolf, F. Alexander, Fiona K. Hamey, Mireya Plass, Jordi Solana, Joakim S. Dahlin, Berthold Göttgens, Nikolaus

910 Rajewsky, Lukas Simon, and Fabian J. Theis. 2019. “PAGA: Graph Abstraction Reconciles Clustering with

911 Trajectory Inference through a Topology Preserving Map of Single Cells.” *Genome Biology* 20 (1): 59.

912 Xu, Zhe, Chaochao Wu, Fang Xie, Gordon W. Slys, Nikola Tolic, Matthew E. Monroe, Vladislav A. Petyuk, et al.

913 2015. “Comprehensive Quantitative Analysis of Ovarian and Breast Cancer Tumor Peptidomes.” *Journal of*

914 *Proteome Research* 14 (1): 422–33.

915 Yang, Jing, Parker Antin, Geert Berx, Cédric Blanpain, Thomas Brabletz, Marianne Bronner, Kyra Campbell, et al.

916 2021. “Author Correction: Guidelines and Definitions for Research on Epithelial-Mesenchymal Transition.”

917 *Nature Reviews. Molecular Cell Biology* 22 (12): 834.

918 Yu, Min, Aditya Bardia, Ben S. Wittner, Shannon L. Stott, Malgorzata E. Smas, David T. Ting, Steven J. Isakoff, et

919 al. 2013. "Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal  
920 Composition." *Science* 339 (6119): 580–84.

921 Zavadil, Jiri, and Erwin P. Böttinger. 2005. "TGF-Beta and Epithelial-to-Mesenchymal Transitions." *Oncogene* 24  
922 (37): 5764–74.

923 Zhang, Yun, Joana Liu Donaher, Sunny Das, Xin Li, Ferenc Reinhardt, Jordan A. Krall, Arthur W. Lambert, et al.  
924 2022. "Genome-Wide CRISPR Screen Identifies PRC2 and KMT2D-COMPASS as Regulators of Distinct EMT  
925 Trajectories That Contribute Differentially to Metastasis." *Nature Cell Biology* 24 (4): 554–64.

926 Zhu, Xiaoming, Li Wei, Yangqiu Bai, Sen Wu, and Shuangyin Han. 2017. "FoxC1 Promotes Epithelial-  
927 Mesenchymal Transition through PBX1 Dependent Transactivation of ZEB2 in Esophageal Cancer."  
928 *American Journal of Cancer Research* 7 (8): 1642–53.

929 Aggarwal, Vaishali, Catalina Ardila Montoya, Vera S. Donnenberg, and Shilpa Sant. 2021. "Interplay between  
930 Tumor Microenvironment and Partial EMT as the Driver of Tumor Progression." *iScience* 24 (2): 102113.

931 Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert  
932 Hulselmans, Florian Rambow, et al. 2017. "SCENIC: Single-Cell Regulatory Network Inference and  
933 Clustering." *Nature Methods* 14 (11): 1083–86.

934 Aird, Katherine M., and Rugang Zhang. 2013. "Detection of Senescence-Associated Heterochromatin Foci  
935 (SAHF)." *Methods in Molecular Biology* 965: 185–96.

936 Ansieau, Stéphane, Jeremy Bastid, Agnès Doreau, Anne-Pierre Morel, Benjamin P. Bouchet, Clémence Thomas,  
937 Frédérique Fauvet, et al. 2008. "Induction of EMT by Twist Proteins as a Collateral Effect of Tumor-  
938 Promoting Inactivation of Premature Senescence." *Cancer Cell* 14 (1): 79–89.

939 Bach, Karsten, Sara Pensa, Marta Grzelak, James Hadfield, David J. Adams, John C. Marioni, and Walid T.  
940 Khaled. 2017. "Differentiation Dynamics of Mammary Epithelial Cells Revealed by Single-Cell RNA  
941 Sequencing." *Nature Communications* 8 (1): 2128.

942 Bach, Karsten, Sara Pensa, Marija Zarocsinceva, Katarzyna Kania, Julie Stockis, Silvain Pinaud, Kyren A. Lazarus,  
943 et al. 2021. "Time-Resolved Single-Cell Analysis of Brca1 Associated Mammary Tumorigenesis Reveals  
944 Aberrant Differentiation of Luminal Progenitors." *Nature Communications* 12 (1): 1502.

945 Bankhead, Peter, Maurice B. Loughrey, José A. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D.  
946 Dunne, Stephen McQuaid, et al. n.d. "QuPath: Open Source Software for Digital Pathology Image  
947 Analysis." <https://doi.org/10.1101/099796>.

948 Berger, Ashton C., Anil Korkut, Rupa S. Kanchi, Apurva M. Hegde, Walter Lenoir, Wenbin Liu, Yuexin Liu, et al.  
949 2018. "A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers." *Cancer Cell* 33  
950 (4): 690–705.e9.

951 Bianchini, Giampaolo, Justin M. Balko, Ingrid A. Mayer, Melinda E. Sanders, and Luca Gianni. 2016. "Triple-  
952 Negative Breast Cancer: Challenges and Opportunities of a Heterogeneous Disease." *Nature Reviews.*  
953 *Clinical Oncology* 13 (11): 674–90.

954 Bloushtain-Qimron, Noga, Jun Yao, Eric L. Snyder, Michail Shipitsin, Lauren L. Campbell, Sendurai A. Mani, Min  
955 Hu, et al. 2008. "Cell Type-Specific DNA Methylation Patterns in the Human Breast." *Proceedings of the*  
956 *National Academy of Sciences of the United States of America* 105 (37): 14076–81.

957 Buj, Raquel, Kelly E. Leon, Marlyn A. Anguelov, and Katherine M. Aird. 2021. "Suppression of p16 Alleviates the  
958 Senescence-Associated Secretory Phenotype." *Aging* 13 (3): 3290–3312.

959 Campisi, Judith, and Fabrizio d'Adda di Fagagna. 2007. "Cellular Senescence: When Bad Things Happen to Good  
960 Cells." *Nature Reviews. Molecular Cell Biology* 8 (9): 729–40.

961 Cancer Genome Atlas Network. 2012. "Comprehensive Molecular Portraits of Human Breast Tumours." *Nature*  
962 490 (7418): 61–70.

963 Cmarko, D., J. Smigova, L. Minichova, and A. Popov. 2008. "Nucleolus: The Ribosome Factory." *Histology and*  
964 *Histopathology* 23 (10): 1291–98.

965 Collado, Manuel, and Manuel Serrano. 2010. "Senescence in Tumours: Evidence from Mice and Humans." *Nature Reviews Cancer.* <https://doi.org/10.1038/nrc2772>.

966 Di Micco, Raffaella, Valery Krizhanovsky, Darren Baker, and Fabrizio d'Adda di Fagagna. 2021. "Cellular  
967 Senescence in Ageing: From Mechanisms to Therapeutic Opportunities." *Nature Reviews. Molecular Cell*  
968 *Biology* 22 (2): 75–95.

970 Engebraaten, Olav, Hans Kristian Moen Vollan, and Anne-Lise Børresen-Dale. 2013. "Triple-Negative Breast  
971 Cancer and the Need for New Therapeutic Targets." *The American Journal of Pathology.* <https://doi.org/10.1016/j.ajpath.2013.05.033>.

972 <https://doi.org/10.1016/j.ajpath.2013.05.033>.

973 Gao, Bowen, Ying Qu, Bingchen Han, Yoshiko Nagaoka, Makoto Katsumata, Nan Deng, Shikha Bose, Liting Jin,  
974 Armando E. Giuliano, and Xiaojiang Cui. 2017. "Inhibition of Lobuloalveolar Development by FOXC1

975 Overexpression in the Mouse Mammary Gland." *Scientific Reports* 7 (1): 14017.

976 Gao, Ruli, Alexander Davis, Thomas O. McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei-Ching Tsai, et al. 2016.

977 "Punctuated Copy Number Evolution and Clonal Stasis in Triple-Negative Breast Cancer." *Nature Genetics*

978 48 (10): 1119–30.

979 Gosselin, Kevin, Adeline Durand, Justine Marsolier, Adeline Poitou, Elisabetta Marangoni, Fariba Nemati,

980 Ahmed Dahmani, et al. 2019. "High-Throughput Single-Cell ChIP-Seq Identifies Heterogeneity of

981 Chromatin States in Breast Cancer." *Nature Genetics* 51 (6): 1060–66.

982 Haldipur, Parthiv, Gwendolyn S. Gillies, Olivia K. Janson, Victor V. Chizhikov, Divakar S. Mithal, Richard J. Miller,

983 and Kathleen J. Millen. 2014. "Foxc1 Dependent Mesenchymal Signalling Drives Embryonic Cerebellar

984 Growth." *eLife* 3 (December). <https://doi.org/10.7554/eLife.03962>.

985 Hanahan, Douglas, and Robert A. Weinberg. 2016. "The Hallmarks of Cancer." *Oxford Textbook of Oncology*.

986 <https://doi.org/10.1093/med/9780199656103.003.0001>.

987 Harada, Takamasa, Joe Swift, Jerome Irianto, Jae-Won Shin, Kyle R. Spinler, Avathamsa Athirasala, Rocky

988 Diegmiller, P. C. Dave P. Dingal, Irena L. Ivanovska, and Dennis E. Discher. 2014. "Nuclear Lamin Stiffness Is

989 a Barrier to 3D Migration, but Softness Can Limit Survival." *The Journal of Cell Biology* 204 (5): 669–82.

990 Herranz, Nicolás, and Jesús Gil. 2018. "Mechanisms and Functions of Cellular Senescence." *Journal of Clinical*

991 *Investigation*. <https://doi.org/10.1172/jci95148>.

992 Ito, Takahiro, Yee Voan Teo, Shane A. Evans, Nicola Neretti, and John M. Sedivy. 2018. "Regulation of Cellular

993 Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-

994 Dependent Pathways." *Cell Reports* 22 (13): 3480–92.

995 Jonkers, Jos, Ralph Meuwissen, Hanneke van der Gulden, Hans Peterse, Martin van der Valk, and Anton Berns.

996 2001. "Synergistic Tumor Suppressor Activity of BRCA2 and p53 in a Conditional Mouse Model for Breast

997 Cancer." *Nature Genetics*. <https://doi.org/10.1038/ng747>.

998 Keenan, Alexandra B., Denis Torre, Alexander Lachmann, Ariel K. Leong, Megan L. Wojciechowicz, Vivian Utti,

999 Kathleen M. Jagodnik, Eryk Kropiwnicki, Zichen Wang, and Avi Ma'ayan. 2019. "ChEA3: Transcription

1000 Factor Enrichment Analysis by Orthogonal Omics Integration." *Nucleic Acids Research* 47 (W1): W212–24.

1001 Koppelstaetter, Christian, Gabriele Schratzberger, Paul Perco, Johannes Hofer, Walter Mark, Robert Ollinger,

1002 Rainer Oberbauer, et al. 2008. "Markers of Cellular Senescence in Zero Hour Biopsies Predict Outcome in

1003 Renal Transplantation." *Aging Cell* 7 (4): 491–97.

1004 Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo.

1005 2015. "The Molecular Signatures Database Hallmark Gene Set Collection." *Cell Systems*.

1006 <https://doi.org/10.1016/j.cels.2015.12.004>.

1007 Lim, Elgene, François Vaillant, Di Wu, Natasha C. Forrest, Bhupinder Pal, Adam H. Hart, Marie-Liesse Asselin-

1008 Labat, et al. 2009. "Aberrant Luminal Progenitors as the Candidate Target Population for Basal Tumor

1009 Development in BRCA1 Mutation Carriers." *Nature Medicine* 15 (8): 907–13.

1010 Liu, Xiaoling, Henne Holstege, Hanneke van der Gulden, Marcelle Treur-Mulder, John Zevenhoven, Arno Velds,

1011 Ron M. Kerkhoven, et al. 2007. "Somatic Loss of BRCA1 and p53 in Mice Induces Mammary Tumors with

1012 Features of Human BRCA1-Mutated Basal-like Breast Cancer." *Proceedings of the National Academy of*

1013 *Sciences of the United States of America* 104 (29): 12111–16.

1014 Lv, Zhi-Dong, Hai-Bo Wang, Xiang-Ping Liu, Li-Ying Jin, Ruo-Wu Shen, Xin-Gang Wang, Bin Kong, Hui-Li Qu, Fu-

1015 Nian Li, and Qi-Feng Yang. 2017. "Silencing of Prrx2 Inhibits the Invasion and Metastasis of Breast Cancer

1016 Both In Vitro and In Vivo by Reversing Epithelial-Mesenchymal Transition." *Cellular Physiology and*

1017 *Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*

1018 42 (5): 1847–56.

1019 Marra, Antonio, Dario Trapani, Giulia Viale, Carmen Criscitiello, and Giuseppe Curigliano. 2020. "Practical

1020 Classification of Triple-Negative Breast Cancer: Intratumoral Heterogeneity, Mechanisms of Drug

1021 Resistance, and Novel Therapies." *NPJ Breast Cancer* 6 (October): 54.

1022 Marsolier, Justine, Pacôme Prompsy, Adeline Durand, Anne-Marie Lyne, Camille Landragin, Amandine

1023 Trouchet, Sabrina Tenreira Bento, et al. 2022. "H3K27me3 Conditions Chemotolerance in Triple-Negative

1024 Breast Cancer." *Nature Genetics* 54 (4): 459–68.

1025 Martins, Filipe C., Subhajyoti De, Vanessa Almendro, Mithat Gönen, So Yeon Park, Joanne L. Blum, William

1026 Herlihy, et al. 2012. "Evolutionary Pathways in BRCA1-Associated Breast Tumors." *Cancer Discovery* 2 (6):

1027 503–11.

1028 Milanovic, Maja, Dorothy N. Y. Fan, Dimitri Belenki, J. Henry M. Däbritz, Zhen Zhao, Yong Yu, Jan R. Dörr, et al.

1029 2018. "Senescence-Associated Reprogramming Promotes Cancer Stemness." *Nature* 553 (7686): 96–100.

1030 Molyneux, Gemma, Felipe C. Geyer, Fiona-Ann Magnay, Afshan McCarthy, Howard Kendrick, Rachael Natrajan,

1031 Alan MacKay, et al. 2010. "BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors

1032 and Not from Basal Stem Cells." *Cell Stem Cell*. <https://doi.org/10.1016/j.stem.2010.07.010>.

1033 Moon, Kevin R., David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim,

1034 et al. 2019. "Visualizing Structure and Transitions in High-Dimensional Biological Data." *Nature*

1035 *Biotechnology* 37 (12): 1482–92.

1036 Onitilo, Adedayo A., Jessica M. Engel, Robert T. Greenlee, and Bickol N. Mukesh. 2009. "Breast Cancer Subtypes

1037 Based on ER/PR and Her2 Expression: Comparison of Clinicopathologic Features and Survival." *Clinical*

1038 *Medicine & Research* 7 (1-2): 4–13.

1039 Pal, Bhupinder, Yunshun Chen, François Vaillant, Bianca D. Capaldo, Rachel Joyce, Xiaoyu Song, Vanessa L.

1040 Bryant, et al. 2021. "A Single-Cell RNA Expression Atlas of Normal, Preneoplastic and Tumorigenic States in

1041 the Human Breast." *The EMBO Journal* 40 (11): e107333.

1042 Paluvai, Hari Krishnareddy, Eros Di Giorgio, and Claudio Brancolini. 2020. "The Histone Code of Senescence."

1043 *Cells*. <https://doi.org/10.3390/cells9020466>.

1044 Patel, A. P., I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, et al. 2014. "Single-

1045 Cell RNA-Seq Highlights Intratumoral Heterogeneity in Primary Glioblastoma." *Science*.

1046 <https://doi.org/10.1126/science.1254257>.

1047 Pervolarakis, Nicholas, Quy H. Nguyen, Justice Williams, Yanwen Gong, Guadalupe Gutierrez, Peng Sun, Darisha

1048 Jhutti, et al. 2020. "Integrated Single-Cell Transcriptomics and Chromatin Accessibility Analysis Reveals

1049 Regulators of Mammary Epithelial Cell Identity." *Cell Reports* 33 (3): 108273.

1050 Polak, Paz, Jaegil Kim, Lior Z. Braunstein, Rosa Karlic, Nicholas J. Haradhavala, Grace Tiao, Daniel Rosebrock, et

1051 al. 2017. "A Mutational Signature Reveals Alterations Underlying Deficient Homologous Recombination

1052 Repair in Breast Cancer." *Nature Genetics* 49 (10): 1476–86.

1053 Probst, Aline V., and Geneviève Almouzni. 2008. "Pericentric Heterochromatin: Dynamic Organization during

1054 Early Development in Mammals." *Differentiation*. <https://doi.org/10.1111/j.1432-0436.2007.00220.x>.

1055 Ray, Tania, Terry Ryusaki, and Partha S. Ray. 2021. "Therapeutically Targeting Cancers That Overexpress FOXC1:

1056 A Transcriptional Driver of Cell Plasticity, Partial EMT, and Cancer Metastasis." *Frontiers in Oncology*.

1057 <https://doi.org/10.3389/fonc.2021.721959>.

1058 Remark, Romain, Taha Merghoub, Niels Grabe, Geert Litjens, Diane Damotte, Jedd D. Wolchok, Miriam Merad,

1059 and Sacha Gnjatich. 2016. "In-Depth Tissue Profiling Using Multiplexed Immunohistochemical Consecutive

1060 Staining on Single Slide." *Science Immunology*. <https://doi.org/10.1126/sciimmunol.aaf6925>.

1061 Roupakia, Eugenia, Georgios S. Markopoulos, and Evangelos Kolettas. 2021. "Genes and Pathways Involved in

1062 Senescence Bypass Identified by Functional Genetic Screens." *Mechanisms of Ageing and Development*

1063 194 (March): 111432.

1064 Samarasinghe, Kusal T. G., and Craig M. Crews. 2021. "Targeted Protein Degradation: A Promise for

1065 Undruggable Proteins." *Cell Chemical Biology* 28 (7): 934–51.

1066 Scully, R., and D. M. Livingston. 2000. "In Search of the Tumour-Suppressor Functions of BRCA1 and BRCA2."

1067 *Nature* 408 (6811): 429–32.

1068 Sedic, Maja, Adam Skibinski, Nelson Brown, Mercedes Gallardo, Peter Mulligan, Paula Martinez, Patricia J.

1069 Keller, et al. 2015. "Haploinsufficiency for BRCA1 Leads to Cell-Type-Specific Genomic Instability and

1070 Premature Senescence." *Nature Communications* 6 (June): 7505.

1071 Selbert, S., D. J. Bentley, D. W. Melton, D. Rannie, P. Lourenço, C. J. Watson, and A. R. Clarke. 1998. "Efficient

1072 BLG-Cre Mediated Gene Deletion in the Mammary Gland." *Transgenic Research* 7 (5): 387–96.

1073 Severino, V., N. Alessio, A. Farina, A. Sandomenico, M. Cipollaro, G. Peluso, U. Galderisi, and A. Chambery.

1074 2013. "Insulin-like Growth Factor Binding Proteins 4 and 7 Released by Senescent Cells Promote

1075 Premature Senescence in Mesenchymal Stem Cells." *Cell Death & Disease* 4 (November): e911.

1076 Stefansson, Olafur Andri, Jon Gunnlaugur Jonasson, Kristrun Olafsdottir, Holmfridur Hilmarsdottir, Gudridur

1077 Olafsdottir, Manel Esteller, Oskar Thor Johannsson, and Jorunn Erla Eyfjord. 2011. "CpG Island

1078 Hypermethylation of BRCA1 and Loss of pRb as Co-Occurring Events in Basal/triple-Negative Breast

1079 Cancer." *Epigenetics: Official Journal of the DNA Methylation Society* 6 (5): 638–49.

1080 Timms, Kirsten M., Victor Abkevich, Elisha Hughes, Chris Neff, Julia Reid, Brian Morris, Saritha Kalva, et al. 2014.

1081 "Association of BRCA1/2 defects with Genomic Scores Predictive of DNA Damage Repair Deficiency among

1082 Breast Cancer Subtypes." *Breast Cancer Research*. <https://doi.org/10.1186/s13058-014-0475-x>.

1083 Watson, Christine J., and Walid T. Khaled. 2008. "Mammary Development in the Embryo and Adult: A Journey

1084 of Morphogenesis and Commitment." *Development* 135 (6): 995–1003.

1085 Wolf, F. Alexander, Fiona K. Hamey, Mireya Plass, Jordi Solana, Joakim S. Dahlin, Berthold Göttgens, Nikolaus

1086 Rajewsky, Lukas Simon, and Fabian J. Theis. 2019. "PAGA: Graph Abstraction Reconciles Clustering with

1087 Trajectory Inference through a Topology Preserving Map of Single Cells." *Genome Biology* 20 (1): 59.

1088 Xu, Zhe, Chaochao Wu, Fang Xie, Gordon W. Slys, Nikola Tolic, Matthew E. Monroe, Vladislav A. Petyuk, et al.

1089 2015. "Comprehensive Quantitative Analysis of Ovarian and Breast Cancer Tumor Peptidomes." *Journal of*  
1090 *Proteome Research* 14 (1): 422–33.

1091 Yang, Jing, Parker Antin, Geert Berx, Cédric Blanpain, Thomas Brabletz, Marianne Bronner, Kyra Campbell, et al.  
1092 2021. "Author Correction: Guidelines and Definitions for Research on Epithelial-Mesenchymal Transition."  
1093 *Nature Reviews. Molecular Cell Biology* 22 (12): 834.

1094 Zavadil, Jiri, and Erwin P. Böttinger. 2005. "TGF-Beta and Epithelial-to-Mesenchymal Transitions." *Oncogene* 24  
1095 (37): 5764–74.

1096 Zhang, Yun, Joana Liu Donaher, Sunny Das, Xin Li, Ferenc Reinhardt, Jordan A. Krall, Arthur W. Lambert, et al.  
1097 2022. "Genome-Wide CRISPR Screen Identifies PRC2 and KMT2D-COMPASS as Regulators of Distinct EMT  
1098 Trajectories That Contribute Differentially to Metastasis." *Nature Cell Biology* 24 (4): 554–64.

1099 Zhu, Xiaoming, Li Wei, Yangqiu Bai, Sen Wu, and Shuangyin Han. 2017. "FoxC1 Promotes Epithelial-  
1100 Mesenchymal Transition through PBX1 Dependent Transactivation of ZEB2 in Esophageal Cancer."  
1101 *American Journal of Cancer Research* 7 (8): 1642–53.

1102