



HAL
open science

H3K27me3 conditions chemotolerance in triple-negative breast cancer

Justine Marsolier, Pacôme Prompsy, Adeline Durand, Anne-Marie Lyne, Camille Landragin, Amandine Trouchet, Sabrina Tenreira Bento, Almut Eisele, Sophie Foulon, Léa Baudre, et al.

► **To cite this version:**

Justine Marsolier, Pacôme Prompsy, Adeline Durand, Anne-Marie Lyne, Camille Landragin, et al.. H3K27me3 conditions chemotolerance in triple-negative breast cancer. *Nature Genetics*, 2022, 54, pp.459-468. 10.1038/s41588-022-01047-6 . hal-03851953

HAL Id: hal-03851953

<https://hal.science/hal-03851953>

Submitted on 21 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.



Distributed under a Creative Commons Attribution 4.0 International License

1 **H3K27me3 is a determinant of chemotolerance**
2 **in triple-negative breast cancer**

3
4 Justine Marsolier^{1,2,*}, Pacôme Prompsy^{1,2,*}, Adeline Durand^{1,2#}, Anne-Marie Lyne^{3#}, Camille
5 Landragin^{1,2}, Amandine Trouchet^{1,2}, Sabrina Tenreira Bento³, Almut Eisele³, Sophie Foulon⁴, Léa
6 Baudre^{1,2}, Kevin Grosselin^{4,5,6}, Mylène Bohec⁷, Sylvain Baulande⁷, Ahmed Dahmani², Laura Sourd², Eric
7 Letouzé⁸, Anne-Vincent Salomon^{9,10}, Elisabetta Marangoni², Leïla Perié^{3,‡}, Céline Vallot^{1,2,‡}

8
9 Affiliations:

10 1 CNRS UMR3244, Institut Curie, PSL University, Paris, France

11 2 Translational Research Department, Institut Curie, PSL University, Paris, France

12 3 CNRS UMR168, Institut Curie, PSL University, Sorbonne University, Paris, France

13 4 CNRS UMR8231, ESPCI Paris, PSL University, Paris, France

14 5 HiFiBio SAS, Paris, France

15 6 Current Affiliation: Broad Institute of MIT and Harvard, Cambridge MA, USA

16 7 Genomics of Excellence (ICGex) Platform, Institut Curie, PSL University, Paris, France

17 8 Functional Genomics of Solid Tumors laboratory, Centre de Recherche des Cordeliers, Sorbonne
18 University, Inserm, USPC, Paris Descartes University, Paris Diderot University, Paris, France

19 9 Pathology-Genetics-Immunology Department, Institut Curie, PSL Research University, Paris, France.

20 10 INSERM U934, Institut Curie, PSL Research University, Paris, France

21
22 * These authors contributed equally to this work

23 # These authors contributed equally to this work

24 ‡ co-last authors

25
26 Correspondence to celine.vallot@curie.fr

27 **Summary**

28

29 **Residual disease, formed by cancer cells persistent to therapy, remains one of the major clinical**
30 **challenges towards full cure^{1,2}. In triple-negative breast cancers, persistence to chemotherapy**
31 **results in the highest risk of recurrence among all breast cancer subtypes³. If there is now consensus**
32 **that the drug-tolerant state is defined by non-genetic features^{4,5}, understanding the mechanisms**
33 **driving tumor cells to drug-tolerance stands as a pre-requisite to efficiently fight recurrence. Here,**
34 **we show that the repressive histone mark H3K27me3 is a determinant of cell fate at the onset of**
35 **chemotherapy exposure, monitoring epigenomes, transcriptomes and lineages with single-cell**
36 **resolution. We show that the persister expression program is primed with both H3K4me3 and**
37 **H3K27me3 in unchallenged cells, H3K27me3 being the lock to its transcriptional activation. We**
38 **further establish that H3K27me3 controls cell fate upon chemotherapy exposure: alleviating the**
39 **genome from H3K27me3 enhances the potential of each cancer cell to tolerate chemotherapy insult.**
40 **Conversely, we demonstrate that preventing H3K27me3 demethylation simultaneously to**
41 **chemotherapy inhibits the transition to drug-tolerant state, and delays tumor recurrence *in vivo*.**
42 **Our results highlight how chromatin landscapes shape the potential of cancer cells to respond to**
43 **initial therapeutic insult.**

44

45 **Introduction**

46 Emergence of resistant phenotypes from initially responding or partially-responding tumors has been
47 modeled as a multi-step process in cancer^{6,7}. Initially, post drug insult, only a pool of persister cells -
48 also called drug-tolerant persister cells (DTPs) - manage to tolerate the cancer treatment and survive.
49 These cells constitute a reservoir of cells from which drug-resistant cells, actively growing under cancer
50 treatment, can ultimately emerge^{6,8,9}. In triple-negative breast cancer (TNBC), both genetic and
51 transcriptomic mechanisms have been proposed to drive cancer evolution towards chemoresistance,
52 through combined selective and adaptive modes of evolution¹⁰. The recent identification of a multi-
53 clonal reversible drug-tolerant state post neo-adjuvant chemotherapy in patient-derived xenografts
54 (PDX)⁵ suggested that the earliest steps of chemoresistance in TNBC are not driven by genetic
55 alterations, but rather by non-genetic plasticity in multiple cancer cells. Similarly, in other cancer types,
56 drug-tolerant states have been identified solely characterized by changes in transcriptomic and
57 epigenomic features in response to targeted therapies or chemotherapies^{4,11,12}.

58 Genetic history of many cancer types has been extensively modelled thanks to both bulk and single-
59 cell approaches^{10,13}. In contrast, little is known about the epigenomic heterogeneity and dynamics of
60 acquisition of epigenetic alterations. While recent studies have focused on the evolution of DNA
61 methylation^{14,15} - among the most stable epigenetic locks to gene expression - contribution to tumor
62 evolution of more versatile epigenetic modifications has remained poorly understood. Single-cell
63 methods to map repressive and permissive histone modifications, key players of cellular plasticity,
64 have emerged only recently^{16,17}, enabling the study of epigenomic diversity within biological systems.
65 Studying cases of acquired resistance in luminal and triple-negative cancer in PDX with single-cell
66 chromatin profiling, we had identified the first examples of epigenomic clones within tumors, solely
67 defined by H3K27me3 landscapes¹⁶. We had detected resistant-like H3K27me3 landscapes in initial
68 untreated tumors, proposing histone modifications as a molecular player of tumor evolution
69 processes. Here, we focused on the earliest steps of cancer treatment prior to the emergence of
70 resistant populations, to identify the initial epigenomic mechanisms driving drug tolerance. Thanks to
71 a characterization of epigenomic and transcriptomic evolution under chemotherapy at single-cell
72 resolution, we show that H3K27me3 epigenomes, and not H3K4me3, are a proxy of cancer cell
73 evolution upon initial chemotherapy insult. We discovered that the persister expression program is
74 actually in a bivalent chromatin configuration in untreated cells, and locked by H3K27me3. Tracing cell
75 lineages together with transcriptomics, we further reveal that H3K27me3 is a key determinant of cell
76 fate upon initial chemotherapy exposure in TNBC: depriving cells of H3K27me3 enhances the potential
77 of each tumor cell to tolerate chemotherapy. Finally, preventing cells under chemotherapy insult to
78 demethylate H3K27me3 residues inhibits their potential to tolerate chemotherapy, and delays tumor
79 recurrence *in vivo*.

80

81 **Deriving persister cells from patient tumor models.** We focused on mechanisms of drug tolerance in
82 residual TNBC, for which patients have the poorest outcome. Resistance to adjuvant chemotherapy
83 cannot be easily studied as biopsies are not routinely performed when the disease progresses (Fig. 1a).
84 To circumvent these limitations, we modeled, *in vivo* and *in vitro*, phenotypes of drug-response
85 observed in patients. *In vivo*, we first treated three patient-derived xenograft (PDX) models - PDX_95,
86 PDX_39 & PDX_172 - established from patients with residual TNBC^{16,18}, with Capecitabine, the standard
87 of care for residual breast tumors (Fig. 1b, Extended Figure 1). After the first round of chemotherapy
88 treatment, mice displayed a pathological complete response (pCR), but tumors eventually recurred
89 ('recurrent') and mice were treated again with chemotherapy, to which tumors responded to various

90 extents, some maintaining constant tumor volume under treatment ('resistant') (Fig. 1b). These
91 recurrent tumors potentially arose from persister cells, surviving initial chemotherapy treatment⁸. We
92 isolated persister cells by pooling the fat pad from mice with pCR (from 4 to 14, Extended Fig. 1a, 1f &
93 1l). To phenocopy a clinical situation of partial response, we also generated 'residual' tumors for one
94 model (PDX_95, n=2) (Extended Fig. 1a) by administering half the dose of Capecitabine.

95 *In vitro*, we treated an initially chemosensitive TNBC cell line (MDA-MB-468), with Fluorouracil (5-FU)¹⁹,
96 the active metabolite of Capecitabine, which is not converted *in vitro*. We drove independently three
97 pools of cells to chemotolerance and subsequently to chemoresistance with prolonged 5-FU treatment
98 (>15 weeks, Fig. 1e). After 3 weeks, only few cells survived drug insult (0.01% of the initial population,
99 Extended Fig. 2a), out of which a minority (36%) divided after 10-15 additional days (Extended Fig. 2a-
100 b); this group of cells, surviving initial chemotherapy, are referred to as 'persister' cells. Over 15 weeks,
101 populations of resistant cells emerged, with doubling times comparable to untreated cells and an IC50
102 to 5-FU over 4-fold higher than untreated population (Extended Fig. 2b).

103

104 **Identification of a pool of basal persister cells *in vivo* and *in vitro*.** To characterize transcriptomic
105 evolution of untreated cells towards chemotolerance and subsequently chemoresistance, we
106 performed single-cell RNA-seq (scRNA-seq) in both cell lines and PDX models (Fig. 1c, 1f, Extended Fig.
107 1 & 2). *In vivo*, scRNA-seq was mandatory to identify the rare human persister cells among the vast
108 majority of stromal mouse cells. Out of the fat pad, we typically isolated hundreds of persister cells
109 per mouse. *In vivo*, persister cells from different mice grouped within one or two expression clusters
110 (Extended Fig. 1b, 1g-h & 1m-n). *In vitro*, diverse cell populations (res #1, 2 and 3) with distinct
111 expression programs, originated from the pool of persister cells, all grouped within common clusters
112 across experiments (scRNA-based clusters R2/R4, Extended Fig. 2c). *In vivo* and *in vitro*, persister cells
113 recurrently activated a set of common pathways compared to untreated cells (Fig. 1c-f, Extended Fig.
114 1c-d, 1i-j, 1o-p Extended Fig. 2d-e). Originating from *KRT5*-expressing cancer cells, persister cells
115 recurrently activated sets of genes further establishing basal cell identity, such as *KRT14* (Fig. 1c, 1f,
116 Extended 1j & 1p). Compared to untreated cells, persister cells *in vivo* and *in vitro* also showed an
117 activation of genes associated with the Epithelial-to-Mesenchymal Transition (EMT, Fig.1c-f, Extended
118 Fig. 1c-d, 1j & 1p, Extended Fig. 2d & 2f) - such as *TAGLN*, an actin-binding protein, previously shown
119 to promote metastasis through EMT²⁰, and *NNMT*, characteristic of the metabolic changes that
120 accompany EMT²¹⁻²³. Persister cells also activated genes involved in the TNFalpha/NF-KB pathway. Part
121 of the persister expression program remained patient-specific: in PDX_95 and *in vitro* persister cells
122 showed for example an activation of the TGF- β pathway with the expression of multiple players
123 including ligands and receptors (Fig.1c & 1f, SI Tables 1-4).

124 *In vivo*, we showed that persister and residual tumor cells actually clustered together (Fig. 1c, cluster
125 R4), thereby sharing common expression features, suggesting similar mechanisms of chemotolerance
126 independent of the residual burden. Yet we detected a higher number of cells in G0/G1 within persister
127 populations than in residual or untreated tumors (Extended Fig. 1e-k-q), a characteristic recapitulated
128 *in vitro* (Extended Fig. 2g) and in line with previous reports^{1,4,11,24}. *In vitro*, we identified two clusters of
129 persister cells (clusters R2 and R4), that differ by their expression of additional EMT markers such as
130 *CDH2* (Fig. 1f) and *TWIST1*. Early individual non-cycling persisters (day 33) solely belonged to cluster
131 R2/CDH2- whereas dividing persisters could either belong to R2/CDH2- or R4/CDH2+ (Fig.1f). Overall,
132 we identified both *in vivo* and *in vitro* a reservoir of persister basal cells with EMT markers and
133 activated NF-KB pathway. NF-KB signaling pathway and EMT associated-genes were proposed as
134 potential drivers of chemoresistance in various tumor types, including lung^{25,26}, pancreatic²⁷,

135 breast^{28,29}. Here we pinpoint NF-KB and EMT pathway activation as the earliest common molecular
136 events at the onset of chemotolerance in TNBC, defining a common Achilles' heel to target
137 chemotolerant cells before they phenotypically diversify.

138

139 **Cell fate bias under chemotherapy.** To follow clonal evolution under therapeutic stress, we had
140 initially introduced unique genetic barcodes in untreated MDA-MB-468 cells prior to our experiments
141 (Extended Fig. 3a). We leveraged our previous barcoding method³⁰ to allow robust detection of
142 barcodes within scRNA-seq data (Fig. 1g), as shown by the number of cells with a detected lineage
143 barcode (Extended Fig. 3b). In addition, we verified that barcode frequencies detected in scRNA-seq
144 data recapitulated those detected in bulk, confirming the sensitivity of barcode detection in scRNA-
145 seq data (Extended Fig. 3c). By combining detection of lineage barcode and expression programs at
146 single-cell resolution, we were able to monitor clonal evolution over the course of the treatment and
147 within each expression clusters (Fig. 1g). If non-cycling persisters (day 33) were multi-clonal (62% of
148 unique barcodes), R6 and R8 chemoresistant clusters were constituted of few clones (3 and 4
149 respectively). According to bulk data (Extended Fig. 3d-e), barcode diversity in the cell population
150 decreased with time both under 5-FU and DMSO exposure. Across experiments and time points, the
151 fraction of unique barcodes within the CDH2-/R2 persister cluster was significantly higher than in
152 CDH2+/R4 cluster (average of 41% versus 8% unique barcode, $p=1.6e-2$, Fig. 1g), demonstrating that if
153 the drug-tolerant state is multi-clonal, only rare persister cells switch to the CDH2+ state.

154 To test if the lineages that persist were selected within the untreated population, we next compared
155 barcode frequencies between the starting population and the 5-FU or DMSO-treated cells using
156 additional bulk experiments (Experiment #3, Extended Fig. 3d-g). If surviving cells had no particular
157 predisposition then they should resemble a random draw of the initial untreated population (day 0,
158 Extended Fig. 3f). In contrast to DMSO-treated cells, barcode frequencies of the 5-FU treated cell
159 deviated from this random scenario (Extended Fig. 3g, $r=0.68$ and $r=0.29$ respectively), indicating that
160 a fraction of lineages present in the untreated population have a predisposition to tolerate
161 chemotherapy. Using our single-cell barcoding dataset, we were able to identify within the untreated
162 populations, future persisters ($n=143$, for which the lineage barcode is found both in persisters and
163 untreated cells) and compare them to 'non-persisting' cells ($n=201$, Extended Fig. 3 h-j). The only
164 transcriptomic differences between those two cell populations were the over-expression of *S100A2*, a
165 calcium binding protein, and *LDHB*, a lactate dehydrogenase, in future persister cells, prior to
166 treatment. These results suggest that a difference in the metabolism of cells could be an indicator of
167 their potential for persistence.

168

169 **Dynamics of genomic and epigenomic features.** To hamper the chemo-driven clonal evolution of
170 cancer cells, we next investigated the molecular basis of such rapid phenotypic evolution. Using whole-
171 exome sequencing (Extended Fig. 4a), we first analyzed mutations, copy-number alterations (CNA) and
172 related mutational signatures acquired by persister and resistant cell populations since the onset of 5-
173 FU treatment. We could not identify any recurrent mutations across experiments (Extended Fig. 4b),
174 or any CNA (amplifications or homozygous deletions) or recurrent mutations affecting known driver
175 genes of breast cancer in any population¹³. Only a minor fraction of mutations found in persister cells
176 were attributed to 5-FU exposure (mutational signature 17³¹) in contrast to resistant cell populations
177 where over 50% of acquired mutations are associated to 5-FU ($p<10^{-10}$, Extended Fig.4c). These results
178 indicated that chemo-related mutations are acquired over a timeframe that is not compatible with the
179 rapid phenotypic evolution seen in persister cells. Finally, computing cancer cell fractions for each

180 mutation, we confirmed that persister populations are extensively multi-clonal (Extended Fig. 4d), in
181 line with the lineage barcoding results.

182 We next investigated changes in epigenomes during chemotherapy treatment. Using single-cell
183 profiling (scChIP-seq), we observed that H3K27me3 epigenomes faithfully captured the evolution of
184 cell states with chemotherapy (Fig. 2a, Extended Fig. 5a and SI Table 4). Persister cells shared a
185 common H3K27me3 epigenome (cluster E1, Fig. 2b, Extended Fig. 5b), in contrast to resistant cells split
186 in clusters E1 and E3. In comparison to untreated cells, cells from cluster E1 showed recurrent
187 redistribution of H3K27me3 methylation, the highest changes ($|\log_2FC| > 2$ and adjusted $p\text{-value} < 10^{-1}$)
188 occurring specifically at transcription start sites (TSS) and gene bodies (Fig. 2c) and corresponding to a
189 loss of H3K27me3 enrichment (75 regions with $\log_2FC < -2$, and 2 regions with $\log_2FC > 2$). This depletion
190 was associated with the highest changes in gene expression observed by scRNA-seq (Fig. 2d and
191 Extended Fig. 5c) and to the transcriptional de-repression of 22% of *persister* genes – defined as genes
192 overexpressed in persister versus untreated cells – such as *TGFB1*, *FOXQ1*, a transcription factor driving
193 EMT (Fig. 2e, SI Table 4). These epigenomic changes were not necessarily kept in resistant cell
194 populations (Extended Fig. 5d), suggesting the existence of transient epigenomic features in persister
195 cells.

196 In the untreated cells, two epigenomic subclones (E2 & E4 clusters, Fig. 2b) were recurrently identified
197 indicative of an epigenomic heterogeneity in this population (n=3 experiments). In contrast to cells
198 from cluster E4 (median correlation score $r = -0.34$), a fraction of cells within cluster E2 shared
199 epigenomic similarities with cells from E1 (Fig. 2f, 49/381 cells over $r = 0.20$, $p < 2.2e-16$). Yet these
200 similarities did not affect persister genes (Extended Fig. 5e), and cells remained discernible from the
201 pool of persister cells (no cells from E2 with r score over median r in E1, $p < 2.2e-16$). This suggests that
202 cells from E2 could fuel the persister population when exposed to chemotherapy, with the need of
203 chemo-induced chromatin changes to achieve tolerance. In addition, we also detected rare cells with
204 a persister epigenomic signature, but in only one of our three experiments (60/976 cells – Extended
205 Fig. 5b), suggesting that spontaneous transition to H3K27me3 drug-tolerant state rarely occurs in the
206 absence of chemotherapy.

207
208 **Persister expression program is locked by H3K27me3.** To test whether H3K27me3 enrichment was
209 the lock to the persister expression program in untreated cells, we treated cancer cells with the EZH2
210 inhibitor (EZH2i-1) UNC1999³², to deplete H3K27me3 from cells, in the absence of chemotherapy.
211 EZH2i-1 treatment phenocopied drug-tolerant state to chemotherapy as expression fold-changes
212 induced by EZH2i-1 were specifically correlated to those induced by chemotherapy exposure at early
213 time points in persister cells (Fig. 2g, Extended Fig. 5f, $r = 0.71$ versus $r = 0.31$ with changes in resistant
214 cells). Furthermore, we observed that EZH2i-1 was sufficient to lead to the activation of 62% of
215 *persister* genes with depletion of H3K27me3 upon 5-FU treatment (23/37 genes), suggesting that
216 H3K27me3 was the sole lock to their activation (Fig. 2h and Extended Fig. 5g). EZH2i-1 was also
217 sufficient to lead to the over-expression of 60% of *persister* genes independently of any H3K27me3
218 enrichment in untreated cells (78/131 genes), such as *KRT14*, suggesting that these genes might be
219 targets of H3K27me3-regulated *persister* genes. To further test this hypothesis, we explored the
220 existence of potential ‘master’ transcription factors (TFs) within the *persister* genes. Using CheA3³³, we
221 identified three transcription factors, *FOXQ1*, *FOSL1* and *N2RF2*, that respectively target 34%, 42% and
222 29% of the *persister* genes (Extended Fig. 5h-i). All 3 are H3K27me3-regulated *persister* genes, with a
223 loss of H3K27me3 upon 5-FU (Extended Fig. 5j) and re-expressed by EZH2i-1 treatment, proposing
224 H3K27me3-regulated TF as potential drivers of the persister expression program.

225
226 **Priming of the persister expression program.** As we observed H3K27me3 changes upon 5-FU
227 treatment precisely at TSS, we further explored the evolution of chromatin modifications at TSS,
228 focusing on H3K4me3, a permissive histone mark shown to accumulate over TSS with active
229 transcription. In contrast to single-cell H3K27me3 epigenomes which were sufficient to separate cell
230 states along treatment (Fig 2a), individual H3K4me3 epigenomes of untreated and persister cells were
231 indiscernible (Fig. 3a-b). Sparse H3K4me3 enrichment was already observed in untreated cells at the
232 TSS of *persister* genes ($p < 10^{-15}$, compared to a set of non-expressed genes, Extended Fig. 6a-b). In
233 individual persister cells, H3K4me3 enrichment was significantly more synchronous at the TSS of
234 *persister* genes with more persister genes simultaneously marked by H3K4me3 in the same cell than
235 in untreated cells ($p = 4.0 \times 10^{-2}$, Extended Fig. 6c-d). Based on these results, we reasoned that in untreated
236 cells H3K4me3 and H3K27me3 could either accumulate on the same TSS but in different cells, or
237 H3K4me3 could accumulate together with H3K27me3 in a subset of cells on the same TSS.

238 To test whether H3K4me3 could co-exist with H3K27me3 in the same individual cells prior to
239 chemotherapy exposure, we performed successive immunoprecipitation of H3K27me3 with H3K4me3
240 (or vice-versa) or H3K27me3 (or H3K4me3) with isotype control (IgG) on mono-nucleosome chromatin.
241 In MDA-MB-468 cells, we detected $n = 1,547$ transcription start sites significantly enriched in DNA
242 immunoprecipitated with both H3K27me3 and H3K4me3, compared to the control (H3K27me3/IgG)
243 precipitated fraction (peak-ratio > 0.15 , q-value $< 10^{-3}$, Extended Fig. 6e-h). We found that bivalent
244 chromatin in untreated cells was detected at TSS of genes associated to mammary stem cell and EMT
245 pathways, as well as various developmental pathways (e.g Hedgehog pathway, Extended Fig. 6h-i). The
246 majority of H3K27me3-regulated *persister* genes (26 out of 37), including the 3 candidate master TFs,
247 were found in a bivalent chromatin configuration in the untreated cell population (e.g *TGF- β 1*, *FOXQ1*,
248 *FOSL1*, Fig. 3b-c, Extended Fig. 6d and SI Table 4). We next studied bivalent chromatin landscapes in
249 additional TNBC cell lines (HCC38 & BT20, Extended Fig. 6i), in our 3 PDX models and $n = 9$ patient
250 samples - 3 of which correspond to the tumors used for PDX derivation (Patient_95, Patient_39 &
251 Patient_172) (Fig. 3d-f, Extended Fig. 7h). We confirm the existence of bivalent programs in all patient
252 tumors, and show that paired PDX models faithfully recapitulate bivalency of patient samples
253 (Extended Figure 7a-f). Bivalent chromatin is also found at genes implicated in EMT and mammary
254 stem cell identity, and developmental pathways (Fig. 3f, Extended Fig. 7h) and at candidate master TFs
255 of the persister programs - *FOXQ1*, *KLF4* or *TFCP2L1* - (Fig. 3d-e, Extended Fig. 7g). With a continuum
256 of epigenomic datasets, from patients to PDX, we demonstrate that cancer cells display a set of
257 bivalent TSS, which could lead to the rapid activation of an entire persister expression program upon
258 therapeutic stress.

259
260 **Depriving cells of H3K27me3 enhances chemotolerance.** To further test if H3K27me3 was a lock to
261 the emergence of persister cells under chemotherapy exposure, we next depleted H3K27me3 from
262 epigenomes prior to chemotherapy treatment in three TNBC cell lines (MDA-MB-468, BT20 and HCC38,
263 Extended Fig. 8). Using EZH2i-1, in addition to an inactive isomer (UNC2400³⁴) and a second EZH2i
264 (GSK126³⁵ referred to as EZH2i-2), we showed that erasing H3K27me3 - without perturbing EZH2
265 protein levels - increased the number of persister cells with both EZH2 inhibitors, while the inactive
266 isomer had no effect (Fig. 4a, Extended Fig. 8). By comparing bulk lineage barcode frequencies across
267 conditions, we further demonstrated that EZH2 inhibitors were actually affecting cell fate under
268 chemotherapy exposure, by rescuing the biased lineage frequency induced by 5-FU treatment (Fig. 4b,
269 Extended Fig. 8c). We observed that co-treatment with EZH2i-1 and EZH2i-2 increased correlation

270 scores between lineage frequencies in 5-FU and DMSO treated population, whereas co-treatment with
271 the inactive isomer had no effect on lineage frequencies (Fig. 4b, Extended Fig. 8c). Performing
272 combined single-cell transcriptomics and lineage tracing as in Figure 1, we then compared the fraction
273 of unique barcodes across expression clusters under 5-FU exposure with or without EZH2i-1, termed
274 “EZH2i-1 persister” and “persister” respectively (Fig. 4c-e). We observed that co-treating cells with
275 EZH2i-1 significantly increased the fraction of unique lineage barcodes in CDH2+ persister populations,
276 identified in Figure 1 (R3 cluster, $p < 10^{-4}$ and over 2 fold-increase, Fig. 4e). Altogether, our results
277 showed that H3K27me3 depletion with EZH2 inhibitors rescued the biased lineage frequency observed
278 under chemotherapy treatment, and enabled a wider variety of cells to switch to the CDH2+ drug-
279 tolerant state. Overall, depleting H3K27me3 from untreated cells not only launched a persister-like
280 expression program, but it also enhanced the potential of each cancer cell to tolerate chemotherapy.

281

282 **Preventing H3K27me3 demethylation inhibits chemotolerance.** Finally, we tested our ability to inhibit
283 the emergence of persister cells by preventing the depletion of H3K27me3 under chemotherapy
284 exposure using a KDM6A/B – “Lysine demethylase 6A/B” inhibitor (KDM6A/Bi - GSK-J4³⁶)
285 simultaneously to chemotherapy. We tested *in vitro* and *in vivo* the ability of GSK-J4 to reduce the pool
286 of persister cells upon chemotherapy exposure, reasoning that a reduced number of persister cells
287 would increase the delay to recurrence. *In vitro*, in opposition to EZH2i, co-treatment with KDM6A/Bi
288 led to a decrease in the number of persisters at day 21 and further completely abolished the growth
289 of persister cells under 5-FU at day 60, whereas it had no effect on untreated cancer cells (Fig. 5a,
290 Extended Fig. 9a-c). Interestingly, administrating KDM6i once persister cells have emerged, rather than
291 in combination to chemotherapy, is inefficient (Fig. 5b, Extended Fig. 9d), demonstrating that the
292 switch from untreated to drug-tolerant state, rather than the drug-tolerant state itself, was sensitive
293 to KDM6i. These results were confirmed in two additional TNBC cell lines, BT20 and HCC38 (Extended
294 Fig. 9e-f). *In vivo*, our objective was to test the potential of KDM6i to limit the emergence of persister
295 cells, when administered simultaneously to chemotherapy. We compared the disease-free survival
296 time of mice treated with Capecitabine alone (n=25) or in combination with KDM6i (n=25). The
297 administration of KDM6i did not inhibit tumor progression in absence of chemotherapy nor was toxic
298 for the mice (Fig. 5c). However administered in combination to chemotherapy, it significantly increased
299 the delay to recurrence (Fig. 5d, $p = 4.0 \times 10^{-2}$) in comparison to chemotherapy alone. Our *in vitro* and *in*
300 *vivo* results together suggest that a fraction of cancer cells could need to actively demethylate
301 H3K27me3 residues to tolerate chemotherapy. These results are consistent with a mechanism where
302 persister genes would be repressed by H3K27me3 in untreated cells, and primed with stochastic
303 H3K4me3 in a subset of cells, with the loss of H3K27me3 unlocking the transition to tolerance.

304

305 Discussion

306 In conclusion, our study shows that H3K27me3 landscapes are determinants of cell fate upon
307 chemotherapy exposure in TNBC. We demonstrate that, prior to treatment, cells display bivalent
308 chromatin landscapes priming the persister expression program with H3K4me3 and H3K27me3. In
309 other words, genes are ready to be activated with H3K4me3, but are repressed with H3K27me3 that
310 remains the lock to the activation of the persister expression program. Using EZH2 inhibitors and
311 lineage tracing strategies, we further demonstrate that, depleting H3K27me3 from the genome
312 rescues the cell fate bias normally observed upon chemotherapy insult; cells have an equal probability
313 of surviving initial chemotherapy insult. Persister cells could be cells without H3K27me3 or the one
314 releasing the H3K27me3 lock, or a mixture of both phenomena as shown here: co-treating cells with a

315 H3K27me3 demethylase inhibitor together with 5-FU, we reduced, but not totally abrogated the
316 number of persister cells. We propose that combining chemotherapy with histone demethylase
317 inhibitors at the onset of chemotherapy exposure will delay recurrence by decreasing the pool of
318 persister cells. Several studies had started to interrogate which epigenetic modifiers could regulate
319 expression programs of persister or resistant cells^{11,35,37,38}. Here, we show that even prior to treatment,
320 the epigenome is already a key player, with a priming of the persister program. Our findings highlight
321 how chromatin landscapes can shape the potential of cancer cells for chemotolerance.
322 Our results that depletion of H3K27me3 with EZH2 inhibitors - at the onset of treatment - enhances
323 chemotolerance leaves open questions regarding the role of H3K27me3 landscapes in cancer
324 evolution, and the usage of EZH2i in cancer treatment. Depending on the context and the timing of
325 administration, EZH2i might have drastically different effects, and such subtleties should be carefully
326 considered before therapy combination. In contrast to our observations at the onset of treatment,
327 TNBC cancer cells with long-term acquired chemo-resistance have been shown to display DNA
328 hypomethylation and large H3K27me3 over transposable elements, hence a vulnerability to EZH2i³⁹.
329 EZH2i were also recently shown to lead to MHC Class I upregulation in cancer cells, thereby showing
330 beneficial immunotherapeutic effects^{40,41}. If such cell plasticity represents a therapeutic opportunity,
331 our results also show that EZH2i could also lead, in some contexts, to the activation of a set of genes
332 driving drug-persistence.
333 Isolated examples of bivalent promoters had been found in tumor cells^{42,43}, here we exhaustively map
334 bivalent promoters genome-wide, revealing epigenomic priming of mammary stem cell genes and
335 signaling pathways of known resistance pathways in TNBC⁴⁴, including Hedgehog, WNT, TGF- β , ATP-
336 binding cassette drug transporters pathways. Such epigenomic priming is reminiscent of
337 developmental bivalency priming mechanisms⁴⁵ found in stem cells prior to differentiation and
338 appears key for the rapid activation of the genes upon therapeutic stress. Remains to be understood,
339 how only a minority of bivalent genes are targeted by gene reactivation upon chemotherapy exposure
340 - which could be associated to the nature of the treatment itself - and whether such priming
341 mechanisms could be shared across cancer types. Determining the precise addressing mechanisms
342 that target H3K27me3 and H3K4me3 writers and readers to TSS of cancer-type bivalent genes under
343 therapeutic stress, will be instrumental in the future to identify dedicated co-factors which could serve
344 as therapeutic targets to further limit the emergence of persister cells.
345
346

347 **References**

348 1. Vallette, F. M. *et al.* Dormant, quiescent, tolerant and persister cells: Four synonyms for the same target in cancer. *Biochemical*
349 *Pharmacology* **162**, 169–176 (2019).
350 2. Shen, S., Vagner, S. & Robert, C. Persistent Cancer Cells: The Deadly Survivors. *Cell* **183**, 860–874 (2020).
351 3. Li, X. *et al.* Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast
352 cancer. *Breast Cancer Res. Treat.* **161**, 279–287 (2017).
353 4. Sharma, S. V. *et al.* A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell* **141**, 69–80 (2010).
354 5. Echeverria, G. V. *et al.* Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-
355 tolerant state. *Sci. Transl. Med.* **11**, eaav0936 (2019).
356 6. Ramirez, M. *et al.* Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun* **7**,
357 10690 (2016).
358 7. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–
359 435 (2017).
360 8. Cortazar, P. *et al.* Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis.
361 *The Lancet* **384**, 164–172 (2014).
362 9. Hata, A. N. *et al.* Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor
363 inhibition. *Nat Med* **22**, 262–269 (2016).
364 10. Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-
365 893.e13 (2018).
366 11. Liao, B. B. *et al.* Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell* **20**,
367 233–246.e7 (2017).
368 12. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell* **174**, 843–855.e19 (2018).
369 13. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
370 14. Mazor, T. *et al.* DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in
371 Brain Tumors. *Cancer Cell* **28**, 307–317 (2015).
372 15. Gaiti, F. *et al.* Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569**, 576–580 (2019).
373 16. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.*
374 **51**, 1060–1066 (2019).
375 17. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).
376 18. Marangoni, E. *et al.* A New Model of Patient Tumor-Derived Breast Cancer Xenografts for Preclinical Assays. *Clinical Cancer*
377 *Research* **13**, 3989–3998 (2007).
378 19. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3**, 330–
379 338 (2003).
380 20. Chen, Z. *et al.* TGF- β -induced transgelin promotes bladder cancer metastasis by regulating epithelial-mesenchymal transition and
381 invadopodia formation. *EBioMedicine* **47**, 208–220 (2019).
382 21. Ulanovskaya, O. A., Zuhl, A. M. & Cravatt, B. F. NNMT promotes epigenetic remodeling in cancer by creating a metabolic
383 methylation sink. *Nat. Chem. Biol.* **9**, 300–306 (2013).
384 22. Shaul, Y. D. *et al.* Dihydropyrimidine Accumulation Is Required for the Epithelial-Mesenchymal Transition. *Cell* **158**, 1094–1109
385 (2014).
386 23. Liang, L., Zeng, M., Pan, H., Liu, H. & He, Y. Nicotinamide N-methyltransferase promotes epithelial-mesenchymal transition in
387 gastric cancer cells by activating transforming growth factor- β 1 expression. *Oncol Lett* (2018) doi:10.3892/ol.2018.7885.
388 24. Oren, Y. *et al.* *Cycling cancer persister cells arise from lineages with distinct transcriptional and metabolic programs.*
389 <http://biorxiv.org/lookup/doi/10.1101/2020.06.05.136358> (2020) doi:10.1101/2020.06.05.136358.
390 25. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance.
391 *Nature* **527**, 472–476 (2015).
392 26. Ryan, S.-L. *et al.* Targeting NF- κ B-mediated inflammatory pathways in cisplatin-resistant NSCLC. *Lung Cancer* **135**, 217–227 (2019).
393 27. Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic
394 cancer. *Nature* **527**, 525–530 (2015).
395 28. Zeng, D. *et al.* Inhibition of Notch1 reverses EMT and chemoresistance to cisplatin *via* direct downregulation of MCAM in triple-
396 negative breast cancer cells. *Int. J. Cancer* **147**, 490–504 (2020).
397 29. Godwin, P. *et al.* Targeting Nuclear Factor-Kappa B to Overcome Resistance to Chemotherapy. *Front. Oncol.* **3**, (2013).
398 30. Eisele, A. S. *et al.* *Erythropoietin directly affects single hematopoietic stem cell differentiation after transplantation.*
399 <http://biorxiv.org/lookup/doi/10.1101/2020.04.20.050146> (2020) doi:10.1101/2020.04.20.050146.
400 31. Christensen, S. *et al.* 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat Commun* **10**, 4571
401 (2019).
402 32. Konze, K. D. *et al.* An Orally Bioavailable Chemical Probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* **8**,
403 1324–1334 (2013).
404 33. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research* **47**,
405 W212–W224 (2019).
406 34. Chagraoui, H. *et al.* SCL/TAL1 cooperates with Polycomb RYBP-PRC1 to suppress alternative lineages in blood-fated cells. *Nat*
407 *Commun* **9**, 5375 (2018).

408 35. Göllner, S. *et al.* Loss of the histone methyltransferase EZH2 induces resistance to multiple drugs in acute myeloid leukemia. *Nat*
409 *Med* **23**, 69–78 (2017).
410 36. Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response.
411 *Nature* **488**, 404–408 (2012).
412 37. Vinogradova, M. *et al.* An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat Chem Biol* **12**, 531–
413 538 (2016).
414 38. Hinohara, K. *et al.* KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance.
415 *Cancer Cell* **34**, 939-953.e9 (2018).
416 39. Deblois, G. *et al.* Epigenetic Switch–Induced Viral Mimicry Evasion in Chemotherapy-Resistant Breast Cancer. *Cancer Discov* **10**,
417 1312–1329 (2020).
418 40. Burr, M. L. *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and
419 Enables Immune Evasion in Cancer. *Cancer Cell* **36**, 385-401.e8 (2019).
420 41. Mahadevan, N. R. *et al.* Intrinsic immunogenicity of small cell lung carcinoma revealed by its cellular plasticity. *Cancer Discov*
421 candisc.0913.2020 (2021) doi:10.1158/2159-8290.CD-20-0913.
422 42. Hahn, M. A. *et al.* Loss of the Polycomb Mark from Bivalent Promoters Leads to Activation of Cancer-Promoting Genes in Colorectal
423 Tumors. *Cancer Research* **74**, 3617–3629 (2014).
424 43. Chaffer, C. L. *et al.* Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. *Cell*
425 **154**, 61–74 (2013).
426 44. Nedeljković, M. & Damjanović, A. Mechanisms of Chemotherapy Resistance in Triple-Negative Breast Cancer—How We Can Rise
427 to the Challenge. *Cells* **8**, 957 (2019).
428 45. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326
429 (2006).
430
431
432

433 **Methods**

434 **Experimental approaches:**

435 **Patient samples and PDX models.** Patient samples used in this study (n=9) originated from patient
436 treated at Institut Curie with residual triple-negative breast cancers post-neoadjuvant chemotherapy,
437 who gave informed consent for the profiling. In this study, we used three xenograft models generated
438 from three different residual triple-negative breast cancers post-neoadjuvant chemotherapy (HBCx95
439 called PDX_95, HBCx39 called PDX_39 and HBCx172 called PDX_172 in the manuscript, see Extended
440 Table2) established previously at Curie Institute with informed consent from the patient^{46,47}. Female
441 Swiss nude mice were purchased from Charles River Laboratories and maintained under specific-
442 pathogen-free conditions. Mouse care and housing were in accordance with institutional guidelines
443 and the rules of the French Ethics Committee (project authorization no. 02163.02).

444 Fig. 1b: Five mice were not treated and kept as controls (termed “*untreated*”) and twenty-seven mice
445 were treated orally with Capecitabine (Xeloda; Roche Laboratories) at a dose of 540 mg/kg, 5 d/week
446 for 6 to 14 weeks. Relative tumor volumes (mm³) were measured as described previously¹⁸. Eight mice
447 were sacrificed after the first round of chemotherapy to study “*residual*” tumors (2 mice) or “*persisters*”
448 (6 mice) human tumor cells. Seven mice with “*recurrent*” tumors (tumor volume between 200 and 600
449 mm³) were treated with a second round of Capecitabine to which they responded or not. “*Resistant*”
450 refers to a tumor which maintains a constant volume under this second round of treatment.

451 Extended Fig. 1f: Three mice were not treated and kept as controls (termed “*untreated*”) and fourteen
452 mice were treated orally with Capecitabine at a dose of 540 mg/kg, 5 d/week for 7 weeks and sacrificed
453 to study “*persisters*” human tumor cells.

454 Extended Fig. 1l: Six mice were not treated and kept as controls (termed “*untreated*”) and four mice
455 were treated orally with Capecitabine for 7 weeks and sacrificed to study “*persisters*” human tumor
456 cells.

457 Fig. 5c: Five mice were treated intraperitoneally with DMSO, five mice were treated intraperitoneally
458 with GSK-J4 alone at a dose of 50mg/kg, 5d/week for 25 days. Twenty-five mice were treated orally
459 with Capecitabine at a dose of 540 mg/kg, 5 d/week for 36 days and twenty-five mice were co-treated
460 with Capecitabine and GSK-J4 for 36 days. Tumor volumes (mm³) were measured to follow recurrence.

461 Fig. 5d: Disease-free survival was defined as the number of days between the observation of a
462 complete response (relative tumor volume *RTV* compared to volume at onset of treatment < 0.2) after
463 the first round of Capecitabine treatment, and the appearance of a recurrent tumor (*RTV* > 3).

464 Statistical analysis was performed using a log-rank test.

465 Before downstream analysis (scChIP-seq, scRNA-seq or sequential ChIP-seq), control and treated
466 tumors were digested for 2h at 37°C with a cocktail of Collagenase I (Roche, Ref: 11088793001) and
467 Hyaluronidase (Sigma-Aldrich, Ref: H3506). Cells were then individualized at 37°C using a cocktail of
468 0.25% Trypsin-Versen (Thermo Fisher Scientific, Ref: 15040-033), Dispase II (Sigma-Aldrich, Ref: D4693)
469 and DNase I (Roche, Ref: 11284932001) as described previously⁴⁸. Then, eBioscience red blood cell
470 lysis buffer (Thermo Fisher Scientific, Ref: 00-4333-57) was added to the cell suspension to remove red
471 blood cells. To increase the viability of the final cell suspension, dead cells were removed using the
472 Dead Cell Removal Kit (Miltenyi Biotec, Ref:130-090-101).

473

474 **Cell lines, culture conditions and drug treatments.** MDA-MB-468 cells were cultured in DMEM (Gibco-
475 BRL, Ref: 11966025), supplemented with 10% heat-inactivated fetal calf serum (Gibco-BRL, Ref: 10270-
476 106). HCC38 and BT20 cell lines were cultured in RPMI 1640 (Gibco-BRL, Ref: 11875085), supplemented
477 with 10% heat-inactivated fetal calf serum. All cell lines were cultured in a humidified 5% CO₂

478 atmosphere at 37 °C, and were tested as mycoplasma negative. GSKJ4 (KDM6A/B inhibitor, Sigma, Ref:
479 SML0701), GSKJ5 (GSK-J4 inactive isomer, Abcam, Ref: ab144397), UNC1999 (EZH2 inhibitor, Abcam,
480 Ref: ab146152), UNC2400 (UNC1999 inactive isoform, Tocris, Ref: 4905) and GSK126 (EZH2 inhibitor,
481 Sigma, Ref:) were used at indicated concentrations. Cells were treated with 5 μM of 5-FU (Sigma, Ref:
482 F6627) alone or in combination with KDM6A/Bi or EZH2i for indicated days. For EZH2i, cells were
483 pretreated with UNC1999, UNC2400 or GSK126 for 10 days before the addition of 5-FU for an
484 additional 21 days (Figure 4 and Extended Figure 8).

485

486 **Colony forming assay.** TBNC cells were plated in 6 multi-well plates at a density of 200,000 cells per
487 well and treated with the indicated drugs for 60 days (MDA-MB-468, Fig. 5a/b and Extended Fig. 9b)
488 or 56 days (BT20) or 50 days (HCC38) (Extended Fig. 9). Cultures were incubated in humidified 37 °C
489 incubators with an atmosphere of 5% CO₂ in air, and treated plates were monitored for growth using
490 a microscope. At the time of maximum foci formation, colony formation was evaluated after a staining
491 with 0.5% Crystal Violet (Sigma, ref: C3886).

492

493 **Cell proliferation, doubling time and IC₅₀.** MDA-MB-468, HCC38 and BT20 cells were stained with
494 Trypan Blue (Invitrogen, Ref: T10282) exclusion test, and counted using a Countess automated cell
495 counter (Invitrogen, Ref: C10228) at indicated time of treatment (Fig. 4a and Extended Fig.8a/d/f).

496 Doubling time (Extended Fig.2b) was calculated using this formula:

497
$$\text{“DoublingTime} = \text{duration} * \log(2) / (\log(\text{Final Concentration}) - \log(\text{Initial Concentration}))\text{”}$$

498 For untreated condition and resistant condition, cell numbers were evaluated on cell population during
499 10 days (n=3). For persister condition, cells were counted manually under the microscope at day 13
500 and day 30 of treatment. Doubling time of 5-FU dividing persister cells was studied from single cell to
501 confluent colony by assaying cell number every 4 days during 27 days (n= 9 single cells).

502 MDA-MB-468 untreated and chemoresistant cells were plated in 96 multi-well plates at a density of
503 10,000 cells per well and treated with increased concentration of 5-FU (1μM to 0.5M) for 72h. Cell
504 cytotoxicity was assayed with XTT kit (Sigma, Ref: 11465015001) and IC₅₀ was calculated as the
505 concentration of 5-FU that is required to obtain 50% of cell viability (Extended Fig. 2b).

506 Cells were classified as ‘persister’, ‘growing persister’ or ‘resistant’ based on a combination of 2
507 biological markers : (i) doubling time under 5-FU, (ii) IC₅₀ to 5-FU (Extended Figure 2b, Fig. 1e):

- 508 - ‘persister’ correspond to non-dividing cells (infinite doubling time) or cells dividing with a
509 doubling time significantly higher than resistant cells under 5-FU
- 510 - ‘resistant’ correspond to cells with a doubling time comparable to untreated cells and a
511 significant higher IC₅₀ to 5-FU compared to untreated cells

512 It should be noted that due to low number of persister cells (0.01% of initial population), we could not
513 measure the IC₅₀ to 5-FU.

514 For Extended Fig. 9d cells were treated with 5-FU at 5uM and indicated concentrations of GSK-J4 or
515 GSK-J5 (D-5 indicated 5 days pre-treatment with GSK-J4 or GSK-J5 before 5-FU treatment (D0), D0
516 indicated co-treatment 5-FU and GSK-J4 or GSK-J5, D10 and D30 indicated that treatment with GSK-J4
517 or GSK-J5 started 10 days or 30 days respectively after the onset of 5-FU treatment (D0). The number
518 of persister cells were counted manually under the microscope at day 42 (n=3).

519 The GraphPad PRISM 9 was used for statistics and the results represent the mean ± sd of three
520 independent experiments. Statistical analysis was performed using the Bonferroni test for multiple
521 comparisons between samples (Fig. 4a, Extended Fig. 8a/d/f, Extended Fig. 9b/d and Extended Fig.2b-

522 right) or one-tailed Mann-Whitney test for the comparison between two conditions (Extended Fig. 2b-
523 left).

524

525 **Western blotting.** In Extended Fig. 8b/e/g, DMSO- and EZH2i-treated cells were lysed at 95°C for 10
526 minutes in Laemmli buffer (50 mM Tris-HCl [pH 6.8], 2% SDS, 5% glycerol, 2 mM DTT, 2.5 mM EDTA,
527 2.5 mM EGTA, 4 mM Sodium Orthovanadate, 20 mM Sodium Fluoride, protease inhibitors,
528 phosphatase inhibitors) and proteins concentrations were measured using a Pierce BCA protein Assay
529 Kit (Thermo Fisher Scientific, Ref: 23225/23227). 10 µg of proteins were then separated on a 4-15%
530 Mini-PROTEAN TGX Stain-Free Gel (Bio-Rad, Ref: 4568085) at 160V. After transfer, the membrane was
531 blocked for 1 h at room temperature in PBS pH 7.4 containing 0.1% Tween-20 and 1% milk (Regilait).
532 Incubation anti-H3K27me3 (Dilution: 1:2000, Cell Signaling, Ref: 9733) or EZH2 (Dilution: 1:2000, Cell
533 Signaling, Ref: 5246) or Tubulin (Dilution: 1:1000 , Thermo Fisher Scientific, Ref: 31460) primary
534 antibodies diluted in PBS pH 7.4, 0.1% Tween-20 were performed at 4°C overnight. Following 2h
535 incubation at room temperature with an anti-rabbit or mouse peroxidase-conjugated secondary
536 antibody (Dilution: 1:10000, Thermo Fisher Scientific, Ref: 31460 or Ref: 31430) diluted in PBS pH 7.4,
537 0.1% Tween-20, antibody-specific labeling bands were revealed (Bio-Rad, ChemiDoc MP) using a
538 SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, Ref: 34579).

539

540 **Lentivirus packaging and cell transduction.** Lentivirus was produced by transfecting the barcode
541 plasmids pRRL-CMV-GFP-BCv2Ascl and p8.9-QV and pVSVG into HEK293T cells as previously
542 described³⁰. MDA-MB-468 cells from ATCC were infected at passage 11 with lentivirus produced from
543 the barcode library (pRRL-CMV-GFP-BCv2Ascl) which includes 18206 different barcodes of 20bp of a
544 random stretch, at a low multiplicity of infection (MOI 0.1) to minimize the number of cells marked by
545 multiple barcodes. Three weeks after transduction, cells were sorted for GFP expression to select cells
546 with barcode insertion, and used for drug treatment.

547

548 **Single-cell RNA-seq.** For each single cell suspension (DMSO-D0-#1, 5-FU-D33-#1, 5-FU-D214-#1, 5-FU-
549 D67-#2, 5-FU-D171-#2, 5-FU-D50-#3, 5-FU-D77-#3 and 5-FU-D202-#3) or PDX dissociated cells
550 (PDX_95, PDX_39 or PDX_172, untreated and persister cells), approximately 3,000 cells were loaded
551 on a Chromium Single Cell Controller Instrument (Chromium Single Cell 3'v3, 10X Genomics, Ref: PN-
552 1000075) according to the manufacturer's instructions. Samples and libraries were prepared according
553 to the manufacturer's instructions. Libraries were sequenced on a NovaSeq 6000 (Illumina) in PE 28-
554 8-91 with a coverage of 50,000 reads/cell.

555

556 **Bulk lineage barcode library preparation and sequencing.** Lineage barcodes are recovered by isolating
557 genomic DNA from cells of interest (NucleoSpin Tissue, Mini kit for DNA from cells and tissue,
558 Macherey Nagel, Ref: 740952.50). From the isolated genomic DNA, barcodes are amplified with three
559 nested PCR steps as described in³⁰ (see Extended Table 1 for primer sequence). In short, after a first
560 specific PCR for the common region of the lineage barcodes, the amplified material was prepared for
561 sequencing by addition of the illumina sequencing adaptaters and indexing and purification.
562 Sequencing was done in order to obtain 50 reads, on average, per barcoded cell.

563

564 **Bulk ChIP-seq.** ChIP experiments were performed as previously described¹⁶ on 3x10⁶ MDA-MB-468
565 cells (DMSO-D67-#2, DMSO-D77-#3, DMSO-D113-#4, 5-FU-D67-#2, 5-FU-D77-#3, 5-FU-D113-#4) using
566 an anti-H3K27me3 antibody (Cell Signaling Technology, Ref: 9733 - C36B11). Sequencing libraries were

567 prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB, Ref: E7645S) according to the
568 manufacturer's instructions. Libraries were sequenced on a NovaSeq 6000 (Illumina) in SE50 mode.

569
570 **Single-cell ChIP-seq.** Cells (DMSO-D60-#1, DMSO-D77-#3, DMSO-D131-#5, 5-FU-D33-#1, 5-FU-D67-#2,
571 5-FU-D171-#2, 5-FU-D147-#3, 5-FU-D131-#6) were labeled by 15 min incubation with 1 μ M CFSE
572 (CellTrace CFSE, ThermoFisher Scientific, Ref: C34554). Cells were then resuspended in PBS
573 supplemented with 30% Percoll, 0.1% Pluronic F68, 25 mM Hepes pH 7.4 and 50 mM NaCl. Cell
574 encapsulation, bead encapsulation and 1:1 droplet fusion was performed as previously described¹⁶,
575 see Extended Table 1 for the sequence of bead barcodes. Immunoprecipitation with H3K27me3
576 antibody (Cell signaling, Ref: 9733 - C36B11) or H3K4me3 antibody (Cell signaling, Ref: 9751-C42D8),
577 DNA amplification and library were performed as in¹⁶. Libraries were sequenced on a NovaSeq 6000
578 (Illumina) in PE100, with 4 dark cycles on Read 2, with a coverage of 100,000 reads/cell.

579
580 **Quantitative chromatin profiling with chromatin indexing.** Chromatin isolation, indexing,
581 immunoprecipitation and library preparation was adapted from⁴⁹. Briefly, 50,000 MDA-MB-468 were
582 lysed and digested with MNase for 20min at 37°C in the following buffer: 46mM Tris-HCl pH 7.4,
583 0.154M NaCl, 0.1% Triton, 0.1% NaDoc, 4.65mM CaCl₂, 0.47x Protease Inhibitor Cocktail (Roche, Ref:
584 11873580001) and 0.09u/uL MNase (Thermo Scientific, Ref: EN0181). Fragmented nucleosomes were
585 then ligated for at least 24h at 16°C to double-stranded barcoded adapters containing 8bp barcodes
586 to combine samples: Pac1-T7-Read2-8bpBarcode-linker-Pac1 (Extended Table 1). Next, 5 indexed
587 chromatin samples (DMSO, 5-FU, UNC, 5-FU + UNC, GSK-J4) were pooled, each containing a different
588 8-bp barcode, to perform anti-H3K27me3 ChIP (Cell Signaling, Ref: 9733 - C36B11) on 250,000 cells in
589 total in each pool. ChIP and DNA amplification was carried out as for scChIP-seq¹⁶ and a sequencing
590 library was produced for both IP and input pools and sequenced on NovaSeq 6000 (Illumina) in PE100
591 mode.

592
593 **Sequential ChIP-seq.** Primary ChIP experiments were performed as described previously¹⁶ on 10x 10⁶
594 untreated MDA-MB-468, BT20 or HCC38 cells or untreated PDX_95, PDX_39 or PDX_172 tumor
595 dissociated cells using the anti-H3K27me3 antibody (Cell Signaling, Ref: 9733 - C36B11 – MDA-MB-468)
596 or anti-H3K4me3 antibody (Cell Signaling, Ref: 9751-C42D8 – MDA-MB-468-bis, BT20, HCC38 and PDX
597 models). After washes, samples were eluted twice at 37°C for 15 min under agitation in an elution
598 buffer (50mM Tris-Hcl pH8, 5mM EDTA, 20mM DTT, 1% SDS) as in. Samples were diluted 10 times to
599 decrease SDS and DTT concentration. 10% of the eluted chromatin was kept as primary ChIP.
600 Secondary ChIP, re-ChIP, was performed overnight on the rest of the primary immuno-precipitated
601 chromatin using an anti-H3K4me3 antibody (MDA-MB-468) or anti-H3K27me3 (MDA-MB-468-bis,
602 BT20, HCC38 and PDX models) or using an anti-IgG antibody (Cell signaling, Ref: 3900 – all samples) as
603 a control, to determine the background level of the re-ChIP experiment. After washes, samples were
604 eluted twice at 65°C for 15 min under agitation in 0.1M NaHCO₃ and 1% SDS as in⁵⁰. After reverse
605 crosslinking and DNA clean-up, 3 to 15 ng of immunoprecipitated DNA were used to prepare the
606 sequencing libraries using the NEBNext Ultra II DNA Library Prep Kit (NEB, Ref: E7645S) according to
607 the manufacturer's instructions. Libraries were sequenced on a NovaSeq 6000 (Illumina) in SE100
608 mode. For MDA-MB-468, we verified that the two ways (H3K27me3->H3K4me3 and H3K4me3-
609 >H3K27me3) yielded similar results. We found a significant overlap of the 1,547 and 2,490 bivalent
610 genes obtained with the two ways (p=2.2e-16, Ext. Fig. 6g) and found that the enriched pathways were
611 strongly correlated (Pearson's r = 0.81, Ext. Fig. 6h).

612
613 **CUT&Tag on frozen tumor samples.** CUT&Tag was performed as in Kaya-Okur et al. with minor
614 modifications on 50,000 to 100,000 nuclei with 1:50 antibody (Cell Signaling Antibodies : Anti-
615 H3K27me3, Ref: 9733- C36B11, Anti-H3K4me3, Ref: 9751- C42D8)^{17,51}. All washes were performed in
616 a volume of 500µL and all centrifugations were done using a swinging bucket centrifuge at 1300g,
617 4min, at 4°C for nuclei preparation and 600g, 8min, 4°C for subsequent steps. Nuclei were extracted
618 and permeabilized from 10-20mg frozen tumor tissues by incubating samples 10min on ice in 6mL ice-
619 cold NE1 buffer (20mM HEPES pH7.2, KCl 10mM, spermidine 0.5mM, glycerol 20%, BSA 1%, NP-40 1%,
620 digitonin 0.01%, proteases inhibitor 1x) after mechanical dissociation. Following antibody incubation
621 and tagmentation, samples were incubated for 1h at 55°C with max speed agitation with 3uL SDS10%
622 and 2,5uL 20mg/mL proteinase K. After DNA extraction (Qiagen, Ref: 139046 MaXtract High density),
623 PCR amplification (with 17 cycles, 20s at 63°C combined annealing/extension step) of the sequencing
624 libraries was performed and profiles were checked on the Agilent TapeStation using High-sensitivity
625 D1000 reagents. CUT&Tag libraries were sequenced on a NovaSeq 6000 (Illumina) in PE50 mode.

626
627
628 **Whole exome sequencing.** Genomic DNA from samples (DMSO-D0, DMSO-D147-#3, DMSO-D171-#5,
629 DMSO-D131-#6, 5-FU-D67-#2, 5-FU-D153-#2, 5-FU-D50-#3, 5-FU-D147-#3, 5-FU-D171-#5 and 5-FU-
630 D131-#6) were extracted with NucleoSpin Tissue, Mini kit for DNA from cells and tissue (Macherey
631 Nagel, Ref: 740952.50) and sequenced on a NovaSeq 6000 (Illumina) with a 100X depth.

632
633 **Computational approaches:**

634 **Single-cell RNA-seq analysis.** The scRNA-seq sequencing files were preprocessed using the cellRanger
635 pipeline. For PDX samples, files were aligned against hg19 and mm10 genomes and only cells with a
636 majority of human reads were retained for the analysis. For the MDA-MB-468 human cell line,
637 sequences were aligned against the hg38 genome only. Cells with less than 3,000 cells for MDA-MB-
638 468 or 2,500 for PDX or more than 8,000 detected genes, or more than 100,000 reads were filtered
639 out, as well as cells with a percent of mitochondrial reads greater than 15% or a percentage of spike in
640 greater than 5%. Normalization, dimensionality reduction and Louvain clustering was done using
641 monocle3 (v0.2.2)⁵² keeping the first 50 Principal Components (PC). Cell cycle was determined for each
642 cell using Seurat (v3.1.5)⁵³. For MDA-MB-468 datasets, we removed from subsequent analysis clusters
643 with less than 0.5% of the total cells (150 cells) (clusters R1, R5, R7, R9, R11 and R12). Differentially
644 expressed genes were obtained by comparing raw gene expression values using edgeR GLM statistical
645 model⁵⁴. For the PDX model, persister cells were compared to cells from the untreated tumor; for
646 MDA-MB468 cells, cells from cluster R2 (persister cells) were compared to cells from cluster R10
647 (untreated population). Genes were considered significantly overexpressed if the fold change was
648 higher than 3 and the adjusted p-value less than 0.01. For MDA-MB-468 samples, as the number of
649 cells was high, a subset of 500 cells per cluster was subsampled from each cluster for the differential
650 analysis and downstream steps. Intra-cluster correlation scores were calculated using Pearson's
651 correlation score, with a random subsampling of n=500 cells per cluster.

652
653 **Single-cell lineage barcode extraction from 10X datasets.** To detect the lentiviral inserted barcodes in
654 the 10x sequencing data we used custom R scripts. To avoid running scripts on all reads we: 1. used
655 samtools to extract all unmapped reads from the 10x output bam file, 2. used awk to take reads with
656 either a 3' or 5' 20bp match to the constant flanking region of the barcode allowing one mismatch, 3.

657 retained only reads with both a 10x cell barcode and 10x UMI. We then located the 20bp match to one
658 of the constant flanking regions allowing one mismatch (but filtered out reads where a mismatch was
659 in the first or last base to ensure the barcode is at the expected position). We then further required a
660 4bp exact match on the other side of the barcode, and then extracted the 20bp viral barcode, read
661 name, 10x cell barcode and 10x UMI.

662 To assign one viral barcode to each 10x cell barcode, we determined a consensus viral barcode for
663 each UMI. For each position in the barcode we returned the most frequently observed base and the
664 proportion of reads supporting this consensus. Barcodes associated with a 10x CB-UMI pair were
665 filtered if the proportion of reads supporting a position was ≤ 0.5 at ≥ 3 positions. Next, one viral
666 barcode for each 10x cell barcode was taken as a consensus across all remaining CB-UMI pairs for each
667 10x cell barcode. We took the most frequently observed base for each position and the barcode was
668 filtered if the proportion of UMIs supporting a position is ≤ 0.5 at ≥ 3 positions. Finally, we checked
669 whether the viral barcodes were in our barcode library and excluded them if not. UMAPs were colored
670 according to lineage identity, for cells each color corresponding to a unique viral barcode. For
671 comparison with bulk datasets, pseudo-bulk barcode frequencies were computed and normalized to
672 10,000 total barcodes/sample, as for bulk.

673

674 **Bulk barcode pre-processing.** The analysis pipeline was performed as previously published⁵⁵. In brief,
675 using R-3.4.0 (R Development Core Team (2019) <http://www.R-project.org>), raw reads were first
676 filtered for perfect to the input index- and common-sequences using XCALIBR
677 (<https://github.com/NKI-GCF/xcalibr>) and filtered against the barcode reference list. Correlation
678 between technical (PCR) replicates (Extended Fig. 3j) was used as quality control: samples were then
679 normalized and filtered for a Spearman correlation between replicates higher than 0.6 and barcodes
680 present in only one of the two replicates were set to zero. The mean of the replicates was used for
681 downstream analysis.

682

683 **Bulk and single-cell lineage barcode analysis.** All barcode frequencies were transformed with asinh.
684 Shannon indexes to assess barcode diversity were computed as described previously⁵⁶. Normalized
685 frequencies from bulk and single-cell datasets were clustered using hierarchical clustering based on
686 Spearman correlation and Ward method. Frequencies across time points and conditions were
687 compared with a Spearman correlation coefficient and associated p-value. To test whether barcode
688 frequencies within DMSO and 5-FU treated cells correspond to a random sampling of the initial
689 untreated population, we used proportionate sampling PPS to simulate an *in silico* barcode frequency
690 vector from a consensus barcode frequency vector of the initial population – obtained from n=6
691 drawings – and compared simulated and observed frequencies as above. For single-cell datasets, we
692 computed the fraction of unique barcodes as the number of unique detected barcodes over the total
693 number of detected barcodes, for a given cluster or cell population.

694

695

696 **Bulk ChIP-seq analysis and consensus peak annotations.** Raw sequencing files were mapped in single-
697 end mode using bowtie with options '-k 1 -m 1'^{57,58} against the human genome (hg38). PCR duplicates
698 were removed using Picard Mark Duplicates function (GitHub Repository.
699 <http://broadinstitute.github.io/picard/>). In order to define a consensus annotation specific to our
700 MDA-MB-468 model, peaks were first called on each of all bulk MDA-MB-468 samples, both DMSO
701 and 5-FU treated, against their respective inputs using Zerone⁵⁷ with a confidence of 95% and window

702 size of 1,000bp for H3K27me3 mark and 500bp for H3K4me3 mark. Peaks were further merged
703 together when closer than 10,000bp. For H3K4me3 this defined the *consensus peak annotation* with a
704 total of 29,714 peaks. For H3K27me3, peaks were further filtered to refine annotation: (i) only keeping
705 peaks having signal in at least two samples and (ii) removing small peaks (< 2,000bp) with
706 overestimated signal due to window size normalization. A total of 9,568 consensus peaks were found
707 for H3K27me3 landscapes in 5-FU and DMSO treated cells, defining the H3K27me3 *consensus peak*
708 *annotation*.

709

710 **Single-cell ChIP-seq read processing.** The single-cell ChIP-seq sequencing files were preprocessed
711 using our single-cell ChIP-seq dedicated pipeline

712 (https://github.com/vallotlab/scChIPseq_DataEngineering).

713 Each #Read 2 was first splitted into a cell barcode sequence composed of the first 79 nucleotides and
714 the last 22 nucleotides corresponding to genomic DNA. Full #Read 1 and genomic DNA of #Read 2 were
715 mapped in paired-end mode to hg38 whole genome using STAR (v2.6.0c) with parameters '--
716 alignEndsType EndToEnd --outFilterMultimapScoreRange 2 --winAnchorMultimapNmax 1000 --
717 alignIntronMax 1 --peOverlapNbasesMin 10 --alignMatesGapMax 450' for PDX model and against hg38
718 only for MDA-MB-468 cell line, by keeping only reads having no more than one reportable alignments
719 and 2 mismatches. For each barcode (*i.e* cell), reads with identical #Read 1 starting sites were marked
720 as duplicates, probably emerging from reverse-transcription or PCR duplicates. #Read 1 sequences
721 paired with unmapped #Read 2, and falling within the same 50bp-window, were further stacked into
722 one read, as possibly originating from PCR duplicates or from the same nucleosome.

723 For H3K27me3 and H3K4me3 experiments in MDA-MB-468 cells, reads were counted according to two
724 annotations: (i) within *consensus peak annotation* (used in Fig. 2a-d, 2f, Extended Fig. 5b,e), and (ii)
725 within a TSS-based annotation (used in Fig. 2e,g, 3a-b, Extended Fig. 5j, 6a-d), comprising 52,138
726 regions of 10kbp centered around TSS of all transcripts of protein coding and lncRNAs from Gencode
727 v34⁵⁹. For PDX untreated tumors, H3K27me3 and H3K4me3 experiments, reads were counted
728 according to TSS-based annotation. For all experiments and annotations, only cells with a coverage
729 over 1,000 reads were kept for downstream analysis, see Extended Table 3 for sample and cell
730 numbers.

731

732 **Single-cell ChIP-seq filtering, dimensionality reduction and clustering.** QC filtering, dimensionality
733 reduction, and clustering were done using ChromSCape⁶⁰, (<https://github.com/vallotlab/ChromSCape>)
734 with default parameters for H3K4me3 datasets, resulting 1,345 cells with signal over 4,983 TSS. For
735 H3K27me3 datasets, minimum coverage was increased to 3,000 reads/cell and for each sample the
736 number of cells was randomly downsampled at 500 cells per sample to ensure equal contribution of
737 each datasets to dimensionality reduction. The resulting matrix contains 3,576 cells with signal over
738 8,858 peaks.

739 In order to exclude from the subsequent analysis known copy number variation (CNV) regions between
740 samples, CNV regions previously identified using ChromHMM⁶¹ on the input of bulk experiment of
741 MDA-MB-468 samples were used by ChromSCape as regions to exclude from the analysis. Coverage
742 tracks of *metacells* for scChIP-seq were obtained by aggregating the signal of single-cells into
743 cumulative signals in each cluster. We define a group of cells as being more 'synchronous' regarding a
744 set of genes (e.g. persister genes) if they have a significantly higher number of genes with H3K4me3
745 signal, according to a Wilcoxon non-parametric rank test.

746

747 **Differential analysis of H3K27me3 chromatin landscapes genome-wide.** These analyses were done
748 using consensus peak annotations, to assess for genome-wide changes without *a priori* on gene
749 annotation. For each scChIP-seq datasets, for each cluster (E1, E2, E4), pseudo-bulk samples were
750 generated by summing up reads from individual cells provided there were more than 50 cells in the
751 sample in the given cluster. Pseudo-bulk scChIPseq signals are normalized by the total number of reads
752 for all cells of the corresponding population ('persister' or 'untreated'). For each loci, both pseudo-bulk
753 tracks are shown at the same magnification, with the same range for the y-axis to enable comparison
754 between pseudo-bulk tracks.

755 We performed two differential analysis based on counts within the *consensus peak annotation*: (i) one
756 to define the specific chromatin changes in persister cells versus untreated cells (Fig. 2c-d, Extended
757 Fig. 5e), where we compared pseudo-bulks and bulks of persister cells to pseudo-bulks and bulks of
758 untreated cells, and (ii) one to compare chromatin landscapes of subpopulations within the untreated
759 populations (Extended Fig. 5e). For (i), as persister cells grouped within one cluster, we combined n=2
760 pseudo-bulks to n=4 bulk matrices from biological replicates to perform differential analysis using
761 Limma package⁶². Peaks with a log2FC over 1 and under -1 and an adjusted p-value below 0.1 were
762 considered significantly enriched or depleted of H3K27me3 in persister cells. For Fig. 2c, we used a
763 generic hg38 genome gene/TSS annotation that classifies regions into categories, e.g. gene TSS,
764 intergenic or enhancer regions. For each category we test whether this category is significantly more
765 prevalent in differentially enriched peaks between persister and untreated states versus in all peaks.
766 The 'enrichment' metric is the log2(number of differential peaks in the category/total number of peaks
767 in that category). Fisher's exact test was used to compare the localization of depleted H3K27me3 peaks
768 in respect to gene annotation.

769
770 **Epigenomics and transcriptome data comparison.** To integrate epigenomic and transcriptome data
771 (SI_Table 4, Fig. 3c), for each gene we combined both TSS-based and peak-based differential analysis
772 (considering peaks closer than 1kbp to the TSS) with the same thresholds as above. We represented
773 this integration as a donut plot (Fig. 3c), taking into consideration all *persister* genes n=168 and adding
774 independently, for each, information on H3K27me3 status upon 5-FU treatment (depleted in TSS or
775 nearby peak, or unchanged), on bivalent status in untreated cells (see related section for thresholds)
776 and on presence in the top 100 of CheA3 predicted TFs (see below).

777
778 **Gene regulatory networks.** In order to test whether persister genes are co-regulated by master
779 regulators, we ran ChEA3³³ data mining algorithm (from TF-target interaction based on multiple
780 sources, e.g. ENCODE, GTEx co-expression, ReMap ChIP-seq, EnrichR, ARCHS4 co-expression and ChIP-
781 seq from the literature) to find TFs with regulons enriched in persister genes *in vitro* and in PDXs. In
782 order to create a background control to assess the quality of the ranking score given by ChEA3, we ran
783 ChEA3 for 1000 random gene sets, expressed in our scRNA-seq data and of the same size. The inverse
784 of the ChEA3 score of the top ranking 100 TF regulons enriched in our persister genes were significantly
785 greater than for the random background gene sets for all models (one-sided T-test p.values with
786 respectively for MDA-MB-468, PDX_95, PDX_39 and PDX_172: 2.2e-16, 5.8e-08, 6.9e-3 & 6.9e-05).

787
788 **Chromatin indexing analysis.** The bulk chromatin indexing sequencing files were first demultiplexed
789 by matching the first 8 bases of #Read 2 without any mismatches to the 8-bp long index of each sample
790 from a pool of 5 samples. The same demultiplexing was done for the corresponding inputs. Afterwards,
791 mapping and demultiplexing was done as in bulk ChIP-seq (see above). Relative total amounts of

792 immunoprecipitated DNA were determined as the ratio of the number of reads in the IP by the number
793 of reads in the corresponding input for each sample within the pool. Coverage tracks were normalized
794 with this ratio.

795

796 **Sequential ChIP-seq analysis.** Fastq files for primary (ChIP) and secondary (ChIP-reChIP)
797 immunoprecipitation were processed as for bulk ChIP-seq (see above). For the MDA-MB-468 cell line,
798 sequential ChIP-seq was processed with H3K27me3 as primary ChIP then H3K4me3 as secondary ChIP
799 or vice versa. For other *in vitro* models and PDXs only the latter H3K4me3 -> H3K27me3 way was kept
800 as peaks were more easily identifiable in the secondary ChIP profiles. IgG secondary ChIP was used as
801 a negative control.

802 For the H3K27me3 -> H3K4me3 reChIP datasets, peaks were called on the secondary ChIP (H3K4me3
803 or IgG) with MACS2⁶³ using the primary H3K27me3 signal as control and with parameters 'macs2
804 callpeak --call-summits -p 0.1 --nomodel --extsize 300'. Summits closer to each other by 1,000 bp were
805 merged using Bedtools⁶⁴. Only peaks overlapping both the TSS annotation and the H3K27me3
806 consensus peak annotation obtained from ChIP-seq experiments (see above) were kept to focus on
807 TSS chromatin landscapes. A ratio and associated p-value for each peak was calculated as follows:

808 (i) Reads were counted within the 2kbp region around the peak summit ("peak") as well as in
809 the 20kbp region around peak ("locus") in each ChIP-reChIP. A ratio of "peak" / "locus" was calculated
810 in order to control for the relative increase in signal in this region (presence of a peak).

811 (ii) Reads were counted in the 2kbp region around the summits ("peak") as well as in the
812 500kbp region around peak ("area") in the primary ChIP and the ChIP-reChIP to create a contingency
813 table.

814 A Fisher exact test was performed on the contingency table to reject the null hypothesis that the
815 number of reads in "peaks" compared to reads in "area" is greater in the ChIP-reChIP than in the
816 primary ChIP. P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure⁶⁵.
817 In order to choose adequate thresholds for adjusted p-value and "peak"/"locus", we calculated the
818 number of false positives for p-values ranging from 0.1 to 0.001 and ratios ranging from 10 to 25%
819 (Extended Fig. 6f), using H3K27me3/IgG ChIP and ChIP-reChIP as a negative control. An adjusted p-
820 value threshold of 0.001 and peak ratio threshold of 15% was chosen to minimize false positives peaks
821 and resulted in 1,266 bivalent peaks covering 1,547 TSS.

822 For the H3K4me3 -> H3K27me3 reChIP datasets, peaks were first called on primary ChIP using MACS2
823 without control with parameters '--call-summits -p 0.01 --nomodel --extsize 300'. The number of reads
824 in the region 2.5kbp upstream and downstream of each peak were counted in the primary and
825 secondary ChIP. Reads were normalized by total library size. First, the ratio between secondary and
826 primary ChIP were calculated for each peak and then the odd-ratio between each TSS and it's 60
827 closest neighbours were calculated from the ratios. In order for a TSS to be considered bivalent, the
828 odd ratio of a given peak compared to the 60 closest neighbour peaks must be greater than 4. Then, A
829 Fisher exact test was performed on the contingency table to reject the null hypothesis that the number
830 of reads in secondary compared to reads in primary ChIP is greater in the given peak than in it's 60
831 neighbor peaks. P-values were adjusted for multiple testing using the Benjamini-Hochberg
832 procedure⁶⁵. In order to choose adequate thresholds for adjusted p-value and "peak"/"locus", we
833 calculated the number of false positives for p-values ranging from 0.001 to 1e-30, using H3K4me3/IgG
834 ChIP and ChIP-reChIP as a negative control. Adjusted p-value thresholds were always lower than 0.001
835 and were defined so that we obtained the greatest number of bivalent peaks in the
836 H3K4me3/H3K27me3 experiment compared to the H3K4me3/IgG negative control.

837 The comparative coverage tracks were generated by calculating the log₂ ratio of secondary ChIP versus
838 primary ChIP using Deeptools bamCompare and then smoothed. For each loci, H3K27me₃/H3K4me₃
839 and H3K27me₃/IgG or H3K4me₃/H3K27me₃ and H3K4me₃/IgG tracks are shown at the same
840 magnification and with the same range for the y-axis for comparison between tracks.

841

842 **Bulk CUT&TAG of patient tumors to assess bivalency.** Fastq files for H3K27me₃ and H3K4me₃ were
843 processed as for bulk ChIP-seq (see above). For each sample, reads were counted on the TSS
844 annotation and normalized using log₂ RPKM. Then, only the top 15% most covered TSS were kept for
845 H3K4me₃ and H3K27me₃. Bivalent TSS were taken as the intersection of these highly covered TSS.

846

847 **Gene set analysis.** For all gene set analysis, we applied hypergeometric tests to identify gene sets
848 enriched within significantly overexpressed genes (scRNA), genes devoid of H3K27me₃ (scChIP-seq) or
849 bivalent genes (Sequential ChIP-seq, bulk Cut&Tag) from MSigDB v5 database⁶⁶, correcting for multiple
850 testing with the Benjamini-Hochberg procedure. Gene sets with an adjusted p-value below 0.1 were
851 considered significantly enriched. The gene background universe for hypergeometric testing was the
852 entire set of expressed genes for scRNA or the 32,937 genes present in Gencode for scChIP-seq or
853 bivalent gene lists. The tests were performed on all lists, but we display only the relevant following
854 lists: 'c2_curated' related to breast (searching for 'MAMMARY' or 'BREAST'), 'c2_curated' related to
855 'KEGG', 'c5_GO' and 'c7_hallmark', filtering out genetic-event related lists (containing 'AMPLICON').

856 When displaying gene set analysis of multiple samples, we first selected gene sets significantly
857 enriched in 3/3 of PDXs or in vitro samples or at least 7/9 samples for human tumors. Then, gene sets
858 were ranked by the average adjusted p-values, and only the top 5 gene sets of each category were
859 displayed. The dotplots representing pathway enrichment across multiple samples were done using
860 clusteRprofile⁶⁷. The gene ratio stands for the fraction of genes belonging to each pathway. For the
861 PDX and in vitro experiments, the -log₁₀ adjusted p-value of the main models are displayed,
862 respectively PDX_95 and MDA-MB-468. In order to calculate the significativity of overlaps when we
863 compared multiple set of pathways, we used the Exact Test of Multi-set intersections⁶⁸.

864

865 **Whole Exome Sequencing data analysis.** The WES sequencing files were mapped using bwa-mem⁶⁹ to
866 the human genome (hg19). Reads falling in the targeted regions were then filtered based on their
867 mapping quality and PCR duplicates were removed. Local Indel Realignment and Base Score
868 Recalibration was applied to deduplicated reads using GATK⁷⁰. Somatic variants were called with
869 Mutect2⁷¹ using early passage 16 (p16) and a series of patient blood samples as reference (referred as
870 "blood", see Extended Fig. 4a). At this step, only mutations labeled as 'PASS' or 't_lod_fstar' were kept
871 and additional filters were applied based on [http://best-practices-for-processing-hts-
872 data.readthedocs.io/en/latest/mutect2_pitfalls.html](http://best-practices-for-processing-hts-data.readthedocs.io/en/latest/mutect2_pitfalls.html). Using p16 as the "normal" sample allowed to
873 directly obtain somatic variants acquired after treatment or not to 5-FU, but we hypothesized that
874 some germline variants might be wrongly called somatic variants so we filtered out variants that were
875 reported in ExAC Non Finnish European database, effectively removing 231 variants possibly germline.
876 For MDA-MB-468 persisters & resistant samples, mutations also present in one of the untreated
877 samples were removed. Mutations occurring in a breast cancer driver gene list from¹³ was used to
878 annotate mutations falling in driver genes.

879 GAP⁷² was used to calculate with precision absolute copy number and B allele frequencies (BAF) taking
880 depth of coverage and allele frequency from a set of known germline variants from the⁷³ as inputs,
881 and using "blood" as normal sample. Palimpsest⁷⁴ was then used to calculate the Cancer Cell Fraction

882 (CCF) of each mutation in each sample, *i.e.* the proportion of cells in the population bearing the
883 mutation, correcting by purity, BAF and absolute copy number of the segment. Then mutations were
884 classified in either ‘subclonal’ or ‘clonal’ depending on their CCF. Finally, de novo mutational signatures
885 were obtained from the mutations context and matched to a set of known signatures from COSMIC v2
886 (https://cancer.sanger.ac.uk/cosmic/signatures_v2) that were observed in breast cancer (*i.e.*
887 signatures 1, 2, 3, 8, 13, 17, 18, 20, 26 & 30).

888

889 **Data and code availability**

890 All sequencing files were deposited to GEO under a private repository GSE164716. All statistical
891 analysis was performed in R (v4.1) using custom R scripts. Codes for data analysis are available at the
892 following repositories <https://github.com/vallotlab/ChemoPersistence> and
893 https://github.com/TeamPerie/lentiviral_barcode_detection_in10X_data/.

894

895 **Online references**

- 896 46. Cottu, P. *et al.* Acquired Resistance to Endocrine Treatments Is Associated with Tumor-Specific Molecular Changes in Patient-
897 Derived Luminal Breast Cancer Xenografts. *Clinical Cancer Research* **20**, 4314–4325 (2014).
- 898 47. Marangoni, E. *et al.* Capecitabine Efficacy Is Correlated with TYMP and RB1 Expression in PDX Established from Triple-Negative
899 Breast Cancers. *Clin Cancer Res* **24**, 2605–2615 (2018).
- 900 48. Petit, V. *et al.* Optimization of tumor xenograft dissociation for the profiling of cell surface markers and nutrient transporters. *Lab*
901 *Invest* **93**, 611–621 (2013).
- 902 49. van Galen, P. *et al.* A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. *Mol. Cell* **61**, 170–180 (2016).
- 903 50. Desvoyes, B., Sequeira-Mendes, J., Vergara, Z., Madeira, S. & Gutierrez, C. Sequential ChIP Protocol for Profiling Bivalent
904 Epigenetic Modifications (ReChIP). in *Plant Chromatin Dynamics* (eds. Bemer, M. & Baroux, C.) vol. 1675 83–97 (Springer New York, 2018).
- 905 51. Bartosovic, M. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature*
906 *Biotechnology* **33**.
- 907 52. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat*
908 *Biotechnol* **32**, 381–386 (2014).
- 909 53. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
- 910 54. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to
911 biological variation. *Nucleic Acids Research* **40**, 4288–4297 (2012).
- 912 55. Perié, L., Duffy, K. R., Kok, L., de Boer, R. J. & Schumacher, T. N. The Branching Point in Erythro-Myeloid Differentiation. *Cell* **163**,
913 1655–1662 (2015).
- 914 56. Jost, L. *Entropy and diversity*. *Oikos* **113**, 363–375 (2006).
- 915 57. Cuscó, P. & Filion, G. J. Zerone: a ChIP-seq discretizer for multiple replicates with built-in quality control. *Bioinformatics* **32**, 2896–
916 2902 (2016).
- 917 58. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the
918 human genome. *Genome Biol* **10**, R25 (2009).
- 919 59. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* **47**, D766–D773
920 (2019).
- 921 60. Prompsy, P., Kirchmeier, P. & Vallot, C. *ChromSCape : a Shiny/R application for interactive analysis of single-cell chromatin profiles*.
922 <http://biorxiv.org/lookup/doi/10.1101/683037> (2019) doi:10.1101/683037.
- 923 61. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215–216 (2012).
- 924 62. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids*
925 *Research* **43**, e47–e47 (2015).
- 926 63. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
- 927 64. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics* **47**, (2014).
- 928 65. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal*
929 *of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
- 930 66. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression
931 profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).
- 932 67. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*
933 **16**, 284–287 (2012).
- 934 68. Wang, M., Zhao, Y. & Zhang, B. Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep* **5**, 16923 (2015).
- 935 69. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- 936 70. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*
937 **43**, 491–498 (2011).
- 938 71. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*

939 31, 213–219 (2013).
940 72. Popova, T. *et al.* Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP
941 arrays. *Genome Biol* 10, R128 (2009).
942 73. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68–74 (2015).
943 74. Shinde, J. *et al.* Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer.
944 *Bioinformatics* (2018) doi:10.1093/bioinformatics/bty388.

945

946 **Acknowledgments.**

947 Single-cell experiments were performed by the Single-Cell platform of the Institut Curie. We thank
948 Antonin Morillon for critical reading of the manuscript.

949

950 **Fundings.**

951 This work was supported by the ATIP Avenir program, by Plan Cancer, by the SiRIC-Curie program SiRIC
952 Grants #INCa-DGOS-4654 and #INCa-DGOS-Inserm_12554, and by a starting ERC grant from the H2020
953 program #948528-ChromTrace (to CV), and by the Fondation de France #00107944 (to JM). The work
954 was supported by an ATIP-Avenir grant from CNRS and Bettencourt-Schueller Foundation, by the Labex
955 CelTisPhyBio #ANR-11-LABX-0038 and by a starting ERC grant from the H2020 program #758170-
956 Microbar (to LP). High-throughput sequencing was performed by the ICGex NGS platform of the
957 Institut Curie supported by the grants Equipex #ANR-10-EQPX-03, by the France Genomique
958 Consortium from the Agence Nationale de la Recherche #ANR-10-INBS-09-08 ("Investissements
959 d'Avenir" program), by the ITMO-Cancer Aviesan - Plan Cancer III and by the SiRIC-Curie program SiRIC
960 Grant #INCa-DGOS- 4654.

961

962 **Author contributions.**

963 JM, AD, CL, LB, SBT, AE and AT performed experiments. scChIP-seq experiments were conducted
964 together with SF and KG. PDX experiments were performed by EM, LS and AD. MB and SB performed
965 sequencing. PP and CV performed omics data analysis. Lineage barcoding data were analyzed by AML,
966 CV and LP. Whole exome sequencing data were analyzed together with EL. CV, LP and JM conceived
967 and designed experiments. CV, JM, PP and LP wrote the manuscript with input from all authors.

968

969 **Competing interest declaration.**

970 CV is founder and equity holder of One Biosciences.

971

972 **Additional Information.**

973 Correspondence and requests for materials should be addressed to celine.vallot@curie.fr.

974 **Extended legends**

975 **Extended Figure 1: *In vivo* models of chemotolerance in TNBC.** **a.** Graph of the relative tumor volumes
976 over time (days) for PDX_95 for eight mice treated with a first round of Capecitabine. **b.** (Left) UMAP
977 representation of scRNA-seq datasets, colored according to expression clusters. (Right) Histogram of
978 the frequency of each expression cluster in the indicated samples. **c.** Barplot displaying the top 5
979 pathways activated in persister cells. **d.** UMAP representation of scRNA-seq datasets, colored
980 according to log₂ expression signals for persister genes, log₂FC and adjusted p-values are indicated
981 above the graph. **e.** Histogram of the proportion of cells in the different cell cycle phases based on
982 expression of cell cycle in the scRNA-seq datasets. Proportions in each sample were compared to
983 untreated sample using Fisher exact test, p-values are indicated. **f.** Graph of the relative tumor
984 volumes over time (days) for PDX_39 for fourteen mice treated with Capecitabine and three untreated
985 mice. **g.** UMAP representation of scRNA-seq datasets, colored according to sample ID. **h.** UMAP
986 representation of scRNA-seq datasets, colored according to expression clusters. **i.** Barplot displaying
987 the top 5 pathways activated in persister cells. **j.** UMAP representation of scRNA-seq datasets, colored
988 according to log₂ expression signals for persister genes, log₂FC and adjusted p-values are indicated
989 above the graph. **k.** Histogram of the proportion of cells in the different cell cycle phases based on
990 expression of cell cycle in the scRNA-seq datasets. Proportions in persister sample were compared to
991 untreated sample using Fisher exact test, p-value are indicated. **l.** Graph of the relative tumor volumes
992 over time (days) for PDX_172 for four mice treated with Capecitabine and six untreated mice. **m.** UMAP
993 representation of scRNA-seq datasets, colored according to sample ID. **n.** UMAP representation of
994 scRNA-seq datasets, colored according to expression clusters. **o.** Barplot displaying the top 5 pathways
995 activated in persister cells. **p.** UMAP representation of scRNA-seq datasets, colored according to log₂
996 expression signals for persister genes, log₂FC and adjusted p-values are indicated above the graph. **q.**
997 Histogram of the proportion of cells in the different cell cycle phases based on expression of cell cycle
998 in the scRNA-seq datasets. Proportions in persister sample were compared to untreated sample using
999 Fisher exact test, p-value are indicated.

1000

1001 **Extended Figure 2: *In vitro* model of chemotolerance in TNBC.** All the experiments were performed
1002 in MDA-MB-468 cells. **a.** (Left) Histogram representing the percentage of the untreated population
1003 that tolerates 5-FU. (Right) Histogram representation of the percentage of persister cells that can
1004 proliferate actively under chemotherapy treatment. **b.** (Left) Histogram representation of the 5-FU
1005 IC₅₀ of untreated and chemoresistant populations. (Right) Histogram representation of the doubling
1006 time (in days) of MDA-MB-468 untreated, persister and resistant cells. (n=3, Mean ± sd, Anova
1007 Bonferroni's multiple comparisons test). **c.** (Left) UMAP representation of scRNA-seq datasets, colored
1008 according to expression cluster membership. (Right) Histogram representing the frequency of each
1009 cluster in the indicated samples. **d.** Barplot displaying the top 5 pathways activated in MM468 persister
1010 cells. **e.** Dot plot representing -log₁₀(q-value) of gene enrichment studies in PDX_95 versus MDA-MB-
1011 468 (MM468). Linear regression, associated correlation score and p-value are indicated. **f.** UMAP plot
1012 representing scRNA-seq datasets, points are colored according to log₂ gene expression signals for
1013 differentially expressed genes between persister cells from cluster R2 and untreated cells from cluster
1014 R10, log₂FC and adjusted p-values are indicated above the graph. **g.** Histogram of the proportion of
1015 cells in the different cell cycle phases based on expression of cell cycle in the scRNA-seq datasets. For
1016 each experiment, proportions in each sample were compared to the corresponding DMSO sample
1017 using Fisher exact test, p-value are indicated.

1018 **Extended Figure 3 : Tracing lineages in cancer cells under 5-FU treatment.** All the experiments were
1019 performed in MDA-MB-468 cells. **a.** Experimental design showing the infection of cells with a lentivirus
1020 produced from the plasmid barcode library (pRRL-CMV-GFP-BCv2Ascl). Cells were then treated with
1021 indicated drugs and scRNA-seq was performed. **b.** Histogram of the fraction of cells with detected
1022 lineage barcodes in scRNA-seq data for each sample. Numbers above bars are the number of cells with
1023 a lineage barcode. **c.** Clustering of lineage barcode frequencies - detected by bulk and scRNA-seq -
1024 using Pearson's correlation score. The size of the dots is proportional to the correlation score. **d.**
1025 Heatmap showing the frequency of individual lineage barcodes (rows), measured by bulk sequencing
1026 in different samples for experiment #3 (columns) and color coded as indicated. Normalized frequencies
1027 are clustered with hierarchical clustering, with Pearson's correlation score and Ward method. **e.**
1028 Dotplot of Shannon diversity indexes calculated from bulk datasets at each different time points under
1029 5-FU or DMSO treatment. **f.** Dot plot representation comparing normalized barcode frequency in
1030 simulated population versus initial population (D0), based on bulk data from experiment #3.
1031 Correlation scores and associated p-value are indicated. **g.** Dot plot comparing normalized barcode
1032 frequency in the DMSO-treated cells at D50 and D147 (Left) or 5-FU-treated cells at D77 and D147
1033 (Right) compared to the initial population at D0, based on bulk data from experiment #3. Pearson's
1034 correlation scores and associated p-value are indicated. **h.** UMAP representation of lineage-barcoded
1035 cells. Cells in orange are untreated cells having a lineage barcode found in at least one persister cell.
1036 Cells in red are matched persister cells. Cells in grey are cells having a lineage barcode which is not
1037 common between persister and untreated cells. **i.** Volcano plot of differential analysis between
1038 'persisting' and 'non persisting' untreated cells. **j.** Distribution of correlation scores between barcode
1039 frequencies of two replicates.

1040
1041 **Extended Figure 4: Genetic profiling of chemotolerant and resistant cells.** All the experiments were
1042 performed in MDA-MB-468 cells. **a.** (Left) Schematic view of the experimental design used to analyze
1043 the whole exome of untreated, persister and resistant cells. (Right) Graph of the coverage of bases per
1044 sample for MDA-MB-468 untreated, persister and resistant cells (n=4). **b.** Venn diagram of the number
1045 of total mutations identified in chemoresistant populations (n=4). **c.** Histogram representations of the
1046 proportions of mutations associated to each cosmic mutational signature in the untreated, persister
1047 and resistant populations (n=2 experiments). **d.** Histogram representing the cancer cell fraction for
1048 untreated, persister and resistant cells (n=2 experiments).

1049
1050 **Extended Figure 5: Detailed profiling of H3K27me3 landscapes in MDA-MB-468 cells under 5-FU**
1051 **treatment.** All the experiments were performed in MDA-MB-468 cells. **a.** Schematic view of the
1052 experimental design used to analyze chromatin landscapes of persister and resistant cells. All samples
1053 were analyzed at the single cell level except 5-FU-D77-#3 and 5-FU-D113-#4. Samples used for bulk
1054 ChIP analysis are indicated with a black asterix. **b.** Histogram representing the frequency of epigenomic
1055 clusters within each sample. **c.** Scatterplot representing for each differentially enriched H3K27me3
1056 peak, log₂ expression FC versus log₂ enrichment FC for the associated gene. Pearson's correlation
1057 scores and associated p-value are indicated. **d.** Cumulative scH3K27me3 profiles over *TGFB1* and
1058 *FOXQ1* in resistant cells. **e.** Venn diagram representation of the region-based differential analysis
1059 performed to extract regions depleted in H3K27me3 jointly in E2/E1 compared to E4 (scChIP seq
1060 dataset) **f.** Dot plot representing log₂ expression fold-change induced by 5-FU in resistant cells versus
1061 EZH2i-1 induced changes, and compared to the untreated population (D0). Correlation scores and
1062 associated p-value are indicated. **g.** Doughnut plot displaying the fraction of 5-FU persister genes

1063 potentially regulated or not by H3K27me3 and expressed upon EZH2i-1 treatment. After 21 days of co-
1064 treatment, 62% of 5-FU persister genes with H3K27me3 depletion are expressed (23/37 genes). **h.**
1065 Heatmap representation of the targets within persister genes of the three TF identified by ChEA3 that
1066 are part of the persister genes. Blue color stands for target genes while white means the gene is not a
1067 target. **i.** Mean rank of TF enrichment in persister genes obtained by ChEA3 for FOXQ1, FOSL1 and
1068 NF2F2 in persister genes (red line) compared to the average mean rank in 100 sets of randomly picked
1069 genes (green curve). The rank of indicated TFs is calculated using the ChEA3 score on a total of 1632
1070 TFs. **j.** Cumulative scH3K27me3 and scH3K4me3 enrichment profiles over *FOSL1* and *NR2F2* in
1071 untreated and persister MDA-MB-468 cells (D33 - H3K27me3 and D60 - H3K4me3). Log2FC for
1072 H3K27me3 and scRNA are indicated for the comparison of persister and untreated populations.

1073

1074 **Extended Figure 6: Analysis of H3K4me3 and H3K4me3/H3K27me3 enrichment in untreated cells. a.**

1075 Heatmap representation of single-cell H3K4me3 enrichment at H3K27me3-regulated persister genes,
1076 non-expressed protein coding genes and housekeeping genes in untreated cells (D0) and persister cells
1077 (D60). **b.** Violin plots representing the distribution of percentage of cells with H3K4me3 signal across
1078 H3K27me3-regulated persister genes, non-expressed protein coding genes and housekeeping genes.
1079 **c.** Violin plot representing the distribution of percentage of H3K27me3-regulated persister genes with
1080 H3K4me3 signal in untreated cells (D0) and persister cells (D60). One-tailed Mann-Whitney test is used
1081 for the comparison between the two conditions **d.** (Up) Cumulative scH3K4me3 profiles over the
1082 *TGFB1* locus between untreated cells (D0) and persister cells (D60). (Down) H3K27me3->H3K4me3 and
1083 H3K27me3->IgG sequential ChIP-seq profiles of *TGFB1* in the untreated population. Comparative
1084 tracks show enrichment over IgG control with associated odd ratio and adjusted p-value. **e.** Dot plot
1085 of the false positive peaks (H3K27me3 -> IgG) by the number of bivalent peaks (H3K27me3 ->
1086 H3K4me3) at various p-values and peak to region ratios. Used thresholds are indicated in red. **f.** Dot
1087 plot of the false positive peaks (H3K4me3 -> IgG) by the number of bivalent peaks (H3K4me3 ->
1088 H3K27me3) at various p-values and peak to region ratios. Used thresholds are indicated in red. **g.** Venn
1089 diagram of MDA-MB-468 bivalent genes found by sequential ChIP-seq in the H3K4me3-H3k27me3
1090 way, the H3K27me3-H3K4me3 way. The enrichment of the intersection between the two ways is
1091 tested using a Fisher test. **h.** Dotplot of the $-\log_{10}(\text{adjusted p.value})$ of bivalent pathways (as in **g.**) in
1092 the H3K4me3-H3K27me3 and the H3K27me3-H3K4me3 ways. **i.** (Left) Barplot displaying the top 5
1093 pathways enriched in H3K27me3/H3K4me3 bivalent genes identified in untreated cells in MDA-MB-
1094 468, BT20 and HCC38. X-axis corresponds to $-\log_{10}$ adjusted p-values. (Right) Venn diagram displaying
1095 the intersection of the pathways enriched in H3K27me3/H3K4me3 bivalent genes identified in the
1096 untreated cells. P.value corresponds to the significance of the overlap calculated with Exact Test of
1097 Multi-set Intersections

1098

1099 **Extended Figure 7 : Investigating chromatin bivalency *in vivo* and in human tumors. a-c.**

1100 displaying the top pathways enriched in H3K27me3/H3K4me3 bivalent genes identified in the human
1101 tumor sample from Patient_95 or in the corresponding PDX model PDX_95 (**a**), Patient_39/PDX_39 (**b**)
1102 or Patient_172/PDX_172 (**c**). **d-f.** Venn diagram displaying the intersection of the pathways enriched
1103 in H3K27me3/H3K4me3 bivalent genes identified in the untreated cells in the human sample from
1104 Patient_95 and its corresponding PDX model (**d**), Patient_39/PDX_39 (**e**) or Patient_172/PDX_172 (**f**).
1105 **g.** H3K27me3 and H3K4me3 chromatin profiles of *KLF4* in 8 human tumor samples. The percentage of
1106 tumoral cells are indicated for each sample. **h.** Dotplot showing the top pathways enriched in genes
1107 displaying a dual H3K27me3 and H3K4me3 enrichment in the human tumor samples from MSigDB

1108 c2_curated KEGG and c5_GO annotations. Color of the dot corresponds to adjusted p-values and the
1109 size of the dot corresponds to the gene ratio.

1110

1111 **Extended Figure 8 : Modulation of chemotolerance to 5-FU with EZH2i *in vitro*.** **a.** Histogram
1112 representing the number of MDA-MB-468 cells pretreated or not with EZH2i-1 (UNC1999) and after
1113 treatment over 21 days with 5-FU (n=3, Mean \pm sd, p-value correspond to Anova Bonferroni's multiple
1114 comparisons test). **b.** Representative images of immunoblotting of MDA-MB-468 cells treated for 21
1115 days with DMSO or indicated EZH2 inhibitors. EZH2, Tubulin and H3K27me3 are represented. **c.**
1116 Clustering of samples according to lineage barcode frequencies, detected by bulk analysis, using
1117 Pearson's correlation score. MDA-MB-468 were co-treated with DMSO or 5-FU and EZH2i-1 for 21 days
1118 (Up) or pretreated with indicated EZH2i ("EZH2i-1": UNC1999, inactive EZH2i-1: "UNC2400" or EZH2i-
1119 2: "GSK126") for 10 days and then co-treated with DMSO or 5-FU for 21 days (Down). **d/f.** Histogram
1120 representing the number of BT20 (**d**) or HCC38 (**f**) cells pretreated with EZH2i inhibitors and after
1121 treatment over 21 days with 5-FU (n=3, Mean \pm sd, p-value correspond to Anova Bonferroni's multiple
1122 comparisons test). **e/g.** Representative images of immunoblotting of BT20 (**e**) or HCC38 (**g**) cells
1123 treated for 21 days with DMSO or indicated EZH2 inhibitors. EZH2, Tubulin and H3K27me3 are
1124 represented. For gel source data, see Supplementary Figure 1

1125

1126 **Extended Figure 9 : Modulation of chemotolerance to 5-FU with KDM6i *in vitro*.** **a.** Projection of
1127 MDA-MB-468 cells treated with GSK-J4 onto the UMAP scH3K27me3 space. **b.** Histogram representing
1128 the number of MDA-MB-468 cells after treatment over 21 days with DMSO or 5-FU alone or in
1129 combination with KDM6i (GSK-J4) (n=3, Mean \pm sd, p-value correspond to Anova Bonferroni's multiple
1130 comparisons test). **c.** Colony-forming assay of MDA-MB-468 treated over 60 days with DMSO or 5-FU
1131 alone or co-treated with GSK-J4 inactive isomer (GSK-J5). **d.** Histogram representing the number of
1132 MDA-MB-468 cells after treatment over 50 days with DMSO or 5-FU alone or co-treatment with
1133 KDM6A/Bi (GSKJ-4) or its inactive isomer (GSK-J5) added at the indicated days (n=3, Mean \pm sd, p-value
1134 correspond to Anova Bonferroni's multiple comparisons test). **e/f.** Colony-forming assay of BT20 (**e**) or
1135 HCC38 (**f**) cells co-treated with DMSO or 5-FU and indicated concentrations of KDM6i (GSKJ-4) or its
1136 inactive isomer (GSK-J5). The data correspond to 1 of 3 biological replicates.

1137

1138 **Extended Table 1: Summary of all models and technologies.** Details of the models, samples,
1139 technologies used as well as the output of each experiment.

1140

1141 **Extended Table 2: Primer sequences.** Primers for lineage barcode detection, scChIP-seq beads
1142 sequence and chromatin indexing index sequences are indicated.

1143

1144 **Extended Table 3: Summary of single-cell ChIP-seq count tables.** Analyzed samples and corresponding
1145 cell numbers over 1,000 reads are indicated.

1146

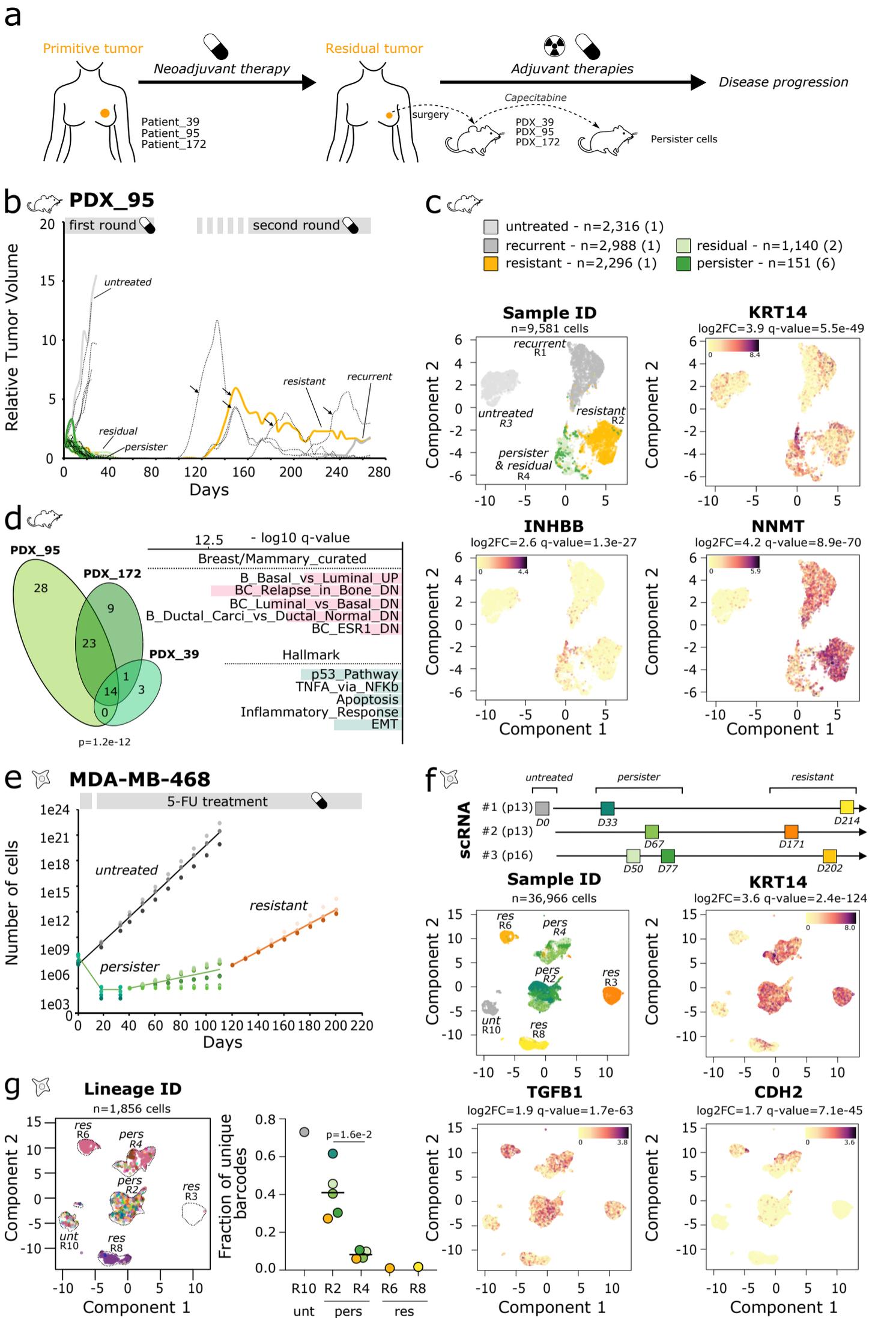


Figure 1: Identification of a pool of basal persister cells in TNBC in vivo and in vitro. **a.** Schematic representation of the standard of care for TNBC patients and the generation of patient-derived xenograft models. **b.** Graph of the relative tumor volumes (RTV) over time (days). Colored growth curves correspond to tumors which have been further studied by scRNA-seq. Black arrows indicate the start of the second round of Capecitabine treatment for the corresponding mice. **c.** (Up) Phenotypes and cell numbers are indicated, with the number of mice used to collect samples in brackets. (Down) UMAP representation of PDX scRNA-seq datasets, colored according to sample of origin (first panel - cluster ID are indicated) or log₂ gene expression signal for differentially expressed genes between persister cells and untreated tumor cells (remaining panels), log₂FC and adjusted p-values are indicated above the graph. **d.** (Left) Venn diagram displaying the intersection of pathways activated in persister cells from the 3 PDX models, among MSigDB c2_curated Breast/Mammary and c7_Hallmark pathways. P-value associated with the intersection is indicated below (exact test of multi-set intersections) (Right) Barplot displaying the top 5 pathways - for each category - activated in persister cells. x-axis corresponds to -log₁₀ adjusted p-values for the model PDX_95. **e.** Graph representation of the cell proliferation of triple negative breast cancer cell line MDA-MB-468 (MM468) treated with 5-FU (green for persister cells, and orange lines for resistant cells) or with DMSO (untreated - grey lines). **f.** (Up) Schematic view of the experimental design. Experiment number and corresponding passage of cells at D0 are indicated. (Down) UMAP representation of MDA-MB-468 cells scRNA-seq datasets, colored according to the sample of origin (first panel - cluster ID are indicated) or log₂ gene expression signal for differentially expressed genes between persister cells (cluster R2) and untreated cells (cluster R10, *KRT14* and *TGFBI* panels) or for a differentially expressed gene between the two persister clusters, i.e. clusters R4 vs R2 (*CDH2* panel). Untreated population (in grey) corresponds to DMSO-D0-#1. **g.** (Left) UMAP representation of scRNA-seq as in 1f, restricted to cells with detected lineage barcode. Cells are colored according to lineage barcode and cluster membership is indicated (Extended Fig. 2c). R1, R2 correspond to RNA-inferred clusters. (Right) Scatter plot of the lineage barcode diversity detected in the scRNA-seq data across clusters and samples. Colors correspond to sample ID as in 1f. (Means are indicated for persister clusters R2 & R4. two-tailed Mann-Whitney test).

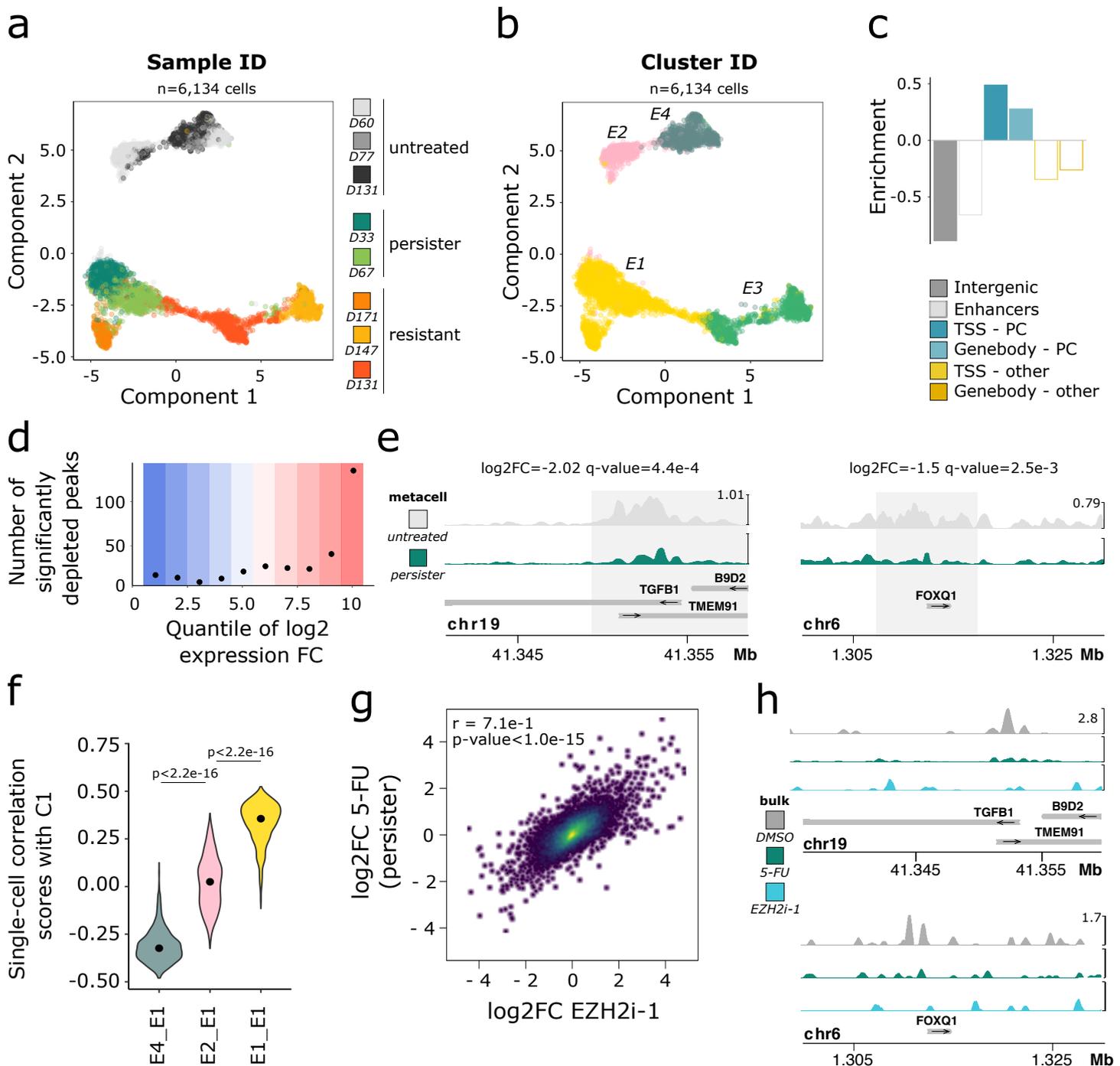


Figure 2: H3K27me3 represents the persister expression program prior to chemotherapy exposure. All experiments were performed in MDA-MB-468 cells. **a.** UMAP representation of scChIP-seq H3K27me3 datasets, cells are colored according to the sample of origin. Persister and resistant samples correspond to 5-FU-treated cells, days of treatment are indicated. **b.** Same as in **a.** with cells colored according to cluster membership. E1, E2 correspond to epigenomic-based clusters. **c.** Enrichment of H3K27me3 significantly depleted peaks in persister cells compared to all peaks across various gene annotation categories (see Methods). Full bars indicate adjusted p-value < 1.0e-2. Empty bars indicate non-significant adjusted p-values. "PC" indicates protein coding genes. **d.** Repartition of H3K27me3 depleted peaks within log₂ expression fold-changes quantiles from scRNA-seq experiments. **e.** Cumulative scH3K27me3 profiles over *TGFB1* and *FOXQ1* in untreated and persister cells (D33). Log₂FC and adjusted p-value correspond to differential analysis of cells from cluster E1 versus cells from clusters E2 + E4. **f.** Violin plot representation of the cell-to-cell inter-correlation scores between cells from clusters E1, E2 or E4 and cells from E1. Pearson's correlation scores were compared using a two-tailed Mann-Whitney test, p-value are indicated above plots. **g.** Dot plot representing log₂ expression fold-change induced by 5-FU or EZH2i-1 at D33 versus D0. Pearson's correlation scores and associated p-value are indicated. **h.** Bulk H3K27me3 chromatin profiles for *TGFB1* and *FOXQ1* in cells treated with DMSO, 5-FU or EZH2i-1 at D33.

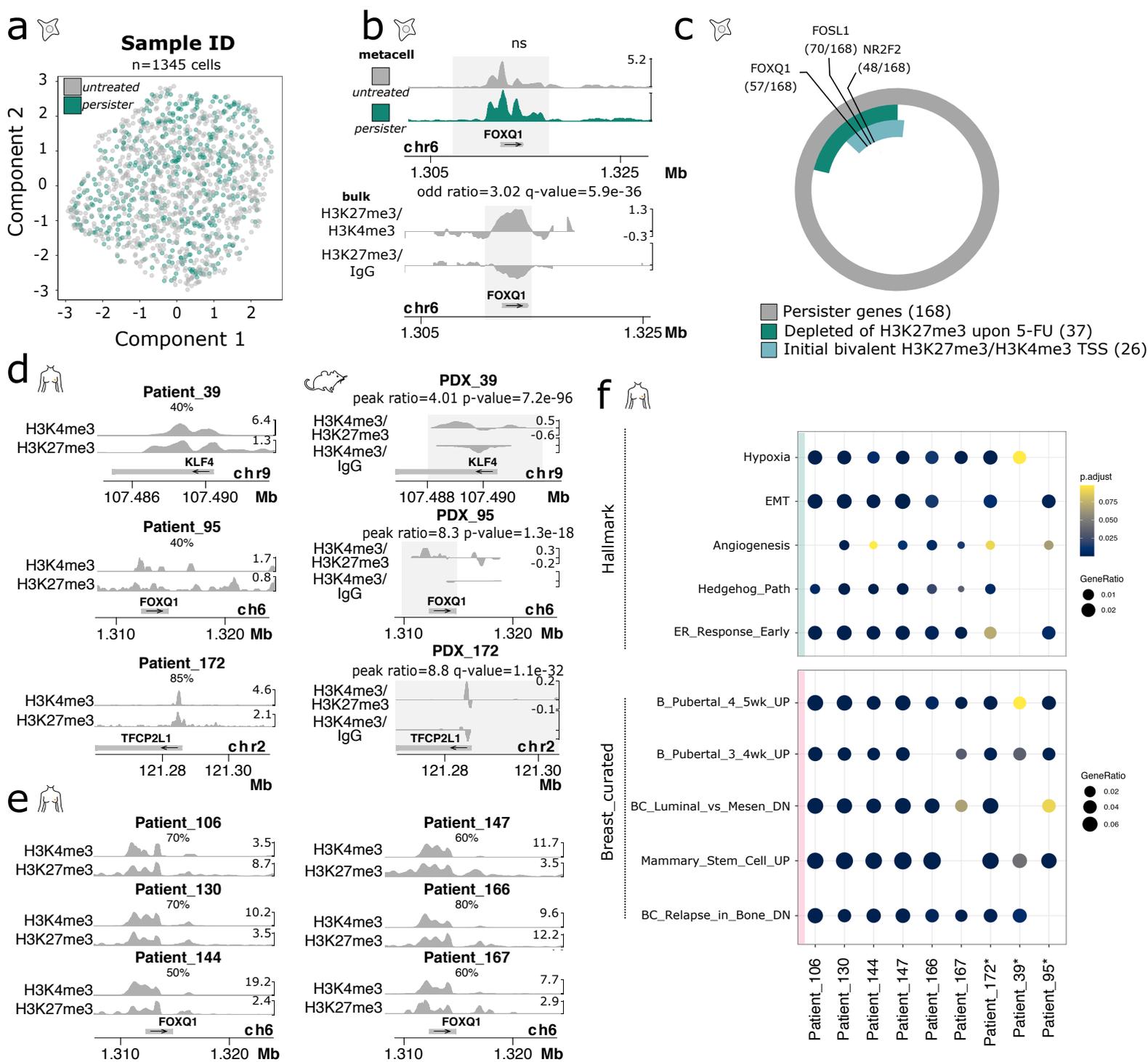


Figure 3: Epigenomes of untreated cells are primed with co-accumulation of H3K27me3 and H3K4me3. **a.** UMAP representation of scChIP-seq H3K4me3 datasets, cells are colored according to the sample of origin - untreated cells (D0) and persister cells (D60). **b.** (Up) Cumulative scH3K4me3 enrichment profiles over **FOXQ1** in untreated cells (D0) and persister cells (D60). 'ns' stands for not significant after differential testing comparing untreated and persister cells. (Down) H3K27me3->H3K4me3 and H3K27me3->IgG sequential ChIP-seq profiles of **FOXQ1** in the untreated population. Comparative tracks show enrichment over IgG control with associated odd ratio and adjusted p-value. **c.** Doughnut chart displaying the fraction of persister genes (n=168) with H3K27me3 loss upon 5-FU treatment or with bivalent chromatin at TSS in untreated cells. Candidate master TFs - among persister genes - are indicated with the number of persister genes potentially regulated by the corresponding TF in parentheses. **d.** (Left) H3K27me3 and H3K4me3 chromatin profiles of human tumor samples for candidate master TF (Patient_39, Patient_95 and Patient_172). The percentage of tumoral cells are indicated for each sample. (Right) H3K4me3->H3K27me3 and H3K4me3->IgG sequential ChIP-seq profiles of untreated population of the three derived PDX_models (PDX_39, PDX_95 and PDX_172). Comparative tracks show enrichment over IgG control with associated odd ratio and adjusted p-value. **e.** H3K27me3 and H3K4me3 chromatin profiles of 6 additional human tumor samples. The percentage of tumoral cells are indicated for each sample. **f.** Dotplot showing the top pathways enriched in genes displaying a dual H3K27me3 and H3K4me3 enrichment in human tumor samples. Color of the dot corresponds to adjusted p-values and the size of the dot corresponds to the gene ratio, i.e. the fraction of bivalent genes belonging to this pathway. Stars indicate human tumor samples used to establish our PDX models

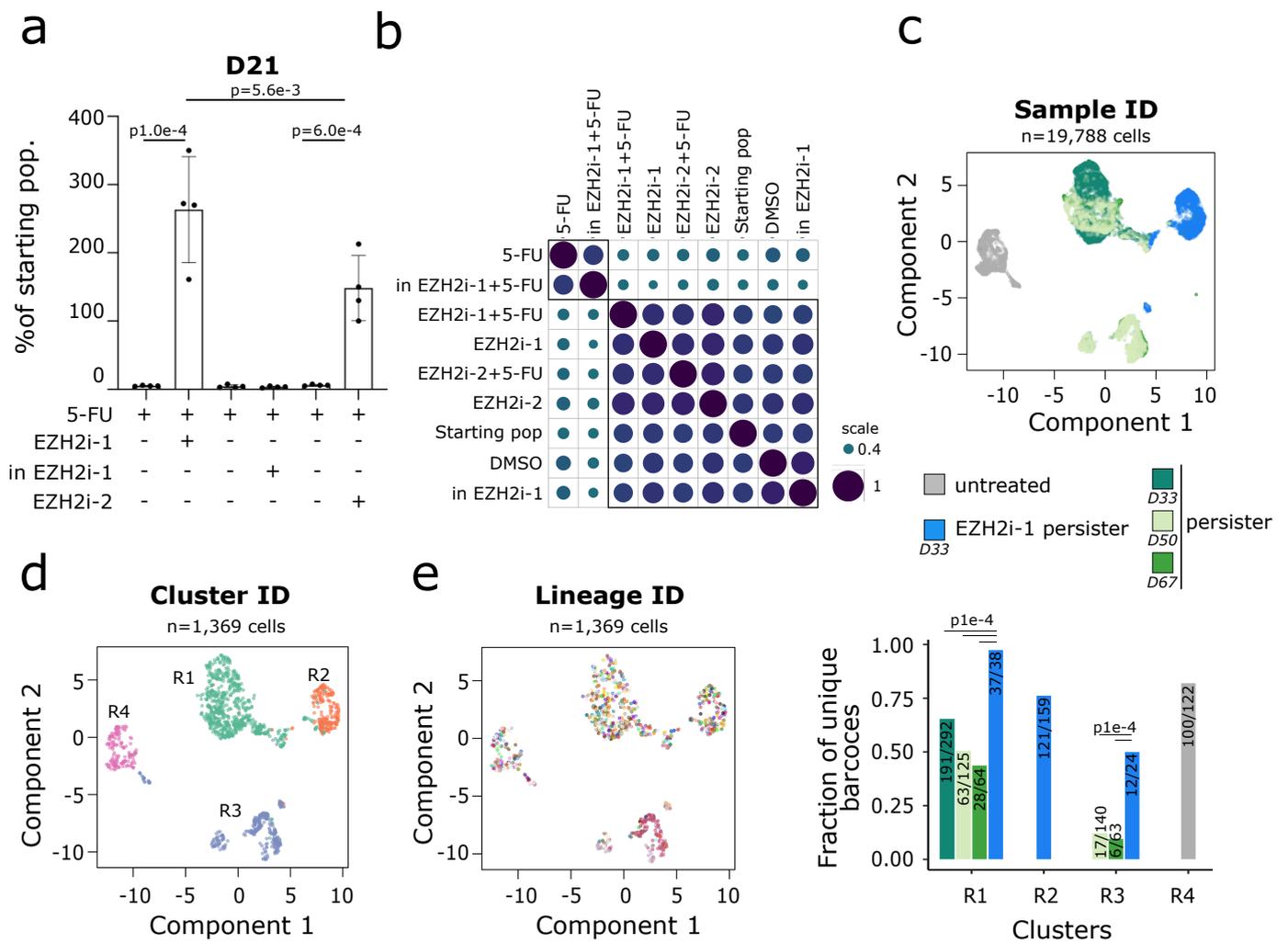


Figure 4: EZH2 inhibition rescues cell fate bias upon chemotherapy exposure. All the experiments were performed in MDA-MB-468 cells. **a.** Histogram representing the number of cells after treatment with 5-FU alone or 5-FU and EZH2i over 21 days, relative to the number of cells at D0. Cells were pre-treated with EZH2i-1, inactive EZH2i-1 or EZH2i-2 for 10 days prior to chemotherapy treatment. (n=3, Mean \pm sd, Anova multiple comparisons test with Bonferroni's correction). **b.** Clustering of samples according to lineage barcode frequencies, detected by bulk analysis, using Spearman correlation score. **c.** UMAP representation of scRNA-seq datasets, colored according to the sample ID of origin. Cells were treated with DMSO (untreated) or with 5-FU alone (persister) or with 5-FU and EZH2i-1 (EZH2i-1 persister). **d.** UMAP representation of scRNA-seq datasets, barcoded cells were selected and colored according to hierarchical clusters **e.** (Left) UMAP representation of scRNA-seq datasets, barcoded cells were selected and colored according to lineage barcode. (Right) Histogram of the lineage barcodes diversity detected in the scRNA-seq data within hierarchical clusters, and across samples. Colors correspond to sample ID as in 4c. Ratio of unique barcodes/total barcodes and p-value are indicated (two-tailed Fisher test).

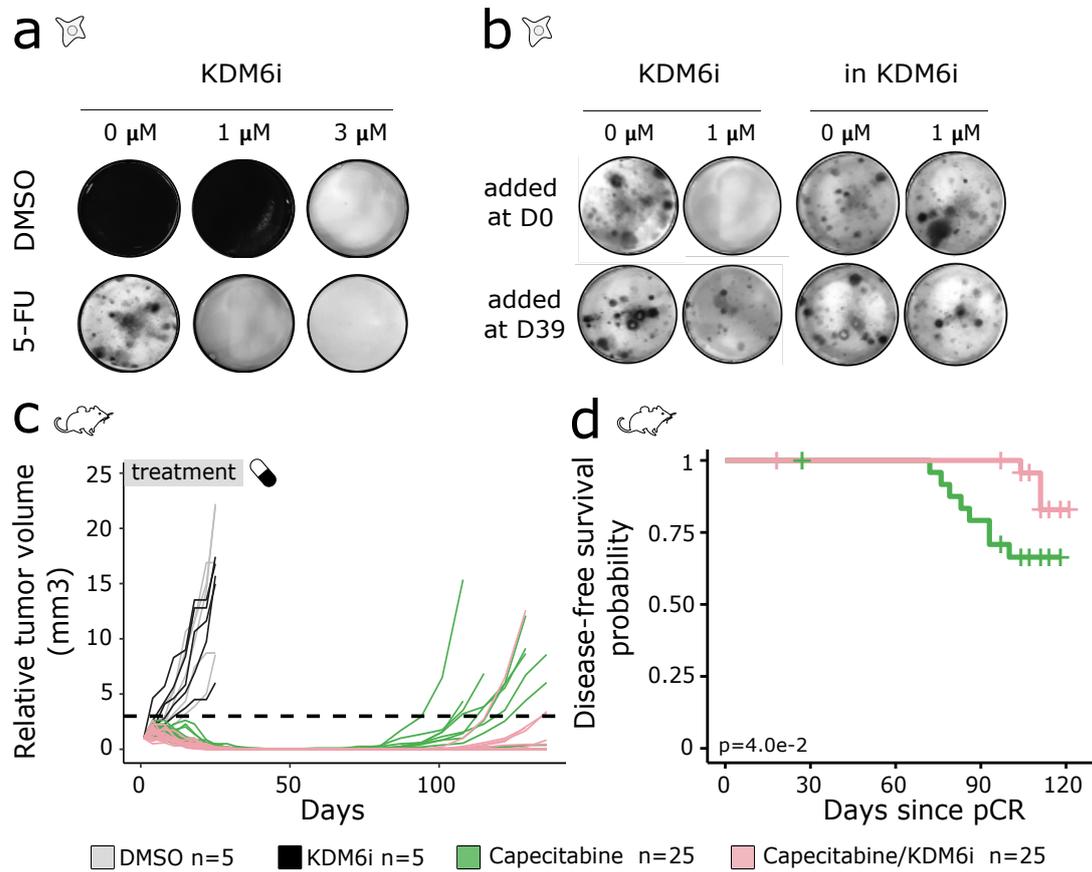
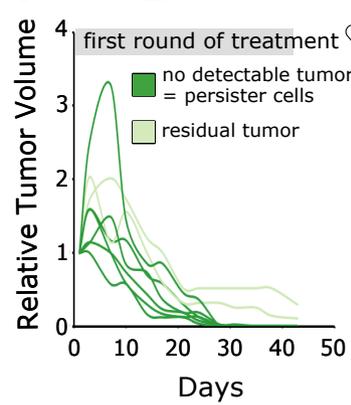


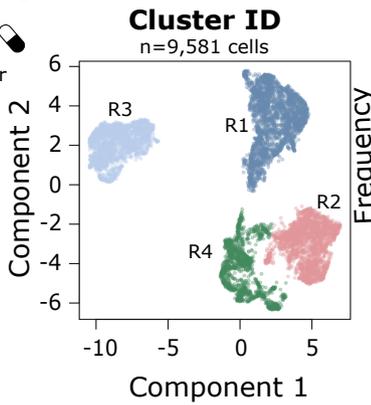
Figure 5: Simultaneous KDM6i and chemotherapy treatment inhibits chemotolerance *in vitro* and delays recurrence *in vivo*. **a.** Colony forming assay at day 60 for 5-FU treated MDA-MB-468 cells in combination with DMSO or indicated concentrations of the KDM6i GSK-J4. **b.** Colony forming assay at day 60 for 5-FU treated MDA-MB-468 cells in combination or not with 1 μM of GSK-J4 or its inactive isomer GSK-J5, either simultaneously - added at D0 - or added at day 39 of chemotherapy treatment. **c.** Relative tumor volumes for n=60 mice treated with either DMSO, GSK-J4, Capecitabine or a combination of Capecitabine and GSK-J4. Dashed line indicates the threshold to detect recurrent tumors, RTV=3. **d.** Kaplan-Meier plot of the overall disease-free survival probability since pathologic complete response pCR to initial treatment (tumor volume<20mm³), number of mice treated and p-value are indicated (log-rank test).

Extended Figure 1

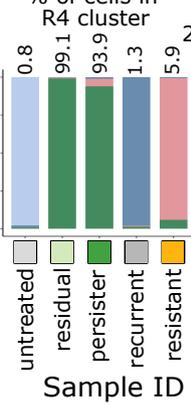
a PDX_95



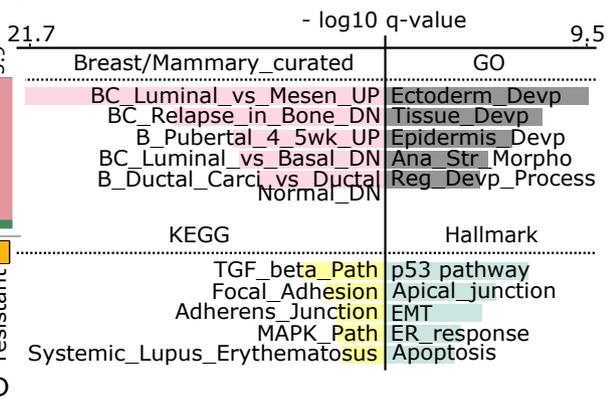
b



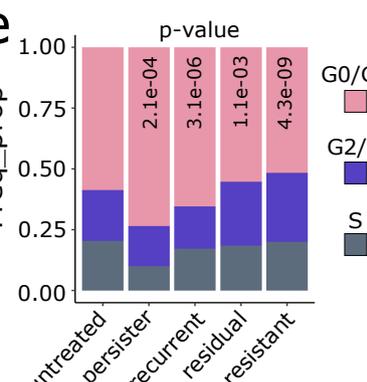
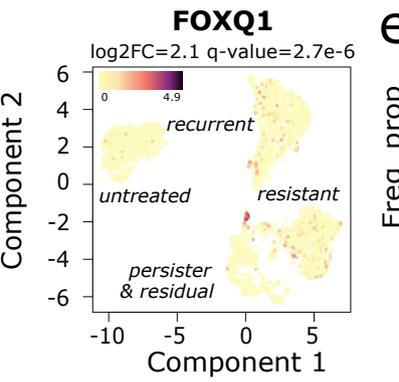
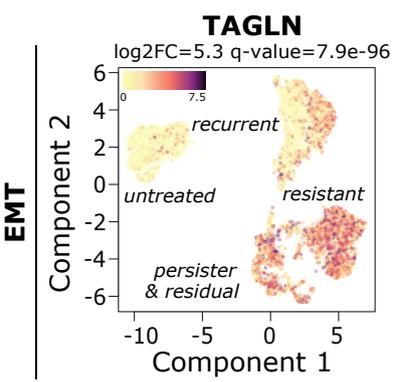
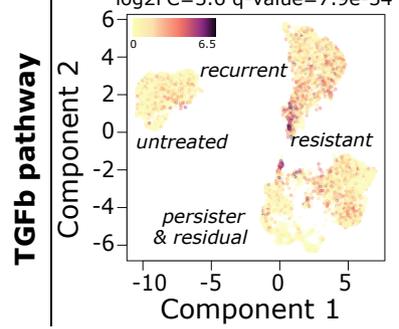
c



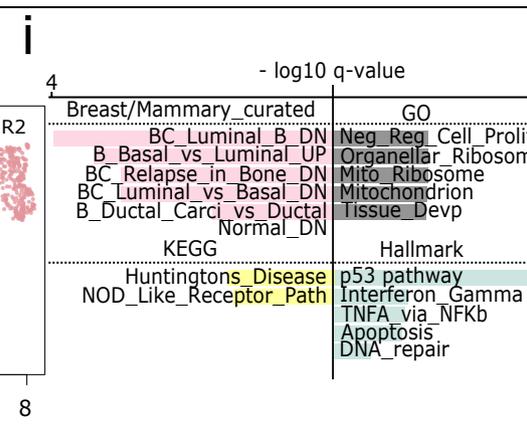
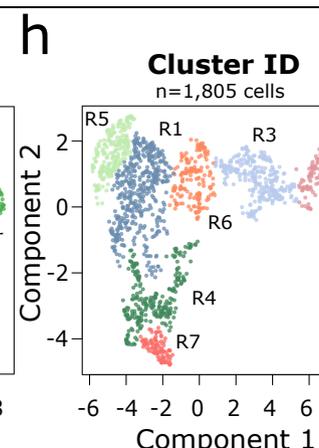
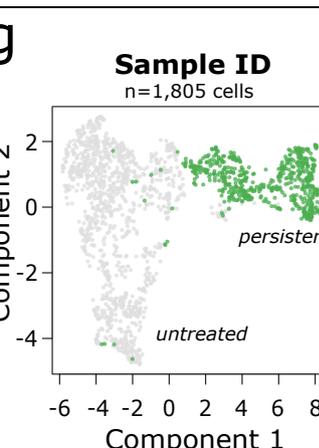
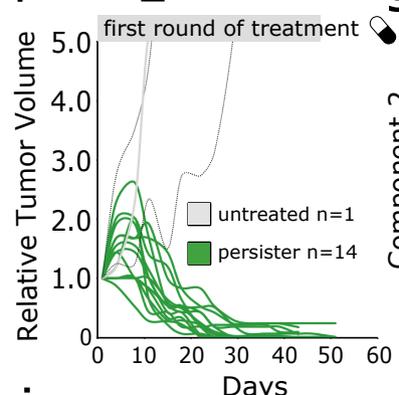
c



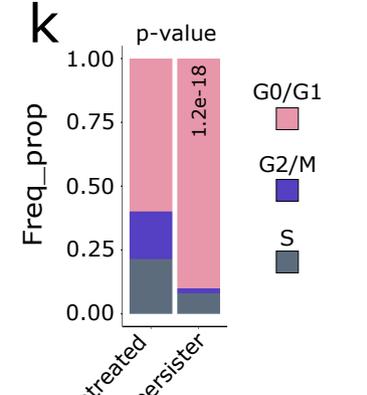
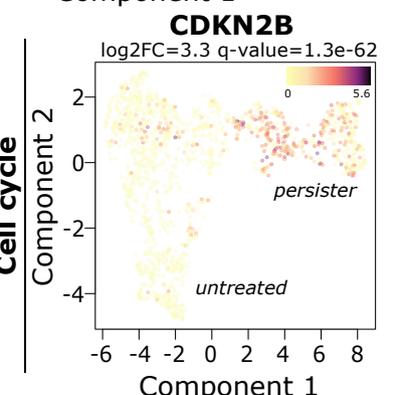
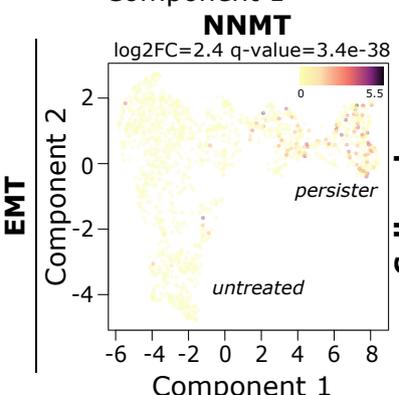
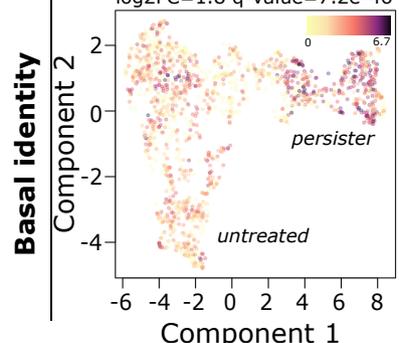
d



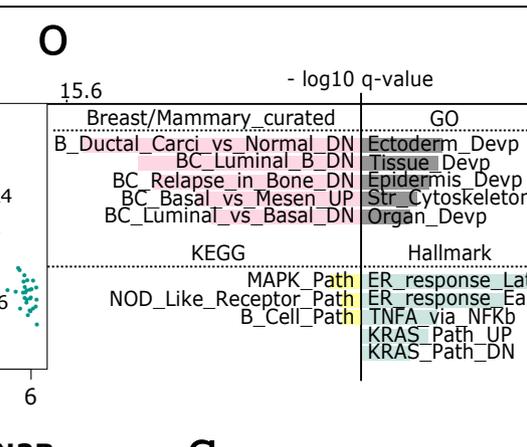
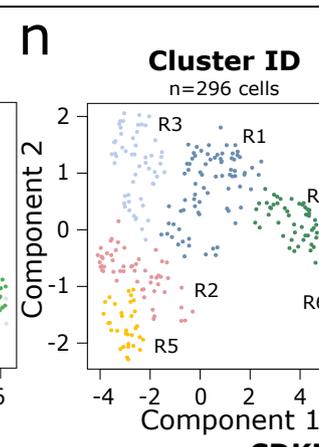
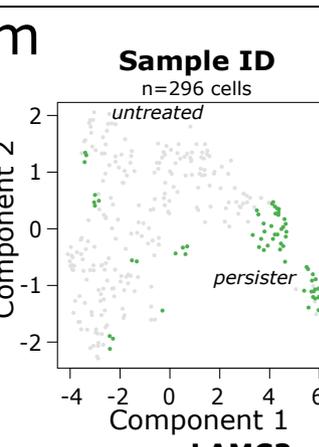
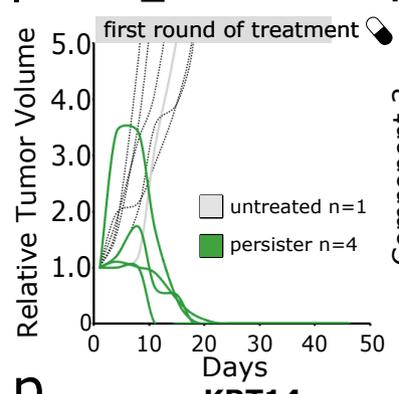
f



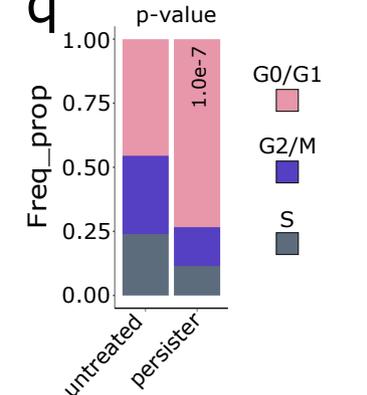
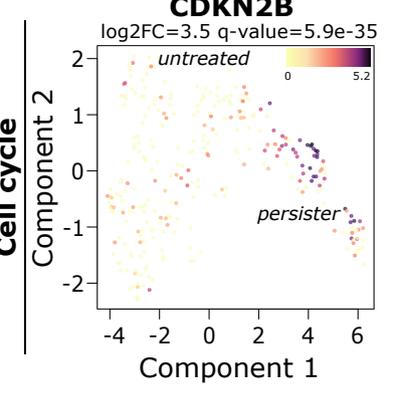
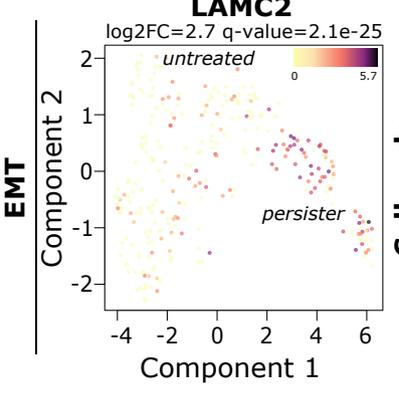
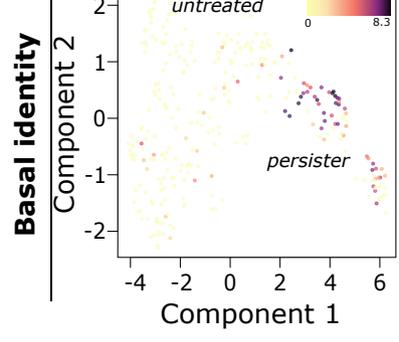
g



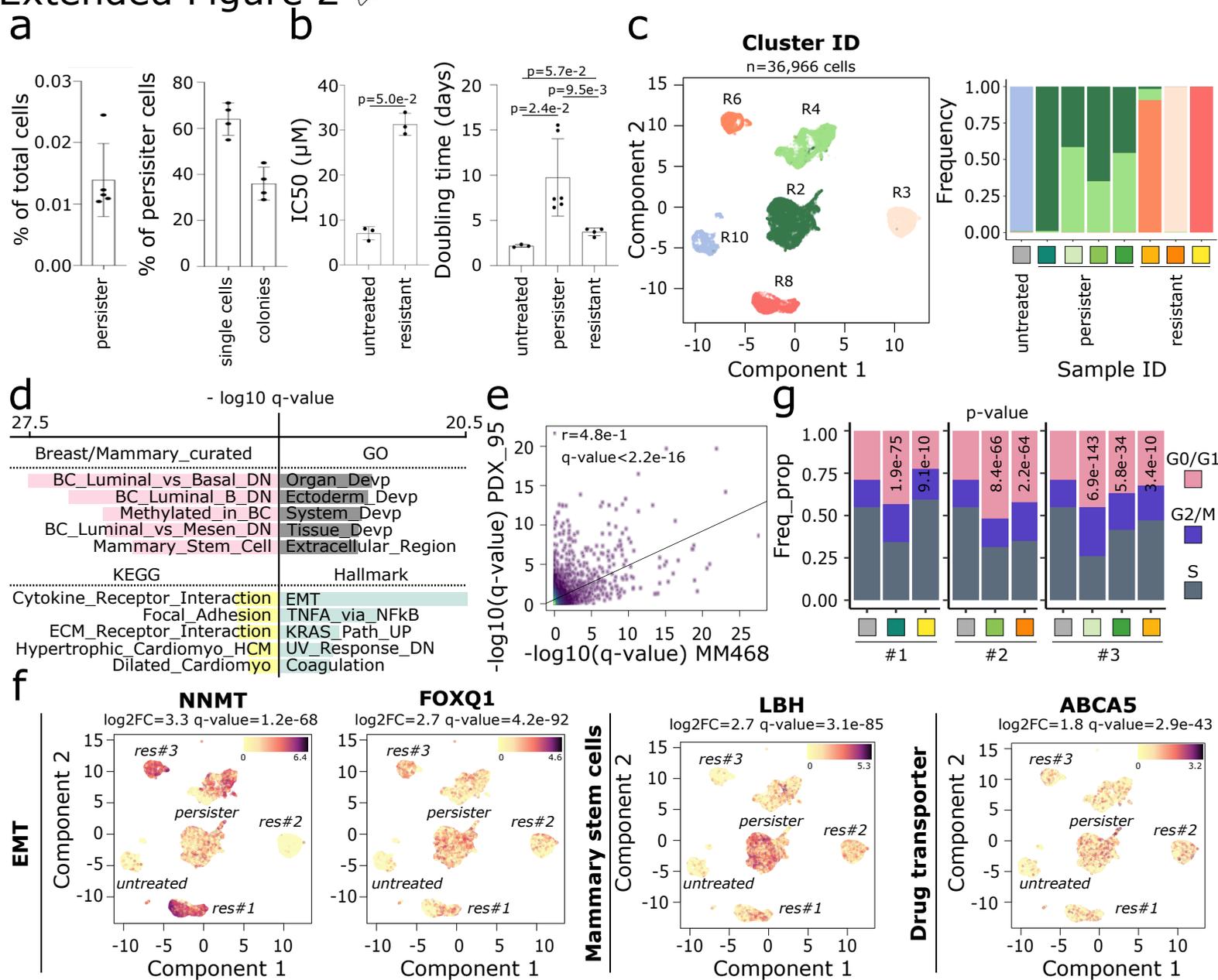
l



o

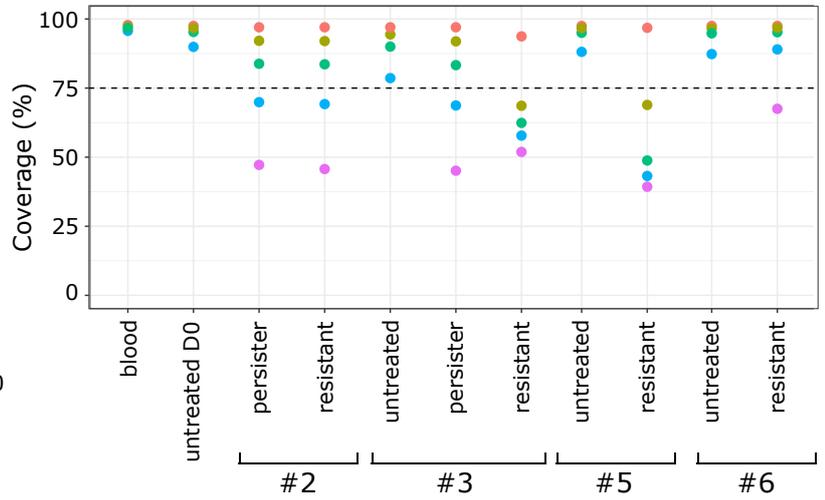
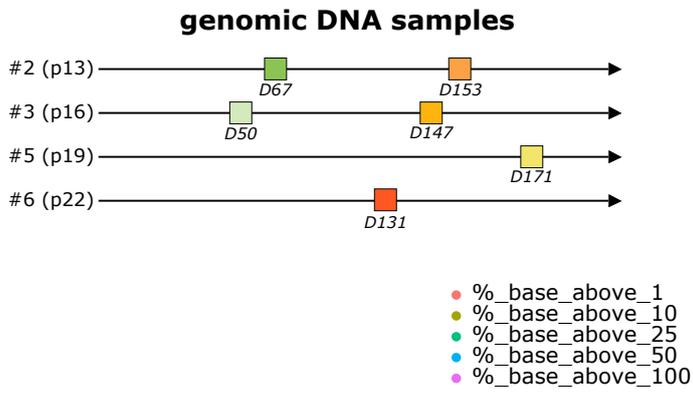


Extended Figure 2

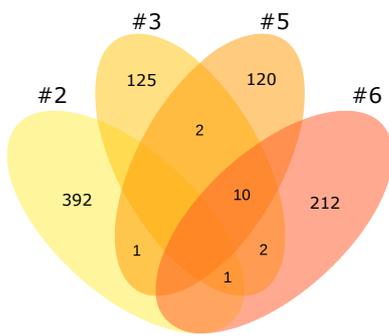


Extended Figure 4

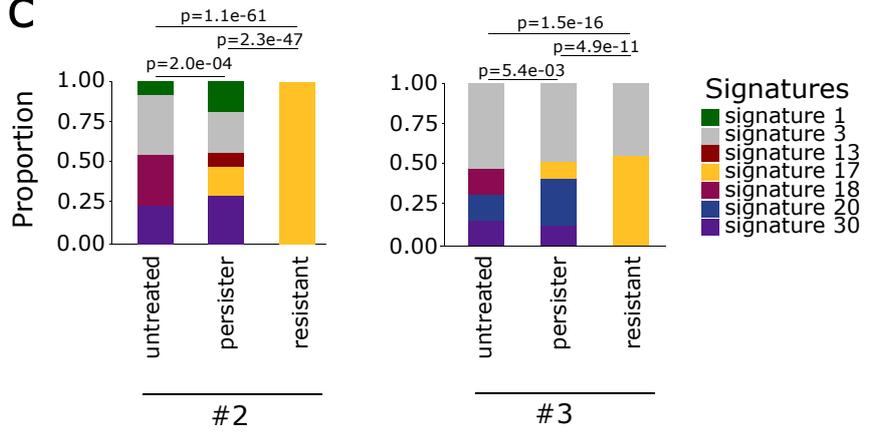
a



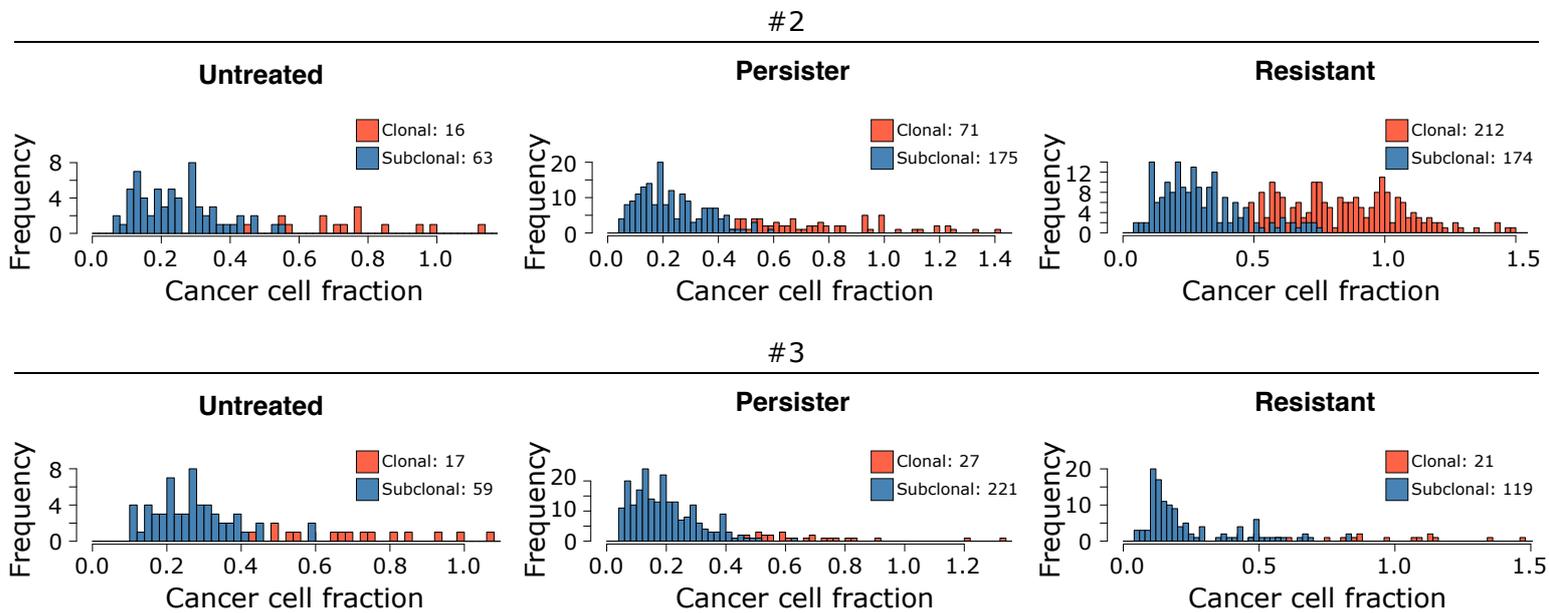
b



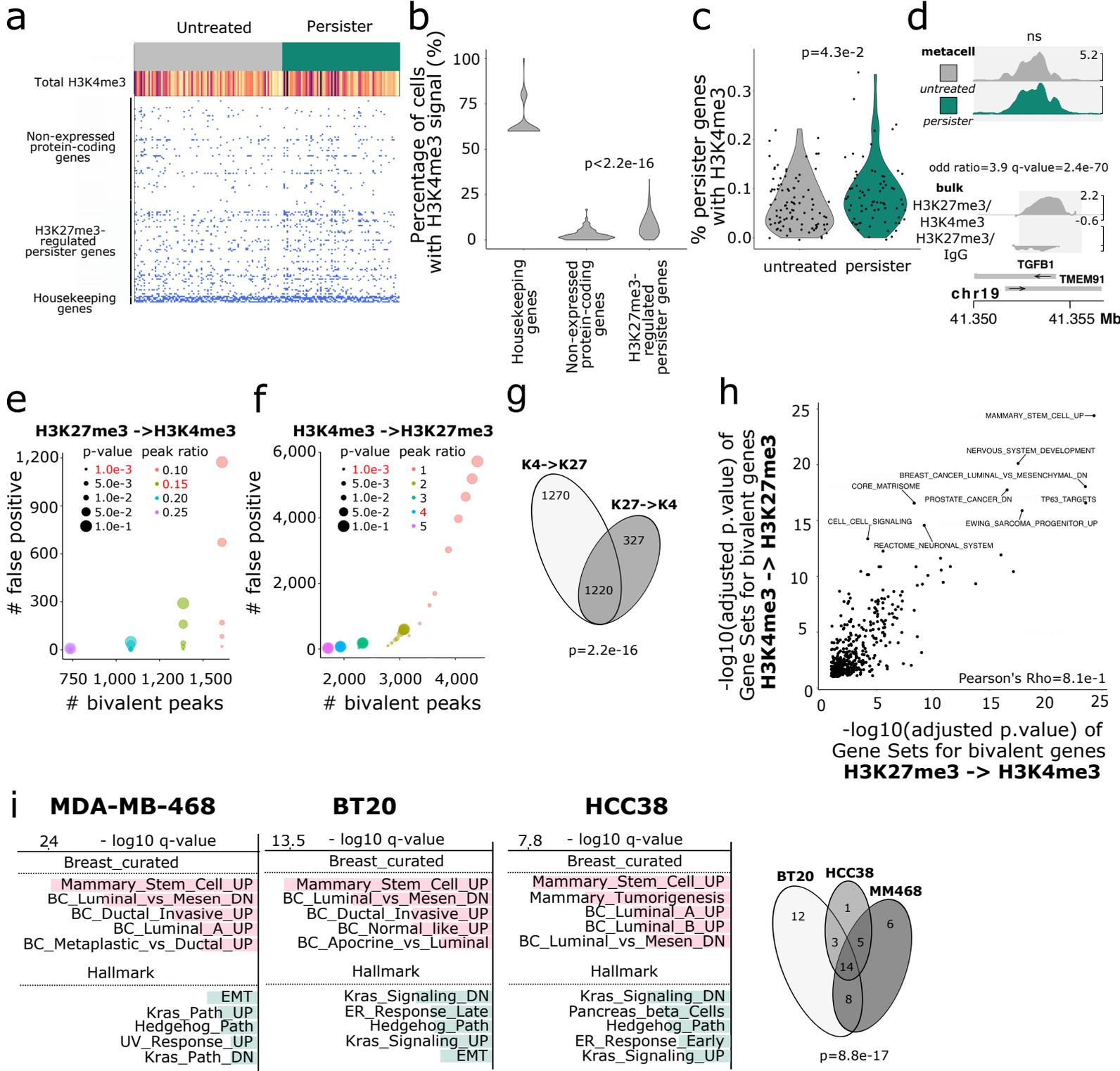
c



d

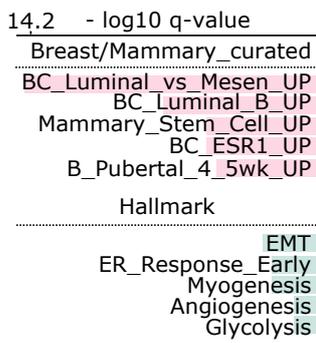


Extended Figure 6

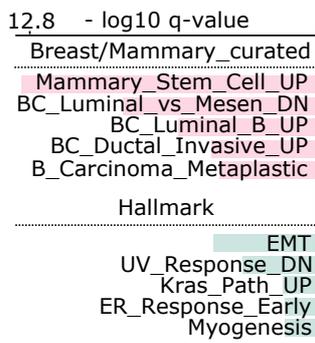


Extended Figure 7

a Patient_95



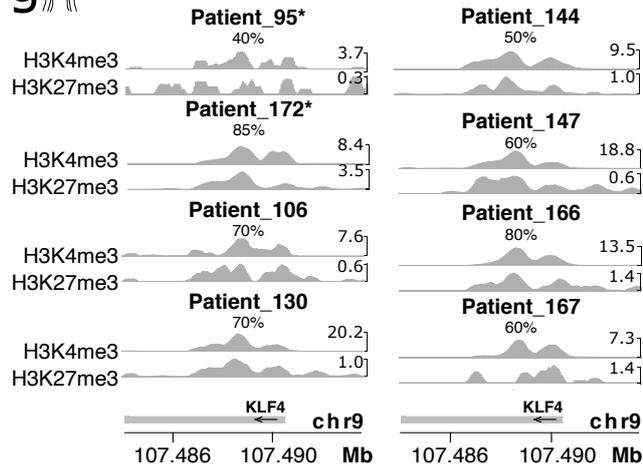
PDX_95



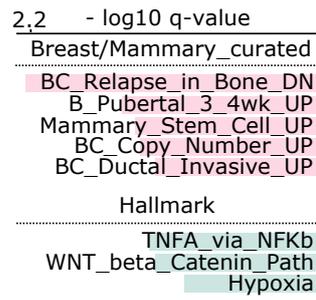
d



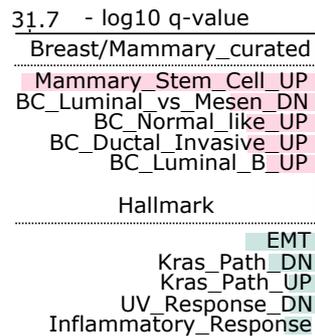
g



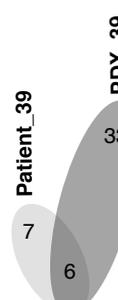
b Patient_39



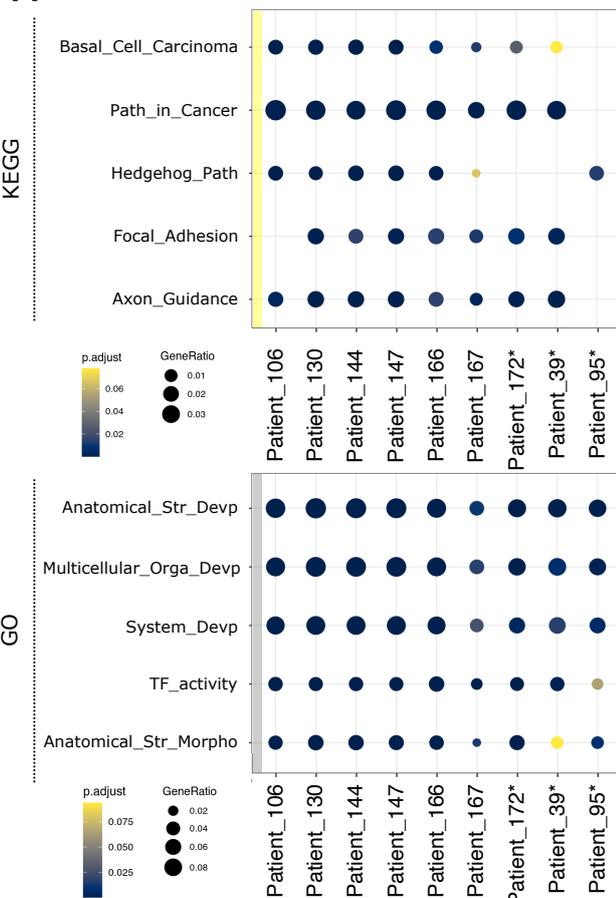
PDX_39



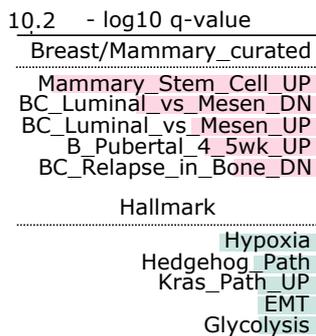
e



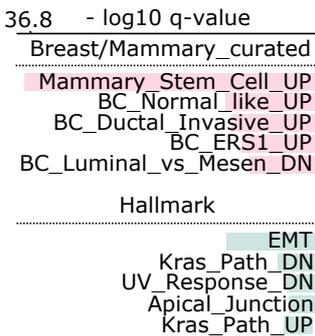
h



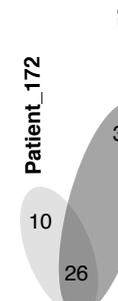
c Patient_172



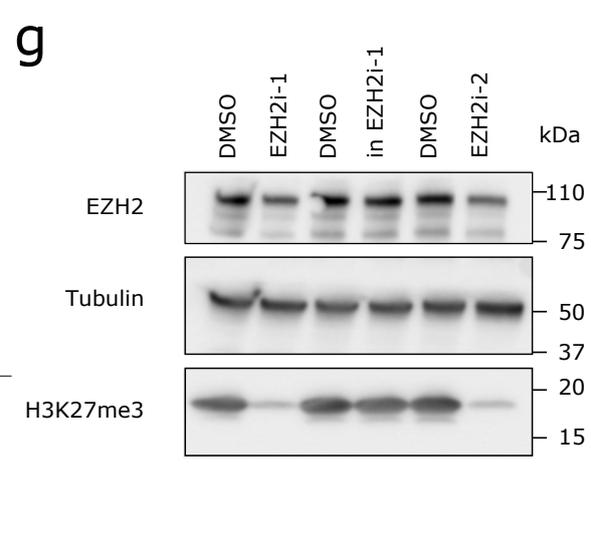
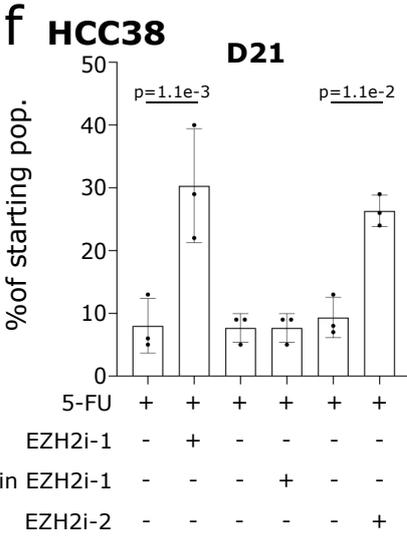
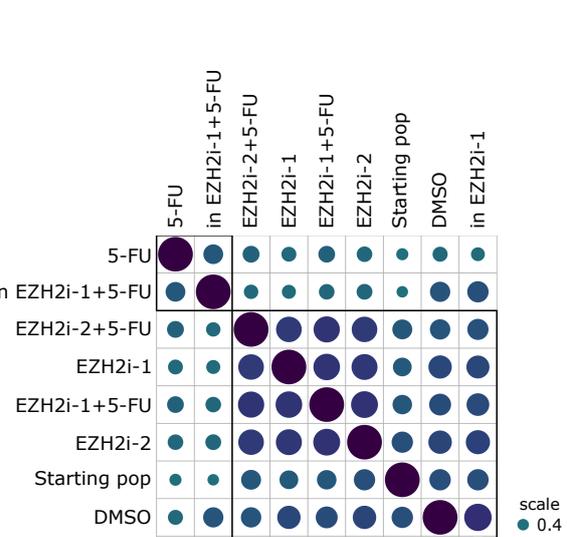
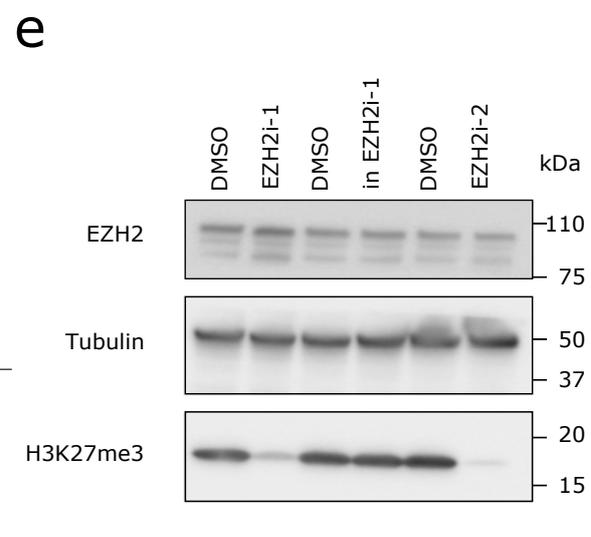
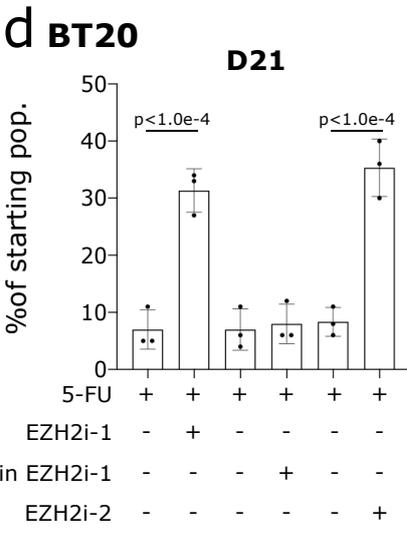
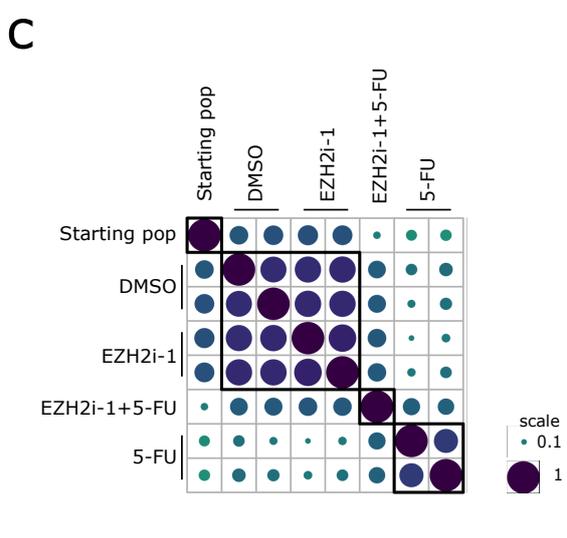
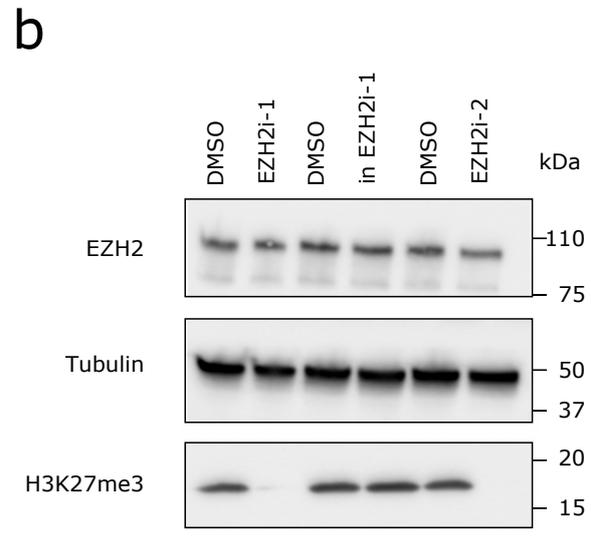
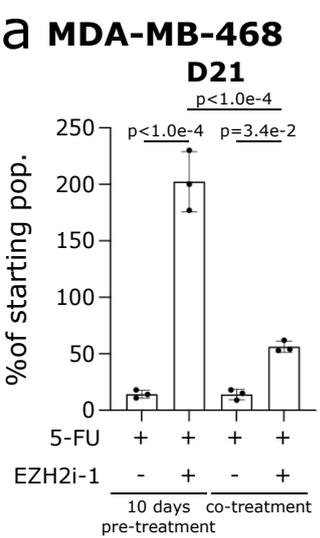
PDX_172



f

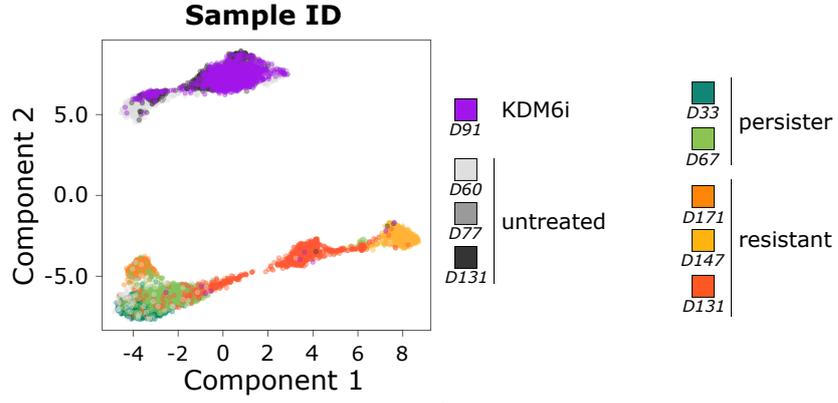


Extended Figure 8

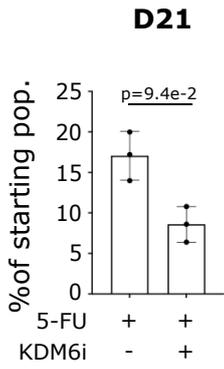


Extended Figure 9

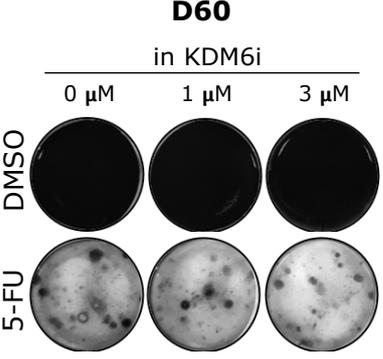
a MDA-MB-468



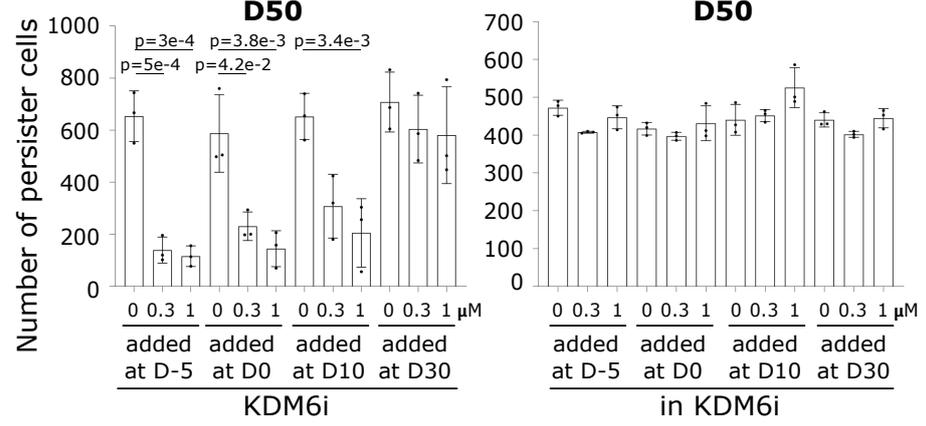
b



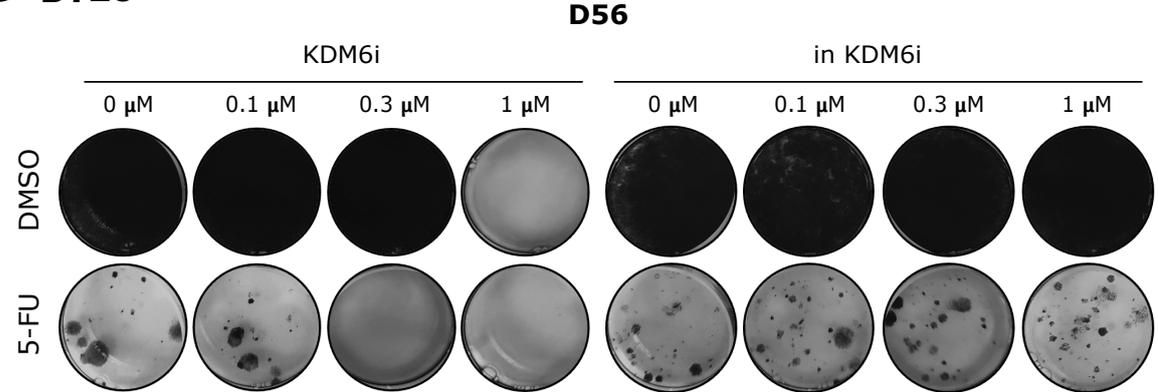
c



d



e BT20



f HCC38

