

# Cycling cancer persister cells arise from lineages with distinct programs

<https://doi.org/10.1038/s41586-021-03796-6>

Received: 1 June 2020

Accepted: 2 July 2021

Published online: 11 August 2021

 Check for updates

Yaara Oren<sup>1,2</sup>, Michael Tsabar<sup>1,3,4,20</sup>, Michael S. Cuoco<sup>1,20</sup>, Liat Amir-Zilberstein<sup>1</sup>, Heidie F. Cabanos<sup>5,6</sup>, Jan-Christian Hütter<sup>1</sup>, Bomiao Hu<sup>7</sup>, Pratiksha I. Thakore<sup>1,18</sup>, Marcin Tabaka<sup>1</sup>, Charles P. Fulco<sup>8,19</sup>, William Colgan<sup>8</sup>, Brandon M. Cuevas<sup>1</sup>, Sara A. Hurvitz<sup>9,10</sup>, Dennis J. Slamon<sup>9,10</sup>, Amy Deik<sup>11</sup>, Kerry A. Pierce<sup>11</sup>, Clary Clish<sup>11</sup>, Aaron N. Hata<sup>5,6</sup>, Elma Zaganjor<sup>12</sup>, Galit Lahav<sup>3</sup>, Katerina Politi<sup>13,14</sup>, Joan S. Brugge<sup>2,15,21</sup> & Aviv Regev<sup>1,16,17,18,21</sup>✉

Non-genetic mechanisms have recently emerged as important drivers of cancer therapy failure<sup>1</sup>, where some cancer cells can enter a reversible drug-tolerant persister state in response to treatment<sup>2</sup>. Although most cancer persisters remain arrested in the presence of the drug, a rare subset can re-enter the cell cycle under constitutive drug treatment. Little is known about the non-genetic mechanisms that enable cancer persisters to maintain proliferative capacity in the presence of drugs. To study this rare, transiently resistant, proliferative persister population, we developed Watermelon, a high-complexity expressed barcode lentiviral library for simultaneous tracing of each cell's clonal origin and proliferative and transcriptional states. Here we show that cycling and non-cycling persisters arise from different cell lineages with distinct transcriptional and metabolic programs. Upregulation of antioxidant gene programs and a metabolic shift to fatty acid oxidation are associated with persister proliferative capacity across multiple cancer types. Impeding oxidative stress or metabolic reprogramming alters the fraction of cycling persisters. In human tumours, programs associated with cycling persisters are induced in minimal residual disease in response to multiple targeted therapies. The Watermelon system enabled the identification of rare persister lineages that are preferentially poised to proliferate under drug pressure, thus exposing new vulnerabilities that can be targeted to delay or even prevent disease recurrence.

To characterize the proliferative dynamics of persisters, we studied the response of PC9 lung cancer cells—which carry an oncogenic mutation in the gene encoding the epidermal growth factor receptor (EGFR)—to osimertinib, a third-generation EGFR tyrosine kinase inhibitor. We treated the cells at maximum effect ( $E_{\max}$ ) concentration (300 nM) (Extended Data Fig. 1a) and used live-cell imaging to quantify the number and timing of division events over the course of treatment, as cells either underwent cell death, arrested or formed visible colonies (Fig. 1a). To ensure that cells from colony-forming lineages are not over-represented in our analysis, we tracked only one cell of each lineage (Methods, Extended Data Fig. 1b).

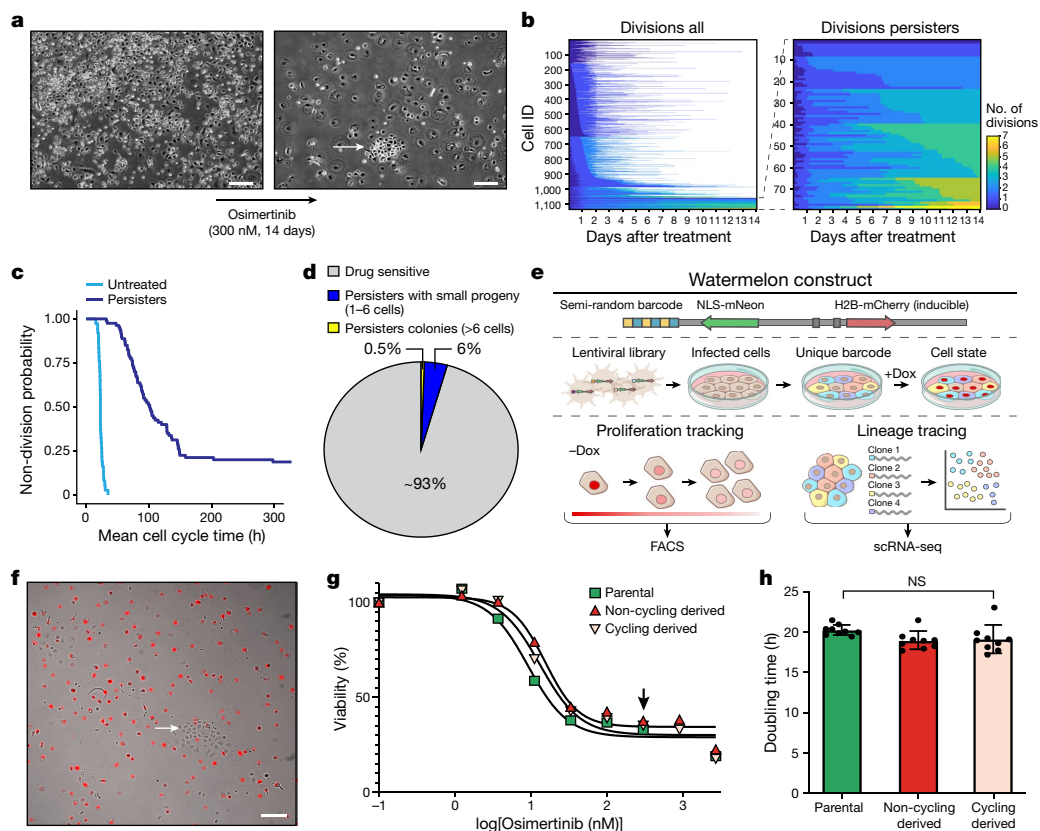
Only 8% of cell lineages gave rise to persisters, defined as cells that were alive at day 14 of drug treatment (Fig. 1b). Cell division profiles of persisters revealed extensive heterogeneity, from persisters that did not divide at all to those that underwent seven divisions (Fig. 1b).

This was in contrast to the untreated parental cell population in which all cells underwent cell division (Fig. 1c). In line with previous reports that colony-forming persisters are rare<sup>3,4</sup>, less than 0.5% of the initial cell population (13% of the persister population) gave rise to multi-cellular persister colonies of more than six cells (Fig. 1d). Thus, persister cells contain a rare proliferative subpopulation that emerges early in the course of drug treatment.

## The Watermelon system

Identifying mechanisms that allow cells to survive drug treatment and regain proliferative capacity requires measuring the cellular and molecular properties of these rare cells before and during drug treatment. To this end, we developed Watermelon, a high-complexity lentiviral barcode library with both green and red fluorescent reporters, for

<sup>1</sup>Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>2</sup>Department of Cell Biology, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Department of Systems Biology, Harvard Medical School, Boston, MA, USA. <sup>4</sup>Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, USA. <sup>5</sup>Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>6</sup>Departments of Medicine, Harvard Medical School, Boston, MA, USA. <sup>7</sup>Department of Pathology, Yale School of Medicine, New Haven, CT, USA. <sup>8</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>9</sup>David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. <sup>10</sup>Jonsson Comprehensive Cancer Center, Los Angeles, CA, USA. <sup>11</sup>Metabolomics Platform, Broad Institute, Cambridge, MA, USA. <sup>12</sup>Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA. <sup>13</sup>Section of Medical Oncology, Department of Pathology, Yale School of Medicine, New Haven, CT, USA. <sup>14</sup>Yale Cancer Center, New Haven, CT, USA. <sup>15</sup>Ludwig Center at Harvard, Boston, MA, USA. <sup>16</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>17</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA. <sup>18</sup>Present address: Genentech, South San Francisco, CA, USA. <sup>19</sup>Present address: Bristol Myers Squibb, Cambridge, MA, USA. <sup>20</sup>These authors contributed equally: Michael Tsabar, Michael S. Cuoco. <sup>21</sup>These authors jointly supervised this work: Joan S. Brugge, Aviv Regev. ✉e-mail: joan\_brugge@hms.harvard.edu; aviv.regev.sc@gmail.com



**Fig. 1 | Persister cells contain a rare proliferative subpopulation.**

**a**, Representative phase-contrast images of a colony of PC9 cells before (left) and after (right) 14 days of treatment with 300 nM osimertinib (Methods). Arrow indicates a colony of cycling persisters. Scale bar, 100  $\mu$ m. **b**, Number of cell divisions for each of 1,135 individual PC9 cell lineages (rows, left) or for the 77 persister lineages (rows, right) tracked by live imaging of a 14-day osimertinib treatment (x axis). White indicates that the lineage perished. **c**, Kaplan–Meier curves for mean cell cycle time of untreated and persister cell lineages from live-cell imaging tracking data. **d**, Proportion of lineages with a large (>6 cells, yellow) or small (1–6 cells, blue) number of progeny or with no progeny (drug sensitive, grey) at end point (14-day osimertinib treatment). **e, f**, The Watermelon system. **e**, The Watermelon vector. A semi-random barcode linked to the NLS–mNeon coding sequence enables lineage tracing. Dox-inducible H2B–mCherry expression facilitates proliferation tracking by

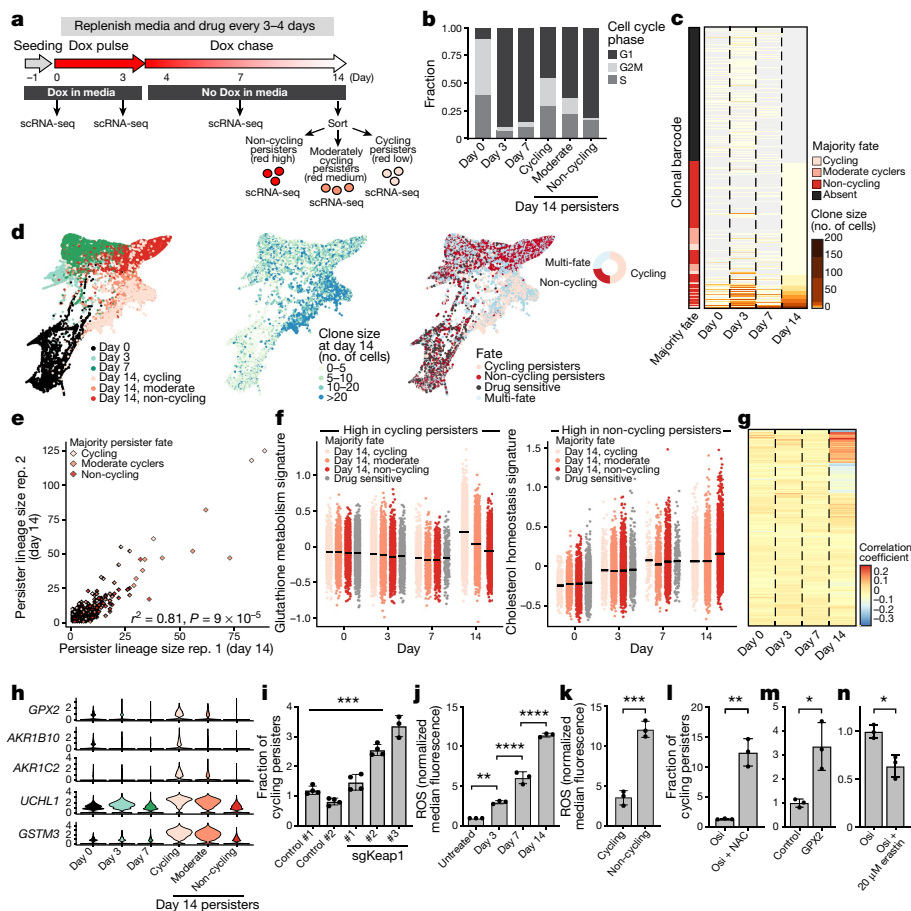
fluorescent dilution. **f**, Merged fluorescence (red) and phase contrast images of Watermelon-PC9 persister cells following 14-day osimertinib treatment during which dox was included for the first 3 days. Arrowhead indicates a colony of cycling persisters. Scale bar 100  $\mu$ m. **g**, Percentage of viable cells (y axis) from populations derived from parental (green), cycling persister (pink) and non-cycling persister (red) cells treated with the indicated concentrations of osimertinib (x axis) for 72 h (Methods). Black arrow indicates the concentration used for establishing persister cells (300 nM). **h**, Mean doubling time (y axis) in populations derived from parental (green), cycling persister (pink) and non-cycling persister (red) cells, grown in drug free media. Data are mean  $\pm$  s.d. of three biologically independent experiments. NS, not significant ( $P > 0.05$ ); two-tailed  $t$ -tests. Images in **a** and **f** are representative of three independent experiments.

simultaneous tracing of clonal lineages as well as the transcriptional and proliferative state of each cell in the population. We achieved lineage tracing by mapping a clone-specific expressed DNA barcode in the 3'-untranslated region of an mNeonGreen protein, and monitored proliferation history by the dilution of a doxycycline (dox)-inducible H2B–mCherry transgene (Fig. 1e, Extended Data Fig. 1c–f). We generated a Watermelon library of more than five million barcodes and used it to transduce 10,000 PC9 cells with a multiplicity of infection (MOI) of less than 0.3 to minimize barcode overlap between clones (Methods).

We tested whether proliferative persisters arise owing to a stable, intrinsically lower sensitivity to EGFR inhibition by comparing the osimertinib sensitivity of cycling and non-cycling persister-derived populations following a short resensitization period. We treated Watermelon-PC9 cells with osimertinib for 14 days, with dox added to the medium for the first three days. Following 11 days of dox chase, during which mCherry was diluted only in proliferating cells, we sorted cycling and non-cycling persister cells (Fig. 1f), and propagated them in drug-free medium for 20 additional passages. Both cycling and non-cycling persister populations reacquired drug sensitivity

rapidly, suggesting that a reversible, rather than a genetic mutational mechanism underlies the ability to cycle under continuous drug treatment (Fig. 1g). Furthermore, cells derived from persister-cycling and persister-non-cycling subpopulations had similar population doubling times, suggesting that the drug-observed phenotypic differences are not due to pre-existing proliferative heterogeneity (Fig. 1h).

To identify mechanisms underlying the ability of persister cells to proliferate, we profiled the expression of 56,419 individual Watermelon-PC9 persister cells by single-cell RNA sequencing (scRNA-seq), at four time points (days 0, 3, 7 and 14) during 14 days of osimertinib treatment (Fig. 2a, Extended Data Fig. 2a–c). On day 14, we sorted persister cells by mCherry expression into three sets: cycling (mCherry<sup>low</sup>), moderately cycling (mCherry<sup>medium</sup>) and non-cycling (mCherry<sup>high</sup>) persisters, and profiled each of these by scRNA-seq separately. We used published signatures<sup>5</sup> to assign each cell to a specific cell cycle phase (Fig. 2b), ascribed cells to lineages on the basis of the expressed barcodes detected by scRNA-seq (Fig. 2c, Extended Data Fig. 2d, Methods), and related cells by their profiles using a force-directed layout embedding (Fig. 2d).



**Fig. 2 | Cycling and non-cycling persisters arise from different cell lineages that express distinct transcriptional programs.** **a**, Experimental scheme. Vertical arrows show time points for scRNA-seq. **b**, Proportion of cells in each cell cycle phase inferred from scRNA-seq. **c**, Lineage size for each lineage barcode (rows) across time points. Lineages are sorted by fate at day 14, and marked by the majority fate of the mCherry populations on day 14 in that lineage (left bar). **d**, Force-directed layout of scRNA-seq profiles coloured by time point and mCherry expression (left), clone size (middle) or fate (right) on day 14 of the lineage the cells belong to. Inset, right: fate distribution at day 14. **e**, Clone size of each persister lineage barcode in two independent experiments seeded from the same barcoded founding population. Colour indicates persister fate of majority of cells in the lineage of combined replicates. *P* value determined by permutation test (Methods). **f**, Signature scores by lineage majority fate. **g**, Correlation coefficient of expression of each gene (row) in

each persister cell at a given time point (column) and the cells' corresponding lineage size at day 14. **h**, Distribution of expression levels ( $\log_2(\text{transcripts per million} + 1)$ , *y* axis) across time for 5 of the top 10 genes positively correlated with persister lineage size. **i**, Fraction of cycling persisters following osimertinib treatment in PC9 cells treated with control or single guide RNA directed against *Keap1* (sgKeap1).  $P = 3 \times 10^{-4}$ . **j, k**, Time course of ROS levels in PC9 cells treated with osimertinib (**j**;  $P_{0,3} = 1.3 \times 10^{-3}$ ,  $P_{3,7} < 1 \times 10^{-4}$ ,  $P_{7,14} < 1 \times 10^{-4}$  (subscripts indicate interval in days), one-way ANOVA with Tukey's correction) or in persister subpopulations (**k**;  $P = 4 \times 10^{-4}$ ). **l–n**, Fraction of cycling persisters following osimertinib treatment with or without NAC (**l**;  $P = 1.1 \times 10^{-3}$ ), in PC9 cells transfected with control or a GPX2 expression plasmid (**m**;  $P = 1.5 \times 10^{-2}$ ), or cells treated with or without erastin (**n**;  $P = 1 \times 10^{-2}$ ). Data are mean  $\pm$  s.d. of three biologically independent experiments (**i–n**). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ; two-tailed *t*-tests.

In line with previous reports<sup>3,6</sup>, osimertinib induced G1 arrest on day 3 and day 7 of treatment. At day 14, cycling persisters (as defined at the endpoint by mCherry dilution) had higher expression-based cell cycle scores compared with moderate cyclers and non-cycling cells (55%, 36% and 18% of cells in G2/M or S phase for cycling, moderate cyclers and non-cycling persisters, respectively) (Fig. 2b, Extended Data Fig. 2e), validating the ability of Watermelon to distinguish persisters by proliferative history. Consistent with our imaging-based analysis, fewer than 12% of barcodes that were present in persisters at day 14 were part of a large clonal expansion (defined as at least a fivefold increase during the course of the 14-day assay) (Fig. 2c, Extended Data Fig. 2f). We observed a significant decrease in lineage diversity ( $^2D$ ) at day 14 compared with the other time points ( $P = 2 \times 10^{-58}$  (day 7 to day 14),  $^2D = 829$  (day 14),  $^2D > 2,020$  (day 0–7), Extended Data Fig. 3a, b), most lineages were retained throughout the experiment overall, when accounting for sampling bias (Extended Data Fig. 3c–e). These clonal dynamics are consistent with recent reports that most cell lineages are able to enter a drug tolerant state following treatment<sup>7,8</sup>.

**Persister fate is lineage dependent**

The scRNA-seq profiles indicated a gradual change of cell state following drug treatment, with cycling and non-cycling persister cells following distinct transcriptional trajectories, and a subset of cycling persisters resembling untreated cells (Fig. 2d left; overlapping light pink and black dots). Examination of clone size dynamics during the course of treatment revealed clonal expansion with the largest increase in clone size observed in the cycling persisters (a maximum of 175, 110 and 53 cells per lineage for cycling, moderate cyclers and non-cycling persisters, respectively) (Fig. 2d, middle, Extended Data Fig. 4a). Almost two-thirds of clones had a single fate, giving rise to either only cycling or only non-cycling persisters, and multi-fate lineages were far less frequent than expected by chance (Extended Data Fig. 4b,  $P = 1 \times 10^{-5}$ , permutation test; Fig. 2d, right).

We hypothesized that this restricted cell fate pattern arises because cells are often committed to a given proliferative fate before drug treatment. To test this hypothesis, we compared persister lineage size

in two independent replicate experiments using the same starting Watermelon-transduced PC9 cell population, treated with osimertinib for 14 days, sorted to three subpopulations according to mCherry expression and profiled by scRNA-seq. The sizes of individual clones at day 14 were highly correlated between replicates (Fig. 2e,  $r^2 = 0.81$ ,  $P = 9 \times 10^{-5}$ , permutation test; Extended Data Fig. 4c, d, Methods), suggesting that each persister clone has a distinct reproducible proliferative potential under drug treatment.

### ROS control persister proliferation

To identify cellular expression programs that are associated with the ability of persisters to cycle, we searched for gene signatures that are differentially expressed between the persister subpopulations, but are cell cycle independent. In line with previous reports<sup>9,10</sup>, epithelial-to-mesenchymal transition (EMT) was induced by the EGFR inhibitor; however, the levels of induction were similar in both the cycling and non-cycling persister populations (Extended Data Fig. 5a), suggesting that this program does not underlie the ability of persisters to cycle. By contrast, cycling and non-cycling persisters upregulated the expression of genes from different pathways. Cycling persisters exhibited higher expression of glutathione metabolism and NRF2 signatures than non-cycling persisters, whereas non-cycling persisters had higher expression of cholesterol homeostasis, interferon- $\alpha$  and Notch-signalling signatures (Fig. 2f, Extended Data Fig. 5b, c, Supplementary Table 1). The signatures associated with cycling persister fate were not upregulated before drug treatment, suggesting that cycling persisters arise from cells poised to induce these programs, rather than from a selection of cells that already express them before drug treatment.

Notably, both oestrogen response<sup>11</sup> and ferroptosis<sup>12,13</sup> have been implicated in the survival of drug-tolerant cells, and we indeed consistently observed a decrease in the total number of persister cells following treatment with either the oestrogen receptor inhibitor fulvestrant or a GPX4 inhibitor (RSL3) (which induces ferroptosis). However, these treatments enriched for the fraction of cycling cells in the surviving population (Extended Data Fig. 5d–j), further emphasizing the difference between cycling and non-cycling persisters, and suggesting that alternative strategies may be needed to target relapse-promoting cells.

To further explore expression differences at the lineage level, we correlated gene expression over time with persister clone size at day 14. At early time points, no genes were strongly correlated with persister clone size, whereas at day 14, expression of a subset of genes correlated with persister clone size (day 14 correlation,  $r > 0.25$ , Fig. 2g). (Some weak correlation is observed as early as day 3; Extended Data Fig. 6a–e). Of the top ten correlated genes more highly expressed in cells from larger clones, five are targets of the oxidative stress-induced transcription factor NRF2 (Fig. 2h), further supporting a role for antioxidant defence programs in persister proliferative fate. Consistent with our differential expression analysis and a recent report on the role of NRF2 in minimal residual disease<sup>14,15</sup>, knockout of *KEAP1*, which encodes a negative regulator of NRF2, increases the fraction of cycling persisters (Fig. 2i).

Given the strong relationship between antioxidant expression signatures and persister proliferative capacity, we tested whether prolonged osimertinib treatment induces reactive oxygen species (ROS) and whether alleviating ROS enhances cycling persisters. Analysis of osimertinib-treated cells using CellROX, a fluorescent ROS reporter, revealed a strong time-dependent increase in ROS levels (Fig. 2j). In line with the observed expression differences between the two persister subpopulations, day 14 cycling persisters exhibited less than one-third of the ROS levels measured in the non-cycling population (Fig. 2k). Moreover, when we alleviated ROS by treatment with the ROS scavenger *N*-acetylcysteine (NAC) from day 3 onwards (Fig. 2l) or by stable overexpression of GPX2 (Fig. 2m), the fraction of cycling persisters increased significantly. Conversely,

treatment with erastin, which inhibits the cystine–glutamate antiporter system crucial for glutathione synthesis, decreased the fraction of cycling persisters (Fig. 2n). Together, these results support a role for redox balance in regulating the proliferative capacity of persisters.

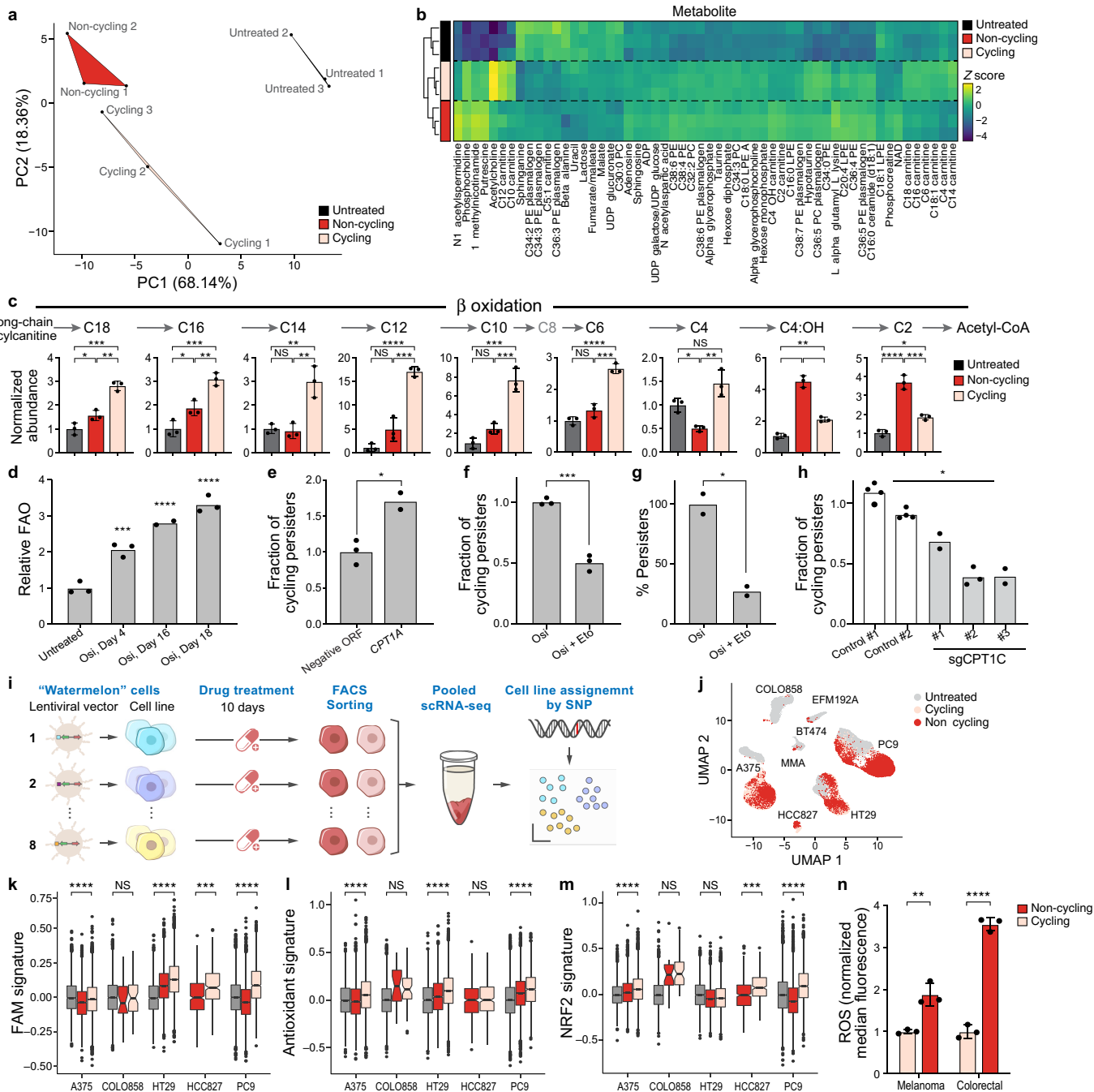
### Metabolic shift in cycling persisters

Because redox balance is tightly linked to metabolism<sup>16</sup>, we performed liquid chromatography–mass spectrometry (LC–MC/MS) metabolic profiling to examine differences between persister subpopulations. Principal component analysis (PCA) over the 229 quantified metabolites separated the cycling persisters, non-cycling persisters and untreated populations (Fig. 3a, Extended Data Fig. 7a–c), with 56 metabolites differentially abundant between the three populations (Fig. 3b). In particular, carnitine-linked fatty acids, which are substrates of mitochondrial  $\beta$ -oxidation, were significantly more abundant in the cycling persisters than in the non-cycling persisters (Fig. 3c). We examined the oxidation of radiolabelled palmitic acid to  $^3\text{H}_2\text{O}$  to assess fatty acid oxidation (FAO), and indeed found a time-dependent increase in FAO with osimertinib treatment (Fig. 3d, Extended Data Fig. 7d, e). Thus, the higher abundance of acylcarnitine species in cycling persisters and the increase in FAO over time with osimertinib treatment both suggest that mitochondrial FAO may contribute to the cycling persister phenotype.

We next tested whether modulating the FAO pathway affects the proliferative capacity of persisters. Overexpression of CPT1A, a rate limiting enzyme that facilitates the transfer of fatty acids into the mitochondria, in Watermelon-PC9 cells resulted in an increase of more than 50% in the fraction of cycling persisters following 14 days of osimertinib treatment (Fig. 3e). Conversely, blocking FAO with etomoxir, the CPT1 inhibitor (albeit with additional targets), at day 3 of osimertinib treatment, when most sensitive cells have already died, reduced the proliferative capacity of persister cells. Co-treatment with etomoxir for 11 days, at a concentration that has minimal inhibitory effects on untreated cells (Extended Data Fig. 7f), significantly reduced the fraction of cycling persisters (Fig. 3f) and the overall fraction of persister cells (Fig. 3g). Concordantly, knocking out *CPT1C*, a gene isoform which is upregulated in human lung tumours<sup>17</sup>, but not *CPT1A*, led to a reduction in the fraction of cycling persister cells (Fig. 3h, Extended Data Fig. 7g). Together, these results support our hypothesis that FAO has a role in establishing the proliferative persister state.

To investigate the generality of these findings to other persister cells, we generated seven additional Watermelon models from EGFR-driven lung cancer (PC9 and HCC827), HER2-driven breast cancer (BT474 and EFM192A) and BRAF-driven melanoma (A375 and COLO858) and colorectal (HT29) cell lines, treated them with clinically relevant kinase inhibitors for 10 days (osimertinib, lapatinib or dabrafenib), sorted them for mCherry levels, and profiled them by scRNA-seq, yielding 50,735 single-cell profiles (Fig. 3i, Supplementary Table 2). In line with previous reports<sup>18,19</sup>, the profiles were grouped by cell identity and not by treatment (Fig. 3j, Extended Data Fig. 8a–e). Across all models, persisters exhibited G1 cell cycle arrest and cycling persisters, when present, had an overall higher fraction of G2M/S phase cells compared with non-cycling persisters, validating the ability of Watermelon to capture this rare proliferative population in multiple models (Extended Data Fig. 8f).

In four out of five models in which a sufficient number (greater than 50) of cycling persister cells were detected, cycling persister cells showed elevated fatty acid metabolism (FAM) signatures compared with their non-cycling counterparts (Fig. 3k). Similar to PC9 cells, A375 melanoma cycling persisters exhibited higher expression of both antioxidant response genes and the NRF2 signature compared with non-cycling cells from the same model (Fig. 3l, m). Conversely, cycling HT29 and HCC827 persister cells upregulated one, but not

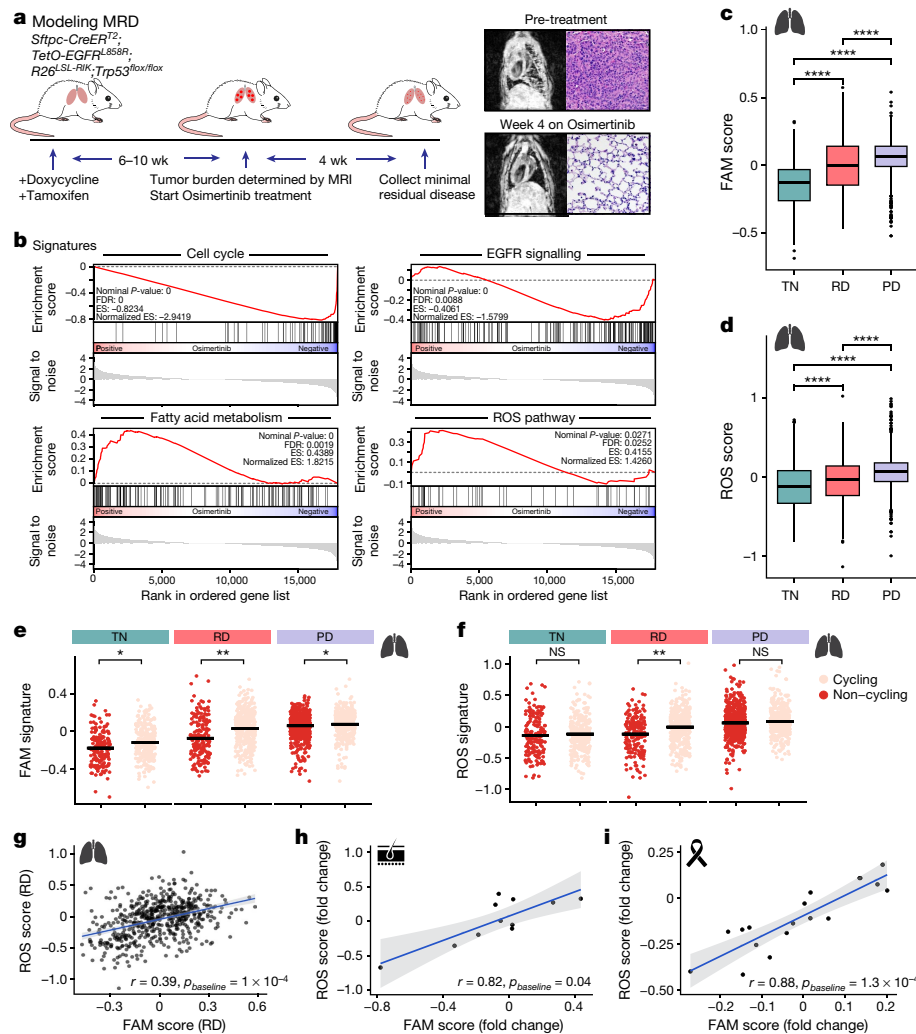


**Fig. 3 | Persister cells shift their metabolism to FAO.** **a**, The first two principal components (PCs) from a PCA of LC-MS metabolite profiles. **b**, Z-scores of differentially abundant metabolites (columns) ( $P < 0.05$ ; one-way ANOVA) between cycling persisters (pink), non-cycling persisters (red) and untreated parental cells (black) (rows). **c**, Mean abundance of  $\beta$ -oxidation metabolites measured by LC-MS. **d**, Mean FAO levels (y axis, relative to mean of untreated cells) measured by  $^3\text{H}$ -palmitic acid oxidation in osimertinib-treated PC9 cells over time.  $P_{0.4} = 8 \times 10^{-4}$ ,  $P_{0.16} < 1 \times 10^{-4}$ ,  $P_{0.18} < 1 \times 10^{-4}$  (subscripts indicate intervals in days); one-way ANOVA with Tukey's correction. **e**, Mean fraction of cycling persisters at day 14 of 300 nM osimertinib treatment of PC9 cells transfected with control or CPT1A-expressing vector.  $P = 1 \times 10^{-2}$ . **f, g**, Mean fraction of cycling persisters (**f**;  $P = 3 \times 10^{-4}$ ) and overall fraction of persister cells (**g**;  $P = 1.6 \times 10^{-2}$ ) at day 14 of treatment with 300 nM osimertinib alone or 300 nM osimertinib and 100  $\mu\text{M}$  etomoxir at days 3–14. **h**, Mean fraction of

cycling persisters for PC9 cells treated with control or a guide RNA directed against *CPT1C* (sgCPT1C),  $P = 3 \times 10^{-2}$ . **i**, Schematic for pooled persister assay. **j**, Uniform manifold approximation and projection (UMAP) of single-cell profiles (dots) from eight Watermelon persister models coloured by sample type. **k–m**, Signature scores of FAM (**k**), antioxidant response (**l**) and NRF2 (**m**) pathways in persister cells from indicated models. Significance based on comparing cycling with non-cycling populations (Methods, Supplementary Table 2). In box plots, centre line shows median, box limits delineate upper and lower quartiles and whiskers extend to 1.5 $\times$  interquartile range past upper and lower quartiles. **n**, Mean ROS levels in melanoma (A375,  $P = 5.1 \times 10^{-3}$ ) and colorectal (HT29,  $P < 1 \times 10^{-4}$ ) persister populations. Data are mean  $\pm$  s.d. of two (**e, g**) or three biologically independent experiments (**c, d, f, h, n**). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ ; two-tailed *t*-tests.

both, of these ROS-related signatures (Fig. 3l, m). Melanoma (A375) and colorectal cancer (HT29) cycling persisters also exhibited significantly lower ROS levels compared with the non-cycling population of

the same cell line (Fig. 3n). Thus, the cycling persister state induced by oncogenic kinase inhibition shared similar metabolic and expression features across cells lines from different tumour types.



**Fig. 4 | Metabolic shift in tumours treated with oncogene-targeted therapy.**

**a**, FAM and ROS pathway signatures increase with treatment in a genetically engineered mouse model. Left, experimental scheme for tumour development and drug treatment. Right, MRI (left image) and haematoxylin and eosin staining (right image) of pre-treatment (top) and post-treatment (bottom) mouse lungs. **b**, Gene set enrichment analysis of cell cycle-, EGFR signalling-, ROS- and FAM-related genes in minimal residual disease samples versus untreated control. **c**, **d**, Signature scores for FAM (**c**;  $P < 2 \times 10^{-16}$  (treatment-naive versus residual disease),  $P < 2 \times 10^{-16}$  (treatment-naive versus progression),  $P = 6 \times 10^{-14}$  (residual disease versus progression)) and ROS (**d**;  $P = 6.7 \times 10^{-7}$  (treatment-naive versus residual disease),  $P < 2 \times 10^{-16}$  (treatment-naive versus progression),  $P = 4.8 \times 10^{-16}$  (residual disease versus progression)) in treatment naive (TN), residual disease (RD) and progressive disease (PD) of human lung adenocarcinoma (457, 557 and 1,088 cells per group, respectively). \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ ; two-sided Wilcoxon test.

### ROS and FAM induced in residual disease

EGFR inhibition also induced a similar shift in vivo, as observed in minimal residual disease in an osimertinib-treated *Trp53*-knockout mouse model<sup>15</sup> carrying the inducible mutant EGFR(L858R) (Fig. 4a, Extended Data Fig. 9a). To establish tumours, we induced lung-specific expression of the human *EGFR* transgene by dox and tamoxifen administration (Methods). Once mice developed lung adenocarcinomas, we treated them with osimertinib until the rate of tumour regression plateaued, indicating minimal residual disease (Fig. 4a, Extended Data Fig. 9b). Bulk RNA sequencing of samples from treated and untreated mice showed that while there was a strong treatment-induced decrease in

Expression data from NCT03433469 trial. In box plots, centre line shows median, box limits delineate upper and lower quartiles and whiskers extend to 1.5 × interquartile range past upper and lower quartiles. **e**, **f**, FAM (**e**;  $P = 1.2 \times 10^{-2}$  (treatment-naive),  $P = 1.1 \times 10^{-3}$  (residual disease),  $P = 1.5 \times 10^{-2}$  (progression)) and ROS (**f**;  $P = 0.9$  (treatment-naive),  $P = 5.3 \times 10^{-3}$  (residual disease),  $P = 0.1$  (progression)) signature scores (y axis) in cells (dots) stratified by cycling (pink) versus non-cycling (red) status across treatment phases of lung adenocarcinoma samples. \* $P < 0.05$  (mixed linear model; Supplementary Methods). **g**, Correlation between ROS and FAM signature scores (x and y axes) in cells (dots) from lung adenocarcinoma residual disease. **h**, **i**, Correlation between ROS and FAM signature scores in melanoma (**h**) and breast cancer (**i**) residual disease samples (dots). ROS (y axis) and FAM (x axis) signature scores in patients with melanoma treated with BRAFi and MEKi (**h**) and patients with HER2<sup>+</sup> breast cancer treated with lapatinib (**i**). Shaded area in **g**–**i** shows 95% confidence interval.

overall cell cycle and EGFR signalling signatures, minimal residual disease cells upregulated the expression of ROS and FAM gene signatures, consistent with our in vitro findings (Fig. 4b, Extended Data Fig. 9c–e).

Finally, we explored the relevance of these metabolic states in patient tumours, by analysing scRNA-seq profiles from human EGFR-driven lung adenocarcinoma tumour samples from a recent study<sup>20</sup> of 30 patients (Supplementary Table 3), spanning treatment-naive tumours, residual disease during targeted therapy response (residual disease), and upon establishment of drug resistance (progression). In line with our in vitro findings, both FAM and ROS pathway signatures were significantly and gradually increased upon treatment from treatment-naive to residual disease to progression (Fig. 4c, d, Extended Data Fig. 10a, b).

The increase in ROS and FAM gene expression was associated with proliferation and was higher in cycling compared with non-cycling persister cells (Fig. 4e, f). Specifically, when we stratified cells as cycling versus non-cycling using the expression of a set of known cell cycle genes<sup>21</sup>, cycling persister cells in the residual disease phase had higher expression of FAM and ROS signatures compared with non-cycling residual disease cells ( $P = 0.005$  (ROS),  $P = 0.001$  (FAM), after controlling for patient and cell complexity; Methods, Fig. 4e, f). Of note, ROS and FAM signatures were more strongly correlated in treated than in treatment-naïve samples ( $r = 0.39$ ,  $P = 10^{-4}$  (versus baseline), calculated by comparing treatment-naïve and residual disease cells; Methods, Fig. 4g, Extended Data Fig. 10c), supporting the importance of oncogene inhibition in linking these pathways.

To test whether similar states are observed in patients in response to other oncogenic kinase inhibitors, we analysed bulk RNA-seq profiles from two additional tumour types: BRAF-driven melanoma collected at baseline or after up to 22 days of treatment with BRAF inhibitor and MEK inhibitor<sup>22</sup>; and HER2-driven breast cancer from patients at baseline or after treatment with lapatinib for 2–3 weeks (Supplementary Table 3). In melanoma, FAM and ROS induction was correlated only upon treatment ( $r = 0.82$ ,  $P = 0.04$  (versus baseline); Fig. 4h, Extended Data Fig. 10d), with 8 out of 11 patients exhibiting induction of at least one of the signatures following treatment. In HER2<sup>+</sup> breast cancer, FAM and ROS signature induction were correlated only upon drug treatment ( $r = 0.88$ ,  $P = 1.3 \times 10^{-4}$  versus baseline; Fig. 4i, Extended Data Fig. 10e), with 50% of patients exhibiting induction of at least one of the signatures following treatment. This supports our model that cycling persisters undergo a metabolic shift in tumours from patients in response to oncogenic kinase inhibitor therapy.

The understanding that non-mutational mechanisms may have a role in tumour relapse has prompted multiple studies focused on identifying factors that contribute to overall persister fitness<sup>23–26</sup>. However, since most persisters remain arrested during drug treatment<sup>3</sup>, factors that contribute to persister-driven relapse are difficult to discern using bulk profiling. The ability of the Watermelon system to simultaneously map the lineage, proliferative history and transcriptional state of individual cells enabled us to identify metabolic and expression adaptations that may facilitate cell cycle re-entry in a rare subset of persister cells. In particular, our finding that some lineages are more poised to undergo a proliferation-promoting adaptive response is in line with studies showing that a subset of cancer cells have a chromatin configuration that renders them more likely to transition to an oncogene-independent state following treatment<sup>27–29</sup>. Together, our results illustrate an approach to understanding not only what underlies the ability of persisters to cycle but also what drives other clinically important persister traits such as time to relapse, thus providing an important step towards the development of therapies that delay disease recurrence.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03796-6>.

- Salgia, R. & Kulkarni, P. The genetic/non-genetic duality of drug ‘resistance’ in cancer. *Trends Cancer* **4**, 110–118 (2018).
- Vallette, F. M. et al. Dormant, quiescent, tolerant and persister cells: four synonyms for the same target in cancer. *Biochem. Pharmacol.* **162**, 169–176 (2019).
- Sharma, S. V. et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
- Hata, A. N. et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.* **22**, 262–269 (2016).
- Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
- Taniguchi, H. et al. AXL confers intrinsic resistance to osimertinib and advances the emergence of tolerant cells. *Nat. Commun.* **10**, 259 (2019).
- Dhimolea, E. et al. An embryonic diapause-like adaptation with suppressed Myc activity enables tumor treatment persistence. *Cancer Cell* **39**, 240–256 (2021).
- Rehman, S. K. et al. Colorectal cancer cells enter a diapause-like DTP state to survive chemotherapy. *Cell* **184**, 226–242 (2021).
- Raof, S. et al. Targeting FGFR overcomes EMT-mediated resistance in EGFR mutant non-small cell lung cancer. *Oncogene* **38**, 6399–6413 (2019).
- Zhu, X., Chen, L., Liu, L. & Niu, X. EMT-Mediated Acquired EGFR-TKI Resistance in NSCLC: Mechanisms and Strategies. *Front. Oncol.* **9**, 1044 (2019).
- Garon, E. B. et al. Antiestrogen fulvestrant enhances the antiproliferative effects of epidermal growth factor receptor inhibitors in human non-small-cell lung cancer. *J. Thorac. Oncol.* **8**, 270–278 (2013).
- Hangauer, M. J. et al. Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature* **551**, 247–250 (2017).
- Friedmann Angeli, J. P., Krysko, D. V. & Conrad, M. Ferroptosis at the crossroads of cancer-acquired drug resistance and immune evasion. *Nat. Rev. Cancer* **19**, 405–414 (2019).
- Fox, D. B. et al. NRF2 activation promotes the recurrence of dormant tumour cells through regulation of redox and nucleotide metabolism. *Nat. Metab.* **2**, 318–334 (2020).
- Foggetti, G. et al. Genetic determinants of EGFR-driven lung cancer growth and therapeutic response *in vivo*. *Cancer Discov.* **11**, 1736–1753 (2021).
- Reczek, C. R. & Chandel, N. S. The two faces of reactive oxygen species in cancer. *Ann. Rev. Cancer Biol.* **1**, 79–98 (2017).
- Zaugg, K. et al. Carnitine palmitoyltransferase 1C promotes cell survival and tumor growth under conditions of metabolic stress. *Genes Dev.* **25**, 1041–1051 (2011).
- Kinker, G. S. et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).
- McFarland, J. M. et al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat. Commun.* **11**, 4296 (2020).
- Maynard, A. et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* **182**, 1232–1251 (2020).
- Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
- Kwong, L. N. et al. Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. *J. Clin. Invest.* **125**, 1459–1470 (2015).
- Viswanathan, V. S. et al. Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature* **547**, 453–457 (2017).
- Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
- Shah, K. N. et al. Aurora kinase A drives the evolution of resistance to third-generation EGFR inhibitors in lung cancer. *Nat. Med.* **25**, 111–118 (2019).
- Havas, K. M. et al. Metabolic shifts in residual breast cancer drive tumor recurrence. *J. Clin. Invest.* **127**, 2091–2105 (2017).
- Liau, B. B. et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. *Cell Stem Cell* **20**, 233–246 (2017).
- Guler, G. D. et al. Repression of stress-induced LINE-1 expression protects cancer cell subpopulations from lethal drug exposure. *Cancer Cell* **32**, 221–237 (2017).
- Yang, C., Tian, C., Hoffman, T. E., Jacobsen, N. K. & Spencer, S. L. Melanoma subpopulations that rapidly escape MAPK pathway inhibition incur DNA damage and rely on stress signalling. *Nat. Commun.* **12**, 1747 (2021).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

### Cell culture

*EGFR*-mutant non-small-cell lung cancer cells PC9, HCC827 (from the laboratory of A.N.H.), *HER2*-amplified breast cancer cell line BT474 and *BRAF*-mutant colorectal cell HT29 line were cultured in RPMI-1640 medium (ThermoFisher Scientific) supplemented with 10% fetal bovine serum (FBS). A375 (*BRAF*-mutant) melanoma line was grown in DMEM (ThermoFisher Scientific) supplemented with 10% FBS. EFM192A (*HER2*-amplified) breast cancer line was grown in RPMI-1640 medium (ThermoFisher Scientific) supplemented with 20% fetal bovine serum (FBS). COLO858 (*BRAF*-mutant) cell line was grown in phenol red-free RPMI 1640 supplemented with 5% FBS and 1% sodium pyruvate (Gibco). MMACSF (*BRAF*-mutant) was grown in DMEM/F12 supplemented with 5% FBS and 1% sodium pyruvate. All cell lines were grown in medium supplemented with penicillin and streptomycin (Thermo Scientific). All breast cancer cell lines were obtained from the Broad Institute's cancer platform. Melanoma and colorectal cell lines used in this study were obtained from the Harvard Medical School Laboratory of Systems Pharmacology. Each cell line was maintained in a 5% CO<sub>2</sub> atmosphere at 37 °C. Cell line identities were confirmed by short tandem repeat fingerprinting and all were found to be negative for mycoplasma using the Universal Mycoplasma Detection Kit (ATCC).

### Persister cell derivation and treatments

Persister cells were derived using an IC90 drug concentration from treatment of *EGFR*-mutant non-small-cell lung cancer, *BRAF*-mutated melanoma or *HER2*-amplified breast cancers with 300 μM osimertinib, 1 μM dabrafenib or 1 μM lapatinib, respectively, for 10 days for the breast cancer and melanoma cells and 14 days for the lung cells. Fresh medium with drug was added every 3–4 days.

For Watermelon vector induction, dox was added to the medium two days before drug treatment and was maintained in the media for the first three days of drug treatment. Unless otherwise noted, any additional drugs were added to three days treated cells that were maintained under constant drug exposure throughout subsequent treatment by replenishing the medium every 3–4 days with fresh drug. For resensitization experiments, persister populations were derived by sorting day 14 osimertinib-treated PC9-watermelon cells based on mCherry expression following 11 days of dox chase. Following sorting, persister populations were propagated in drug-free media for 20 passages before starting the assay. Cell numbers following treatment were determined by imaging plates and quantifying nuclei using the Acumen Cellista plate cytometer (TTP Labtech).

### Live-cell imaging

Cells were grown in poly-D-lysine-coated glass bottom plates (MatTek Corporation) and imaged using a Nikon Eclipse TE-2000 inverted microscope with a 10× Plan Apo objective and a Hamamatsu Orca ER camera, equipped with environmental chamber controlling temperature, atmosphere (5% CO<sub>2</sub>) and humidity. For long-term live-cell imaging experiments, drug-containing medium was replaced every day to prevent changes in drug concentration due to evaporation. Images were acquired using the MetaMorph Software. For the initial 24 h, images were acquired every 10 min, for 24–120 h images were acquired every 20 min and later images were acquired hourly.

Single-cell tracking was performed as described<sup>30</sup>. In brief, we used a semi-automated MATLAB-based method to track and annotate cell fates. The method relies on identification of cell centroids using intensity and shape information of a constitutively expressed nuclear marker (H2B-mCherry), centroid linkage using nearest-neighbour, and user correction and annotation of cell fate events (tracking software available at: <https://github.com/balvahal/p53CinemaManual>). A single cell of each progeny was tracked throughout the time course or until cell death.

Non-persisters were tracked from the beginning of the movie. Only one representative from each progeny was tracked, and following

division one of the sisters was randomly selected for tracking. To track persister cells, all cells that appeared in the last frame were first manually tracked from the last frame to the first frame. After determining a day 0 initiating cell, to prevent over-representation of large progenies in the analysis, for each lineage, only one path was tracked forward in a given time point. To determine the number of progeny in each persister lineage, cells were manually tracked and counted and cell death was manually recorded. Cell cycle time for untreated cells was measured between first and second mitotic events. For persisters, cell cycle time was measured from the first mitosis occurring after the first 72 h following treatment to any subsequent mitosis and averaged, in order to avoid counting cell cycle events occurring before osimertinib uptake.

For doubling-time assays, cells were seeded in 96-well plates at a density of  $3 \times 10^3$  cells per well and measured using an InCuCyte ZOOM live-cell imaging system (Essen BioScience).

### Cloning of the Watermelon library

A lentivirus backbone was constructed containing: TRE3GS promoter for H2B-mCherry and hPGK promoter for expression of Tet-On rtTA element, NLS-mNeon and a polyadenylated lineage barcode cassette. We prepared the vector backbone by digesting 20 μg of vector with SbfI (New England Biolabs (NEB)) overnight at 37 °C followed by gel purification using 1% E-Gel Ex (ThermoFisher Scientific). The cut backbone was extracted with the QIAprep miniprep Kit (Qiagen) and the resulting DNA was eluted in 100 μl of H<sub>2</sub>O and purified with 0.70× AMPure XP SPRI beads.

The double-stranded lineage barcodes were generated by annealing two DNA primers, a 90-bp-long oligonucleotide containing a semi-random 30-bp-long barcode sequence (that is, 15 repeats of A/T (W)–G/C (S)) and a flanking primer pair for barcode amplification and performing a single cycle extension reaction (primer list is in Supplementary Table 4). The resulting oligonucleotides were run on 2% E-Gel Ex and purified with 2.5× AMPure XP SPRI beads.

To generate a Watermelon library with a diverse lineage barcode pool, the lineage oligonucleotides were mixed together with the backbone fragment in a 5:1 molar ratio together with an equal volume of Gibson master mix (Gibson Assembly Cloning Kit, NEB), incubated at 50 °C for 4 h, cleaned with 0.75× AMPure XP SPRI, eluted in 15 μl H<sub>2</sub>O and electroporated into Endura competent cells (Lucigen). We expanded the cells in LB liquid culture supplemented with carbenicillin (Sigma-Aldrich) for 16 h at 30 °C and purified the pooled library plasmid with the Endotoxin-Free Plasmid Maxiprep Kit (Qiagen). Library complexity was estimated by sequencing the library plasmid pool at a depth of approximately 68 million reads.

### Lentivirus production

Lentiviral particles were produced by transfecting 293 T cells with dVPR and VSVG packaging plasmids, using the X-tremeGene transfection reagent (Sigma-Aldrich) according to the manufacturer's instructions. Media was replaced with DMEM medium (ThermoFisher Scientific) supplemented with 20% FBS 20 h after transfection, and media containing virus particles were collected 48 h after transfection. For open reading frame (ORF) overexpression and guide screen, virus particles were concentrated using Amicon 100-kDa 15-ml columns (Millipore) in a cold centrifuge at 1,000g to a final volume of 500 μl per virus. Virus was aliquoted and stored at –80 °C until use.

### Watermelon cell line construction

Parental cell lines were transduced using the watermelon virus and cells were spin infected using 16 μg ml<sup>-1</sup> polybrene at 2,500 rpm for 30 min at 30 °C. After a 24 h incubation with virus, medium was changed and 72 h after infection, cells were sorted for mNeon expression. To ensure that the majority of cells were labelled with a single barcode per cell for Watermelon lentiviral infection, we used a target multiplicity of infection (MOI) of at most 0.3, corresponding to less than 30%



# Article

mNeon-expressing cells 72 h after infection. Sorted cell populations with 10,000 cells each were expanded in culture for three passages, aliquoted to  $2 \times 10^6$  cells per vial and stored in liquid nitrogen.

## Single-cell capture for PC9 time-course experiment

Watermelon cells were thawed, expanded in dox ( $2 \mu\text{g ml}^{-1}$ )-containing medium for 96 h and mCherry positive cells were sorted using a MoFlo Astrios Cell Sorter (Beckman Coulter) and re-plated. Following 96 h of recovery, the cells were seeded into six-well plates at 300,000 cells per well and were given 24 h to attach before adding 300 nM osimertinib. Dox was continuously added to the media until day 3 of drug treatment. Cells were collected at day 0 (untreated), 3, 7 and 14 of drug treatment. To obtain cell suspension for single cell profiling, cells were scraped from the well, washed and resuspended in FACS buffer (0.5% BSA in phosphate-buffered saline), and filtered through a 40- $\mu\text{m}$  strainer. To delineate the differences between persister populations, day 14 cells were gated based on mCherry expression. Following sorting, the cells were spun down and approximately 9,000 single cells per sample were loaded to the Chromium Controller (10X Genomics). scRNA-seq libraries were generated using the 10X Genomics Chromium Single Cell 3' Kit v2 and the 10X Chromium Controller (10X Genomics) according to the standard v2 protocol. The resulting 3' scRNA-Seq libraries were pooled together and sequenced with a HiSeq (Illumina, R2 read length 98 base pairs). To increase lineage barcode capture, targeted sequencing of the barcode area was performed using the whole transcriptome amplification product generated as a part of the v2 protocol as a PCR template (primer list in Supplementary Table 4). Targeted libraries were gel purified and sequenced with a MiSeq (Illumina).

## scRNA-seq data analysis

**Read alignment and data processing.** Reads were mapped to the GRCh38 human transcriptome using Cell Ranger 2.1.0 (10X Genomics), and TPM was calculated for each gene in each filtered cell barcodes sample. TPM values were then divided by 10 (TP100K), since the complexity of our single-cell libraries is estimated to be on the order of 100,000 transcripts. For each cell, we quantified the number of genes expressed and the proportion of the transcript counts derived from mitochondrial encoded genes. Cells with either <1,000 or more than 4,200 detected genes or >0.1 mitochondrial fraction were excluded from further analysis. Finally, the resulting expression matrix was filtered to remove genes detected in fewer than three cells. All the above steps were done using the Seurat v2 R package<sup>31</sup>.

**Detection of differentially expressed gene signatures.** To identify cellular programs that are associated with the ability of persisters to cycle, we searched for differentially expressed, cell cycle independent, gene signatures that show the largest difference between the persister sub-population. First, we identified genes that are differentially expressed (had an adjusted *P* value lower than 0.001 and a  $|\log_2(\text{fold change})| > 0.2$ ) in the cycling persisters after regressing out known cell-cycle genes based on a published gene list<sup>5</sup> using the MASTDETest implemented in Seurat. Next, we used hypergeometric to test which gene signatures were enriched in this gene set. This resulted in 37 gene signatures (false discovery rate *q*-value <0.001). Finally, for each of the 37 gene signatures, we calculated an overall expression signature score per cell, as previously described<sup>32</sup> and the mean overall expression signature score per sample. Finally, signatures in which the mean signature level was lower for persister cells compared with untreated day 0 cells, were filtered out (see Supplementary Table 1 for final signature scores).

**Calculating overall expression signature scores.** Given a gene signature and a gene expression matrix, we first binned the genes to 10–50 expression bins (depending on data complexity) according to their average expression across the cells or samples. For each gene signature, we sampled 100 random compatible signatures for normalization. The

final reported score is computed by subtracting the 'real' mean signature score from the randomized one. For more detailed description of the method please refer to ref.<sup>32</sup>.

## Metabolite profiling

Polar cell extracts were profiled using negative and positive ionization mode using LC–MS methods. Negative ionization mode data were acquired using an ACQUITY UPLC (Waters) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX). Positive ionization data were acquired using an LC–MS system composed of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp) coupled to a Q Exactive hybrid quadrupole orbitrap mass spectrometer (ThermoFisher Scientific). For both modes, cell-sorted samples were extracted using 200  $\mu\text{l}$  80% methanol containing 0.5  $\text{ng } \mu\text{l}^{-1}$  inosine-<sup>15</sup>N<sub>4</sub>, 0.5  $\text{ng } \mu\text{l}^{-1}$  thymine-d<sub>4</sub>, and 1  $\text{ng } \mu\text{l}^{-1}$  glycocholate-d<sub>4</sub> as internal standards (Cambridge Isotope Laboratories).

For the negative extraction, 90  $\mu\text{l}$  of each sample was centrifuged (10 min, 9,000g, 4 °C) and the supernatants (10  $\mu\text{l}$ ) were injected directly onto a 150  $\times$  2.0 mm Luna NH2 column (Phenomenex, Torrance CA). The column was eluted at a flow rate of 400  $\mu\text{l min}^{-1}$  with initial conditions of 10% mobile phase A (20 mM ammonium acetate and 20 mM ammonium hydroxide (Sigma-Aldrich) in water (VWR)) and 90% mobile phase B (10 mM ammonium hydroxide in 75:25 v/v acetonitrile/methanol (VWR)) followed by a 10-min linear gradient to 100% mobile phase A. The ion spray voltage was –4.5 kV and the source temperature was 500 °C. Raw data were processed using MultiQuant 3.0.3 software (AB SCIEX) for automated peak integration.

For the positive extraction, 80  $\mu\text{l}$  of each sample was dried down using the turbovap (TurboVap LV, Caliper Life Sciences), Each sample was resuspended in 8  $\mu\text{l}$  of water and then crash with 72  $\mu\text{l}$  extraction solution 74.9:24.9:0.2 v/v/v acetonitrile/methanol/formic acid containing stable isotope-labelled internal standards (valine-d8, Sigma-Aldrich; and phenylalanine-d8, Cambridge Isotope Laboratories). Extracts were vortexed for 1 min, samples were spun at 10,000 rcf for 10 min at 4 °C and the resulting supernatant was moved to autosampler vials. Raw data were processed using TraceFinder software (Thermo Fisher Scientific) and Progenesis QI (Nonlinear Dynamics).

## Metabolomics data analysis

Relative abundance metabolite data from positive and negative ionization mode extractions were analysed in part using the MetaboAnalystR package<sup>33</sup>. First, metabolites containing any missing values across samples were removed from the dataset. For metabolites that were detected in both positive- and negative-ionization extractions, we only considered values from the negative extraction as it used a triple quad for the analysis, which often yields higher accuracy. Prior to statistical analysis, the values of each metabolite across samples were log-normalized and mean-centred. Significantly different metabolites detected by ANOVA were clustered by complete linkage hierarchical clustering and visualized in heat map format using the pheatmap R package. UMAP representation of metabolomics data was generated using the umap R package with local connectivity of 2 and *k* = 3.

## FAO measurements

For FAO assays, cells in 6-well plates were treated with 100  $\mu\text{M}$  etomoxir as indicated. Pulsing was performed in serum-free medium containing 1 mM carnitine with 0.75  $\mu\text{Ci}$  [9,10(*n*)-<sup>3</sup>H] palmitic acid (GE Healthcare) for 2 h. The medium was collected and eluted in columns packed with DOWEX 1X2-400 ion-exchange resin (Sigma) to analyse the released <sup>3</sup>H<sub>2</sub>O. <sup>3</sup>H<sub>2</sub>O was measured in counts per minute (CPM) and normalized to total cellular protein using BCA Protein Assay Kit (ThermoFisher Scientific).

## ROS measurements

To measure relative levels of ROS, drug treated and control cells were stained with 5  $\mu\text{M}$  CellROX Deep Red Reagent (Thermo Fisher) for

30 min at 37 °C, washed three times with PBS, trypsinized and analysed using MoFlo Astrios Cell Sorter (Beckman Coulter). To ensure a sufficient number of cycling cells for analysis, HT29 and A375 Watermelon models were profiled on day 10 and 21, respectively.

### Pooled persister experiment

Watermelon cell lines were generated as described above for PC9 cells. Sorted cell populations, 1,000 cells each, were expanded in culture for three passages, aliquoted to  $2 \times 10^6$  cells per vial and stored in liquid nitrogen. To derive persister cells, we treated EGFR-driven lung (PC9, HCC827), HER2-driven breast cancer (BT474, EFM192A), BRAF-driven melanoma (A375, COLO858, MMACSF), and colorectal (HT29) cell lines with osimertinib, lapatinib and dabrafenib, respectively. After 10 days of continuous drug treatment, cells were sorted based on mCherry intensity. Persister populations derived from different cell lines were combined to generate a cycling and a non-cycling cell pool. These pools together with the drug-naïve samples were subjected to 10X single-cell capture and profiling, as described above. To assess reproducibility, for a subset of models, we derived persisters from multiple vials of the same cell line and subjected them to cell capture and profiling (Extended Data Fig. 8e).

### Pooled persister analysis

Pooled scRNA-seq experiments were demultiplexed using Demuxlet<sup>34</sup>. First, allelic fractions for each cell line were estimated across a reference set of 1,000,000 SNPs<sup>19</sup> from bulk RNA-seq data using FreeBayes<sup>35</sup>. The FreeBayes settings ‘pooled-continuous’ and ‘report-monomorphic’ were used and a pseudo-count of 1 was added to the reference and alternate allele read counts. Then, Demuxlet was run using the PL tag and a fixed mixing proportion, alpha, of 0.5.

### ORF assays

Purified ORF expression vectors for CPT1A and GPX2 were ordered from Genecopoeia (CPT1A: EX-A1436-Lv156, GPX2: EX-A3079-Lv156). PC9 cells harbouring the Watermelon construct were transfected with individual vectors and treated with drug as previously described. We confirmed increased expression of each ORF by quantitative PCR with reverse transcription (RT–qPCR). RT–qPCR primers were designed to: (1) capture both endogenous and plasmid sequences of each gene; (2) span only gene introns; (3) produce amplicons no larger than 200 bp; and (4) have melting temperatures of  $60 \pm 1$  °C (primers listed in Supplementary Table 4). For each sample, two replicates of 20,000 cells each were lysed in RLT buffer (Qiagen 79216) and RNA was isolated using Dynabeads MyOne Silane beads (ThermoFisher Scientific 37002D). Next, samples were treated with TURBO DNase (ThermoFisher Scientific AM2238), and subsequently purified again with silane beads. Reverse transcription was performed using SuperScript II Reverse Transcriptase (ThermoFisher Scientific 18064022) according to the manufacturer’s protocol. The resultant cDNA was diluted 1:10 and quantitative PCR was performed with SYBR Green Master Mix (ThermoFisher Scientific 4368706).

### Genetically engineered mouse model experiments

Tumour-bearing mice were treated with osimertinib at 25 mg kg<sup>-1</sup> for 5 consecutive days each week (5 days on–2 days off) when their tumour size reached around 300–500 mm<sup>3</sup>. The treatment lasted for four weeks before mouse lungs were excised. Tumour burden was measured by MRI at least once before treatment and once before the mice were euthanized (further details in Supplementary Information).

### Statistical analyses

Statistical tests and graphing of data were performed with GraphPad Prism v.7.0a and R v. 3.6.1 unless noted otherwise. Unless otherwise noted, *P* values were calculated using unpaired, two-tailed *t*-tests assuming unequal variance. Multiple hypothesis correction was done using Holm’s method.

### Reporting summary

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

### Data availability

RNA-seq data have been deposited in the NCBI Genome Expression Omnibus (GEO) under the accession code GSE150949. The Watermelon library and plasmid are available on Addgene (Addgene IDs 155257 and 155258). Source data are provided with this paper.

### Code availability

Code used in this study is available from <https://github.com/yaaraore/watermelon> (additional code is available upon request).

30. Reyes, J. et al. Fluctuations in p53 signaling allow escape from cell-cycle arrest. *Mol. Cell* **71**, 581–591 (2018).
31. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
32. Jerby-Arnon, L. et al. A cancer cell program promotes t cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997 (2018).
33. Chong, J., Yamamoto, M. & Xia, J. MetaboAnalystR 2.0: from raw spectra to biological insights. *Metabolites* **9**, E57 (2019).
34. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
35. Erik Garrison, G. M. Haplotype-based variant detection from short-read sequencing. Preprint at <https://arxiv.org/abs/1207.3907> (2012).

**Acknowledgements** We thank the Broad Cytometry Facility (P. Rogers, S. Saldi, C. Otis and N. Pirete); L. Gaffney and A. Hupalowska for help with figure preparation and artwork; A. Martinez Gakidis for scientific editing; the Yale Center for Genome Analysis (F. Lopez-Giraldez and D. Zhao); and the Yale Center for Research Computing for data processing, guidance, RNA-seq and use of the research computing infrastructure, specifically the Ruddle Cluster. Y.O. is supported by the Hope Fund for Cancer Research, Grillo-Marxuch Postdoctoral Fellowship and the Rivkin Scientific Scholar Award. M. Tsabar was supported by the American Cancer Society–New England Pay-if Group Postdoctoral Fellowship, PF-18-126-01-DMC. H.F.C. was supported by a V Foundation Scholar Award (D2015-027). P.I.T. is supported by an NIH F32 Postdoctoral Fellowship from National Institute of Allergy and Infectious Disease (1F32AI138458-01). Funding for breast cancer patient study came from grants from Sanofi and GlaxoSmithKline. A.N.H. was supported by NIH/NCI K08 CA197389. S.A.H. was supported in part by NCI/NIH CA016042 as well as the Marni Levine Memorial Research Award. B.H. received support from a Lo Graduate Fellowship for Excellence in Stem Cell Research. Additional support came from R01CA121210, R01CA120247 and P50CA196530. Yale Cancer Center Shared Resources used in this article were in part supported by NIH/NCI Cancer Center Support Grant P30 CA016359. The work was supported by the Klarman Cell Observatory, the NHRRI Center for Cell Circuits, and Howard Hughes Medical Institute (A.R.) as well as by the Breast Cancer Research Foundation–BCRF-16-020 and the Sheldon and Miriam Adelson Medical Research Foundation (J.S.B.).

**Author contributions** Study design, data interpretation and preparation of the manuscript: Y.O., J.S.B. and A.R. Execution of experiments: Y.O., M. Tsabar, H.F.C., M.S.C., L.A.-Z., E.Z., P.I.T., B.H., A.D., K.A.P. and B.M.C. Computational and statistical analysis: Y.O., J.-C.H., M.S.C., W.C., M. Tabaka and M. Tsabar. Resources: M. Tabaka, C.P.F., S.A.H., D.J.S., G.L., C.C., A.N.H. and K.P.

**Competing interests** A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and was a scientific advisory board member of ThermoFisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov until 31 July 2020. From 1 August 2020, A.R. has been an employee of Genentech. J.S.B. is a consultant for Agios Pharmaceuticals, eFFECTOR Therapeutics, and Frontier Medicines. Y.O., A.R. and J.S.B. are inventors on US patent application 16/563,450 filed by the Broad Institute to express barcode libraries as described in this manuscript. C.P.F. is now an employee of Bristol Myers Squibb. K.P. is co-inventor on a patent licensed to Molecular MD for EGFR(T790M) mutation testing (through MSKCC). K.P. has received Honoraria/Consulting fees from Takeda, NCCN, Novartis, Merck, AstraZeneca, Tocagen, Maverick Therapeutics, Dynamo Therapeutics, Halda and research support from AstraZeneca, Koltan, Roche Boehringer Ingelheim and Symphogen. A.N.H. a consultant for Nuvalent and is supported by Novartis, Pfizer, Amgen, Blueprint Medicines, Lilly, Roche/Genentech, Nuvalent, Relay Therapeutics. S.A.H. has contracted research with Ambrx, Amgen, AstraZeneca, Arvinas, Bayer, Daiichi-Sankyo, Genentech/Roche, Gilead, GSK, Immunomedics, Lilly, MacroGenics, Novartis, Pfizer, OBI Pharma, Pieris, PUMA, Radius, Samumed, Sanofi, Seattle Genetics, Dignitana, Zymeworks, Phoenix Molecular Designs and Lilly, and stock options in NK Max.

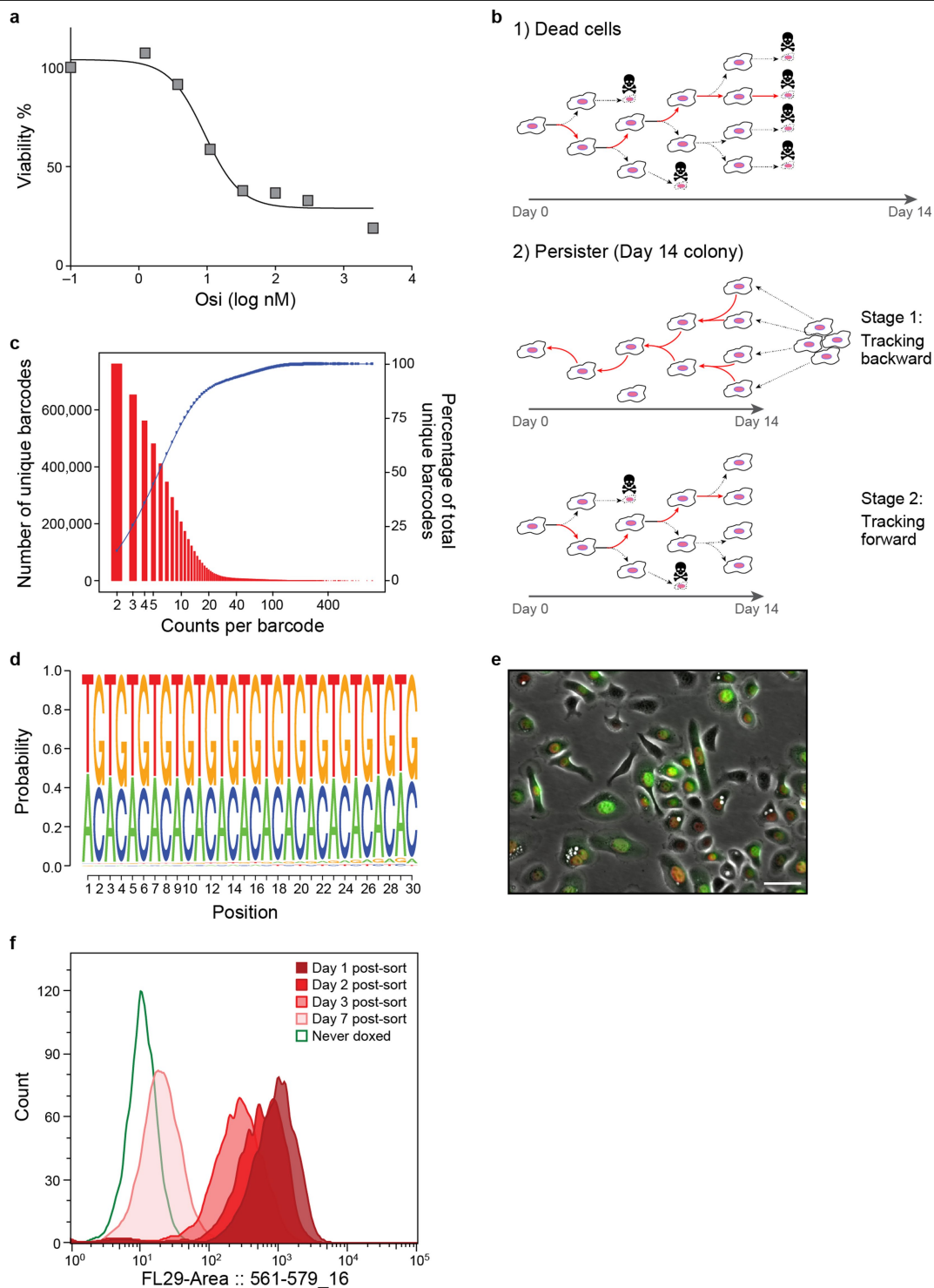
### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03796-6>.

**Correspondence and requests for materials** should be addressed to J.S.B. or A.R.

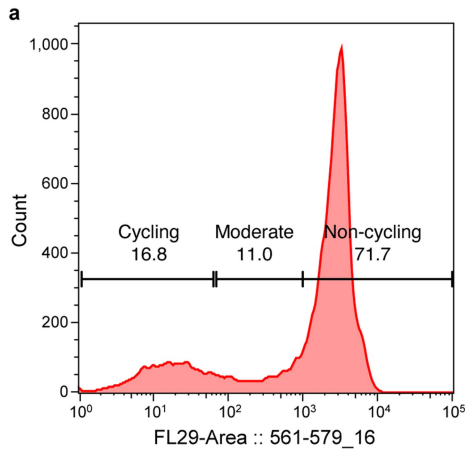
**Peer review information** Nature thanks Navdeep Chandel, Ken Duffy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



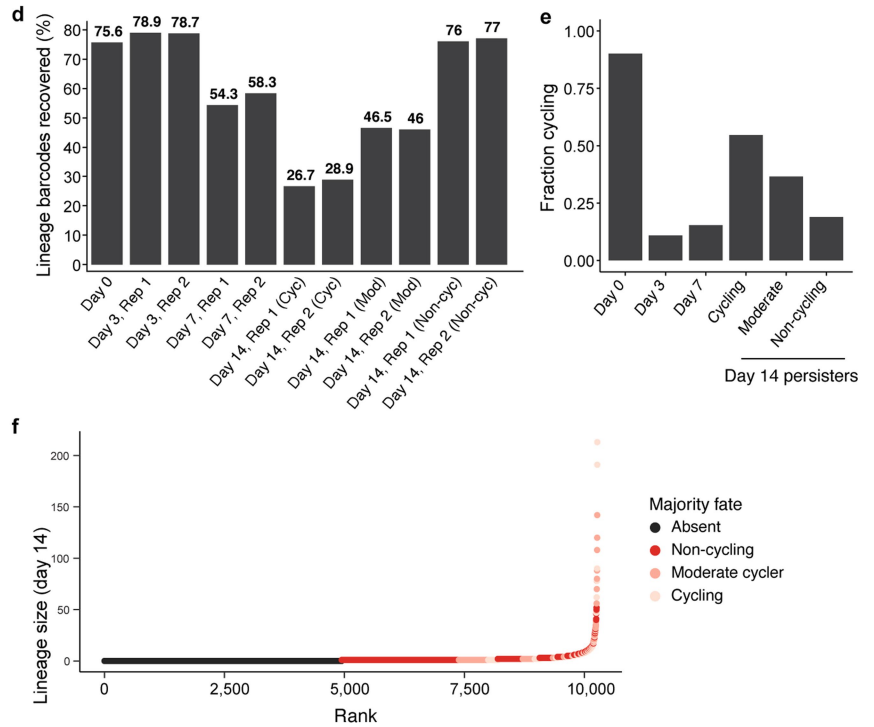
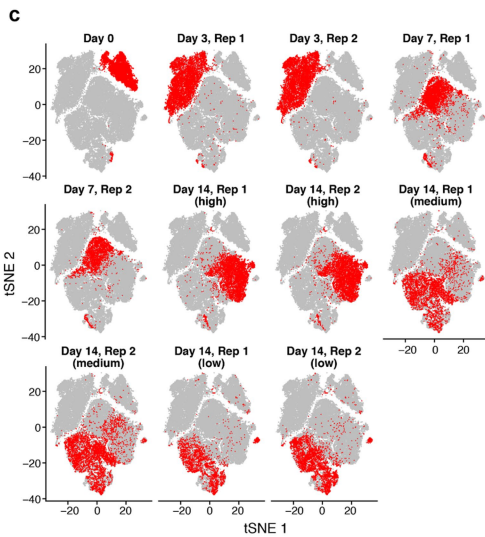
**Extended Data Fig. 1 | Lineage detection efficacy and fluorescent dilution capacity of the Watermelon library.** **a**, Viability of PC9 cells treated osimertinib. % of viable PC9 cells (y axis) after 72 h of treatment with osimertinib at different concentrations (x axis). **b**, Schematic of tracking. (1). Tracking of non-persister cells. Red arrows: cell that was tracked in the lineage. (2). Tracking of persister cells. Colonies detected at day 14, and the time lapse is viewed and tracked from day 14 backward to day 0 to detect the common initiator cell. After the initiator cell is detected, this initiator is tracked forward as in **a** (red line indicates the tracked cell). **c**, Watermelon library complexity. Distribution of number of unique lineage barcodes (y axis, red bars) in a Watermelon plasmid library sequenced at a depth of  $\sim 68 \times 10^6$  reads. Blue curve:

cumulative wealth distribution of unique barcodes. **d**, Watermelon library sequence diversity. Sequence logo of nucleotide composition at each position (x axis) relative to the beginning of the barcode sequence of 5,472,944 unique lineage barcodes detected in the Watermelon library. **e**, PC9-Watermelon cell line grown in dox containing media. A merge of the green, red and bright field channels is shown. Scale bar  $20 \mu\text{m}$ . **f**, Fluorescence dilution of H2B-mCherry over time reports proliferative history. Distributions of mCherry fluorescence level (x axis) for  $n = 3000$  cells analysed by flow cytometry at each time point (colour legend) from cells transduced with the Watermelon vector, exposed to dox for 48 h, sorted for red positive cells and seeded in separate wells at  $t = 0$  (Methods). Data are representative of two independent experiments.



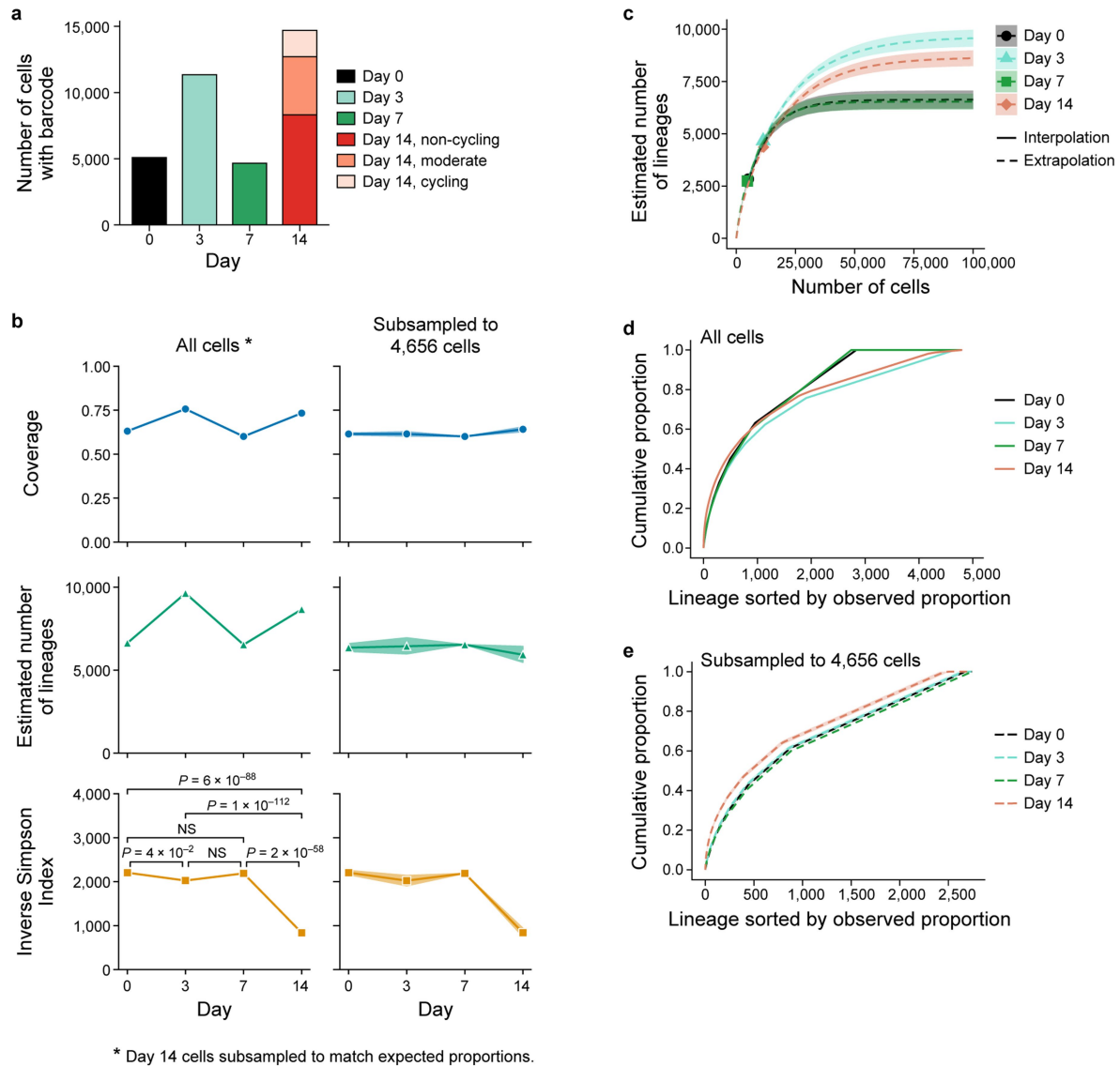
**b**

Sample	Post-filter cell number
Day 0	6778
Day 3 replicate 1	7093
Day 3 replicate 2	7273
Day 7 replicate 1	4770
Day 7 replicate 2	3515
Day 14 cycling (red low) replicate 1	2690
Day 14 moderate (red med) cyclers replicate 1	4221
Day 14 non-cycling (red high) replicate 1	5380
Day 14 cycling (red low) replicate 2	4153
Day 14 moderate (red med) cyclers replicate 2	5097
Day 14 non-cycling (red high) replicate 2	5449



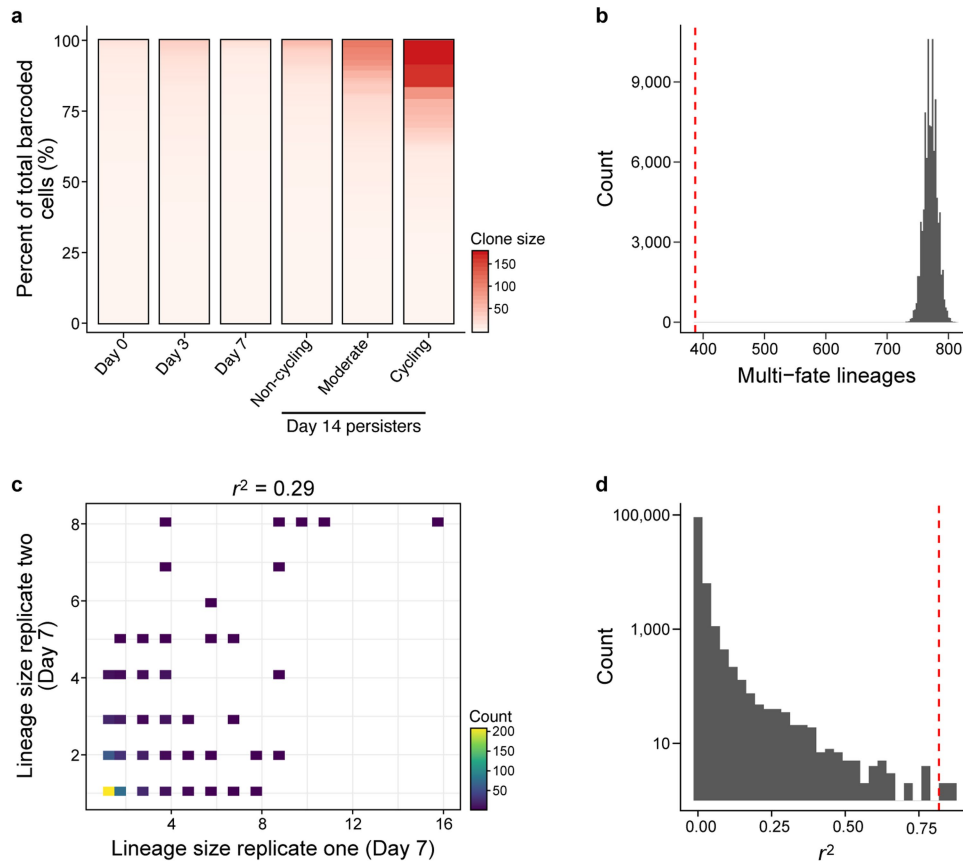
**Extended Data Fig. 2 | scRNA-seq along a time course of osimertinib treated PC9-Watermelon cells. a**, Sorting strategy. Distribution of mCherry fluorescence level (x axis) in Watermelon-PC9 cells gates at day 14 of osimertinib treatment, marked by representative sorting gates used to sort three persisters subpopulations: cycling, moderate cyclers and non-cycling. **b**, Number of high-quality cells profiled in each sample. **c**, Changes in expression profiles following treatment. *t*-stochastic neighbourhood embedding (tSNE) of 56,419 PC9-Watermelon cell profiles (dots), coloured (red)

by the labelled time point. **d**, Assignment of cells to lineages by lineage barcode. Percent of cells (y axis) at each time point/subpopulation (x axis) that have a detected lineage barcode. **e**, Identification of cycling cells. Percent of cells (y axis) at each time point/subpopulation (x axis) that express either the G2/M or S phase signature. **f**, Majority fate. Clone size on day 14 (y axis) of each persister lineage barcode inferred from scRNA-seq ordered by ascending rank order (x axis) and coloured by majority fate based on flow sample provenance of the captured cells.



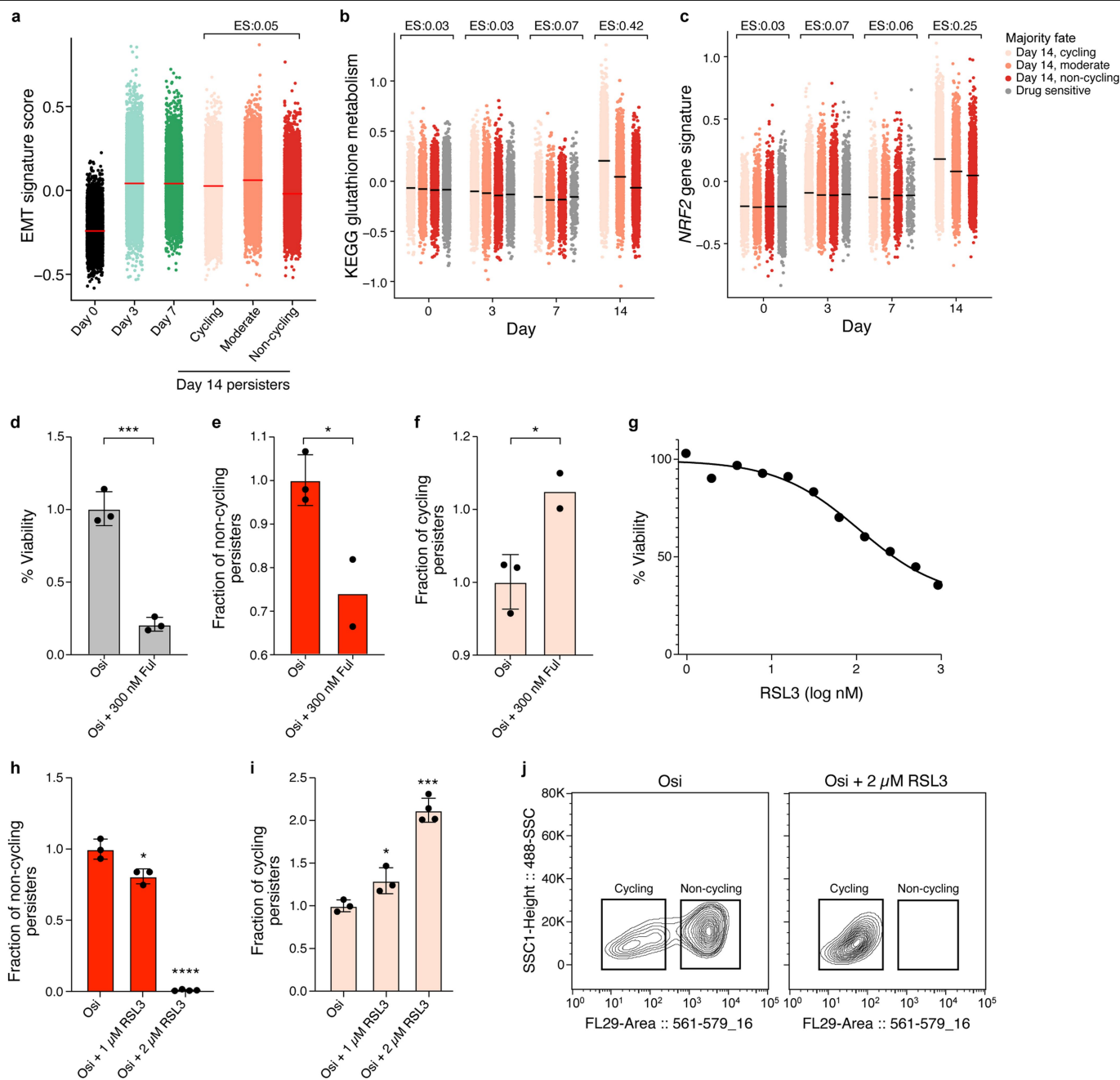
**Extended Data Fig. 3 | Estimates of lineage diversity.** **a**, Difference in number of cells profiled per time point. Number of cells (y axis) with captured lineage barcode at each day (x axis). Day 14 cells are partitioned by the three mCherry populations (legend). **b–e**, Species diversity estimators can be biased by coverage. Estimated sample coverage (cumulative proportion of all lineages in the total population that were observed, top, y axis, Methods), estimated number of lineages in the population (middle, y axis, Methods), and estimated inverse Simpson Index, also known as Hill number of order 2 (bottom, y axis, Methods) at each time point (x axis), computed from all cells with barcodes (left) or subsampled without replacement to match the smallest number of cells per time point, 4,656 cells on day 7 (right). Confidence bands (shaded area) indicate the empirical pointwise 95% coverage confidence interval over 1,000 subsampling repetitions. Since standard species richness estimators are not suited for the analysis of estimated proportions from stratified sampling, we randomly subsampled 8,320, 1,949, and 1,276 day 14 cells without replacement from the cycling and moderate cyclers and non-cycling population, respectively (left panel, for unsorted population proportions see Extended Data Fig. 2a). P-values obtained by (asymptotic) two-sided Welch's

t-test with bootstrap estimated standard errors, Holm-corrected with level 5% (Methods,  $n = 5,087$ , day 0,  $n = 11,348$ , day 3,  $n = 4,656$ , day 7,  $n = 11,545$ , day 14 subsampled). **c**, Alternative estimates of number of lineages with rarefaction. Rarefaction curves for the expected observed number of different lineages (y axis) at varying hypothetical sample sizes (x axis) for each time point (coloured lines). Actual number of observed lineages: marker; Interpolated results: solid lines; Extrapolation beyond the observed number: dashed lines. Day 14 cells were subsampled as done for the estimation of the number of lineages in the right-hand side panels of **b**. Shaded areas: confidence bands at 95% confidence level. **d, e**, Estimated cumulative proportion (eCDF) of lineages in the total population (y axis) sorted in decreasing order of estimated lineage proportion (x axis) for each time point (coloured lines) when estimating the proportion from all sequenced cells with barcodes (**d**) or subsampled to 4,656 cells (**e**) as in **b**. Subsampling (**b, e**) and rarefaction (**c**) facilitate comparison between different time points since estimators of population diversity are strongly biased by sample size. Confidence bands indicate the empirical pointwise 95% coverage confidence interval over 1000 repetitions of the subsampling.



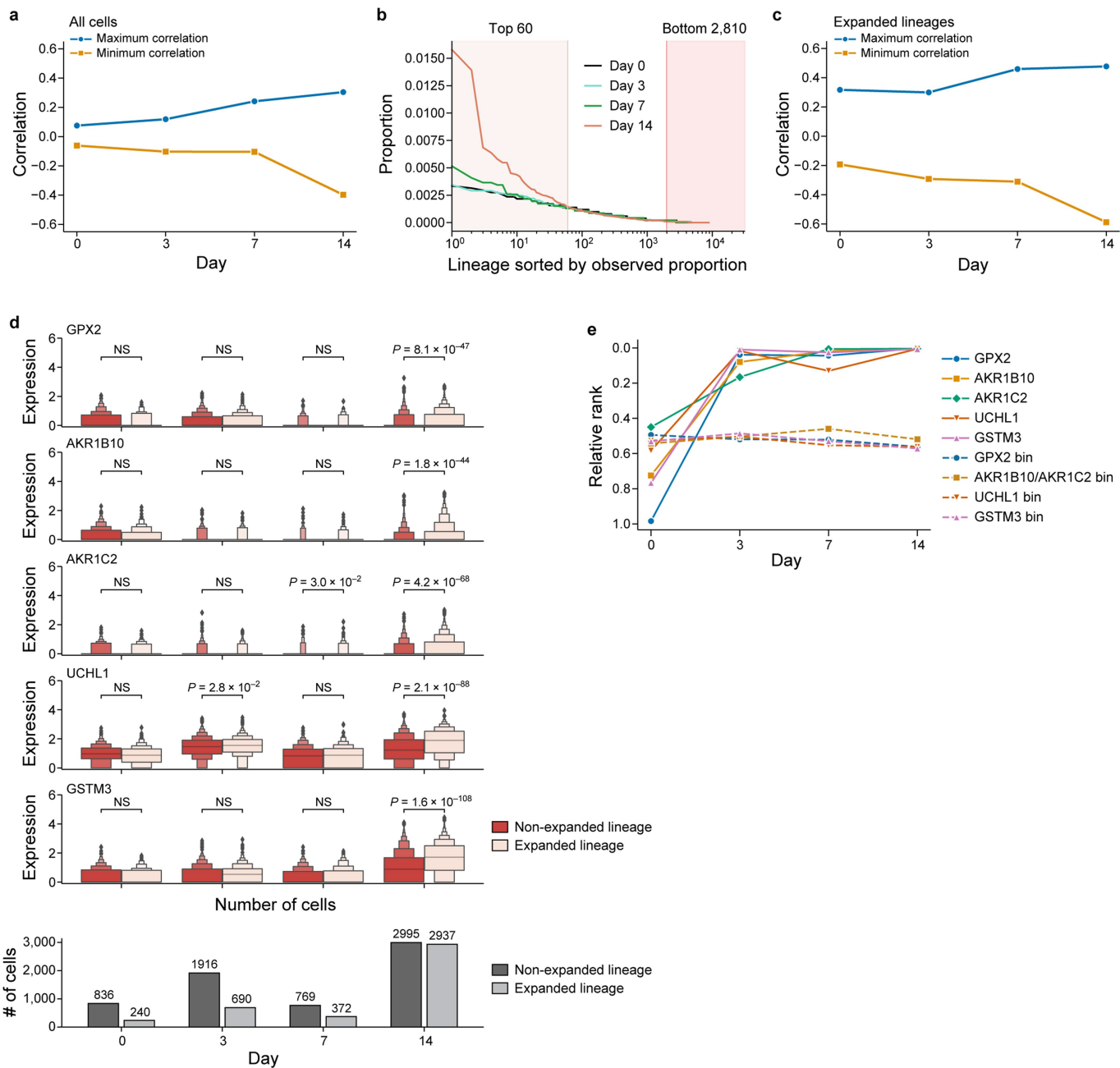
**Extended Data Fig. 4 | Lineage fate analysis.** **a**, Single cell-derived clone size by sample. In each sample, detected barcodes were sorted in descending order by the sum of their counts. Each unique lineage barcode was accounted as a separate clone. **b**, The number of observed multi-fate lineages is significantly smaller ( $P=1 \times 10^{-5}$ ) than expected by chance. Distribution of the number of multi-fate lineages (x axis) in simulated data. Red line: observed number of

multi-fate lineages. **c, d**, Clone size reproducibility is significantly higher than expected by chance. **c**, Clone size on day 7 of each persister lineage barcode inferred from scRNA-seq (x, y axes) in each of two independent experiments seeded from the same barcoded founding cell population. Top: linear correlation coefficient. **d**, Distribution of  $r^2$  values (x axis) in simulated day 14 data. Red line: observed  $r^2$  between the two replicates at day 14.



**Extended Data Fig. 5 | Differences in transcriptional programs and drug response in cycling and non-cycling persisters.** **a**, EMT signature expression is similar in cycling and non-cycling persisters. Distribution of expression levels of EMT (y axis,  $\log_2(\text{TPM}+1)$ ) across time points and subpopulations. Effect size (ES): difference between the mean signature score of cycling and non-cycling persisters. **b, c**, Higher expression of glutathione metabolism and NRF2 signatures in cycling *vs.* non-cycling persisters. Signature score (y axis) of glutathione metabolism (**b**) and NRF2 pathway (**c**) signatures in cells profiled at each time point (x axis) stratified by their lineage majority fate at day 14 (colour legend). Effect size (ES) indicates difference between the mean signature score of cycling and non-cycling persisters. **d-f**, Persister populations show differential sensitivity to fulvestrant. Effect of fulvestrant co-treatment

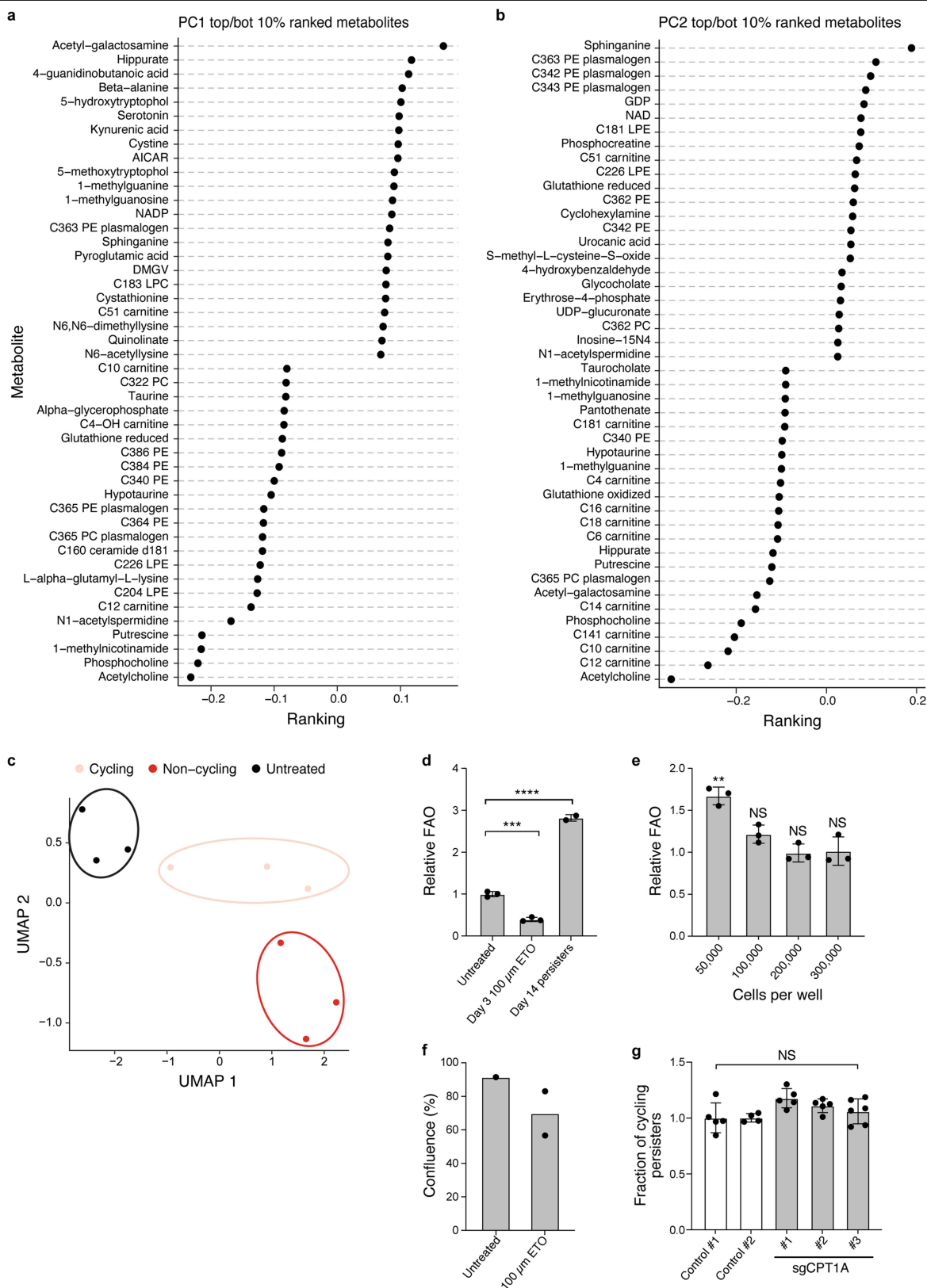
(300nM fulvestrant during days 14-20 of 300nM osimertinib) on overall survival (**d**) non-cycling (**e**) and cycling cells (**f**). **g-i**, Persister populations show differential sensitivity to RSL3. Effect of different doses of RSL3 co-treatment (during days 3-14 of 300nM osimertinib) on overall survival (**g**) non-cycling (**h**) and cycling cells (**i**). **j**, Co-treatment with 2  $\mu\text{M}$  of RSL3 shifts surviving persister cells to cycling. Distributions of mCherry fluorescence level (x axis) for cells analysed by flow cytometry with and without RSL3 co-treatment (panel legend). Error bars are mean  $\pm$  s.d. of two (**e, f**) or three biologically independent experiments (**d, h, i**). \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ ; two-tailed *t*-tests compared to osimertinib only condition (**d-f, h-i**).



**Extended Data Fig. 6 | Correlation between clone size at day 14 and gene expression.** **a**, Gene expression is progressively more predictive of persister lineages over time. For each time point (x axis), maximum (blue) and minimum (orange) over the correlation coefficients (y axis) of each gene and the lineage size at day 14 (also see Fig. 2g). **b**, Choosing expanded and non-expanded lineages for gene expression comparisons. Cut-offs (vertical lines) for highly expanded and non-expanded lineages on day 14 based on the estimated proportion of each lineage in the population (y axis), sorted by decreasing proportion (x axis, log scale) for each time point (coloured lines). **c**, For each time point (x axis), maximum (blue line) and minimum (orange line) over the correlation coefficients (y axis) of each gene and the lineage size at day 14 as in **a**, but restricted to cells of highly expanded lineages as in **b**. **d**, Genes with top correlation to lineage expansion. Top five rows: distribution of gene expression of top correlated genes (log normalized counts, y axis) at each time point (x axis), comparing cells from non-expanded (red) and expanded (pink)

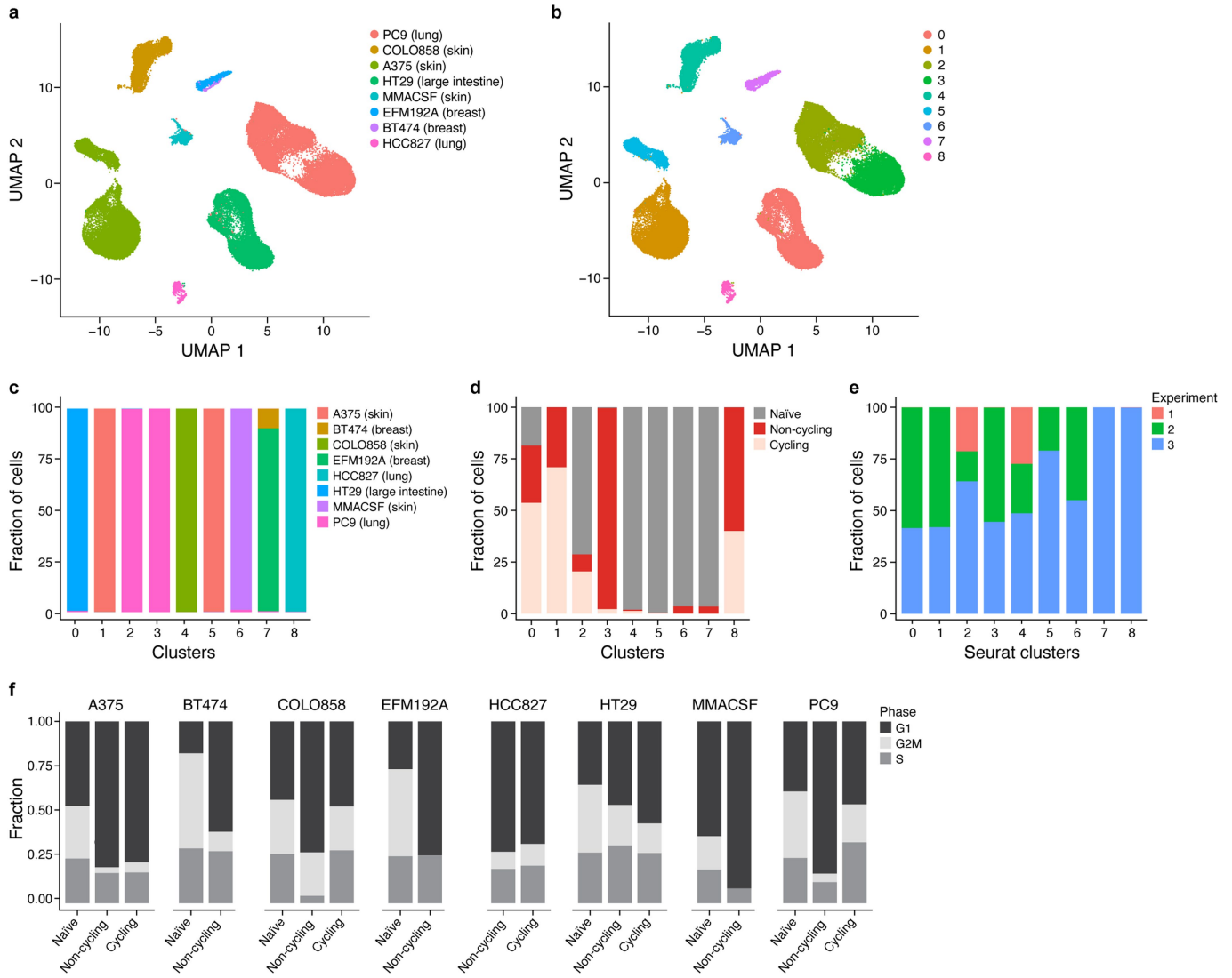
lineages, as defined in **b**. Bottom row: numbers of cells (y axis) per time point in non-expanded, (dark grey) and expanded (light grey) lineages. Distributions are visualized as enhanced box plots indicating median (grey bar) and geometric progression of quantiles (progressively decreasing box widths for 75th, 87.5th, 93.75th, 96.875th, etc. percentiles, and analogously for 25th, 12.5th, 6.25th, 3.125th, etc. percentiles, labelling up to 1.5625% of the data as outliers). Bonferroni-Holm adjusted P values, determined by a two-sided Mann-Whitney U-test with continuity correction, or no significance (NS,  $P > 5\%$ ). **e**, Increase in correlation of top correlated genes as early as day 3. For each time point (x axis), rank of selected genes' (coloured solid lines) correlation with the lineage size at Day 14 among all genes (y axis), normalized to lie between 0 and 1, and average relative correlation rank of genes with similar mean expression as determined by grouping genes by their mean log-normalized expression over all time points combined into 20 bins (coloured dashed lines) (Methods).





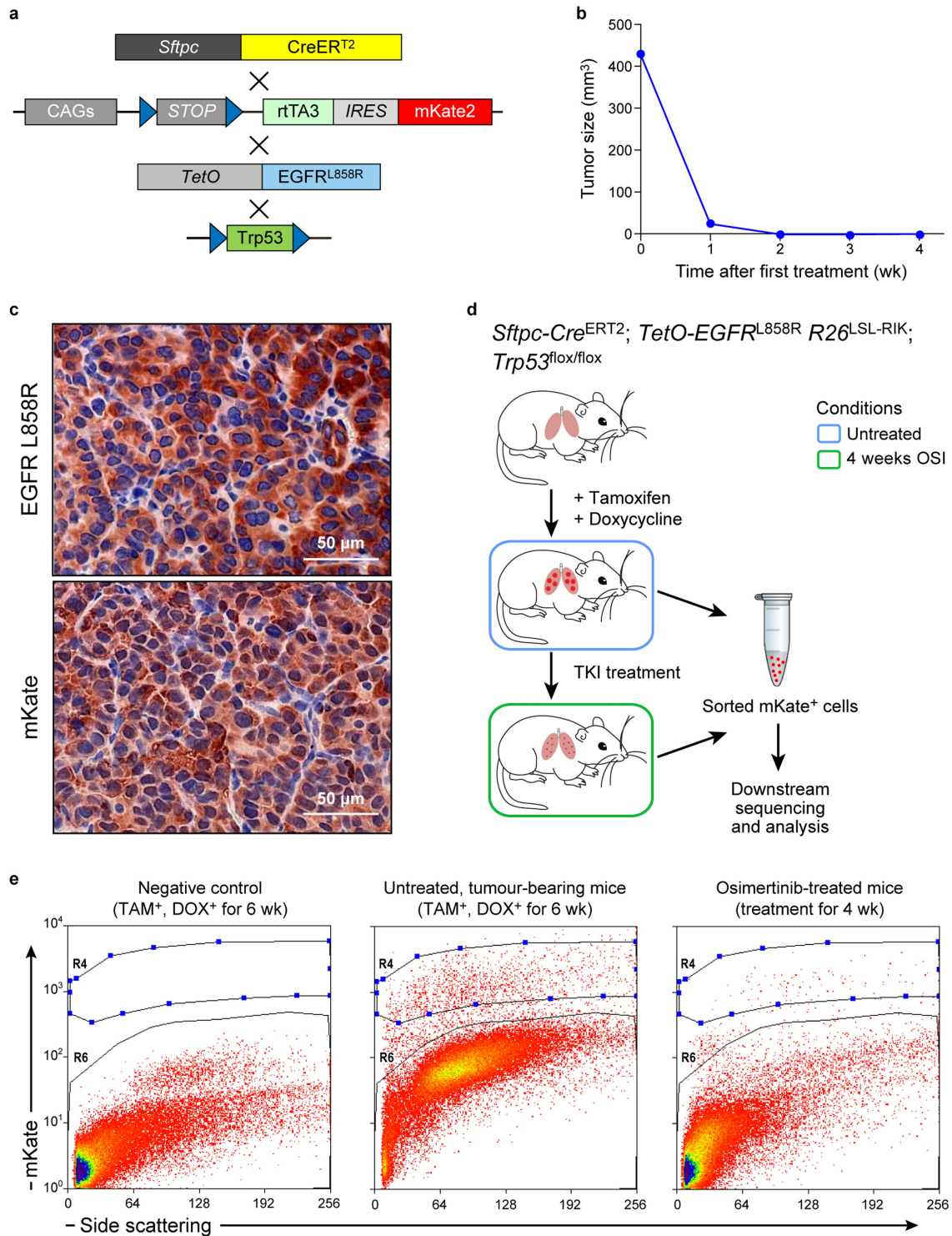
**Extended Data Fig. 7 | Metabolite profiles of cycling persisters, non-cycling persisters and untreated parental cells and FAO measurements.** **a, b**, PCA loadings (x axis) for the top 46 metabolites (y axis) associated with PC1 (**a**) and PC2 (**b**). **c**, UMAP representation of metabolomics data. **d**, Mean FAO level (y axis, relative to mean of the untreated controls) measured by 3H-palmitic acid oxidation in PC9-Watermelon cells either untreated, treated only with 100 μm etomoxir for 3 days, or treated with 300nM osimertinib for 14 days. **e**, Mean FAO levels (y axis, relative to cells seeded at 300,000 per well, as used for the

osimertinib time course) in PC9-Watermelon cells seeded at different densities (x axis) 24 h before measurement. two tailed *t*-tests; \*\* $P < 0.01$ ; NS – not significant (compared to 300,000 cells per well). **f**, Mean confluence (y axis) of PC9-Watermelon cells treated with 100μm Etomoxir for 14 days (Methods). **g**, Mean fraction of cycling persisters for control and sgCPT1A PC9 cells. Error bars are mean +/- s.d. of two (**f**) or three biologically independent experiments (**d, e, g**). \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ ; NS, not significant ( $P > 0.05$ ); two-tailed *t*-tests (**e-g**).



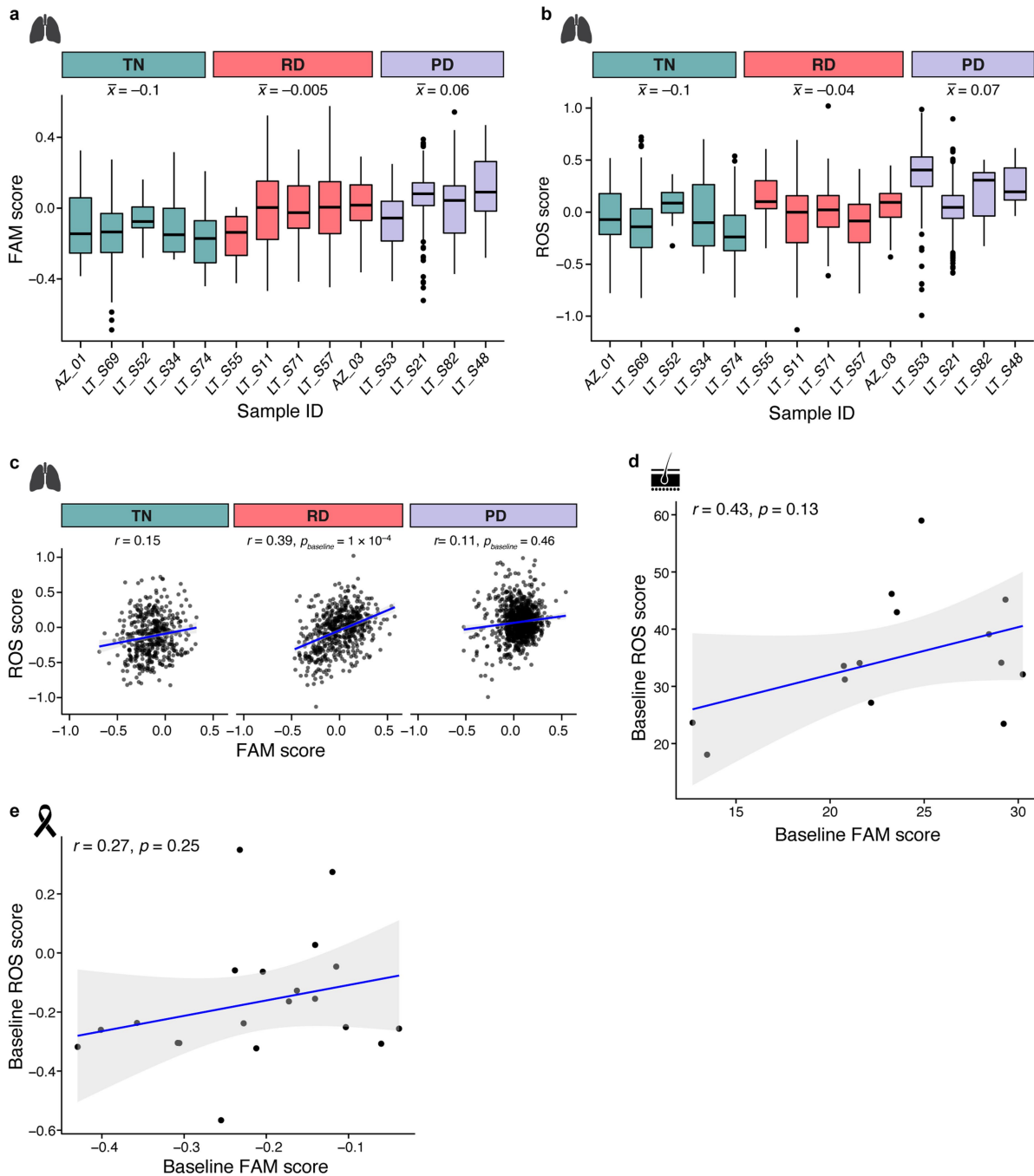
**Extended Data Fig. 8 | Single cell analysis of multiple Watermelon persister models. a, b,** UMAP representation of cells coloured by cell line (a) and cluster identity (b). **c–e,** Fraction of cells (y axis) in each cluster (x axis) coloured by cell

line (c), treatment (d) and experimental replicate (e). **f,** Proportion of cells (y axis) in each cell cycle phase (coloured stacked bars) based on cell cycle scores inferred from scRNA-seq data across cell line models.



**Extended Data Fig. 9 | Modelling minimal residual disease using an engineered transgenic mouse model. a**, Transgenic composition of the EGFR<sup>L858R</sup> genetically engineered mouse model used in this study. **b**, Tumour burden in a representative osimertinib-treated mouse as measured by MRI. **c**, IHC staining for EGFR L858R mutant and mKate in treatment-naive mouse

lung tumours. **d**, Experimental schema for isolation and molecular profiling of mKate<sup>+</sup> cells. **e**, Flow gating scheme for mKate<sup>+</sup> cells. R4 cells from untreated and mice bearing minimal residual disease were sorted and subjected to sequencing. mKate negative cells (R6). Images in **c** are representative of three independent experiments.



**Extended Data Fig. 10 | Changes in expression of metabolic programs in patient tumours. a, b,** Increase in FAM and ROS pathway signatures in drug-treated human lung adenocarcinoma. Distribution of expression scores of FAM (**a**) and ROS (**b**) signatures in cells from individual EGFR-driven lung adenocarcinoma tumours (with more than 10 cells) across different treatment time points (x axis). Box plots are represented by centre line, median; box limits, upper and lower quartiles; whiskers extend at most  $1.5 \times$  interquartile

range past upper and lower quartiles.  $\bar{x}$ : Mean signature level for time point. For number of cells per patient see Supplementary Table 3. **c–e,** Correlation between ROS (y axis) and FAM (x axis) signature scores in (**c**) treatment-naive (TN), residual disease (RD) and progressive disease (PD) human lung adenocarcinoma, (**d**) treatment-naive melanoma, and (**e**) treatment-naive breast cancer. Significance based on bootstrap test (**c**, Methods) and t distribution (**d, e**), 95% confidence interval (**d, e**, shaded area).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

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

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

Raw gene expression data and processed counts are available on GEO, accession no. GSE150949. Figure 4 and Extended Data Figure 6 were generated using raw data from NCT01072175 trial (melanoma), NCT00769470 trial (breast) and NCT03433469 (lung).

## Field-specific reporting

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

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size

Data exclusions

Replication

Randomization

Blinding

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a Involved in the study

Antibodies

Eukaryotic cell lines

Palaeontology and archaeology

Animals and other organisms

Human research participants

Clinical data

Dual use research of concern

### Methods

n/a Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

Authentication

Mycoplasma contamination

Commonly misidentified lines (See [ICLAC](#) register)

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

Outcomes

## Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

Sample preparation

PC9 cells were trypsinized, resuspended in PBS and sorted live

Instrument

Astrios EQ Cell Sorter

Software

Astrios EQ Cell Sorter proprietary software and flowJo v10

Cell population abundance

Purity was high based on sequencing and function assay

Gating strategy

For all flow experiments live and singlet gates were used:  
Live: x-FSC1-Height  
y-SCC1-Height  
Singlet: x-SCC1-Height  
y-SCC1-Width  
Additional gates are described in Supplementary Information

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