

Untreated controls reveal that pre-treatment prediction of cancer cell fate mostly reflects growth fitness

Rabeya Noor

`rabeyanoor.dev@gmail.com`

Daffodil International University

Khairum Jahan Roza

Kumudini Women's Medical College and Hospital

Research Article

Keywords: Drug tolerance, Lineage tracing, Single-cell RNA-seq, Sister cells, Cell fate prediction, Non-genetic resistance

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11337829/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Untreated controls reveal that pre-treatment prediction of cancer cell fate mostly reflects growth fitness

Rabeya Noor^{1*} and Khairum Jahan Roza²

¹Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia, Savar, Dhaka 1216, Bangladesh

²Kumudini Women's Medical College and Hospital, Mirzapur, Tangail, Bangladesh

*Correspondence: rabeyanoor.dev@gmail.com

Email addresses: RN, rabeyanoor.dev@gmail.com; KJR, kj.roza.dr@gmail.com

ORCID iDs: 0009-0005-1045-5927 (RN); 0009-0007-2839-0900 (KJR)

Abstract

Background: Some cancer cells survive treatment without genetic change, and many studies suggest they can be recognised in advance from their gene expression. Sister-cell lineage tracing tests this. One daughter of a dividing cell is sequenced before treatment and the other treated. However, a cell whose sister survives may simply grow well. How much of the predictable signal reflects drug tolerance rather than growth fitness is unknown.

Results: Using the identically processed untreated arm of the ReSisTrace experiment, we compared each drug-arm predictor with a fitness model trained on untreated cells. In 32,642 pre-treatment ovarian cancer cells, a linear model predicted the sister's fate with an area under the receiver operating characteristic curve (AUROC) of 0.72–0.75, against 0.50–0.60 for published methods on olaparib; a graph neural network and two single-cell foundation models did no better. The fitness model, which never saw a drug, reached 0.70–0.73. The drug-specific gain was about 0.02 AUROC, positive in every arm, not reliably separable from zero, and at most about 0.05. The shared signal marked a proliferative, low-stress state that a label-free score largely recovered. In a second lineage system, pre-treatment clone size

outpredicted expression under all six treatments, and in 11 patients the fitness state was depleted, not enriched, after chemotherapy.

Conclusions: Most of what a pre-treatment transcriptome predicts about a cancer cell's fate is growth fitness rather than drug tolerance. Such claims should be tested against an untreated arm processed in the same way; our package `persist.fitadj` does this for any lineage-tracing design.

Keywords: Drug tolerance, Lineage tracing, Single-cell RNA-seq, Sister cells, Cell fate prediction, Non-genetic resistance

Background

Cancer therapy often fails because some cells survive it, and survival does not always need a mutation. Sharma et al. [1] described a small, reversibly drug-tolerant subpopulation in several cancer cell lines. Such persister cells can later acquire genetic resistance [2, 3], and non-genetic resistance is now recognised as a broad class of adaptive mechanisms [4, 5]. A central question is whether the cells that survive were already different before the drug was given. If they were, their state could be measured, and perhaps targeted, in advance.

Several studies suggest that they were. In melanoma, rare cells expressing resistance markers at high levels were more likely to become resistant [6]; such fluctuations last for several divisions [7], can be traced back to rare drug-naïve precursors [8], largely decide which resistant type a clone forms [9], and can be modulated before treatment [10]. Clones resistant to one of six treatments were more likely to resist others [11], and time-lapse lineages suggest that survival is biased by states inherited over two to three generations [12]. Other studies point the other way. In colorectal cancer, all cells rather than a rare subpopulation may enter a persister state [13]; in organoids, resistance arose from transient plasticity [14]; and sisters diverge quickly in their response to death signals [15, 16]. Part of a cell's fate is therefore decided after treatment begins.

Lineage barcodes read out by single-cell RNA sequencing link a cell's state to the fate of its relatives at scale [17]. They have revealed primed fate in haematopoiesis [18, 19], clonal dominance in leukaemia [20, 21] and drug-tolerant lineages in lung and breast cancer [22, 23]. Most of these designs label whole clones, however, and a large clone is more likely to be seen again simply because it has more cells. ReSisTrace avoids this [24]. Barcoded cells divide once, one sister is sequenced before treatment and the other is treated, so each sequenced cell is labelled by its own sister's fate. The design also includes an untreated control arm that goes through the same synchronisation, splitting and re-plating steps, which the authors added to assess the confounding effect of growth fitness. They found transcriptomic differences in this arm too, but did not build classifiers or report how well fate could be predicted.

Computational tools for single-cell drug response, such as scDEAL [25], SCAD [26], Beyondcell [27] and CaDRReS-Sc [28], were built for a different problem. A recent benchmark of nine such methods used the ReSisTrace olaparib arm as ground truth [29]. When pre-treatment cells were compared with post-treatment survivors, the best method reached a mean area under the receiver operating characteristic curve (AUROC) of 0.869. When only pre-treatment cells with sister-derived labels were used, most methods fell to 0.50–0.55 and the best, scDEAL, to about 0.6. Neither this benchmark nor the methods it evaluated compared their predictions with the untreated arm, so a score on such a task cannot say whether a method has learned tolerance of the drug or general growth fitness.

We therefore ask three questions on the ReSisTrace data (Fig. 1a). How much does a cell's pre-treatment transcriptome say about its sister's fate? How much of that signal is specific to the drug, rather than the growth advantage that the control arm was designed to reveal? And what is the best performance one could hope for, given that the label belongs to a different cell? We compare each predictor with a model trained on the untreated arm at equal training size, bound the achievable performance from sister pairs, check both estimators on synthetic data with a known answer, and test the answers in a second lineage system, two further

lineage data sets, bulk tumours and patients. We find that most of the predictable signal is growth fitness, and we release the evaluation as a reusable Python package.

Results

Study design and terms used

We use a few terms throughout. An *arm* is one condition of the experiment, namely carboplatin, olaparib, natural killer (NK) cells or the untreated control, each run in two replicates. A pre-treatment cell is *pre-resistant* if its treated sister was recovered after treatment and *pre-sensitive* otherwise. Cells that share a lineage barcode form a *family*, and every evaluation kept whole families on one side of the split, so a model is never tested on a relative of a cell it was trained on. Performance is reported as AUROC, where 0.5 is chance and 1 is perfect. The *fitness model* is the model trained on the untreated control arm. In that arm a sister is recovered without any drug, only by growing back after handling, so the model captures growth fitness rather than drug tolerance.

We used the authors' processed labels. After excluding 4,369 cells without a detected lineage label, 32,642 pre-treatment cells remained, 29,540 pre-sensitive and 3,102 pre-resistant, exactly the counts in the original report (Table S15 in Additional file 2). Pre-resistant cells made up 7.8–11.2% of labelled cells in the drug arms and 14.7–14.9% in the control arm (Fig. 1b). The cells formed 29,437 families, and within every family all cells carried the same label.

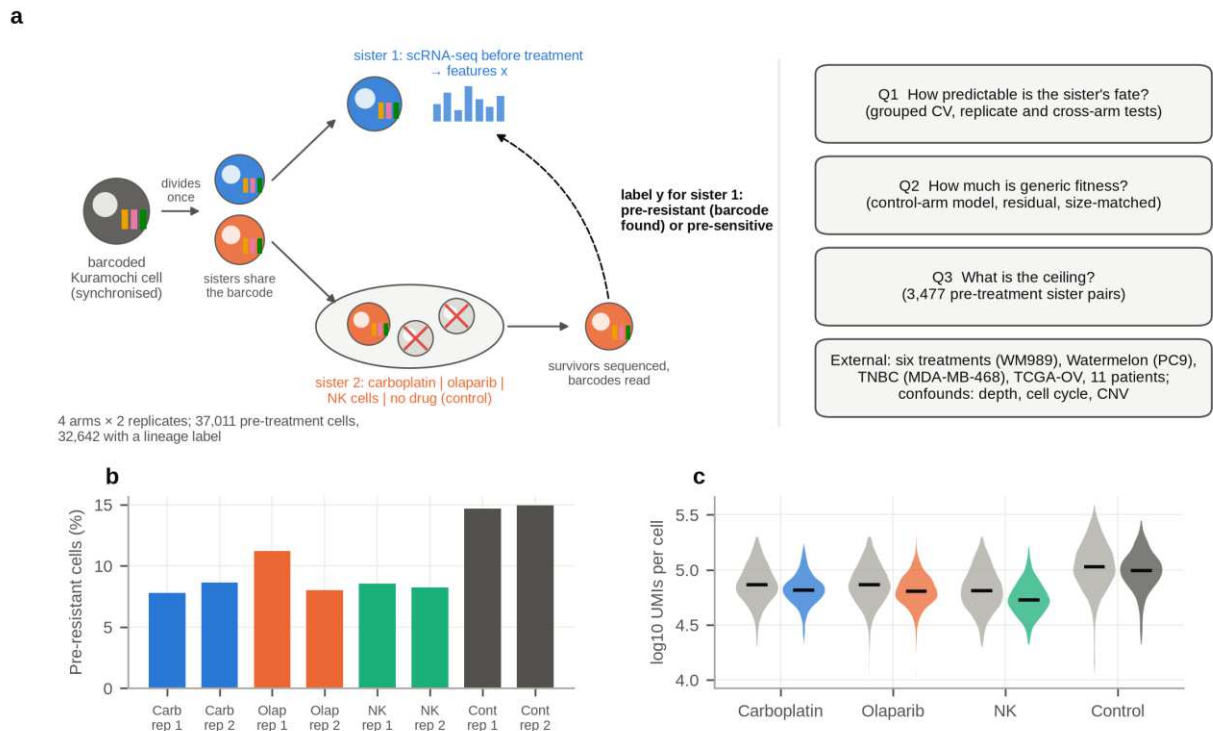


Fig. 1 Study design and data. **a**, ReSisTrace design and the questions addressed in this study. **b**, Fraction of labelled pre-treatment cells that are pre-resistant in each sample. **c**, Unique molecular identifiers (UMIs) per cell for pre-sensitive (grey) and pre-resistant (coloured) cells in each arm; bars mark medians.

A cell's state before treatment predicts its sister's fate

A logistic regression on 50 principal components of gene expression, obtained by principal component analysis (PCA), reached AUROC 0.747 for carboplatin, 0.744 for olaparib, 0.723 for NK cells and 0.708 for the control (Fig. 2a, Table S1 in Additional file 2). The signal passed three checks. When the model was retrained on labels shuffled between families within each sample, AUROC fell to around 0.50, and none of 100 such refits reached the observed value in any arm ($p < 0.01$, the smallest p-value 100 permutations allow; Table S2 in Additional file 2). Training on one replicate and testing on the other gave 0.70–0.76 (Fig. 2b). Restricting the analysis to cells without a sister in the pre-treatment sample, as the original authors did, gave slightly higher values (carboplatin 0.754, olaparib 0.753, NK 0.737, control 0.720).

Technical factors explain only part of the signal. Pre-resistant cells had fewer UMIs than pre-sensitive cells in every arm, including the control (Fig. 1c), and a model using only technical

covariates (sequencing depth, number of detected genes, mitochondrial fraction and cell-cycle scores) reached 0.59–0.65 (Fig. 2a). We therefore ran the main analyses in three preprocessing variants, keeping these covariates, removing cell-cycle scores, or removing cell cycle, depth and mitochondrial content from the expression data before computing the principal components. Removing cell-cycle scores changed nothing (0.707–0.748); removing depth and mitochondrial content as well lowered AUROC only modestly (carboplatin 0.723, olaparib 0.727, NK 0.680, control 0.704). The principal components are computed without labels but on all cells, test cells included; fitting gene selection, scaling and the principal components on the training cells of each fold only changed AUROC by at most 0.008 in either direction (Table S31 in Additional file 2).

This simple model does much better than published methods. On the same olaparib task, under the protocol of the published benchmark (all pre-resistant cells against an equal number of randomly drawn pre-sensitive cells, three draws), it reached 0.739–0.756 with ordinary five-fold cross-validation and 0.734–0.747 with families kept together (Fig. 2c), against the 0.50–0.60 reported by Shen et al. [29]. Keeping families together lowered AUROC by less than 0.01, so the gap is not caused by sisters leaking between training and test sets; it comes from training directly on the lineage labels. Table S20 in Additional file 2 lists the inputs, labels and protocol behind each value; the published methods were not rerun here.

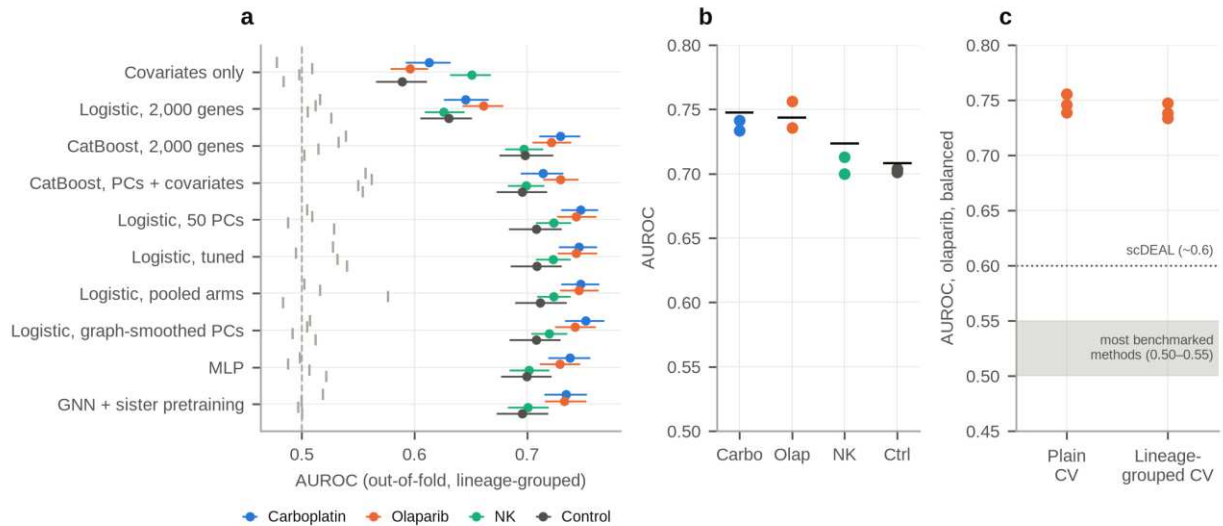


Fig. 2 The pre-treatment state predicts a sister's fate. **a**, Out-of-fold AUROC (lineage-grouped five-fold cross-validation) for each model and arm; error bars are 95% family-bootstrap intervals; grey ticks show the same models retrained on family-level shuffled labels (one to five permutations per model). **b**, Replicate hold-out AUROC (dots) for the PCA-logistic model; bars show within-arm AUROC. **c**, Olaparib under the benchmark protocol of Shen et al. [29], with ordinary and lineage-grouped cross-validation (three balanced draws each); the shaded band and dotted line mark the values reported for existing methods in that study.

More complex models do not predict better

No more flexible model found more signal (Table 1). Tuning the regularisation strength and the number of components changed AUROC by at most 0.002, and adding labelled cells from the other arms by at most 0.004. Gradient boosting was slightly worse (0.70–0.73), and linear models on 2,000 individual genes were worse still (0.63–0.66). A graph neural network pretrained to place sisters close together reached 0.696–0.734, below logistic regression in every arm, and no setting in a sweep of its hyperparameters did better (Fig. S1 in Additional file 1, Tables S1 and S18 in Additional file 2; Supplementary Note 1 in Additional file 1). Embeddings from two single-cell foundation models, Geneformer [30, 31] and scGPT [32], reached 0.671–0.683 and 0.673–0.693 in the drug arms (Table S17 in Additional file 2). Models ranging from a linear classifier to a 104-million-parameter transformer thus recover the same signal and no more. The limit lies in the information, not in the model.

150 **Table 1 Models and representations used in this study.**

Model	What it is	Input	Result (AUROC unless stated)
Logistic regression (main model)	Linear classifier with balanced class weights	50 principal components of expression	0.708–0.747 within each arm
Tuned or pooled logistic regression	Main model with tuned settings, or with labelled cells from the other arms added	50 principal components	Changed AUROC by at most 0.002 (tuned) and 0.004 (pooled)
Logistic regression on genes	Main model applied to single genes	2000 highly variable genes	0.63–0.66
Technical-covariate model	Logistic regression on technical factors only	Sequencing depth, detected genes, mitochondrial fraction, cell-cycle scores	0.59–0.65
Gradient boosting (CatBoost)	Ensemble of decision trees	Principal components and technical covariates	0.70–0.73
Multilayer perceptron	Small neural network, with or without sister-pair pretraining	50 principal components	0.700–0.738 without and 0.708–0.742 with pretraining
Graph neural network	Neural network that also reads each cell's 15 nearest neighbours, pretrained so that sisters lie close together	50 principal components and a cell–cell graph	0.696–0.734
Geneformer V2-104M_CLcancer	Transformer with 104 million parameters, pretrained on 104 million human and about 14 million cancer transcriptomes; used without fine-tuning	Raw counts as a ranked list of genes, giving a 768-dimensional embedding	0.671–0.683 (drug arms)
scGPT (whole-human)	Transformer pretrained on 33 million human cells; used without fine-tuning	Raw counts of 5005 genes, giving a 512-dimensional embedding	0.673–0.693 (drug arms)
Harmony, Pearson residuals, scVI	Batch-corrected components, components of analytic Pearson residuals, and the latent space of a variational autoencoder	Expression of the same 2000 genes	Same conclusions as the main model (Table S17 in Additional file 2)
Fitness model	Main model trained on the untreated control arm	50 principal components	0.70–0.73 on drug-arm labels
Other-arm model	Main model trained on the other two drug arms	50 principal components	Gain over the fitness model of 0.004 to 0.006
Label-free density score	Mean distance to the 15 nearest neighbours; uses no labels	50 principal components	0.63 for NK cells to 0.70 for carboplatin
Sister ceiling	Bivariate-normal latent-liability model	Model scores of pre-treatment sister pairs	About 0.84 (drug arms) and 0.76 (control) if the treated sister were scored
Cox regression	Proportional-hazards survival model	Signature scores of 303 ovarian tumours from The Cancer Genome Atlas (TCGA)	Hazard ratios (Fig. 7c)

151 AUROC, area under the receiver operating characteristic curve. Ranges are over the four
152 arms, or over the three drug arms where stated, from lineage-grouped five-fold cross-
153 validation. Methods give full settings for every model.

154 ***Most of the signal is not specific to the drug***

155 This is the central test of the paper. If the pre-treatment state encoded tolerance of a particular
156 drug, a model trained on one drug should transfer poorly to an unrelated drug, and the fitness

model, which never saw any drug, should transfer worst of all. Neither was the case. All twelve transfers between arms gave AUROC 0.69–0.74 (Fig. 3a). The fitness model scored 0.725 on carboplatin, 0.713 on olaparib and 0.697 on NK labels, within 0.02–0.03 of the models trained on those arms (Fig. 3b), and drug-arm models predicted the control labels with 0.69–0.70, close to the control model's own 0.708. The scores of the drug and fitness models were strongly correlated (Spearman 0.80–0.82).

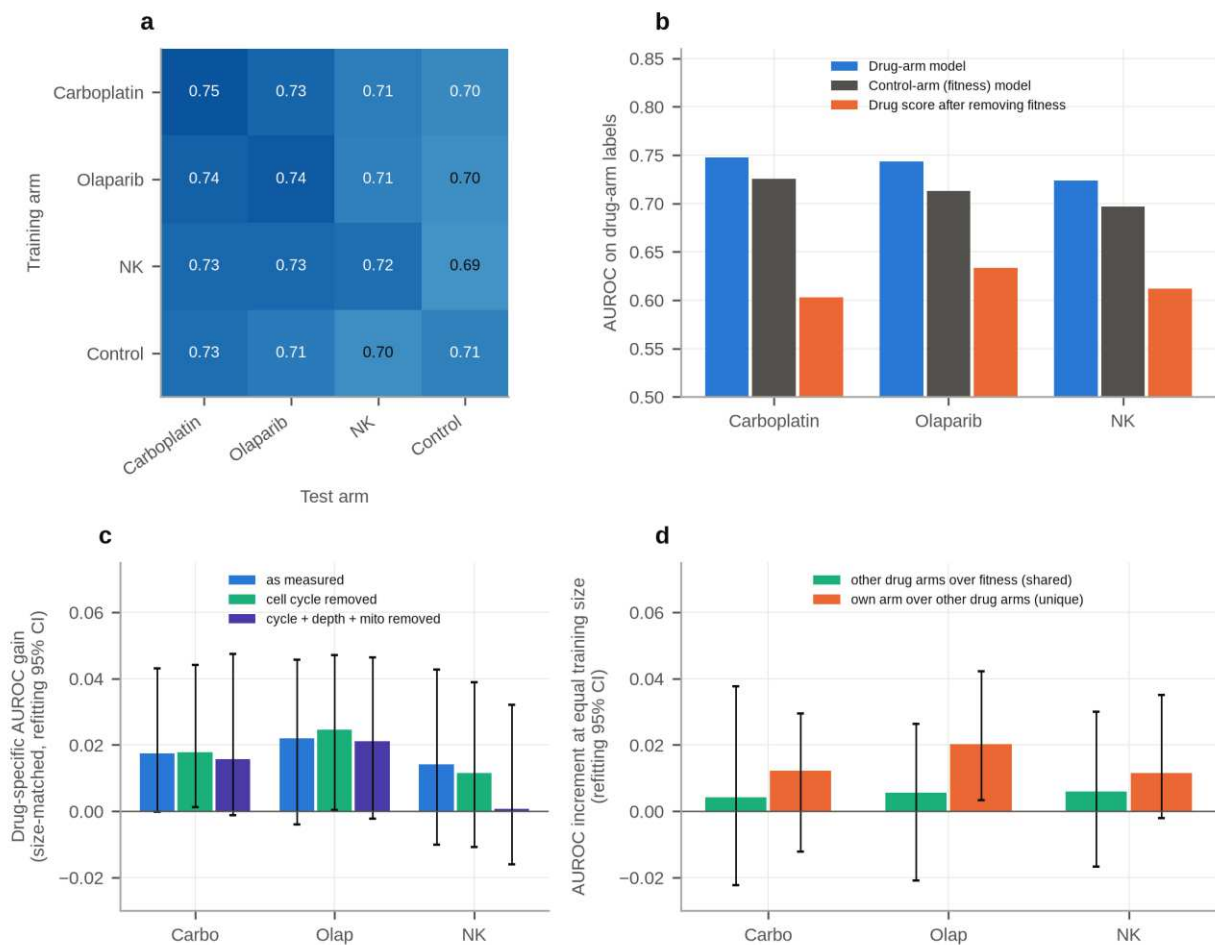


Fig. 3 Most of the signal is shared with the untreated control. **a**, AUROC of PCA-logistic models trained on one arm and tested on another; the diagonal shows within-arm cross-validation. **b**, For each drug arm, AUROC of the drug-arm model, of the control-arm (fitness) model applied to drug-arm cells, and of the drug score after within-sample regression on the fitness score. **c**, Drug-specific AUROC gain over the fitness model with training size matched (Δ AUROC), under three preprocessing variants; error bars are 95% intervals from 200 bootstrap replicates that refit the whole procedure (Table S27 in Additional file 2). Narrower

intervals that only resample the evaluation are listed in Table S19 in Additional file 2. **d**, With every model trained at the control arm's size, the AUROC gain on each drug arm's labels of a model trained on the other two drug arms over the fitness model (green, shared with other treatments) and of the arm's own model over the other-arm model (orange, unique to the drug); the two add up to the drug-specific gain. Error bars are refitting 95% bootstrap intervals (Table S28 in Additional file 2).

A drug-specific part does exist, but it is small. We defined the *drug-specific gain* as how much better a model trained on a drug arm's own labels predicts those labels than the fitness model does. Because the control arm has fewer labelled cells (4,183) than the drug arms (8,942–10,123), we trained the drug models on random sets of whole families of the same size as the control arm, repeated ten times (Methods). The gain was 0.014 for carboplatin, 0.025 for olaparib and 0.017 for NK cells (Fig. 3c, Table S19 in Additional file 2). Adding the drug model's score to a model that already contained the fitness score improved the fit in every arm (likelihood-ratio $p < 10^{-25}$), and the part of the drug score not explained by the fitness score still reached AUROC 0.60–0.63. The drug score therefore carries information that the fitness score does not, but it adds only one to three points of AUROC to a signal of 0.72–0.75.

The experiment is too small to say for certain that this gain is real. A design calculation shows that, with 3,879 labelled control families, it can detect a gain of about 0.025 to 0.028 with 80% power, and every estimate falls below that. When we repeated the whole procedure inside each of 200 bootstrap samples, refitting both models each time, the estimates, with 95% confidence intervals (CIs), were 0.017 for carboplatin (–0.0002 to 0.043), 0.022 for olaparib (–0.004 to 0.046) and 0.014 for NK (–0.010 to 0.043) (Table S27 in Additional file 2). Across all nine combinations of arm and preprocessing variant, the gain exceeded 0.01 in eight and was close to zero in the ninth (NK with all technical covariates removed, 0.001), and whether an interval cleared zero depended on the random training sets, which is what an effect at the edge of detection looks like. The same intervals give an upper limit, which is the more useful result here. The gain is at most 0.043 to 0.046 AUROC, less than a fifth of the 0.22 to 0.25 by

which the drug-arm models exceed chance. The data therefore rule out a drug-specific component larger than about 0.05 AUROC.

The result did not depend on how expression was represented. With Harmony-corrected components [33], analytic Pearson residuals [34], scVI [35], Geneformer or scGPT in place of the principal components, drug-arm models reached 0.671–0.746, the fitness model 0.627–0.722, and the drug-specific gain was 0.013–0.049 (Table S17 in Additional file 2). The largest gains came from scGPT, whose fitness model was also the weakest; a weaker fitness baseline leaves more of the shared signal to be counted as drug-specific.

Is the remainder tied to the particular drug, or to surviving treatment in general? A model trained on the other two drug arms, at the same training size, separates the two, because it has learnt from treated cells but never from this drug (Fig. 3d, Table S28 in Additional file 2). It improved on the fitness model by only 0.004 for carboplatin and 0.006 for olaparib and NK cells, whereas each arm's own model added a further 0.012, 0.020 and 0.012. By these point estimates, two thirds to four fifths of the gain beyond fitness was unique to the drug. The split is less certain than the gain itself; the drug-unique part excluded zero only for olaparib (0.020, 95% CI 0.003–0.042), the only targeted agent of the three.

A label-free score recovers much of the shared signal

A clue to what the shared signal is came from gradient boosting, which, trained on shuffled labels, still scored 0.550–0.562 on the true labels in every arm, whereas logistic regression scored 0.500–0.501. A flexible model can apparently pick up something about how cells are arranged in expression space that is related to the labels. We therefore tested a score that uses no labels at all, namely how crowded a cell's neighbourhood is, measured as the mean distance to its 15 nearest neighbours within its own sample. Pre-resistant cells lay in denser, more typical regions, and this density score alone reached AUROC 0.70 for carboplatin, 0.69 for olaparib, 0.63 for NK cells and 0.67 for the control (Fig. S2 in Additional file 1, Table S4 in Additional file 2). Cells in common states therefore tend to have sisters that are recovered, with or without a drug. The effect is specific to the sister design. In the two clone-level data

225 sets described below, the same score reached only 0.53 and 0.57 (Table S13 in Additional
226 file 2).

227 ***The fitness state is proliferative; the drug-specific state cycles more slowly***

228 Comparing pre-resistant with pre-sensitive cells within each sample gave 1,128–1,265
229 replicate-consistent differentially expressed genes in the drug arms and 639 in the control arm
230 (Table S7 in Additional file 2). Most were shared, and only 57 (carboplatin), 221 (olaparib) and
231 140 (NK) genes differed significantly from the control comparison (Table S8 in Additional file
232 2).

233 The shared fitness component is a proliferative, low-stress state. In the control arm, cells
234 whose sisters were recovered had higher MYC-target, mTORC1, E2F and fatty-acid-
235 metabolism programmes, and lower TNF- α /NF- κ B, TGF- β , interferon, p53 and apoptosis
236 programmes (Fig. 4a, Table S9 in Additional file 2). This fits an advantage in re-plating and
237 regrowth. The drug-specific component points the other way. After the control comparison was
238 subtracted, E2F targets, mitotic spindle and MYC targets were lower in pre-resistant cells in
239 all three drug arms. Among cells that are otherwise fit, those whose sisters survived a drug
240 were relatively less proliferative. A pre-specified panel of 28 marker genes gave the same
241 picture without signs of senescence (Fig. S3 in Additional file 1, Table S16 in Additional file 2,
242 Supplementary Note 2 in Additional file 1). Few genes were unique to a single drug (8 for
243 carboplatin, 4 for olaparib and 48 for NK cells; Table S29 in Additional file 2), and neither of
244 the two pathway hypotheses that we fixed in advance held after correction, so the drug-unique
245 part is spread thinly across many genes rather than concentrated in a known mechanism.

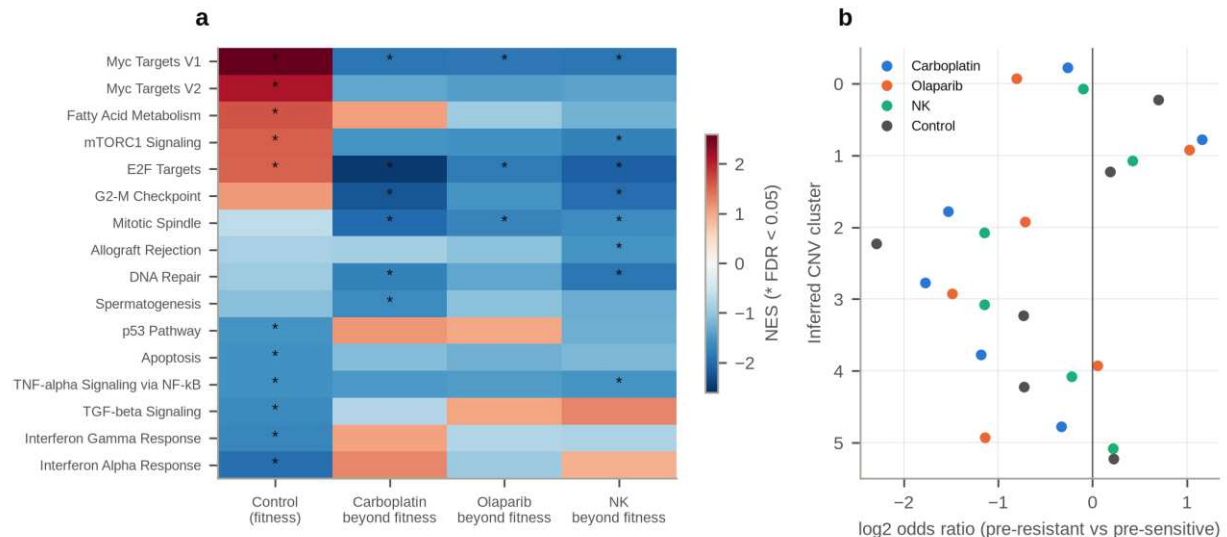


Fig. 4 Biology of the shared and drug-specific components. a, Normalised enrichment scores of Hallmark gene sets for the control contrast (fitness) and for each drug contrast after subtracting the control contrast; asterisks mark false discovery rate (FDR) < 0.05. **b**, Odds ratios for inferred copy-number clusters among pre-resistant versus pre-sensitive cells in each arm.

Copy-number differences do not explain the signal. Inferred copy-number clusters reproduced the subclone enrichment reported in the original study [24] in the drug arms but not in the control arm (Fig. 4b, Table S11 in Additional file 2), predicted the labels poorly (AUROC 0.54–0.60), and expression added strongly to them in every arm (Table S12 in Additional file 2, Supplementary Note 3 in Additional file 1). Scored against the untreated arm, the original study's findings also look different. Its mRNA-surveillance programmes were partly present without any drug, and its link between pre-resistance and homologous-recombination deficiency (HRD) was as strong in the untreated arm as in the drug arms (Table S30 in Additional file 2, Supplementary Note 4 in Additional file 1).

Sister discordance caps predictability

The label belongs to the sister, not to the sequenced cell, so even a perfect model of the sequenced cell has a ceiling. We estimated it from 3,477 pairs of sisters from 2,939 families in which both sisters were sequenced before treatment. Sisters were similar but not identical (mean intraclass correlation 0.75 over principal components 1–10, against 0.00 for random

pairs from the same sample; Fig. 5a, Table S6 in Additional file 2), and the scores of the principal-component model correlated 0.63–0.81 between them. We combined this correlation with the observed AUROC in a latent-liability model, allowing 10% of pre-sensitive labels to be wrong, within the error range estimated by the original authors [24]. If the same score could be applied to the treated sister itself, it would reach about 0.84 in the drug arms and 0.76 in the control arm (Fig. 5b). Reading the wrong cell therefore costs at most 0.09–0.12 AUROC in the drug arms and 0.05 in the control arm (Fig. 5c, Table S5 in Additional file 2). Simulations with a known answer confirmed that this estimate can only err upwards (Fig. S5 and Supplementary Note 7 in Additional file 1), so these are upper bounds. Even with the treated cell in hand, a large part of fate would remain unexplained, in line with live-cell studies showing that fate is only partly inherited [15, 16].

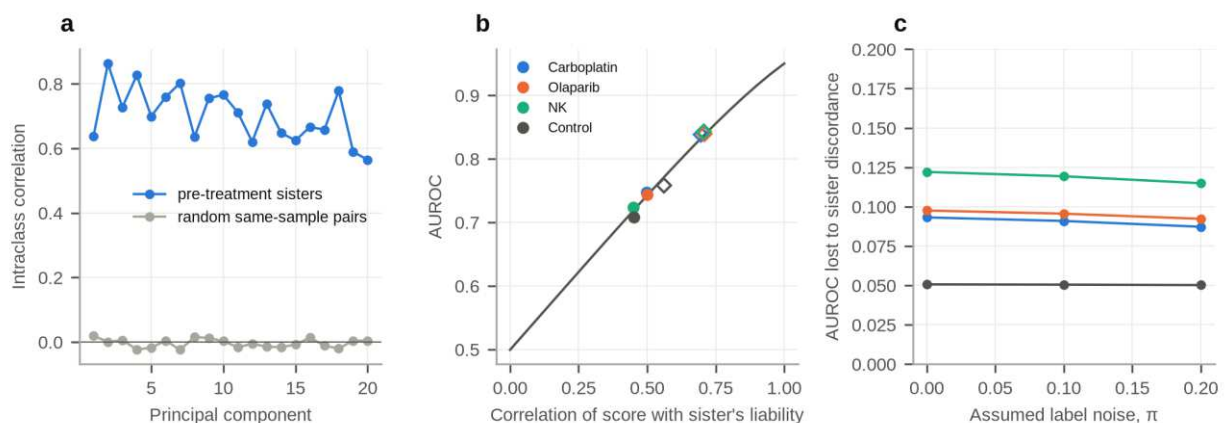


Fig. 5 Sister discordance caps predictability. **a**, Intraclass correlation of principal-component scores between pre-treatment sisters and between random same-sample pairs. **b**, Expected AUROC as a function of the correlation between score and sister liability under the bivariate-normal model with label noise $\pi = 0.1$ (line), with each arm's observed PCA-logistic AUROC (dots) and the AUROC implied if the treated sister's own state were scored (diamonds). **c**, AUROC lost to sister discordance as a function of assumed label noise π .

In a second lineage system, clone size predicts better than cell state

ReSisTrace is the only design we know of with an untreated arm carried through the whole protocol. A related question can still be asked wherever clone size before treatment is known,

because a lineage that has grown larger before any drug is, by definition, one that grows well. Schaff et al. [11] barcoded WM989 melanoma cells, profiled an untreated sample, and split the remaining cells across six treatments (dabrafenib, trametinib, CoCl₂, acidic medium, cisplatin and doxorubicin). We obtained 7,165 untreated cells from 1,359 lineages, of which 30.6% (trametinib) to 78.3% (acidic medium) lay in a lineage detected again after treatment. Clone size won in every treatment. It predicted detection with AUROC 0.68–0.90, against 0.54–0.66 for a logistic model on 50 principal components (Fig. 6a). Adding expression to clone size raised AUROC by at most 0.04, and for four treatments not at all. Once the part of the expression score explained by clone size was removed, it fell to 0.44–0.59 (Fig. 6c; Table S22 in Additional file 2). The two targeted drugs were the treatments in which clone size mattered least (0.70 and 0.68) and expression added most (0.04 and 0.03), matching the weak clone-size correlations the original authors reported for exactly these two. Predictions transferred only between similar treatments (Fig. 6b), and an independent preregistered reanalysis of the same data also found that clone abundance dominates [36] (Supplementary Note 5 in Additional file 1).

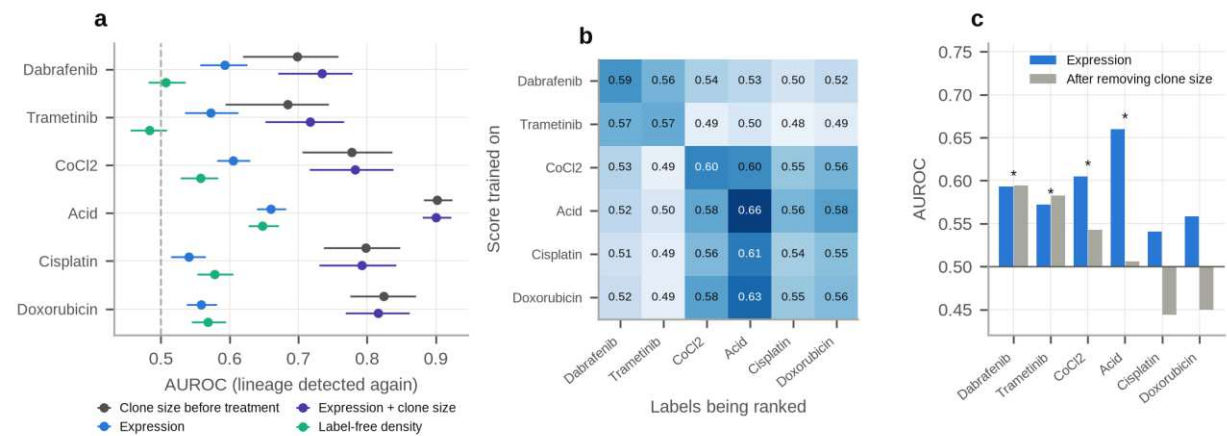


Fig. 6 A second lineage system with six treatments [11]. a, AUROC for predicting whether an untreated cell's lineage is detected again after each treatment, using clone size before treatment, expression (50 principal components), both, and the label-free density score; lineage-grouped cross-validation, error bars are 95% lineage-bootstrap intervals. **b**, AUROC when the out-of-fold score trained on one treatment's labels is used to rank another treatment's

labels. **c**, Expression AUROC before and after residualising on log clone size; asterisks mark treatments in which expression added to clone size (likelihood-ratio $p < 0.05$).

Other lineage data, tumours and patients show the same pattern

The same ordering held in two further clone-level data sets. In the Watermelon data (PC9 lung cancer cells treated with osimertinib [22]), 5,087 barcoded cells from 2,993 lineages were labelled by whether their lineage was observed again at day 14. Clone size alone predicted this label with AUROC 0.76, and expression alone 0.61 (Fig. 7a). In triple-negative breast cancer cells treated with afatinib [23], 1,541 untreated barcoded cells from 448 lineages were labelled by whether their lineage was among the authors' drug-tolerant expanded persisters after 40 days. Clone size gave 0.88, expression 0.64, and both together 0.88 (Fig. 7b).

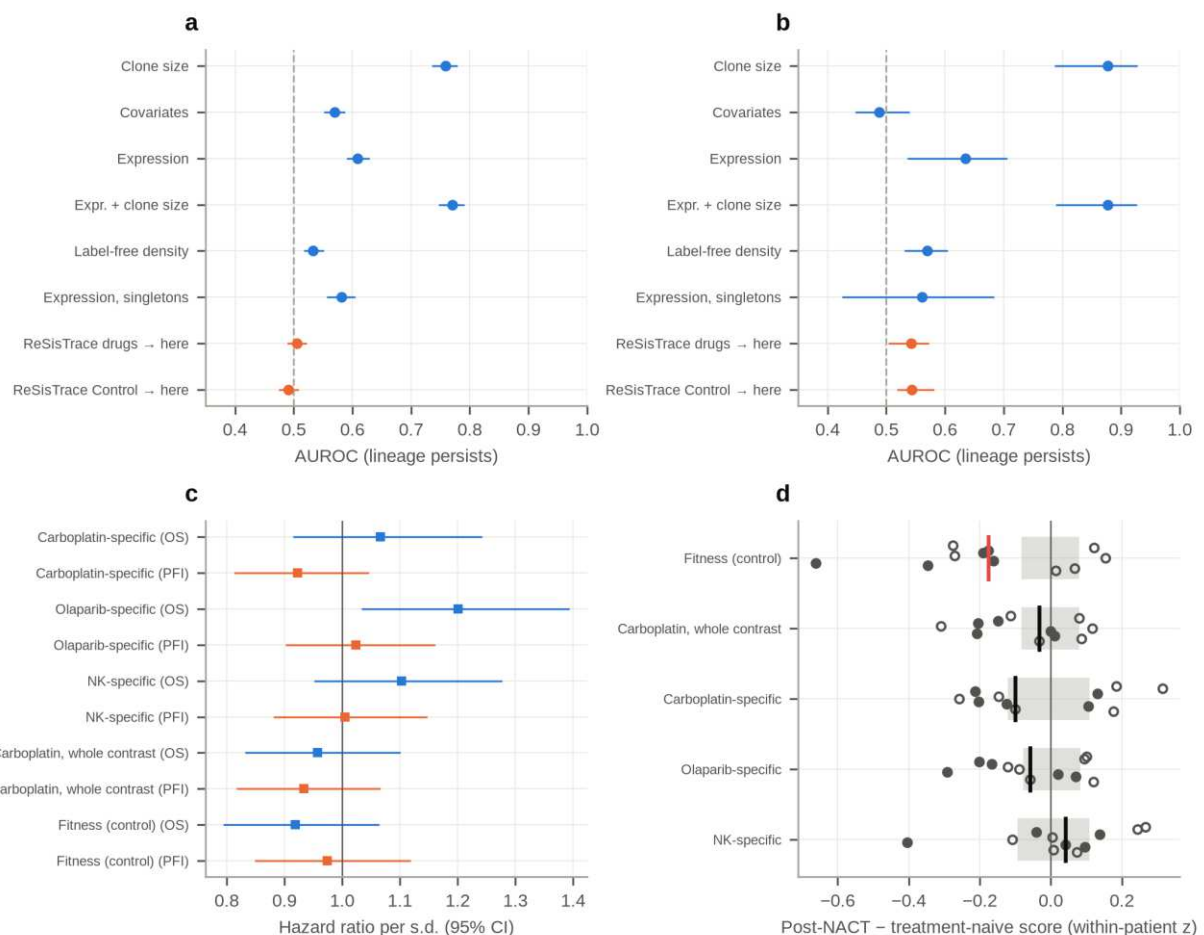


Fig. 7 External and clinical checks. a,b, AUROC for predicting lineage persistence from untreated cells in Watermelon (**a**) and triple-negative breast cancer (**b**), using clone size, technical covariates, expression, the label-free density score, expression within singleton

lineages, and zero-shot transfer of ReSisTrace models; error bars are 95% lineage-bootstrap intervals. **c**, Hazard ratios per standard deviation of signature score for overall survival and progression-free interval in TCGA-OV; the fitness, whole-contrast and drug-specific signatures are defined in Methods (Differential expression and pathways). **d**, Change in signature score from treatment-naïve to post-treatment malignant cells in 11 patients with high-grade serous ovarian cancer receiving neoadjuvant platinum–taxane chemotherapy [37]; one dot per patient, filled where both samples came from the same anatomical site; the vertical line is the median (red for the fitness signature, the only one with Wilcoxon $p < 0.05$) and the grey band spans the mean ± 2 s.d. of the medians of 200 expression-matched random gene sets.

In 303 primary ovarian tumours from TCGA, none of five signatures predicted overall survival or progression-free interval after correction for multiple testing (Fig. 7c, Table S14 in Additional file 2). In patients, the fitness state was depleted after chemotherapy. Zhang et al. [37] sequenced malignant cells from 11 patients with high-grade serous ovarian cancer before and after neoadjuvant platinum–taxane chemotherapy (8,806 malignant cells). If treatment selected the ReSisTrace carboplatin states, post-treatment cells should score higher on them; they did not (Wilcoxon $p = 0.24$ and 0.83). The fitness signature moved the other way. It was lower after treatment in 7 of 11 patients (Wilcoxon $p = 0.042$), lower than for all 200 matched random gene sets (two-sided $p = 0.005$), and lower in all five patients whose two samples came from the same anatomical site (Fig. 7d, Table S21 in Additional file 2). The proliferative, low-stress state that predicts recovery in a culture dish is therefore depleted, not enriched, in tumours after chemotherapy. Post-treatment samples also differ in ways a sister design does not capture, so this is a consistency check rather than a validation.

Without an untreated arm, the drug-specific gain cannot be measured reliably

Almost no study has an untreated arm, so we asked whether something built from the treated arms alone can stand in for it. Neither substitute we tried worked (Fig. S4 in Additional file 1, Table S26 in Additional file 2, Supplementary Note 6 in Additional file 1). Measured against the label-free density score, the drug-specific gain was two to three times too large (0.043 –

0.090); measured against a model trained on the other drug arms, it nearly vanished (0.0045–0.0065). A substitute can show that a pre-resistance signal is mostly fitness, but it cannot say by how much.

Discussion

We asked whether a cancer cell's transcriptome before treatment predicts whether its sister survives, and how much of that prediction is about the drug. The pre-treatment state is predictive, and a simple model reaches AUROC 0.72–0.75, far above published methods on the same task. But most of the signal is not about the drug. A model trained on cells that never saw a drug predicts drug-arm fates nearly as well, a score that uses no labels recovers much of the same signal, and the shared component is a proliferative, low-stress state. The drug-specific remainder makes biological sense, pointing to slower cycling among otherwise fit cells, and it is positive in every arm. But it is worth only one to three points of AUROC, and this experiment cannot reliably separate it from zero once the uncertainty of fitting is counted.

Put simply, a common view is that a few cancer cells are already equipped, before any drug is given, with a special ability to survive that drug. Our results point to a more ordinary explanation for most of what can be predicted in advance. The cells whose sisters survived were, on the whole, cells that were fit before treatment. Handling in the laboratory is itself a strain, because cells are synchronised, split and re-plated, and fitter cells tend to recover from it whether or not a drug follows. That is why a model trained only on untreated cells predicts survival under a drug almost as well as a model trained on the drug arm itself.

This matters because, without an untreated arm, a model that picks out cells with a growth or recovery advantage will look like a resistance predictor, and will even appear to generalise across drugs. Transfer between drugs is often read as evidence for a shared tolerance programme; here it reflects a shared fitness programme. This is relevant to recent reports of multi-treatment resistance [11], whose design had no drug-free arm carried through the same culture. The same care applies to benchmarks. Our simple model reached about 0.74 on the

olaparib task on which transfer-learning methods scored 0.50–0.60 [29], but most of that signal comes from the fitness axis that the control arm reveals. A benchmark that scores fate prediction on drug-arm labels alone rewards fitness and tolerance equally; including a model trained on the untreated arm as a reference would show what a method has actually learned.

The pattern is not a quirk of one protocol. A second system, with six treatments, a different cancer type and a different barcoding scheme, gave the same ordering. How much a lineage had grown before treatment mattered more than what its cells were expressing. In all three clone-level data sets, clone size predicted whether a lineage would be seen again much better than expression did, so expression-based claims from clone-level labels need clone size as a baseline. The patient cohort points the same way from the other direction. Cells that survived platinum-based chemotherapy were lower on the fitness signature than the same patients' untreated cells. A state that marks recovery in a culture dish need not mark survival in a patient.

What should future studies do? The clearest answer is to carry an untreated arm through the same protocol, and to report the gain of drug-arm models over a model trained on that arm, with training size matched, next to the raw AUROC. Where that is impossible, we found no reliable substitute. Studies should also expect a ceiling. Because sisters are not identical, even a perfect model of the sequenced cell loses up to 0.09–0.12 AUROC in the drug arms. A methodological lesson came from the synthetic experiments. Bootstrap intervals that do not refit the model covered the truth well below their nominal rate, so anyone putting a bootstrap around a cross-validated score should refit inside each replicate. We provide these steps as a small package for any lineage-tracing design with an untreated arm.

This study has limitations. ReSisTrace labels are noisy, because cells whose sister was lost are counted as pre-sensitive, and the ceiling handles this only through an assumed noise level; it also rests on a simplified Gaussian liability model, so it is approximate and an upper bound. Our copy-number clusters were inferred from expression, not measured from DNA. Differential expression treated the cells of a sample as independent; because 90% of families

consist of a single labelled cell, sisters add little dependence, and all classifier evaluations kept families together. All main analyses use one ovarian cancer cell line, the external checks rely on clone-level labels, bulk tumours and 11 patients, and the second system has no drug-free arm, so clone size stands in for fitness there. Finally, we used both foundation models as frozen feature extractors rather than fine-tuning them. A different design might do better, but the near-identical performance of very different models suggests that the information, not the model, is the limit.

Conclusions

A cancer cell's transcriptome before treatment predicts whether its sister survives, well above chance and far better than transfer-learning methods suggested on the same data. Most of that signal, however, is not about the drug. It is shared with cells that were never treated, and it tracks a proliferative, low-stress state. In a second lineage system, its counterpart, clone size, outweighs expression under every treatment. The drug-specific remainder is small, consistent in sign and below what this experiment can resolve. Studies that claim a pre-existing state predicts drug tolerance should therefore carry an untreated arm through the same protocol and report the gain over it, with training size matched. The package provided here does this for any design with such an arm.

Methods

Data. ReSisTrace count matrices and cell metadata were downloaded from the Gene Expression Omnibus (GEO; GSE223003 [38]; 16 samples covering four arms, two replicates and pre/post time points). Only pre-treatment cells were modelled. Labels (drugSens), within-sample sister families (sisters) and families linking pre-resistant cells to their surviving sisters (prer_r_group) were used as provided by the authors. In the authors' pipeline, a pre-treatment cell is pre-resistant if its lineage label matches a post-treatment cell perfectly, and pre-sensitive otherwise (including partial matches). We defined families as connected

427 components over shared sisters or prer_r_group identifiers within each sample. Cells had
428 already been quality-filtered by the authors, and we applied no further filtering.

429 **Preprocessing.** Samples were aligned to the union of 34,353 genes, with genes absent from
430 a sample set to zero counts. Counts were normalised to 10,000 per cell and log-transformed
431 with Scanpy [39]. Cell-cycle scores used the gene lists of Tirosh et al. [40]. Highly variable
432 genes were ranked by normalised dispersion within each sample, and the 2,000 with the best
433 median rank were kept after excluding mitochondrial and ribosomal genes and genes detected
434 in fewer than three cells. Expression was z-scored within each sample and clipped at ± 10 . For
435 the regress and regress_all variants, covariates were regressed out within each sample,
436 namely cell-cycle scores for regress, plus log counts, log detected genes and mitochondrial
437 fraction for regress_all. Fifty principal components were computed on all pre-treatment
438 cells. This step uses no labels. As a check, within-arm cross-validation (with the same folds)
439 and the fitness model were repeated with gene selection (among the 5,005 stored genes),
440 per-sample scaling and the principal components fitted on the training cells only and then
441 applied unchanged to the test cells (Table S31 in Additional file 2).

442 **Evaluation.** We used three designs, namely (i) five-fold stratified group cross-validation within
443 each arm, with replicates pooled and families as groups; (ii) training on one replicate and
444 testing on the other; (iii) training on one arm and testing on each other arm. AUROC and the
445 area under the precision–recall curve (AUPRC) were computed on pooled out-of-fold
446 predictions, with 95% confidence intervals from 500 family-level bootstrap resamples. Label-
447 shuffled controls permuted training labels at the family level within each sample, which
448 preserves both the number of positives per sample and the fact that sisters share a label. The
449 exact permutation test refitted the model on 100 such permutations using the same folds, and
450 the reported p is $(1 + \#\{b: A_b \geq A_{\text{obs}}\})/101$.

451 **Models.** Logistic regression (scikit-learn [41]) used balanced class weights and $C = 0.05$
452 unless tuned. The tuned variant chose C from $\{0.005, 0.02, 0.05, 0.2, 1\}$ and the number of
453 components from $\{10, 20, 30, 50\}$ by three-fold grouped inner cross-validation. The pooled

variant added all labelled cells from arms other than the test arm to the training set. CatBoost [42] used 800 iterations, depth 6, learning rate 0.05 and balanced class weights. The graph neural network read each cell's 15 nearest neighbours within its sample and was optionally pretrained with a contrastive loss over sister pairs; its architecture and training are given in the Supplementary Methods (Additional file 1).

Alternative representations. The 50 principal components were replaced by (i) Harmony-corrected components (harmonypy, batch = sample), (ii) 50 principal components of analytic Pearson residuals ($\theta = 100$) computed per sample on raw counts of the same 2,000 genes (Scanpy), (iii) a 30-dimensional scVI latent space (scvi-tools, batch = sample, up to 100 epochs with early stopping), (iv) Geneformer embeddings [30], using the cancer-tuned V2 model with 104 million parameters (Geneformer-V2-104M_CLcancer [31]), and (v) scGPT embeddings [32] from the whole-human checkpoint. Neither foundation model was fine-tuned, so their embeddings do not see the labels; the embedding procedure is given in the Supplementary Methods (Additional file 1). All representations were z-scored within each sample, and the logistic and fitness analyses were rerun unchanged.

Fitness analysis. Throughout, fitness is defined operationally as what predicts a sister's recovery without any drug, that is, the score of a PCA-logistic model trained on the lineage labels of the untreated control arm. Formally, let $\mathcal{A}$ be a learning algorithm, D_k a training sample of k labelled families and $f_{\mathcal{A},D}$ the resulting score. Writing $A(f)$ for the AUROC of a score f against the drug arm's labels, the quantity we estimate is

$$\Delta_k = \mathbb{E}_{D_k^{\text{drug}}} \left[A \left(f_{\mathcal{A}, D_k^{\text{drug}}} \right) \right] - \mathbb{E}_{D_k^{\text{ctrl}}} \left[A \left(f_{\mathcal{A}, D_k^{\text{ctrl}}} \right) \right],$$

the expected advantage of training on the drug arm's own labels over training on untreated labels, at a common training size k . Both expectations are over the draw of training families, which is why k is fixed at the labelled control arm's size and why the interval must resample that draw rather than only the cells used for evaluation. For each drug arm, the fitness score was the prediction of this model trained on all labelled control cells. We report the drug model's

out-of-fold AUROC; the fitness model's AUROC on the drug-arm labels; the AUROC of the drug logit after within-sample linear regression on the fitness logit; a one-degree-of-freedom likelihood-ratio test for adding the drug logit to a model containing the fitness logit; and the size-matched gain (Δ AUROC). For the latter, the drug model was fitted by the same lineage-grouped five-fold cross-validation, but each training fold was reduced to a random set of whole families containing as many cells as the labelled control arm (4,183), and Δ AUROC was averaged over 10 random reductions. The intervals quoted in Results refit the fitness model and the reduced drug model inside each of 200 family-level bootstrap replicates (scripts/28_delta_refit_bootstrap.py); a faster interval that only re-scores fitted models is in Table S19 in Additional file 2, and both are described in the Supplementary Methods (Additional file 1).

Fitness, general treatment and the particular drug. Three PCA-logistic models were compared on each drug arm's labels, all trained on random sets of whole families totalling as many cells as the labelled control arm. These were the fitness model, a model trained on the pooled labels of the other two drug arms, which never sees the arm being scored, and the arm's own model, scored out of fold. The general-treatment gain is the AUROC of the second minus that of the first, the drug-unique gain the AUROC of the third minus that of the second, and the two add up to the drug-specific gain. General treatment here means whatever the treated arms share and the control arm lacks, which includes differences in handling as well as biology.

Ceiling. Write S_A for a model's score on the sequenced sister, S_B for the same model's score on its treated sibling, and L_B for the latent liability that decides whether the treated sister survives. Sister concordance $c = \text{corr}(S_A, S_B)$ was taken over pre-treatment sister pairs after a rank-based inverse-normal transform of the scores. A fraction π of observed negatives are really positives whose sister was not recovered, so $\text{AUROC}_{\text{obs}} = (1 - \pi) \text{AUROC} + \pi/2$; inverting this gives the noise-corrected value. Under the bivariate-normal model the AUROC of a score against a dichotomised liability is a monotone function g of their correlation, so

507 $\rho_{\text{obs}} = g^{-1}(\text{AUROC})$. Assuming that the sequenced sister bears on its sibling's fate only
508 through the state the two share ($S_A \perp L_B \mid S_B$),

509
$$\rho_{\text{fate}} = \frac{\rho_{\text{obs}}}{c}, \quad \text{AUROC}_{\text{if sister seen}} = g(\rho_{\text{fate}}),$$

510 and the sister cost is the difference between that and the observed AUROC. When the
511 assumption fails, the estimate errs upwards, so the ceiling and the cost are upper bounds, as
512 proved in the Supplementary Methods (Additional file 1).

513 **Typicality score.** For each cell, we computed the mean Euclidean distance to its 15 nearest
514 neighbours in the 50-component space among cells of the same sample, and used its negative
515 as a label-free score. The direction (denser = pre-resistant) is the only choice made with
516 reference to the labels.

517 **Differential expression and pathways.** Wilcoxon tests compared pre-resistant with pre-
518 sensitive cells within each sample. Replicate z-scores were combined by Stouffer's method,
519 and genes with opposite signs in the two replicates were flagged. Drug-specific genes had
520 $\text{FDR} < 0.05$ for the drug contrast and for the difference from the control contrast, $z_{\text{diff}} = (z_{\text{drug}} -$
521 $z_{\text{control}})/\sqrt{2}$, and consistent signs. Preranked gene set enrichment analysis (GSEA [43]) with
522 GSEAPy [44] used the MSigDB Hallmark collection [45] with 1,000 permutations. Signatures
523 took up to 50 genes in each direction with $\text{FDR} < 0.05$, ranked by z-score, after removing
524 mitochondrial and ribosomal genes. The fitness signature used the control contrast; the whole-
525 contrast signature of a drug arm used its drug contrast without reference to the control; and
526 the drug-specific signature of a drug arm used only its drug-specific genes, ranked by z_{diff} .
527 Genes unique to one drug, the re-scoring of published claims and the copy-number clusters
528 are described in the Supplementary Methods (Additional file 1).

529 **Second lineage system.** Matrices for the nine samples of Schaff et al. [11] were taken from
530 GEO (GSE279162 [46]), and lineage barcodes were assigned to cells as described in the
531 Supplementary Methods (Additional file 1). A lineage counted as surviving a treatment if any
532 cell in that treatment's sample carried it. Clone size is the number of untreated cells with that

lineage. Expression was log-normalised (10,000 counts per cell); the 2,000 genes with the highest dispersion were z-scored and reduced to 50 principal components. Logistic models were evaluated with five-fold cross-validation grouped by lineage, with 95% intervals from 300 lineage-level bootstrap resamples.

External data. For Watermelon (GEO GSE150949 [47]), barcoded day-0 cells were labelled by the presence of their lineage barcode in any day-14 sample. For the triple-negative breast cancer study (GEO GSE228382 [48]), per-cell lineage annotations are not in GEO, so we exported the authors' Seurat object from figshare [49] to plain files with a short R script, and labelled untreated day-0 cells with the authors' DTPE_lineages annotation. For each data set we evaluated lineage-grouped cross-validation of clone-size, covariate and expression models, and zero-shot transfer of ReSisTrace logistic models, with intervals from 300 lineage-level bootstrap resamples. For the ovarian cancer cohort of TCGA (TCGA-OV), RSEM expression and curated survival endpoints [50] were obtained from UCSC Xena [51]. Five signatures were tested (the fitness signature, the three drug-specific signatures and the whole-contrast carboplatin signature); primary tumours were scored by the mean z-score of up-genes minus down-genes, and Cox models (lifelines [52]) were fitted per standard deviation of score.

Patient data. Count data and cell annotations for 11 patients with high-grade serous ovarian cancer sampled before and after neoadjuvant platinum–taxane chemotherapy were obtained from GEO (GSE165897 [53]; Zhang et al. [37]). Only cells the authors annotated as malignant epithelial cells were used (8,806 of 51,786). Counts were normalised and log-transformed, and each gene was z-scored within patient, so that patient identity cannot drive the comparison. A signature score was the mean z of its up-genes minus the mean z of its down-genes, and each patient contributed the mean score of their post-treatment cells minus that of their treatment-naïve cells. Significance was assessed by a Wilcoxon signed-rank test over the 11 patients and by comparison with 200 random gene sets matched in size and mean expression.

559 **Software.** The evaluation is packaged as the Python module `persist.fitadj` [54], which
560 takes plain arrays, so any representation and any scikit-learn classifier can be used.
561 `fitness_adjusted_report(X, y, groups, arms, control="Control")` returns, for each
562 drug arm, the AUROC of the drug-arm model and of the fitness model, the residual AUROC,
563 the likelihood-ratio test, and the size-matched drug-specific gain with its family bootstrap
564 interval. `family_permutation_null(X, y, groups, samples)` gives the family-level
565 permutation test, and `sister_ceiling(auroc, sister_concordance(scores, pairs),`
566 `prevalence)` gives the AUROC expected if the treated sister itself were scored. The module
567 is tested in `tests/test_fitadj.py`.

568 **Use of large language models.** The authors used a large language model to assist with
569 writing code and text, checked all analyses and text, and take full responsibility for the content.

570 **Declarations**

571 **Ethics approval and consent to participate.** Not applicable. Only publicly available,
572 previously published data were analysed.

573 **Consent for publication.** Not applicable.

574 **Availability of data and materials.** All data sets analysed during the current study are publicly
575 available. They are the ReSisTrace data in the Gene Expression Omnibus (GEO) repository,
576 GSE223003 [38]; the PC9 osimertinib (Watermelon) lineage data, GSE150949 [47]; the triple-
577 negative breast cancer lineage data, GSE228382 [48], with per-cell lineage annotations on
578 figshare [49]; the second lineage system, GSE279162 [46]; the patient cohort, GSE165897
579 [53]; and TCGA-OV through UCSC Xena [51]. All result tables underlying the figures and text,
580 including full differential expression and gene-set enrichment results, are included in this
581 published article (Additional file 2, Tables S1–S31). The code that regenerates every result,
582 figure and table from the public data is available on GitHub [54].

583 **Project name:** SisFate (analysis code and the Python package `persist`, with the evaluation
584 module `persist.fitadj`)

585 **Project home page:** <https://github.com/rabeyanoor/SisFate>

586 **Archived version:** a release will be archived at Zenodo with a DOI on acceptance

587 **Operating system(s):** platform independent (tested on Linux)

588 **Programming language:** Python 3.12; one R script exports a Seurat object

589 **Other requirements:** dependencies are pinned in `uv.lock` (NumPy 2.5, scikit-learn 1.9,
590 Scanpy 1.12, PyTorch 2.14 CPU); the Geneformer and scGPT forward passes used one
591 NVIDIA T4 GPU (Kaggle), and their kernel scripts are included

592 **License:** MIT

593 **Any restrictions to use by non-academics:** none

594 The module `persist.fitadj` takes plain arrays and any scikit-learn classifier, so the fitness-
595 adjusted evaluation (control-arm baseline, size-matched Δ AUROC with a family-level
596 bootstrap interval, family permutation null and sister ceiling) can be applied to other lineage-
597 tracing designs that include an untreated arm. The configuration file
598 (`configs/default.yaml`) fixes the random seed and all hyperparameters, a single script
599 (`scripts/run_all.sh`) regenerates all results on a laptop CPU except the Geneformer and
600 scGPT forward passes, which need a GPU (see above), and every metric is stored together
601 with a hash of the configuration that produced it. The repository also includes a 7 MB example
602 file with the principal components, labels, families and sister pairs of all labelled ReSisTrace
603 cells, and a notebook (`notebooks/quickstart.ipynb`) that reproduces the main model, the
604 fitness model and the sister ceiling from it in a few minutes without downloading the raw data.

605 **Competing interests.** The authors declare that they have no competing interests.

606 **Funding.** This research received no specific grant from any funding agency in the public,
607 commercial or not-for-profit sectors.

Authors' contributions. RN conceived the study, carried out all analyses and wrote the paper. KJR contributed to the biological interpretation of the results. Both authors read and approved the final version.

Acknowledgements. We thank the authors of the ReSisTrace, Watermelon, triple-negative breast cancer, multi-treatment melanoma and ovarian cancer patient studies for making their data publicly available, and The Cancer Genome Atlas Research Network.

References

1. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*. 2010;141:69-80. <https://doi.org/10.1016/j.cell.2010.02.027>.
2. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med*. 2016;22:262-269. <https://doi.org/10.1038/nm.4040>.
3. Ramirez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, et al. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun*. 2016;7:10690. <https://doi.org/10.1038/ncomms10690>.
4. Marine JC, Dawson SJ, Dawson MA. Non-genetic mechanisms of therapeutic resistance in cancer. *Nat Rev Cancer*. 2020;20:743-756. <https://doi.org/10.1038/s41568-020-00302-4>.
5. Soragni A, Knudsen ES, O'Connor TN, Tognon CE, Tyner JW, Gini B, et al. Acquired resistance in cancer: towards targeted therapeutic strategies. *Nat Rev Cancer*. 2025;25:613-633. <https://doi.org/10.1038/s41568-025-00824-9>.
6. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. 2017;546:431-435. <https://doi.org/10.1038/nature22794>.
7. Shaffer SM, Emert BL, Reyes Hueros RA, Cote C, Harmange G, Schaff DL, et al. Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell*. 2020;182:947-959. <https://doi.org/10.1016/j.cell.2020.07.003>.

635 8. Emert BL, Cote CJ, Torre EA, Dardani IP, Jiang CL, Jain N, et al. Variability within rare cell states
636 enables multiple paths toward drug resistance. *Nat Biotechnol.* 2021;39:865-876.
637 <https://doi.org/10.1038/s41587-021-00837-3>.

638 9. Goyal Y, Busch GT, Pillai M, Li J, Boe RH, Grody EI, et al. Diverse clonal fates emerge upon drug
639 treatment of homogeneous cancer cells. *Nature.* 2023;620:651-659. [https://doi.org/10.1038/s41586-](https://doi.org/10.1038/s41586-023-06342-8)
640 [023-06342-8](https://doi.org/10.1038/s41586-023-06342-8).

641 10. Harmange G, Hueros RAR, Schaff DL, Emert B, Saint-Antoine M, Kim LC, et al. Disrupting cellular
642 memory to overcome drug resistance. *Nat Commun.* 2023;14:7130. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-023-41811-8)
643 [023-41811-8](https://doi.org/10.1038/s41467-023-41811-8).

644 11. Schaff DL, White PE, Cote CJ, Watterson GE, Lin KZ, Fasse AJ, et al. Pre-existing cell states
645 predict resistance to multiple treatments. *Cell Genom.* 2026;6:101191.
646 <https://doi.org/10.1016/j.xgen.2026.101191>.

647 12. Iyer A, Alva A, Granada AE, Chakrabarti S. Inheritable cell-states shape drug-persister
648 correlations and population dynamics in cancer cells. *PLoS Comput Biol.* 2025;21:e1013446.
649 <https://doi.org/10.1371/journal.pcbi.1013446>.

650 13. Rehman SK, Haynes J, Collignon E, Brown KR, Wang Y, Nixon AML, et al. Colorectal cancer
651 cells enter a diapause-like DTP state to survive chemotherapy. *Cell.* 2021;184:226-242.
652 <https://doi.org/10.1016/j.cell.2020.11.018>.

653 14. Oliveira EA, Milite S, Fernandez-Mateos J, Cresswell GD, Yara-Romero E, Vlachogiannis G, et al.
654 Epigenetic heritability of cell plasticity drives cancer drug resistance through a one-to-many genotype-
655 to-phenotype paradigm. *Cancer Res.* 2025;85:2921-2938. [https://doi.org/10.1158/0008-5472.CAN-25-](https://doi.org/10.1158/0008-5472.CAN-25-0999)
656 [0999](https://doi.org/10.1158/0008-5472.CAN-25-0999).

657 15. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell
658 variability in TRAIL-induced apoptosis. *Nature.* 2009;459:428-432.
659 <https://doi.org/10.1038/nature08012>.

660 16. Paek AL, Liu JC, Loewer A, Forrester WC, Lahav G. Cell-to-cell variation in p53 dynamics leads
661 to fractional killing. *Cell.* 2016;165:631-642. <https://doi.org/10.1016/j.cell.2016.03.025>.

662 17. Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat
663 Rev Genet. 2020;21:410-427. <https://doi.org/10.1038/s41576-020-0223-2>.

664 18. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional
665 landscapes links state to fate during differentiation. Science. 2020;367:eaaw3381.
666 <https://doi.org/10.1126/science.aaw3381>.

667 19. Wang SW, Herriges MJ, Hurley K, Kotton DN, Klein AM. CoSpar identifies early cell fate biases
668 from single-cell transcriptomic and lineage information. Nat Biotechnol. 2022;40:1066-1074.
669 <https://doi.org/10.1038/s41587-022-01209-1>.

670 20. Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S, et al. Non-genetic
671 determinants of malignant clonal fitness at single-cell resolution. Nature. 2022;601:125-131.
672 <https://doi.org/10.1038/s41586-021-04206-7>.

673 21. Gutierrez C, Al'Khafaji AM, Brenner E, Johnson KE, Gohil SH, Lin Z, et al. Multifunctional
674 barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution
675 and treatment. Nat Cancer. 2021;2:758-772. <https://doi.org/10.1038/s43018-021-00222-8>.

676 22. Oren Y, Tsabar M, Cuoco MS, Amir-Zilberstein L, Cabanos HF, Hütter JC, et al. Cycling cancer
677 persister cells arise from lineages with distinct programs. Nature. 2021;596:576-582.
678 <https://doi.org/10.1038/s41586-021-03796-6>.

679 23. Pellicchia S, Franchini M, Viscido G, Arnese R, Gambardella G. Single cell lineage tracing
680 reveals clonal dynamics of anti-EGFR therapy resistance in triple negative breast cancer. Genome
681 Med. 2024;16:55. <https://doi.org/10.1186/s13073-024-01327-2>.

682 24. Dai J, Zheng S, Falco MM, Bao J, Eriksson J, Pikkusaari S, et al. Tracing back primed resistance
683 in cancer via sister cells. Nat Commun. 2024;15:1158. <https://doi.org/10.1038/s41467-024-45478-7>.

684 25. Chen J, Wang X, Ma A, Wang QE, Liu B, Li L, et al. Deep transfer learning of cancer drug
685 responses by integrating bulk and single-cell RNA-seq data. Nat Commun. 2022;13:6494.
686 <https://doi.org/10.1038/s41467-022-34277-7>.

687 26. Zheng Z, Chen J, Chen X, Huang L, Xie W, Lin Q, et al. Enabling single-cell drug response
688 annotations from bulk RNA-seq using SCAD. *Adv Sci.* 2023;10:2204113.
689 <https://doi.org/10.1002/advs.202204113>.

690 27. Fustero-Torre C, Jiménez-Santos MJ, García-Martín S, Carretero-Puche C, García-Jimeno L,
691 Ivanchuk V, et al. Beyondcell: targeting cancer therapeutic heterogeneity in single-cell RNA-seq data.
692 *Genome Med.* 2021;13:187. <https://doi.org/10.1186/s13073-021-01001-x>.

693 28. Suphailai C, Chia S, Sharma A, Tu L, Da Silva RP, Mongia A, et al. Predicting heterogeneity in
694 clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. *Genome Med.*
695 2021;13:189. <https://doi.org/10.1186/s13073-021-01000-y>.

696 29. Shen L, Sun X, Zheng S, Hashmi A, Eriksson J, Mustonen H, et al. Predicting pre-treatment
697 resistance or post-treatment effect? A systematic benchmarking of single-cell drug response models.
698 *bioRxiv.* 2026. <https://doi.org/10.64898/2026.04.10.717709>.

699 30. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, et al. Transfer learning
700 enables predictions in network biology. *Nature.* 2023;618:616-624. [https://doi.org/10.1038/s41586-](https://doi.org/10.1038/s41586-023-06139-9)
701 [023-06139-9](https://doi.org/10.1038/s41586-023-06139-9).

702 31. Chen H, Venkatesh MS, Gómez Ortega J, Mahesh SV, Nandi TN, Madduri RK, et al. Scaling and
703 quantization of large-scale foundation model enables resource-efficient predictions in network biology.
704 *Nat Comput Sci.* 2026;6:450-463. <https://doi.org/10.1038/s43588-026-00972-4>.

705 32. Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: toward building a foundation
706 model for single-cell multi-omics using generative AI. *Nat Methods.* 2024;21:1470-1480.
707 <https://doi.org/10.1038/s41592-024-02201-0>.

708 33. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate
709 integration of single-cell data with Harmony. *Nat Methods.* 2019;16:1289-1296.
710 <https://doi.org/10.1038/s41592-019-0619-0>.

711 34. Lause J, Berens P, Kobak D. Analytic Pearson residuals for normalization of single-cell RNA-seq
712 UMI data. *Genome Biol.* 2021;22:258. <https://doi.org/10.1186/s13059-021-02451-7>.

713 35. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell
714 transcriptomics. *Nat Methods*. 2018;15:1053-1058. <https://doi.org/10.1038/s41592-018-0229-2>.

715 36. Aviv H. Pretreatment transcriptional state carries condition-specific information about future clonal
716 detection in a lineage-traced melanoma line. *Zenodo*. 2026.
717 <https://doi.org/10.5281/zenodo.22850050>.

718 37. Zhang K, Erkan EP, Jamalzadeh S, Dai J, Andersson N, Kaipio K, et al. Longitudinal single-cell
719 RNA-seq analysis reveals stress-promoted chemoresistance in metastatic ovarian cancer. *Sci Adv*.
720 2022;8:eabm1831. <https://doi.org/10.1126/sciadv.abm1831>.

721 38. Tracing back primed resistance in cancer via sister cells. *Gene Expression Omnibus*. 2023.
722 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE223003>. Accessed 10 Oct 2026.

723 39. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis.
724 *Genome Biol*. 2018;19:15. <https://doi.org/10.1186/s13059-017-1382-0>.

725 40. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the
726 multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352:189-196.
727 <https://doi.org/10.1126/science.aad0501>.

728 41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine
729 learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
730 <https://jmlr.org/papers/v12/pedregosa11a.html>.

731 42. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with
732 categorical features. In: *Advances in Neural Information Processing Systems* 31. 2018.
733 <https://arxiv.org/abs/1706.09516>.

734 43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set
735 enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
736 *Proc Natl Acad Sci USA*. 2005;102:15545-15550. <https://doi.org/10.1073/pnas.0506580102>.

737 44. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment
738 analysis in Python. *Bioinformatics*. 2023;39:btac757. <https://doi.org/10.1093/bioinformatics/btac757>.

739 45. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
740 Signatures Database hallmark gene set collection. *Cell Syst.* 2015;1:417-425.
741 <https://doi.org/10.1016/j.cels.2015.12.004>.

742 46. Pre-existing cell states predict resistance to multiple treatments. *Gene Expression Omnibus.*
743 2024. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE279162>. Accessed 10 Oct 2026.

744 47. Cycling cancer persisters arise from lineages with distinct transcriptional and metabolic programs.
745 *Gene Expression Omnibus.* 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150949>.
746 Accessed 10 Oct 2026.

747 48. Single cell lineage tracing reveals clonal dynamics of anti-EGFR therapy resistance in triple
748 negative breast cancer. *Gene Expression Omnibus.* 2024.
749 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228382>. Accessed 10 Oct 2026.

750 49. Gambardella G, Franchini M. Single cell lineage tracing reveals clonal dynamics of anti-EGFR
751 therapy resistance in triple negative breast cancer. *figshare.* 2024.
752 <https://doi.org/10.6084/m9.figshare.24787458>.

753 50. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated
754 TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell.*
755 2018;173:400-416. <https://doi.org/10.1016/j.cell.2018.02.052>.

756 51. UCSC Xena. TCGA Ovarian Cancer (OV): gene expression (HiSeqV2) and curated survival data.
757 [https://xenabrowser.net/datapages/?cohort=TCGA%20Ovarian%20Cancer%20\(OV\)](https://xenabrowser.net/datapages/?cohort=TCGA%20Ovarian%20Cancer%20(OV)). Accessed 10
758 Oct 2026.

759 52. Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw.* 2019;4:1317.
760 <https://doi.org/10.21105/joss.01317>.

761 53. Longitudinal single-cell RNA-seq data of metastatic ovarian cancer. *Gene Expression Omnibus.*
762 2021. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165897>. Accessed 10 Oct 2026.

763 54. Noor R. SisFate: fitness-adjusted evaluation of pre-treatment cell fate prediction. *GitHub.* 2026.
764 <https://github.com/rabeyanoor/SisFate>.

765 **Additional files**

766 **Additional file 1** (.docx). **Supplementary Notes, Methods and Figures.** Supplementary
767 Notes 1–7 (graph model and hyperparameter sweep; pre-specified marker panel and the
768 insulin-like growth factor axis; inferred copy-number subclones; the original study's pathway
769 and HRD findings revisited; further analyses of the second lineage system; substitutes for an
770 untreated arm; the estimators checked against a known answer), Supplementary Methods,
771 and Supplementary Figures S1–S5 with their legends.

772 **Additional file 2** (.xlsx). **Supplementary Tables S1–S31.** All result tables underlying the
773 figures and text, one table per sheet, including full differential expression and gene-set
774 enrichment results. The first sheet lists every table with a short description. The tables are:

775 **Table S1.** Model performance. Out-of-fold AUROC and AUPRC of all models, data splits and
776 preprocessing variants, with 95% family-bootstrap intervals and shuffled-label controls.

777 **Table S2.** Robustness checks. Singleton-only cells, the exact permutation test (100 refits) and
778 the Shen et al. benchmark protocol.

779 **Table S3.** Fitness confound. Drug-arm model versus the control-arm (fitness) model for each
780 drug arm and preprocessing variant.

781 **Table S4.** Label-free typicality score. AUROC of the local-density score (negative mean
782 distance to the 15 nearest neighbours within sample).

783 **Table S5.** Sister-discordance ceiling. Decomposition of predictability by sister discordance for
784 each model.

785 **Table S6.** Sister similarity of principal components. Intraclass correlation of principal
786 components between pre-treatment sisters and random same-sample pairs.

787 **Table S7.** Differential expression in all arms. Pre-resistant versus pre-sensitive Wilcoxon tests
788 within sample, with replicates combined by Stouffer's method.

789 **Table S8.** Drug versus fitness differential expression. Drug contrast compared with the control
790 contrast for each gene, with the gene class defined in Methods.

791 **Table S9.** Hallmark gene-set enrichment. Preranked GSEA on MSigDB Hallmark gene sets
792 (1,000 permutations).

793 **Table S10.** Signature gene lists. Genes used for TCGA-OV signature scoring.

794 **Table S11.** Copy-number cluster enrichment. Enrichment of inferred copy-number clusters
795 among pre-resistant cells (Fisher's exact test, BH-corrected).

796 **Table S12.** Copy-number confound. Cluster identity versus expression: AUROCs and
797 likelihood-ratio test for expression beyond cluster.

798 **Table S13.** External lineage data sets. Lineage-grouped cross-validation in the Watermelon
799 and TNBC afatinib data, and zero-shot transfer of the ReSisTrace model.

800 **Table S14.** TCGA-OV survival. Cox models per standard deviation of signature score in
801 TCGA-OV primary tumours.

802 **Table S15.** Label audit. Label counts per sample compared with Dai et al. (2024).

803 **Table S16.** Pre-specified marker panel. Drug contrast, control contrast and their difference for
804 the marker panel shown in Fig. S3.

805 **Table S17.** Alternative representations. Main and fitness analyses repeated on alternative
806 representations of the same cells, including Geneformer and scGPT embeddings.

807 **Table S18.** Graph network hyperparameter sweep. One-at-a-time sweep of the graph network
808 hyperparameters.

809 **Table S19.** Drug-specific gain with evaluation-only bootstrap. Size-matched gain over the
810 fitness model (Δ AUROC) with 95% paired family-level bootstrap intervals.

811 **Table S20.** Benchmark comparison. Olaparib pre-treatment task: methods benchmarked by
812 Shen et al. (2026) and the logistic model of this study.

813 **Table S21.** Patient cohort. ReSisTrace signature scores in malignant cells of patients with
814 high-grade serous ovarian cancer before and after neoadjuvant chemotherapy (GSE165897).

815 **Table S22.** Second lineage system: prediction. Lineage-grouped AUROC for clone size,
816 expression, both and the label-free score in six treatments (GSE279162).

817 **Table S23.** Second lineage system: transfer and ceiling. Cross-treatment transfer of out-of-
818 fold scores and the clone-concordance ceiling.

819 **Table S24.** Simulation calibration. Synthetic experiments with a known answer: bias and
820 interval coverage of the drug-specific gain and bias of the sister ceiling.

821 **Table S25.** Experiment planning. Families per arm needed to detect a given drug-specific
822 gain.

823 **Table S26.** Substitutes for an untreated arm. Drug-specific gain measured against the control-
824 arm model and against two baselines that use no control-arm labels.

825 **Table S27.** Drug-specific gain with refitting bootstrap. Gain over the fitness model with
826 intervals from refitting the whole procedure in each bootstrap replicate (the interval quoted in
827 the text).

828 **Table S28.** Fitness, general-treatment and drug-unique signal. AUROC of models trained on
829 the untreated arm, the other drug arms and the arm itself at equal training size.

830 **Table S29.** Drug-unique genes and gene sets. Genes and Hallmark gene sets unique to one
831 drug, and the four pre-specified pathway tests.

832 **Table S30.** Published pathway and HRD claims revisited. The pathway and HRD claims of
833 Dai et al. (2024) re-scored in every arm, including the untreated control.

834 **Table S31.** Preprocessing fitted on training cells only. AUROC with the label-free
835 preprocessing fitted on all cells, with PCA fitted on training cells, and with all steps fitted on
836 training cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.xlsx](#)