Tracing back primed resistance in cancer via sister cells

Jun Dai[1,3], Shuyu Zheng[1,3], Matías M. Falco[1], Jie Bao[1], Johanna Eriksson[1], Fernando Perez-Villatoro[1], Anniina Färkkilä[1], Olli Dufva[2], Khalid Saeed[2], Satu Mustjoki[2], Yinyin Wang[1], Ali Amiryousefi[1], Jing Tang[1]*, Anna Vähärautio[1]*

Author affiliations

[1] Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2] Research Program in Translational Immunology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[3] These authors contributed equally to this work.

* These authors jointly supervised this work: Jing Tang, Anna Vähärautio. Email: jing.tang@helsinki.fi; anna.vaharautio@helsinki.fi

Keywords: lineage tracing, cancer, resistance, sister cells, cell states, scRNA-seq.

Abstract

Exploring non-genetic evolution of cell states during cancer treatments has become attainable by recent advances in lineage-tracing methods coupling cell states to future fates. However, transcriptional changes that drive pre-treatment cells into resistant fates may be subtle, necessitating high resolution analysis. We developed ReSisTrace that uses shared transcriptomic features of synchronised sister cells to predict the states that prime treatment resistance, and allows identification of asymmetric features that drive phenotypic heterogeneity. Applying ReSisTrace in ovarian cancer cells revealed that BRCAness transcriptional signatures are associated with pre-existing vulnerability not only to olaparib and carboplatin treatments, but also to natural killer cell-mediated cytotoxicity. This novel connection between DNA repair defect and susceptibility to natural killer cells was further validated both functionally and in a clinical cohort. The high-resolution analysis by ReSisTrace enables resolving pre-existing transcriptional features of treatment vulnerability, facilitating molecular patient stratification for personalised therapies.

Introduction

Recent studies have highlighted the role of non-genetic heterogeneity driving cancer treatment resistance[1–4]. Mechanistically, such heterogeneity accumulates continuously via inherent stochasticity in gene expression, and is further promoted upon each cell division due to uneven partitioning of biomolecules and organelles[5]. As a result, even isogenic cancer cells manifest variability in responses to drugs, associated with transient changes in their chromatin or expression states before the treatment[1–4]. Recently, new methodologies have been developed to explore these pre-existing resistant states using physical separation of cell populations[6] and predefined markers[7]. Notably, unbiased approaches that utilise barcode-based lineage tracing with single-cell RNA-sequencing (scRNA-seq) readout have identified fate-coupled transcriptomic profiles[8,9,10]. As scRNA-seq consumes the cells being analysed, cell fate predictions in these approaches are based on the assumption that cells from the same progeny can be used as 'proxies' for each other. However, these studies do not control the number of cell divisions, and therefore the progeny size varies from

a couple of generations to millions of cells, causing un-accounted uncertainty to the prediction of primed resistance states.

We addressed these limitations by developing ReSisTrace, which takes full advantage of the transcriptional similarity of sister cells, while removing noise from discordant expression, to enable unbiased prediction of transcriptionally primed resistance in maximal resolution. ReSisTrace also allows exploring the transcriptomic features that are asymmetric between sister cells, and thus increase phenotypic heterogeneity in cancer cell populations. We further controlled for the effects of experimental fitness, and analysed primed resistance in the subclonal context to address the interplay between genetic and non-genetic traits in modulating treatment response. We applied ReSisTrace to study primed resistance against chemotherapy, targeted therapy and, for the first time, also against immune cytotoxicity, all in the context of high-grade serous ovarian cancer (HGSOC). The improved resolution and quality control measures enabled us to detect subtle transcriptomic patterns modulating distinct therapy responses, establishing a new connection between DNA repair defects and vulnerability to anti-cancer innate immunity.

Results

ReSisTrace reveals primed resistance via sister cell inference

To reveal the cell states that are primed to treatment resistance, we allowed uniquely labelled, synchronised cells to divide once, after which half of the cells were analysed by scRNA-seq, while the other half underwent the treatment (**Fig. 1a**). Surviving cells were allowed to recover, and then analysed by scRNA-seq to identify resistant lineages and their expression profiles. We determined the transcriptomes of these resistant lineages at the pre-treatment stage based on sister-cell coupling, and annotated them as pre-resistant cells. Cells with labels missing from post-treatment samples were considered to be pre-sensitive. We applied ReSisTrace in *BRCA2* mutant HGSOC cell line KURAMOCHI[11] against clinically relevant treatments, including conventional chemotherapy (carboplatin), and a PARP inhibitor (PARPi; olaparib) that is effective in *BRCA1/2* mutated HGSOC[12]. Furthermore, we determined the pre-resistant states of cells that survived natural killer (NK) cytotoxicity. To remove the confounding effects of growth fitness in the pre-resistance signatures, we included a non-treatment control. All the assays were performed in two replicates.

We achieved the unique labelling of the cells by using lentiviral constructs with random barcodes, and subsetting cells from a large pool, analogous to the concepts used in unique molecular identifiers (UMIs)[13]. Prior to splitting the sample, we confirmed that the majority of colonies consisted of two cells, while less than 5% had four or more cells (**Extended Data Fig. 1a, Extended Data Fig. 1b left**). Among the lineage labels captured by scRNA-seq in the pre-treatment sample, approximately 90% were represented by only one cell (**Extended Data Fig. 1b right**). Mathematical modelling and computer simulations showed that the pre-resistant cells can be robustly captured with relatively high accuracy that is not sensitive to experimental variables, such as cell loss (**Supplementary Table 1, Methods**). We also found that the consistency between the replicates is high (**Extended Data Fig. 1c**). These results support the validity of our approach, warranting a high-resolution coupling of cell fate with their transcriptomic features.

To validate the use of sister cells as proxies, we confirmed that cells with shared labels, i.e. putative sister cells, have significantly more similar transcriptomes than random pairs of cells (**Fig. 1b)**. Our

results are consistent with a recent study where sister cells also displayed similar mean transcriptional levels, as well as synchronous transcriptional dynamics[14]. We further compared the genes that showed significantly higher similarity within sister cell pairs, termed 'sister-concordant genes', against the other genes, termed 'sister-discordant genes' (**Fig. 1c**). The sister-discordant genes were expressed at lower levels than sister-concordant genes, along with increasing drop-out noise ($P < 2.2 \times 10^{-16}$, two-tailed t-test, **Extended Data Fig. 2a**). To control the discrepancy of gene expression levels and sample sizes (**Extended Data Fig. 2a**), only a subset of sister-concordant and sister-discordant genes were compared using RNA velocity and pathway analysis. We found that sister-discordant genes showed higher relative degradation rates ($P = 4.2 \times 10^{-10}$, two-tailed t-test) and slightly lower splicing rate relative to transcription ($P = 0.008$, two-tailed t-test, **Fig. 1d**, **Extended Data Fig. 2b-c**), suggesting that the recently synthesised transcripts, and to an even higher extent, the transcripts to be degraded tended to show higher expressional variation within the sister cells. While sister-concordant genes were not enriched in any pathway (**Extended Data Fig. 2d**), the discordant genes were highly enriched in pathways related to PI3K/AKT/mTOR signalling, including Inositol phosphate metabolism and Phosphatidylinositol signalling system (**Fig. 1e**). Intriguingly, previous studies have shown that PI3K/AKT/mTOR signalling drives asymmetric cell division in divergent contexts, such as normal immune and epidermal differentiation, as well as cancer[15–17].

Subclonal enrichment contributes to treatment-specific pre-resistance signatures

We next explored the pre-resistant, pre-sensitive, and post-treatment cell populations in the transcriptomic space. A UMAP embedding showed no distinct separation between the pre-resistant and pre-sensitive cells, while the post-treatment cells were clearly shifted compared to the pre-treatment cells (**Fig. 2a** and **Extended Data Fig. 3a**), in line with previous results that did not identify distinct populations of pre-resistant cells[9]. As expected, all the post-treatment samples showed larger clone sizes (**Extended Data Fig. 3b)** and smaller lineage diversity compared to the pre-treatment samples (**Extended Data Fig. 3c)**, where olaparib and carboplatin showed stronger reduction than either control or NK conditions ($P = 3.80 \times 10^{-5}$, paired t-test, n = 8).

To assess the putative role of genetic heterogeneity for the observed population structure, we inferred copy number variations (CNVs) from the scRNA-seq data (**Extended Data Fig. 4a**). We identified five subclones that closely matched the clusters in the UMAP (**Fig. 2b)**. The subclonal identity within each lineage was largely consistent before and after the treatment (**Extended Data Fig. 4b**), further supporting that the sister cells in each lineage truly shared a reproducible genetic background. Variance normalised Euclidean distances of pre-resistant cells were smaller than those within pre-sensitive cells for each subclone in both treatment and control samples (**Extended Data Fig. 5**), suggesting that also fitness within the applied experimental conditions poses a strong selection pressure to the cancer cell population, necessitating the non-treatment control in the analysis of primed resistance.

To assess potential subclonal selection by the treatments, we determined the proportions of each subclone in the pre-sensitive, pre-resistant, and post-treatment populations at different conditions (carboplatin, olaparib, NK, and non-treatment control). Subclone A was significantly enriched in pre-resistant compared to pre-sensitive cells at all treatment conditions, especially for the carboplatin and olaparib treatments, when compared to the control samples ($P = 8.88 \times 10^{-11}$ and $P < 2.2 \times 10^{-16}$, respectively, two-proportion Z-test, **Fig. 2c**). Importantly, the proportions of each subclone were

similar between pre-resistant and post-treatment populations at each treatment condition (**Fig. 2c**, middle and rightmost panels), confirming that future subclonal selection during the treatment is already evident in the pre-resistant cells.

We compared the pre-resistance signature, defined as transcriptional changes between the pre-resistant to pre-sensitive populations (**Supplementary Table 2**), and the subclone signature, defined as transcriptional changes between each subclone against the others. The pre-resistance signatures were adjusted by the non-treatment fitness control. The signatures subclone A were positively correlated with the pre-resistance signatures for all the three treatments (**Fig. 2d**), consistent with the enriched proportion of this subclone (**Fig. 2c**). In contrast, the transcriptional signatures of subclones B, C or D were negatively correlated with the pre-resistance signature, in line with their decreasing proportions during the treatments. Interestingly, the signatures of subclone E were also positively correlated to the pre-resistance signatures in carboplatin and NK treatments, but not in olaparib (**Fig. 2d**), again consistent with observed subclonal enrichment patterns. To assess the biological processes behind these transcriptional changes, we performed a gene ontology (GO) analysis (**Fig. 2e**). Subclone A displayed increased cotranslational endoplasmic reticulum (ER) targeting and nonsense mediated decay (NMD) pathway, with decreased mitochondrial transcription. These processes were also associated with the primed resistance against olaparib, and to lesser extent with carboplatin. Our results are in line with previous studies, as DNA damaging therapies, such as carboplatin and olaparib, can potentiate ER stress to lethal levels[18], while NMD has been shown to protect cells from both DNA damage[19], and immune surveillance[20]. We found that mitochondrial expression patterns were associated with primed sensitivity towards PARPi, in line with a recent study[21], whereas oxidative phosphorylation has been shown to be associated with improved NK killing only in *TP53* mutant context[22], such as HGSOC. Furthermore, decreasing levels of mitosis and chromatin-associated processes were observed in both subclone E and pre-resistant cells for carboplatin and NK treatments. Interestingly, while sensitivity to the NK killing was associated with mitosis also in previous studies[22,23], the cell cycle associations of the two DNA damaging therapies diverged according to their mechanisms of action. Both mitosis and DNA replication, i.e. S phase, were enriched in primed sensitivity against carboplatin that crosslinks DNA regardless of cell cycle phase, thus hitting especially mitotic cells that are the most vulnerable to DNA damage[24]. Instead, PARP inhibition acts through stalling replication forks at each S phase[25], and correspondingly only DNA replication, not mitosis, was associated with PARPi sensitivity. Together, these results support the validity of ReSisTrace, as our findings are in line with previous research, such as distinct cell cycle vulnerabilities of the DNA damaging therapies.

BRCAness associates with vulnerability to innate anti-cancer immunity

Next, we set out to analyse the key DNA repair pathway that controls the response to PARP inhibitors and to a more limited extent, also to carboplatin. PARP inhibitors target cells with homologous recombination deficiency (HRD), also called 'BRCAness', in a synthetic lethal manner[26,27]. We therefore hypothesised that fluctuations in pre-existing DNA repair capabilities could contribute to primed PARP inhibitor resistance. We generated two independent BRCAness associated transcriptional signatures, one from an isogenic cell line pair model and the other from clinical data (**Fig. 3a-c**). We created an isogenic, homologous recombination proficient (HRP) cell line, from *BRCA1* mutated COV362 ovarian cancer cells by restoring the *BRCA1* function. The *BRCA1*-restored cells were significantly more resistant than the parental cells to olaparib at multiple dosages (**Fig. 3c**),

demonstrating the validity of the experimental HRP restored model. To define an experimental signature of BRCAness, we compared the expression profiles of control (i.e HRD, or BRCAness+) and *BRCA1*-restored (i.e HRP, or BRCAness-) cells. In contrast, the clinical BRCAness signature was defined as the deconvoluted[28] gene expression changes from The Cancer Genome Atlas (TCGA) cohort[29] where HGSOC patients were stratified into BRCAness +/- groups based on a genomic classifier called ovaHRDscar[30]. Kaplan-Meier survival analysis confirmed that both transcriptional BRCAness signatures were associated with improved overall survival (OS) in ovarian serous cystadenocarcinoma patients in TCGA cohort (**Fig. 3b, Extended Data Fig. 6a**).

We then compared the BRCAness transcriptional signatures to the pre-sensitive signatures (i.e. the inverse of pre-resistance signatures). The pre-sensitive signatures in olaparib and carboplatin conditions showed positive correlations, which were expected based on their mechanisms of action that target BRCAness (**Fig. 3d** and **Extended Data Fig. 6b-c)**. Surprisingly, the pre-sensitive signature for NK condition showed even higher correlations with both of the BRCAness signatures (**Fig. 3d** and **Extended Data Fig. 6b-c)**. This unexpected finding prompted us to explore whether there is a causal connection between BRCAness and sensitivity to NK cell-mediated killing in ovarian cancer cells. Indeed, we found that the BRCAness+ cells were more sensitive to NK cell-mediated killing than *BRCA1*-restored COV362 cells (**Fig. 3e**, lower-left). We further validated the result in the KURAMOCHI cell line, confirming the functional connection between DNA repair defects and innate anti-tumour immunity in ovarian cancer (**Fig. 3e**, lower-right). Furthermore, we analysed how the presence of intratumoral NK cells derived by a NK transcriptional signature (**Methods, Supplementary Table 3**) modulates patient survival in the context of BRCAness. The analysis of the interaction between the NK signature and BRCAness status showed that the NK signature was associated with improved survival only in the BRCAness+ (HRD) patients (P = 0.018, Cox proportional hazard model), whereas no survival benefits were found in the BRCAness- (HRP) genetic context (**Fig. 3f**). Our discovery suggests that in addition to having increased T cell infiltration[31] and improved immune checkpoint response[32], BRCAness tumours are more responsive to innate immune cytotoxicity and could thus be targeted with upcoming NK based immunotherapies.

Discussion

Fates of cancer cells with recent shared ancestry are closely intertwined, showing aligned drug responses[33] and preferential site of metastasis[34,35]. Cells of the same progeny, at least until a couple of generations apart, also present similar expression profiles[36,37]. However, pre-treatment transcriptional patterns do not fully capture the future fate of cells[37], and cells with extreme fates, such as high-dose drug persistence, may not display evident expression features prior treatment[37]. These studies demonstrate that the transcriptional signal of a primed fate can be both subtle and non-deterministic. Thus, it is crucial to address possible technical confounders and transcriptional noise to detect the subtle signals that diverge the cells to distinct fates. In ReSisTrace, we carefully controlled for fitness-associated patterns in primed resistance, and removed the transcriptional noise from genes whose expression is discordant between the sister cells.

Live cell tracking revealed that sister cells have a clearly higher correlation in marker gene expression than cousins or unrelated cells[6]. When compared to other lineage tracing methods that allow many cell divisions after labelling, the more controlled protocol of ReSisTrace provides a clear advantage of higher resolution. Importantly, the optimal progeny size of a lineage tracing approach is dictated

by the dosing of the treatment. When aiming to characterise the general features of rare persisters that survive a high dose treatment, a larger progeny size is beneficial to increase the chances of detecting each rare pre-resistant lineage. In contrast, for drugs where the clinically relevant concentrations are relatively lower, the higher resolution of ReSisTrace allows detection of the subtle transcriptional features that predispose to resistance. We applied ReSisTrace in the treatment of HGSOC, where neither chemotherapy nor PARP inhibitors are administered as high-dose therapy.

A core assumption behind ReSisTrace is that cells divide mostly symmetrically, and we showed that the transcriptomes of sister cells indeed were highly concordant. However, a minority of the transcripts that are less concordant are expected to create noise for ReSisTrace, and even more so for analogous methods using larger progeny. Our method allows noise removal from the pre-resistant expression patterns, and moreover, comparison of the differences between the sister-concordant and sister-discordant genes. We found that discordant transcripts contained a slightly higher proportion of newly made transcripts, and were clearly less stable. On pathway level, these transcripts were enriched in PI3K/AKT/mTOR signalling. Our results thus support the role of PI3K signalling in passing metabolic cues from the environment to drive asymmetric cell division[39]. Importantly, asymmetric cell division allows to maintain phenotypic heterogeneity even in genetically homogeneous populations, and may thus contribute to suboptimal treatment responses leading to cancer recurrence even without genetic evolution[40].

Non-genetic heterogeneity should ideally be assayed in the context of genetic heterogeneity, as clinical tumours are never completely isogenic due their inherent genomic instability. To improve clinical relevance of ReSisTrace in HGSOC that is an extremely copy number unstable cancer[8], we analysed primed resistance within the subclonal context via scRNA-seq inferred copy number variations. We found that subclonal structures drove global transcriptional clustering, and discovered treatment-specific enrichment of subclones. These enrichment patterns explained a considerable part of the transcriptomic features of the pre-resistant populations, such as those representing changes in proteostasis, mitotic nuclear division, and chromatin conformation. It is important to consider that subclonal enrichment contributes to differential expression of not only CNV affected drivers but also the co-amplified passenger genes[40], thus providing an additional level of noise to the pre-resistance signal and supporting the analysis of pre-resistant states in their subclonal context. Thus, ReSisTrace revealed the interplay between genetic and phenotypic heterogeneity in primed resistance, an aspect that becomes increasingly relevant for patient-derived models such as organoids or xenografts.

Our study is the first analysis of primed resistance against anti-cancer immunity, as previous approaches have studied only kinase inhibitors[41,42] or chemotherapy cocktails[43]. Applied to ovarian cancer cells, ReSisTrace revealed that the primed sensitivity to NK killing had an even stronger association with BRCAness transcriptional signatures than the synthetic lethal drug olaparib. We validated the vulnerability to NK killing in two isogenic cell line pairs. Furthermore, we found that the NK gene signature was associated with improved survival only in the ovarian cancer patients with BRCAness. This observation has two key implications. First, it emphasises the role of cell plasticity - which produces variable levels of BRCAness associated signatures in a *BRCA2* mutant cell line - in heterogeneous drug responses even in a synthetic lethal context. Secondly, DNA damage has been shown to induce NK mediated cell killing via upregulation of ligands activating innate immune receptors, such as those binding NKG2D (encoded by *KLRK1)*[42], and aneuploidy has been shown to activate NK clearance of non-transformed senescent cells via NF-κB signalling[43] but no direct reports

of the effect of BRCAness on the NK cytotoxicity exists. The observation could also bear clinical implications, suggesting that in addition to showing improved T cell-mediated immunosurveillance[44], BRCAness tumours could benefit from the NK based immunotherapies currently being developed. This could also apply for breast, prostate and pancreatic cancers that exhibit BRCAness tumours[45]. Hence, coupling cellular transcriptome to its future fate not only empowers finding of novel molecular features and potential drug targets of pre-resistant cells, but also enables molecular patient stratification.

Methods

Cell culture

COV362 cells (Merck, 07071910) were grown in DMEM (Corning, 15-013-CV) with 10% foetal bovine serum (FBS) (Gibco, 10270106), 1% penicillin–streptomycin (PS) (Gibco, 15140122), and 1% GlutaMAX (Gibco, 35050038). KURAMOCHI (JCRB Cell Bank, JCRB0098) cells were grown in RPMI1640 (Corning, 10-040-CV) with 10% FBS, and 1% PS.

Expansion of NK cells

NK cells were expanded by using K562-mbIL21-41BBL feeder cells as described previously[46]. Briefly, peripheral blood mononuclear cells (PBMCs) were isolated via gradient centrifugation with Ficoll-Paque from buffy coats of healthy donors. Five million PBMCs were suspended together with ten million feeder cells irradiated with 100 Gy, into 40 ml of R10 media supplemented with 10 ng/ml recombinant human IL-2 (R&D Systems, 202-IL-050). Cells were passaged two times per week and feeder cells were added in a 1:1 ratio after 7 days to maintain stimulation. After two weeks of culture, NK cells were purified with the NK Cell Isolation Kit (Miltenyi) and live-frozen. For all the experiments, NK cells from the same donor were used, with 3 days of culture between thawing and the experiments.

Synthesis of lineage barcodes library

The pBA439_UMI20 plasmid was constructed by ssDNA synthesis and NEBuilder® assembly tool. The lineage barcodes part ssDNA_UMI20 (5'-CTGGGGCACAAGCTTAATTAAGAATTCANNNNTGNNNNACNNNNGANNNNGTNNNNCTAGGGCCTAGAGGGCCCGTTTAAAC-3') was ordered from Integrated DNA Technologies. pBA439 was a gift from Jonathan Weissman (Addgene, plasmid # 85967). pBA439 vector was digested with EcoRI-HF (NEB, R3101S) and AvrII (NEB, R0174S), then the larger linear fragment pBA439_linear was recycled and purified by columns (Macherey-Nagel, 740609.50). The ssDNA_UMI20 was assembled with pBA439_linear to construct a pBA439_UMI20 library by the NEBuilder HiFi DNA assembly (NEB, E2621L). For amplification, the pBA439_UMI20 library was electroporated into 25 µl of Lucigen Endura ElectroCompetent Cells (Lucigen, 60242-2) using prechilled 1 mm electroporation cuvettes (BioRad, 1652089) in a BioRad GenePulser I machine set to 10 µF, 600 Ω, and 1800 V. Within seconds after the pulse, 1 ml of 37 °C Recovery Medium (Lucigen, 80026-1) was added and bacteria were grown in round-bottom tubes for 1 h at 37 °C while shaking at 180 r.p.m. Then, 1 ml of the bacterial culture was plated on a 25 × 25 cm bioassay plate containing LB medium with 100 µg/ml ampicillin. Plates were incubated at 32 °C for 22 h, then LB medium was added and cells were scraped off the plate. Bacterial cells were pelleted by 15 min of

centrifugation at 5,000 RCF at 4 °C, and plasmid DNA was extracted with a Maxiprep endotoxin-free kit (Macherey-Nagel, 740424.10).

Lentivirus production and cell transduction

Lentiviral particles were produced by transfecting 293FT cells (Thermo Fisher, R70007) with the pBA439_UMI20 library, along with packaging constructs psPAX2 (Addgene, plasmid # 12260), and pMD2.G (Addgene, plasmid # 12259). Constructs were titred by serial dilution on the 293FT cells. The lentivirus was collected 48 h and 72 h after transfection and aliquoted to store at -80 °C following filtering through a low-protein binding 0.45-µm filter.

Experimental BRCAness +/- models

To clone the *BRCA1/2* plasmids, lentiMPH v2 (Addgene, plasmid #89308) was modified by replacing Hygro selection marker with GFP to generate lentiMPH-GFP. *BRCA1* was amplified from pDEST-mCherry-LacR-BRCA1(Addgene, plasmid #71115), and *BRCA2* was amplified from pCIN *BRCA2* WT (Addgene, plasmid #16245). PCR amplified products (templates and primers presented in **Supplementary Table 4**) were assembled by the NEBuilder HiFi DNA assembly (NEB, E2621L) to generate lenti-EMPTY-GFP, lenti-BRCA1-GFP, and lenti-BRCA2-GFP. Lentivirus were made as described earlier to transduce either COV362 or KURAMOCHI cells to restore the *BRCA1* and *BRCA2* functions.

The COV362 model was validated with a colony formation assay, where COV362 control and *BRCA1* restored cells (10000 cells per well) were exposed to 0, 2.5 or 5 uM olaparib. Colonies were stained with crystal violet after 12 days of incubation.

ReSisTrace experimental workflow

Six million KURAMOCHI cells were transduced with the pBA439_UMI20 lentivirus library at MOI between 0.2 and 0.25, and 2 days after transduction 2 ug/ml puromycin were added for a seven day selection to remove unbarcoded cells. All barcoded cells were equally divided into 6 vials (6 millions cells per vial), 4 of which were cryopreserved and 2 were used for the experiment described below. Cells were synchronised with 2mM thymidine (Merck, T1895), then after releasing and washing out of thymidine, 16,000 cells/well were seeded on 6-well plates. After 48 hours, 4 wells were used for cell counting, and 8 wells were used for the ReSisTrace experiment. For each of the 8 samples, half of the cells were loaded for scRNA-seq as pre-treatment samples, and the remaining half were seeded back for either normal culture (13 days) or treated with carboplatin (Selleckchem, S1215, 1.2 µM, 3 days), olaparib (Selleckchem, S1060, 1.2 µM, 7 days), NK cells (26,000 cells, 1 day). The drug concentrations or number of NK cells were chosen to kill 70-80% of cancer cells compared to untreated cultures. After treatment, cells were recovered for 10, 6 and 7 days respectively, after which about 16,000 cells from each well were loaded for scRNA-seq as post-treatment samples.

Mathematical modelling and simulation of lineage tracing

We determined the expected number of different labels during the sample split. We asked the question: given a total of N cells with K lineage labels prior cell doubling, what is the probability of observing k different labels after the sample split if n cells are selected in one of the two subsamples? For $i = 1, \ldots k$, let $I_i = 1$ if label i is selected at least once, and the total number of labels selected

for the sample is $Y = \sum_1^k I_i$. Assuming sampling without replacement, we have the expectation of $E(Y) = \sum_1^k E(I_i)$, where $E(I_i) = 1 \cdot \Pr(I_i = 1) + 0 \cdot \Pr(I_i = 0)$ by the definition of expectation. By the theory of combinatorics, it follows that $\Pr(I_i = 1) = 1 - \Pr(I_i = 0) = 1 - b/a$, where a is the number of ways to choose n cells from N, and b is the number of ways to choose n cells, of which none has label i. Therefore,

$$E(Y) = K(1 - \frac{\binom{N-N/K}{n}}{\binom{N}{n}})$$

Accordingly, the number of labels that have only one copy is $K - (K - E(Y)) \cdot N/K$. For example, for an initial sample consisting of $N = 100$ cells with $K = 100$ labels, after doubling $N_{\text{post}} = 200$. When splitting the sample into two of equal sizes, the expected number of different labels in one of the resulting samples is 75.13, of which around 50 labels will have exactly one copy.

We also simulated the lineage tracing process by considering additional latent parameters, including the proportion of pre-resistant cells in the sample denoted as $r_{\text{pre-resistant}} \in (0.01, 0.3)$. The range of $r_{\text{pre-resistant}}$ was chosen to reflect the observed cell viability after the treatment. Furthermore, we considered the effect of drop-out rate during the lineage tracing , $d \in (0, 0.9)$. We fixed the initial sample size $N = K = 16{,}000$ as described earlier, and let the $r_{\text{pre-resistant}}$ and d vary. Using the pre-resistant cells as the positive group, we evaluated the true positive rate (TPR, i.e. the ratio of true positive versus total number of positive cells), and the false omission rate (FOR, i.e. the ratio of false negative versus predicted negative cells), and the balanced accuracy (BA, i.e. the average of TPR and true negative rate (TNR, which is 1 as all the pre-sensitive cells should be captured). Confidence intervals were obtained by 100 iterations.

Single-cell RNA-sequencing library preparation

On the 10x Genomics platform, for samples from ReSisTrace experiments, we used the Chromium Next GEM Single Cell 3′ Kit v3.1 (PN-1000268), the Chromium Next GEM Chip G Single Cell kit (PN-1000120), and the Dual Index Kit TT Set A (PN-1000215). For samples from isogenic BRCAness +/- COV362 cells, cells were multiplexed with the 3’ CellPlex Kit Set A (PN-1000261), then 7000 cells were loaded for scRNA-seq. For library preparation, the 3’ Feature Barcode Kit (PN-1000262) and the Dual Index Kit NN Set A (PN-1000243) were additionally used according to the manufacturer’s instructions. The resulting cDNA libraries were quantified by both Agilent BioAnalyzer (Agilent, 5067-4626) and q-PCR (Roche, KK4835), and sequenced by Illumina NovaSeq 6000.

NK cells mediated killing for BRCAness +/- models

On day 1, 8000 cancer cells per well were seeded on 96-well plates, and NK cells were cultured with 10 ng/ml IL2. On day 2, different numbers of NK cells were added into cancer cells. On day 3, cell viability was measured by the CellTiter-Glo® Luminescent Assay (Promega, G9242).

Lineage label mapping and processing

The lineage barcodes sequence (5’-CTGGGGCACAAGCTTAATTAAGAATTCANNNNTGNNNNACNNNNGANNNNGTNNNNC

TAGGGCCTAGAGGGCCCGTTTAAAC-3') was added to the GDCh38.d1.vd1 reference genome with GENCODE v25 annotation, as an individual gene named *pBA439_UMI_20* according to the parental lentiviral library. The Cell Ranger software (v5.0.1) was used to perform read alignment and UMI quantification using the modified reference.

Lineage label sequences for each cell were extracted from the "possorted_genome_bam.bam" file output by cellranger count command, by considering the sequences mapped to gene *pBA439_UMI_20* and the sub-pattern 'CANNNNTGNNNNACNNNNGANNNNGTNNNNCT'. To mitigate the effect of sequencing errors, the directional network-based method from UMI-tools (v1.0.1) was applied on the lineage label sequences. The sequences that differ at a single base from the representative sequence were corrected. In addition, we removed the non-unique lineage label sequences expressed by more than four cells in the pre-treatment sample. We assigned the cells from each pre-treatment sample into two groups: 1) pre-resistant cells with lineage labels matching those of the corresponding post-treatment sample; 2) pre-sensitive cells assigned to lineages that were not detected after the treatment.

Gene expression analysis from scRNA-seq data

We used the Seurat (v4.0.4) to perform the data quality control, normalisation, top variable gene selection, scaling, dimensionality reduction, and DEG analysis. For each ReSisTrace sample, the genes that express in less than 3 cells were removed. Based on the distribution of the UMI counts, number of genes and percentage of mitochondrial transcripts, we filtered out the cells using the thresholds indicated in **Supplementary Table 5**.

We used the LogNormalize method to normalise gene counts, and the FindVariableFeatures function to find the top 2,000 variable genes.The FindMarkers function was used to find the differentially expressed genes by Wilcoxon rank sum test between different groups. To control the experimental noise, when the sisters are not evenly divided during the sample split, we performed the differential analysis only on the cells lacking sisters in the same pre-treatment sample.

Sister cell similarity

We merged the pre-treatment data from all the experiments. Cells with matching labels were defined as sister cells, gaining 2939 lineages with sister cells. We compared the Euclidean distance of the transcriptomes for all the sister pairs versus random cell pairs in the pre-treatment samples, using the two-tailed t-test. Lowly expressed genes whose total normalised expressions are less than 1.25 (10% quantile) were removed. The UMAP coordinates for visualisation of randomly selected sister cells were determined with the first 20 principal components of the top 2,000 variable genes. For extracting the sister concordant genes, we considered only the lineages with two sister cells, and removed the sister cells with the Euclidean distance larger than 14.9 (90% quantile). The genes with $\log_2$ fold changes ($\log_2$FCs) significantly lower in the sister cell pairs compared to random cell pairs were defined as the sister-concordant genes (two-tailed t-test with an FDR-adjusted *P* value threshold of 0.05).

RNA velocity analysis

For follow-up analyses, we selected expression-matched sets of sister-concordant and sister-discordant genes. We generated a histogram plot for the sum of expression in the pooled samples, and selected genes that fell in the most balanced bin (bin for the second smallest expression levels).

We used the velocyto (v0.17.17) to align and quantify the spliced and unspliced RNA for each gene. Genes with RNA counts (both unspliced and spliced) less than 20 were removed. The dynamical model from scVelo (v0.2.3) was used to fit the gene-specific rates of RNA transcription, splicing, and degradation.

Shannon diversity

Shannon diversity of lineage labels for each sample was determined by the vegan R package (v 2.6.2).

Pathway analysis

KEGG pathways for the expression matched sister-concordant and discordant genes were determined by Enrichr (https://maayanlab.cloud/Enrichr/)[47,48]. Gene set enrichment analysis (GSEA) function from the clusterProfiler package (3.18.1) was used for the ordered fold changes (logarithmic scaled) extracted from the fitness corrected pre-resistance differential expression analysis to inspect the enriched gene ontology biological process terms.

Subclone analysis

CNV profiles for the individual cancer cells for each sample were obtained with InferCNV (1.7.1)[49] using a common set of stromal cells from a previously published data[50] as references. The Leiden algorithm was used for determining the underlying subclonal populations based on the CNV profiles. Exceptionally for the Olaparib1 sample, the algorithm was not able to recognise the smallest subclone (subclone E), due to the small sample size. Therefore, the inferCNV analysis was jointly run with the other Olaparib sample for the subclonal label determination. The Euclidean distance normalised by the variance (ie. Mahalanobis distance) for each cell pair within a CNV subclone was determined from the top 2,000 variable genes, using the distance R package (v 0.1.8). The distances for 100,000 random cell pairs were shown as the violin plots.

BRCAness signatures and NK signature

Two independent BRCAness signatures were derived. The clinically-derived BRCAness signature was extracted from The Cancer Genome Atlas Cohort (TCGA) cohort where HGSOC patients were stratified into BRCAness +/- groups based on a genomic scar analysis using ovaHRDscar[48]. For transcriptomic comparison between these groups, we used the *limma* R package (3.46.0)[51] on bulk expression data[51] that had been deconvoluted to retain only cancer cell derived expression using PRISM (0.9)[52]. The experimental BRCAness signature was extracted from the *BRCA1*+/- cell line pair by DEG analysis from the scRNA-seq data as described earlier. The $\log_2$FC between pre-resistant and pre-sensitive cells for each gene was averaged across replicates and then subtracted by the $\log_2$FC of pre-fitness versus pre-non-fitness cells from the non-treatment control samples. We calculated Pearson's correlation coefficients between the fitness-corrected pre-resistance $\log_2$FCs and subclonal $\log_2$FCs using sister-concordant genes only. For the BRCAness signatures, we used $\log_2$FCs of HRD to HRP using sister-concordant genes only, thresholded with indicated *P* values within the BRCAnsess sets, and determined the Spearman correlation to fitness-corrected pre-sensitive $\log_2$FCs.

The NK transcriptional signature was obtained from the coexpressed NK cell markers in our previous analysis of 22 clinical HGSOC tumours[50]. For the estimation of NK cell enrichment within the immune compartment of TCGA cohort, we calculated Mann-Whitney U statistic-like scores from the deconvoluted gene expressions using the *UCell* package (1.3.1)[52]. A Cox proportional hazard

regression model on the overall survival was built in R using the *survival* package (3.3.1)[53], with the formula: *coxph(Surv(time, status) ~ nk_signature:BRCAness_status),* wherein *'nk_signature'* refers to the UCell score described above.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper, Extended Data and/or Supplementary Information. Additional data related to this paper may be requested from the authors. scRNA-seq raw data and counts will be made available in Gene Expression Omnibus (GEO) with accession code (inserted prior publication).

Code availability

The source code for data analysis will be made available at GitHub (https://github.com/TangSoftwareLab/ReSisTrace) and Zenodo (inserted prior publication).

References

1. Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
2. Ramirez, M. *et al.* Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat. Commun.* **7**, 10690 (2016).
3. Hangauer, M. J. *et al.* Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature* **551**, 247–250 (2017).
4. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
5. Thomas, P., Terradot, G., Danos, V. & Weiße, A. Y. Sources, propagation and consequences of stochasticity in cellular growth. *Nat. Commun.* **9**, 4528 (2018).
6. Shaffer, S. M. *et al.* Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. *Cell* **182**, 947-959.e17 (2020).
7. Meyer, M. *et al.* Profiling the Non-genetic Origins of Cancer Drug Resistance with a Single-Cell Functional Genomics Approach Using Predictive Cell Dynamics. *Cell Syst* **11**, 367-374.e5 (2020).
8. Gutierrez, C. *et al.* Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. *Nat Cancer* **2**, 758–772 (2021).
9. Oren, Y. *et al.* Cycling cancer persister cells arise from lineages with distinct programs. *Nature* (2021) doi:10.1038/s41586-021-03796-6.
10. Goyal, Y. *et al.* Pre-determined diversity in resistant fates emerges from homogenous cells after anti-cancer drug treatment. *bioRxiv* (2021) doi:10.1101/2021.12.08.471833.
11. Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4**, 2126 (2013).

12. Ledermann, J. *et al.* Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. *Lancet Oncol.* **15**, 852–861 (2014).

13. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).

14. Phillips, N. E., Mandic, A., Omidi, S., Naef, F. & Suter, D. M. Memory and relatedness of transcriptional activity in mammalian cell lineages. *Nat. Commun.* **10**, 1208 (2019).

15. Lin, W.-H. W. *et al.* Asymmetric PI3K Signaling Driving Developmental and Regenerative Cell Fate Bifurcation. *Cell Rep.* **13**, 2203–2218 (2015).

16. Dey-Guha, I. *et al.* Asymmetric cancer cell division regulated by AKT. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12845–12850 (2011).

17. Dainichi, T. *et al.* PDK1 Is a Regulator of Epidermal Differentiation that Activates and Organizes Asymmetric Cell Division. *Cell Rep.* **15**, 1615–1623 (2016).

18. Mandic, A., Hansson, J., Linder, S. & Shoshan, M. C. Cisplatin induces endoplasmic reticulum stress and nucleus-independent apoptotic signaling. *J. Biol. Chem.* **278**, 9100–9106 (2003).

19. González-Huici, V., Wang, B. & Gartner, A. A Role for the Nonsense-Mediated mRNA Decay Pathway in Maintaining Genome Stability in. *Genetics* **206**, 1853–1864 (2017).

20. Litchfield, K. *et al.* Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat. Commun.* **11**, 3800 (2020).

21. Lahiguera, Á. *et al.* Tumors defective in homologous recombination rely on oxidative metabolism: relevance to treatments with PARP inhibitors. *EMBO Mol. Med.* **12**, e11217 (2020).

22. Belkahla, S. *et al.* The metabolism of cells regulates their sensitivity to NK cells depending on p53 status. *Sci. Rep.* **12**, 3234 (2022).

23. Nolte-'t Hoen, E. N. M. *et al.* Increased surveillance of cells in mitosis by human NK cells suggests a novel strategy for limiting tumor growth and viral replication. *Blood* **109**, 670–673 (2007).

24. Terasima, T. & Tolmach, L. J. X-ray sensitivity and DNA synthesis in synchronous populations of HeLa cells. *Science* **140**, 490–492 (1963).

25. Simoneau, A., Xiong, R. & Zou, L. The cell cycle effects of PARP inhibitors underlie their selectivity toward BRCA1/2-deficient cells. *Genes Dev.* **35**, 1271–1289 (2021).

26. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).

27. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).

28. Häkkinen, A. *et al.* PRISM: Recovering cell type specific expression profiles from individual composite RNA-seq samples. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab178.

29. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).

30. Perez-Villatoro, F. *et al.* Optimized detection of somatic allelic imbalances specific for homologous recombination deficiency improves the prediction of clinical outcomes in ovarian cancer. *bioRxiv* (2021) doi:10.1101/2021.08.19.456809.

31. Clarke, B. *et al.* Intraepithelial T cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. *Mod. Pathol.* **22**, 393–402 (2009).

32. Hsiehchen, D. *et al.* DNA Repair Gene Mutations as Predictors of Immune Checkpoint Inhibitor Response beyond Tumor Mutation Burden. *Cell Rep Med* **1**, (2020).

33. Bhang, H.-E. C. *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).

34. Contreras-Trujillo, H. *et al.* Deciphering intratumoral heterogeneity using integrated clonal tracking and single-cell transcriptome analyses. *Nat. Commun.* **12**, 6522 (2021).

35. Belderbos, M. E. *et al.* Clonal selection and asymmetric distribution of human leukemia in murine xenografts revealed by cellular barcoding. *Blood* **129**, 3210–3220 (2017).

36. Kimmerling, R. J. *et al.* A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat. Commun.* **7**, 10220 (2016).

37. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, (2020).

38. Dey-Guha, I. *et al.* A mechanism for asymmetric cell division resulting in proliferative asynchronicity. *Mol. Cancer Res.* **13**, 223–230 (2015).

39. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).

40. Leary, R. J. *et al.* Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16224–16229 (2008).

41. Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat. Biotechnol.* (2021) doi:10.1038/s41587-021-00837-3.

42. Gasser, S., Orsulic, S., Brown, E. J. & Raulet, D. H. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* **436**, 1186–1190 (2005).

43. Wang, R. W., Viganò, S., Ben-David, U., Amon, A. & Santaguida, S. Aneuploid senescent cells activate NF-κB to promote their immune clearance by NK cells. *EMBO Rep.* **22**, e52032 (2021).

44. Launonen, I.-M. *et al.* Single-cell tumor-immune microenvironment of BRCA1/2 mutated high-grade serous ovarian cancer. *Nat. Commun.* **13**, 835 (2022).

45. Thompson, D., Easton, D. F. & Breast Cancer Linkage Consortium. Cancer Incidence in BRCA1 mutation carriers. *J. Natl. Cancer Inst.* **94**, 1358–1365 (2002).

46. Denman, C. J. *et al.* Membrane-bound IL-21 promotes sustained ex vivo proliferation of human natural killer cells. *PLoS One* **7**, e30264 (2012).

47. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-7 (2016).

48. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).

49. T. Tickle, G. C. Ti, M. Brown, B. Haas. inferCNV of the Trinity CTAT Project (Klarman Cell Observatory, Broad Institute of MIT and Harvard). (2019).

50. Zhang, K. *et al.* Longitudinal single-cell RNA-seq analysis reveals stress-promoted chemoresistance in metastatic ovarian cancer. *Sci Adv* **8**, eabm1831 (2022).

51. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and

microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

52. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
53. Therneau, T. M. & Grambsch, P. M. Modeling Survival Data: Extending the Cox Model. *Statistics for Biology and Health* (2000) doi:10.1007/978-1-4757-3294-8.

Acknowledgements

The results are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This work was supported by The Sigrid Jusélius Foundation (AV and JT), The Cancer Foundation Finland (AV), The Academy of Finland projects 289059 (AV), 319243 (AV), 317680 (JT), 320131 (JT), ERA PerMed JTC2020 PARIS/Academy of Finland project 344697 (AV), European Research Council No. 716063 (JT).

Cartoons in Figures 3a and 3e were created with BioRender.com.

Author contributions

Study design, data interpretation and preparation of the manuscript: AV, JD, JT, SZ, MF, JE. Execution of experiments: JD, JE, JB; NK cell preparation: OD, KS. Computational and statistical analysis: SZ, JT, MF, AA, YW and FP. Resources: AV, JT, SM, AF. All authors wrote the paper.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Requests for material, and correspondence should be addressed to Jing Tang and Anna Vähärautio.

Figures and legends

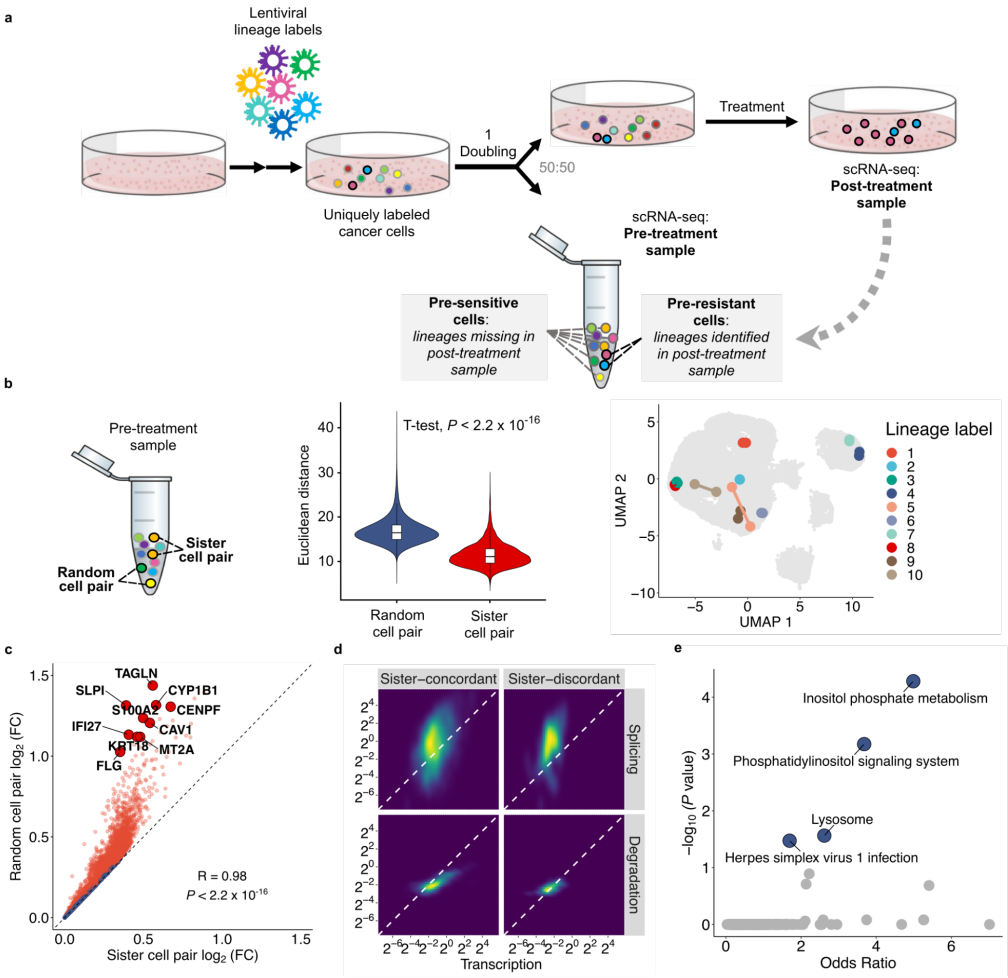


Figure 1: ReSisTrace leverages sister cell similarity to discover cellular states primed for resistance. **a**, Schematic representation of the ReSisTrace approach. Uniquely labelled cells are synchronised and allowed to divide once. Subsequently, the sample is divided in two for scRNA-seq and treatment. Next, barcodes of recovered cells in the post-treatment sample are determined by scRNA-seq and used to identify pre-resistant cells in the pre-treatment sample. **b**, Sister cell similarity. Left, schematic of sister cell pairs and random cell pairs in a pre-treatment sample. Middle, Euclidean distance of sister cell pairs in comparison to random cell pairs ($P < 2.2 \times 10^{-16}$, two-tailed t-test; random cell pairs, n = 6.85×10^{8}; sister cell pairs, n = 3,477). Right, a subset of randomly selected 10 sister cell pairs shown in the Uniform manifold approximation and projection (UMAP). **c**, Gene expression fold changes in sister cell pairs (x axis) versus random cell pairs (y axis). Sister-concordant genes are displayed as red, and sister-discordant genes as blue. **d**, Inferred splicing and degradation rates of 919 sister-concordant and 107 sister-discordant genes. **e**, Pathway enrichment of 1,332 sister-discordant genes.

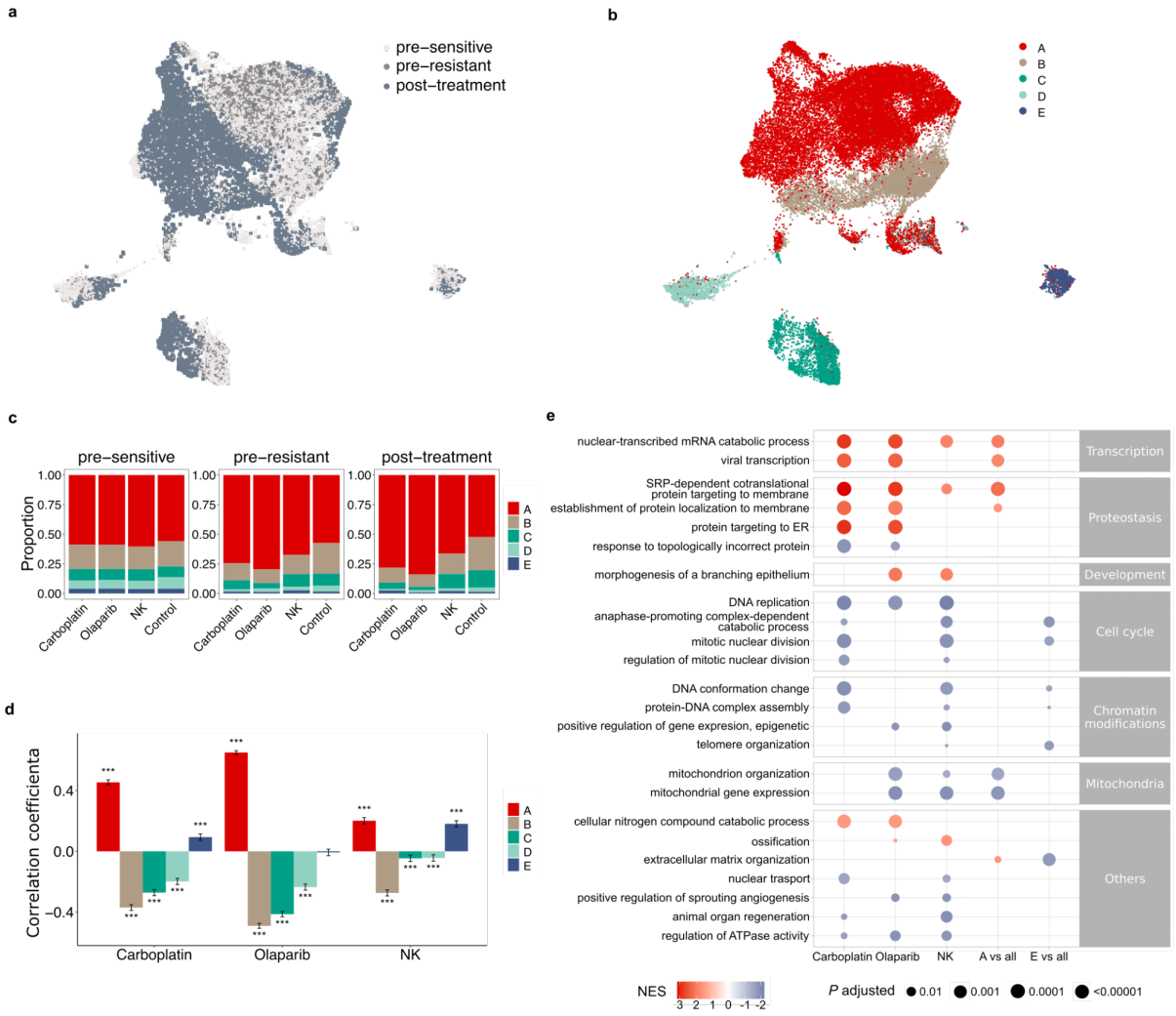


Figure 2: ReSisTrace uncovers the interplay between genetic and transcriptomic features in primed resistance. **a**, **b**, UMAP projection of single cells colored by (**a**) the treatment groups (pre-sensitive n = 19,540, pre-resistant n = 3,102, and post-treatment n = 7,802), or (**b**) the inferred CNV subclones (A, n = 24,903; B, n = 7,904; C, n = 3,824; D, n = 2,482; E, n = 1,331). **c**, Subclonal proportions in pre-sensitive, pre-resistant, and post-treatment populations for different treatments. **d**, Pearson's correlation coefficients between the pre-resistance signatures and the subclone specific signatures. Error bars indicate the 95% confidence intervals of the correlation coefficients. ***P < 0.001; t test; n = 8,419. **e**, Representative gene ontologies enriched in the pre-resistance signatures as well as the signatures of subclones A and E.

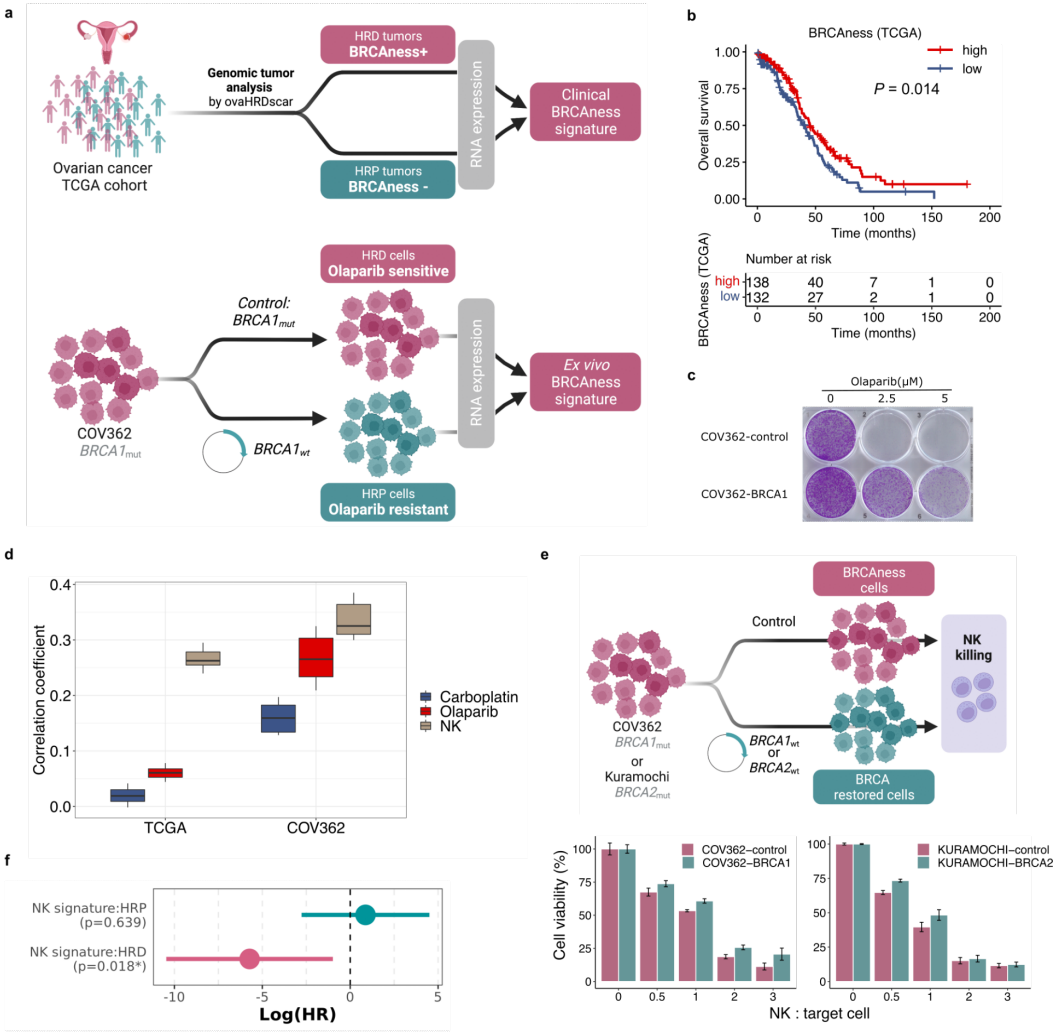
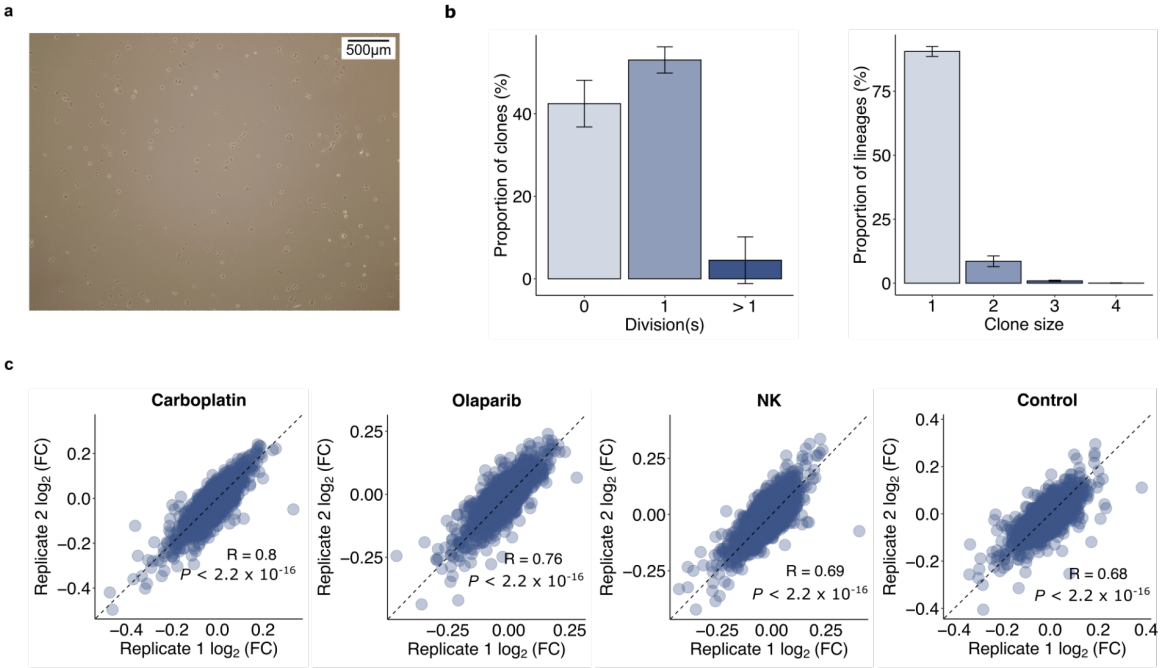
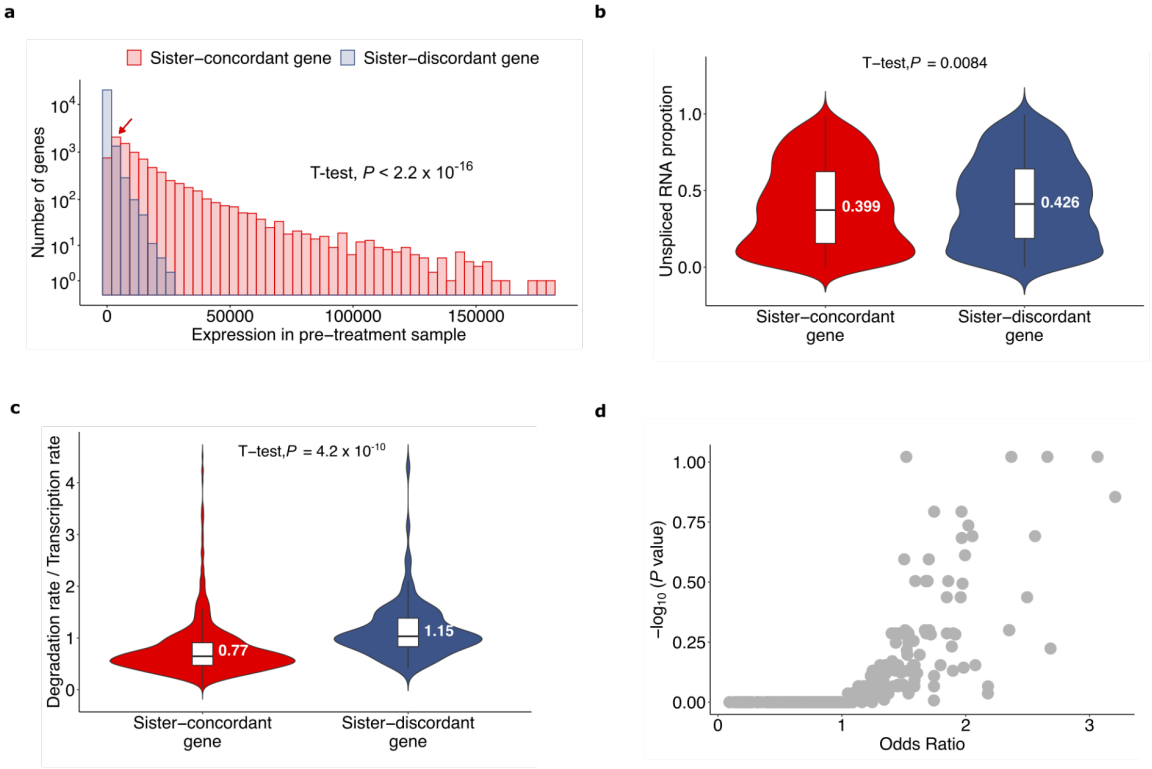


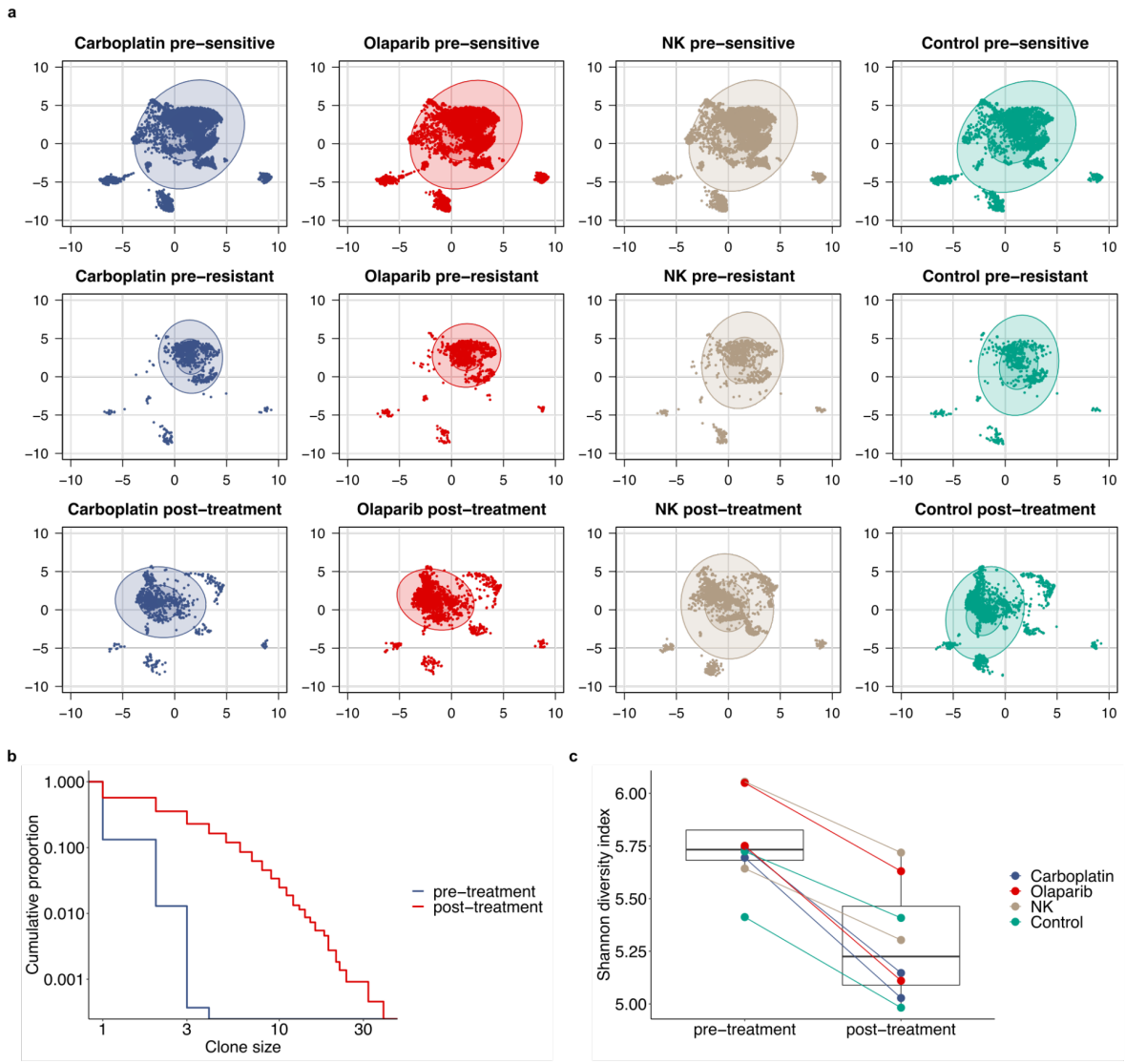
Figure 3: ReSisTrace reveals a novel type of immune vulnerability of BRCAness cancer. a, Schematic representation of the methods to determine clinical (top) and experimental (bottom) BRCAness signatures. Clinical BRCAness transcriptional signature was extracted from The Cancer Genome Atlas (TCGA) expression data after stratification by ovaHRDscar. Experimental BRCAness signature was derived from scRNA-seq analysis of parental (*BRCA1* mutant) COV362 cells against *BRCA1*-restored cells. **b**, Kaplan-Mayer curves for the overall survival in TCGA cohort with ovarian cancer, stratified to high and low groups by the clinical BRCAness signature (P = 0.014, log-rank test). **c**, Colony formation assay validating the differential sensitivity to Olaparib of COV362 control and *BRCA1*-restored cells. **d**, Spearman's rank correlation coefficients between pre-sensitive gene signatures and clinical (left), or experimental (right) BRCAness signature. Quantiles were determined by varying the adjusted P value threshold for filtering the genes in the clinical BRCAness signature ($0.005 < P < 0.05$, n = [207, 879]), and in the experimental BRCAness signature ($5\times10^{-16} < P < 0.05$, n = [211, 946]). **e**, Top, schematic representation for NK cells mediated killing experiments using isogenic BRCAness models; bottom left, measurements of NK cell-mediated killing of COV362 control and *BRCA1*-restored cells; bottom right, KURAMOCHI control and *BRCA2*-restored cells at indicated ratios of NK to target cells. Error bars are mean +/- s.d. of three replicates. **f**, Hazard ratios of overall survival in logarithmic scale for the interaction between NK transcriptional signature, and patients' HRD/HRP status. P values are determined by the Cox regression models. Error bars are 95% confidence intervals.



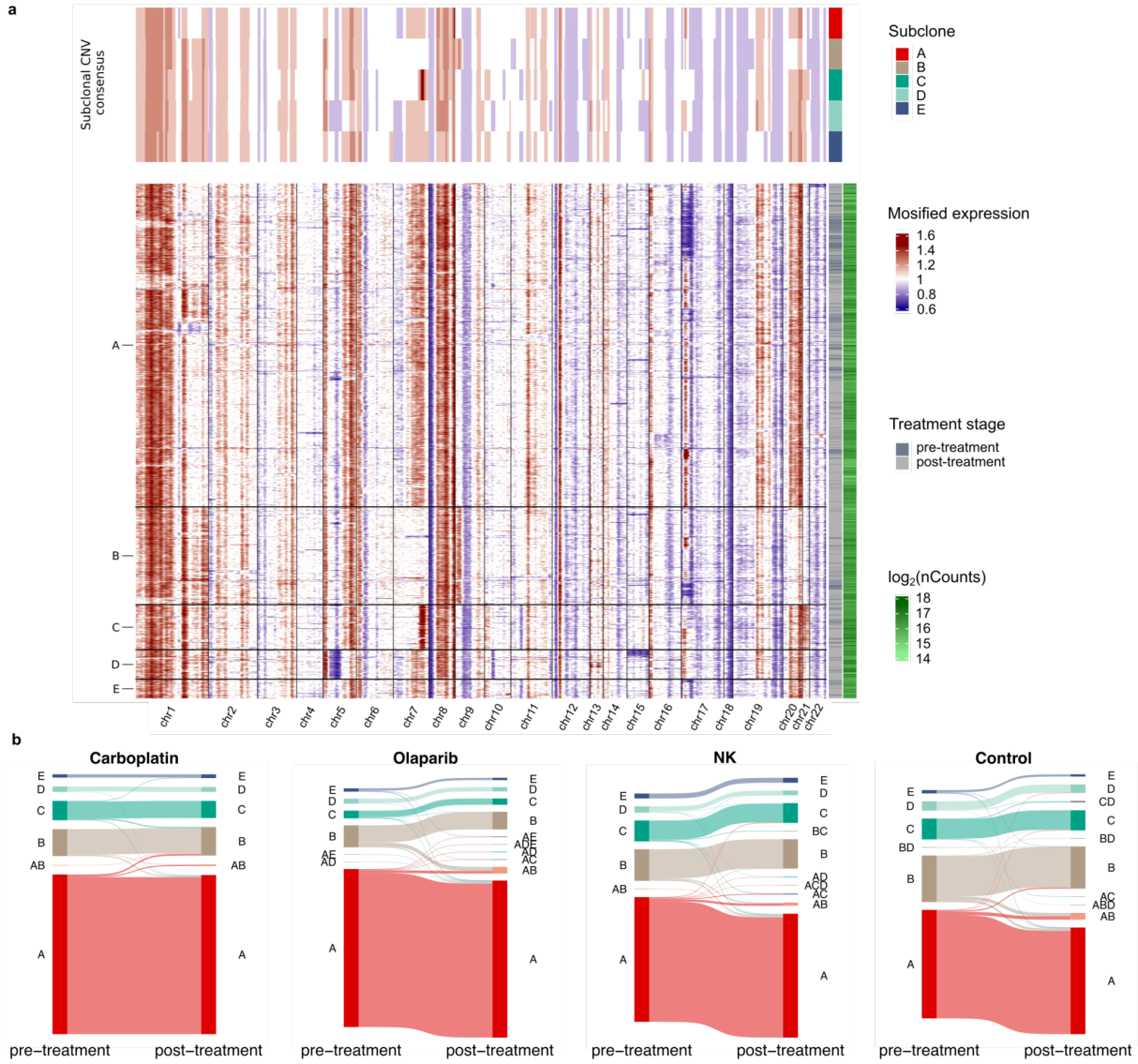
Extended Data Figure 1. Cell doubling, lineage labelling efficiency, and quality assessment of ReSisTrace. a, A representative phase-contrast image of cells after one doubling time of 48 hours. **b**, Proportion of clones for different numbers of divisions (left), and proportion of lineages for different clone sizes (right). Error bars are mean +/- s.d. of six images (left), or of eight samples (4 treatment conditions each with 2 replicates; right). **c**, Consistency of gene expression changes for pre-resistant against pre-sensitive cells between the replicates of each treatment condition (Carboplatin, n = 30,613; Olaparib, n =30,586 ; NK, n = 30,613; Control, n = 27,478).



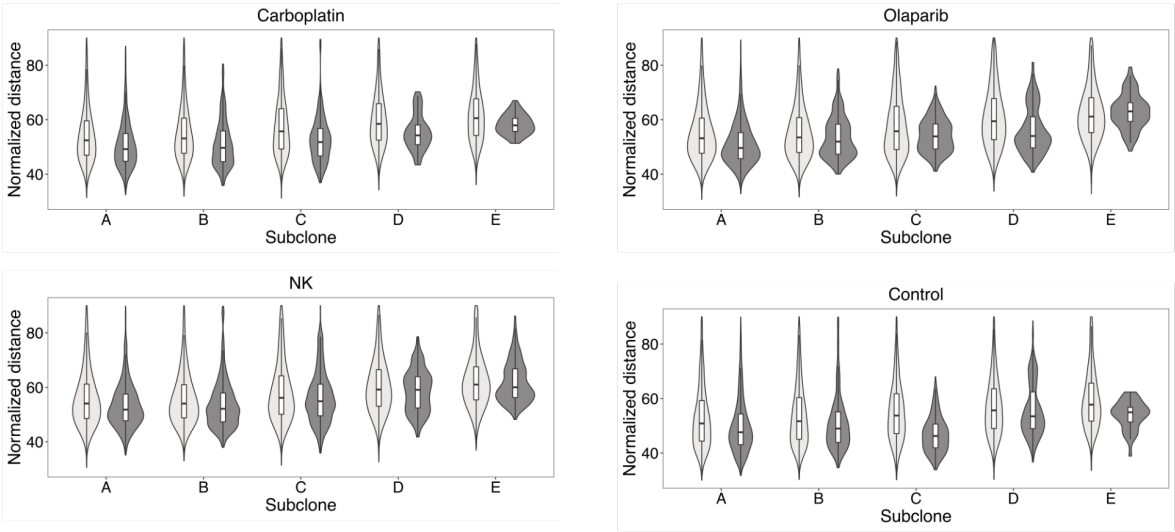
Extended Data Figure 2. Characteristics of sister-concordant and sister-discordant genes. a, Histogram showing the number of genes at different expression levels of sister-concordant (red, n = 8,419) or sister-discordant (blue, n = 21,896) genes in the pre-treatment samples. Red arrow indicates the expression range ([1839.89, 5519.11]) used to determine the subsets of sister-concordant and sister-discordant genes in the RNA velocity and pathway enrichment analyses. **b**, **c**, Average proportion of unspliced RNAs (**b**, sister-concordant, n = 1,916; sister-discordant, n = 1,257) and average RNA degradation to transcription ratio (**c**, sister-concordant, n = 919; sister-discordant, n = 107) of sister-concordant and discordant genes. **d**, Pathway enrichment of sister-concordant genes.



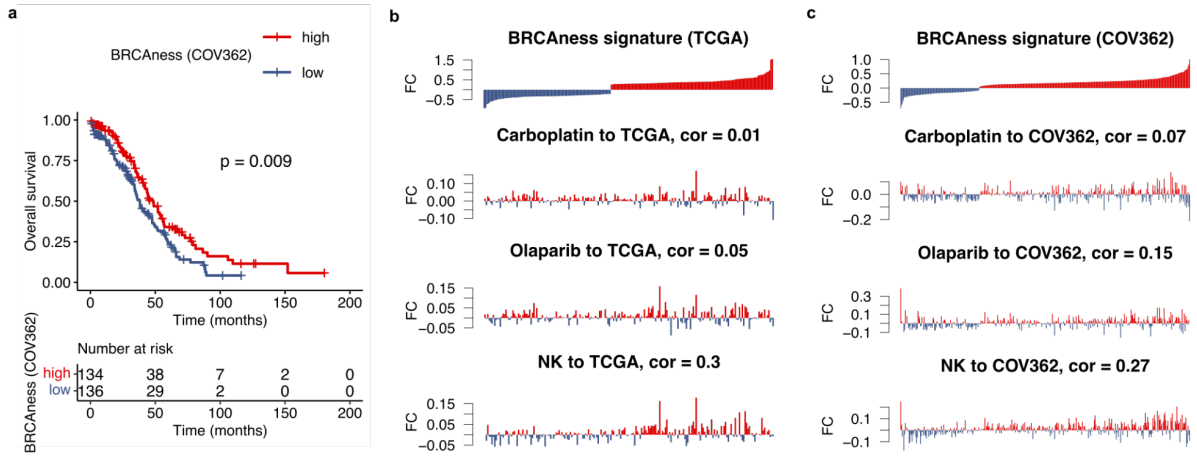
Extended Data Figure 3. Diversity of ReSisTrace populations. a, UMAP projection of pre-sensitive, pre-resistant, and post-treatment cells for each treatment. **b**, Cumulative proportion of lineages with different clone sizes in the pre-treatment (blue) or post-treatment (red) samples. **c**, Shannon diversity index in the pre-treatment or post-treatment samples for each replicate in different treatment conditions.



Extended Data Figure 4. Lineages maintain their subclonal identities during the treatment. a, Heatmap showing the subclonal CNV profiles of the carboplatin samples inferred from their scRNA-seq data. The lower panel shows the CNV profiles for the individual cells. The upper panel shows the consensus CNV profiles for each identified subclone. **b**, Sankey plot for the subclonal dynamics during each treatment. Each subclone is represented by a colored band, the height of which corresponds to the proportion of identified lineages.



Extended Data Figure 5. Subclone-specific distances of pre-sensitive (light grey) and pre-resistant (dark grey) populations.



Extended Data Figure 6. Patient survival and treatment-specific pre-sensitivity associated with the transcriptional BRACAness signatures. a, Kaplan-Mayer curves for the overall survival in TCGA cohort with ovarian cancer, stratified to high and low groups by the experimental BRCAness signature (COV362) ($P = 0.009$, log-rank test). The genes in the signatures were selected by adjusted P value threshold of 1×10^{-8}. **b**, **c**, Waterfall plots for the clinical BRCAness signature (TCGA) with 207 genes (x axis) selected by adjusted P value threshold of 0.005 (**b**), or the experimental BRCAness signature (COV362) with 386 genes (x axis) selected by the adjusted P threshold of 1×10^{-8} (**c**). The y axes show the $\log_2$FC of the genes when comparing HRD to HRP (the top row panel of **b** or **c**), or comparing pre-sensitive to pre-resistant cells from different treatments as indicated by the captions in the other panels.

Supplementary Information

Supplementary Table 1. Mean and standard deviation for the true positive rate (TPR), false omission rate (FOR), and balanced accuracy (BA) for indicated ranges of pre-resistant cell proportions and drop-out rates.

Supplementary Table 2. Pre-resistance gene expression signatures for the treatments, the CNV subclone signatures, and BRCAness signatures

Supplementary Table 3. NK transcriptional signature

Supplementary Table 4. Primers for experimental BRCAness model constructs

Supplementary Table 5. Thresholds for the scRNA-seq data quality control