Using Gene Set Enrichment to Detect Proliferation Confounding in Single-Cell Signatures

Krishay Uthayasegar

Dublin High School

Abstract

Small single-cell RNA sequencing cohorts are increasingly used to search for pre-treatment cell states that predict chemotherapy resistance, but candidate signatures from such cohorts are vulnerable to three failure modes: cell-level pseudoreplication, patient identity masquerading as shared biology, and gene programs that capture proliferation rather than resistance. This study asks one methodological question: can gene set enrichment, supported by patient-level and batch-aware checks, determine whether a candidate resistance signature reflects resistance biology or an analytical artifact? As a test case, we reanalyzed a public paired pre- and post-chemotherapy dataset (GSE205472) from five hormone receptor positive (HR+) breast cancer patients (three resistant, two sensitive; five of the eight patients in the source study). A signature derived from genes upregulated after treatment in resistant patients was dominated by cell-cycle genes. Over-representation analysis showed strong enrichment for the Hallmark G2-M Checkpoint pathway (adjusted $p = 1.4 \times 10^{-9}$), and ranked gene set enrichment analysis of patient-level expression change placed G2-M Checkpoint, E2F Targets, and Mitotic Spindle at the top (false discovery rate < 0.001); after regressing out cell-cycle scores on the full gene matrix, these pathways were no longer enriched (false discovery rate 0.24 to 0.69). The signature scored higher, not lower, in pre-treatment cells from sensitive patients (exact one-sided $p = 1.0$). A signature-independent clustering test appeared positive ($p = 0.069$ by normal approximation), but the smallest exact p-value attainable with three versus two patients is 0.10, and the result rested on one patient contributing six pre-treatment tumor cells; after Harmony correction for patient identity that patient's membership in the post-treatment compartment fell from 100% to 0% (exact $p = 0.30$). The number of genes rising after treatment in all three resistant patients (2,518) did not exceed a label-shuffling null (empirical $p = 0.15$), and in simulations without any true effect, a test treating cells as independent produced false positives in 38% to 46% of cohorts at a nominal 5% when patients differed, whereas the exact patient-level test did not exceed its nominal rate.

These results provide no support for a pre-existing resistance signal in this cohort and show how enrichment testing can reveal what a candidate signature actually measures. We present the analysis as a case study and checklist for small cohorts.

Keywords

Gene set enrichment, Ranked GSEA, Proliferation confounding, Single-cell RNA sequencing, Pseudobulk analysis, Batch effects, Harmony integration, Reproducibility, Pseudoreplication, Chemotherapy resistance, Breast cancer

1. Introduction

Breast cancer is the most frequently diagnosed cancer among women worldwide, with approximately 2.3 million new cases in 2020 (Sung et al., 2021). Hormone receptor positive, HER2-negative disease accounts for about 73% of breast cancers diagnosed in the United States (Howlader et al., 2014), and HR+ tumors are also treated with neoadjuvant chemotherapy (NAC) before surgery to reduce tumor burden (Jia et al., 2024). Pathological complete response (pCR), defined as the absence of residual invasive tumor in the breast and axillary lymph nodes at surgery, is the standard readout of chemotherapy sensitivity in this setting. Rates of pCR are lowest in HR+ disease and can be as low as 0.3% in luminal A tumors, and the molecular mechanisms of this chemoresistance are poorly understood (Haque et al., 2018).

NAC for HR+ breast cancer is a multi-drug regimen delivered over several cycles, typically combining an anthracycline, an alkylating agent such as cyclophosphamide, and a taxane. Because response varies by subtype and by patient within a subtype, identifying molecular features that predict response before treatment begins is a persistent goal of translational research.

A major obstacle is intratumoral heterogeneity. Tumors contain molecularly distinct subpopulations with divergent transcriptional programs and drug sensitivities (Meacham & Morrison, 2013), and bulk RNA sequencing averages expression across millions of cells, masking rare subpopulations that may drive treatment failure (Tanay & Regev, 2017). Two models have been proposed to explain chemoresistance: acquired resistance, in which treatment induces a resistant state, and pre-existing resistance, in which rare cells with intrinsic resistance programs survive because they were present before treatment began. A reversible drug-tolerant state of this kind was described in cancer cell line models (Sharma et al., 2010). Single-cell studies in breast cancer have since reported evidence for both. In triple-negative disease, chemoresistant genotypes were detectable before treatment alongside transcriptional programs acquired during therapy (Kim et al., 2018), and the HR+ dataset reanalyzed here distinguished innate from acquired chemoresistance programs on the basis of copy number (Jia et al.,

2024). Distinguishing the models matters clinically, because pre-existing resistance could in principle be detected and targeted before treatment starts.

Studies that search for pre-treatment resistant states share structural vulnerabilities. Cohorts are small, often fewer than ten patients; the unit of analysis is frequently the cell rather than the patient; and candidate signatures are rarely tested for confounding by generic processes such as cell-cycle activity. Cell-level tests treat thousands of correlated cells from one patient as independent observations, which inflates false positive rates (Zimmerman et al., 2021), and pseudobulk approaches that aggregate cells within each patient address this problem (Squair et al., 2021). Patient identity is a major source of variation that integration methods such as Harmony aim to remove (Korsunsky et al., 2019), although methods differ in how well they separate technical from biological variation (Luecken et al., 2022). Finally, a signature built from post-versus-pre differential expression can capture proliferation instead of resistance, because chemotherapy preferentially eliminates actively cycling cells and higher proliferation is associated with higher pathological response (Denkert et al., 2013). Each of these points is known individually. Whether a particular candidate signature is affected is rarely tested, and worked examples on real data remain useful.

Research question. This study asks one question: can gene set enrichment analysis, supported by patient-level and batch-aware checks, determine whether a candidate pre-treatment resistance signature from a small single-cell cohort reflects resistance biology or an analytical artifact? We use a published paired dataset of HR+ breast cancer (GSE205472) as the test case. The biological hypothesis that motivated the test case, that chemotherapy-resistant tumors contain pre-treatment cells resembling those that survive therapy, is treated here as the claim under test and not as a second aim. We report each check in full, including where it undermines the initial result.

2. Methods

This study is retrospective and exploratory. Because chemotherapy response was known for all five patients before any analysis, the resistance signature was derived using outcome information, and testing it in the pre-treatment samples of the same patients is a test of retrospective self-consistency, not

an independent predictive test. The design can identify candidate patterns for adequately powered, independently validated cohorts; it cannot establish that pre-treatment resistant-like cells exist or can be used to predict response prospectively.

2.1 Dataset and preprocessing

Single-cell RNA sequencing data were obtained from the NCBI Gene Expression Omnibus (accession GSE205472), which contains tumor biopsies from eight HR+ breast cancer patients (Jia et al., 2024). Five of the eight had paired pre- and post-treatment single-cell samples passing quality control; these five patients are the cohort analyzed throughout this paper, because the post-versus-pre comparison requires both timepoints. In the source study, patients received standard neoadjuvant chemotherapy over eight cycles, with per-patient regimens reported there (Jia et al., 2024). Three patients (SC-P04, SC-P05, SC-P07) were chemotherapy resistant and two (SC-P01, SC-P02) were sensitive, based on the clinical resistance labels provided with the dataset. Count matrices were loaded with Scanpy's read_10x_mtx function (Wolf et al., 2018), using gene symbols as features (Scanpy v1.11.1), and sample identity and treatment state were mapped to barcodes by barcode suffix.

Quality control was applied in two stages. Cells expressing fewer than 200 genes and genes detected in fewer than three cells were removed, and cells with more than 20% mitochondrial content were excluded as putative apoptotic or damaged cells. These thresholds removed no cells from the deposited matrix, which appears to have been filtered before deposition, so 91,235 cells entered the analysis (cell counts at each stage are in Table 1). Counts were normalized to 10,000 reads per cell and log transformed with Scanpy's normalize_total and log1p functions. The 2,000 most highly variable genes were used for dimensionality reduction, clustering, tumor scoring, and signature derivation (Sections 3.1 to 3.4.2). A copy of the full log-normalized matrix (all genes) was kept separately and used for the cell-cycle regression, pseudobulk permutation, and ranked enrichment analyses (Sections 3.4.1, 3.4.3, and 3.4.4). Earlier versions of this paper stated that the full matrix was used for scoring; that was incorrect for the main pipeline, and the restriction to highly variable genes is why the first cell-cycle check was inconclusive.

2.2 Dimensionality reduction and clustering

Principal component analysis used the ARPACK solver in Scanpy with 30 components. A k-nearest-neighbor graph was built with k = 15, UMAP embeddings were computed (McInnes et al., 2018), and Leiden clustering was performed (Traag et al., 2019). Clustering used the full multi-patient dataset without batch correction, because all samples were provided as a single pre-merged count matrix.

2.3 Cell type annotation and removal of immune cells

Cell types were annotated with CellTypist v1.3 using the Cells_Adult_Breast reference model and majority voting (Domínguez Conde et al., 2022). Cells assigned to immune labels (T cell, B cell, natural killer, macrophage, monocyte, dendritic, mast, plasma, and granulocyte related) were removed, because immune and non-immune populations form highly distinct groups that dominate the neighbor graph and obscure finer distinctions among epithelial cells.

2.4 Luminal subsetting and re-clustering

Because breast cancer arises predominantly from luminal epithelial cells, the immune-depleted dataset was subset to cells with a luminal epithelial CellTypist label. Principal component analysis, neighbor graph construction, Leiden clustering, and UMAP embedding were recomputed on this subset independently of the full-dataset clustering.

2.5 Quantitative tumor versus stromal labeling

To label tumor versus stromal cells within the luminal compartment without manual cluster assignment, gene set scores were computed for every luminal cell with Scanpy's score_genes function. The tumor score used TSPAN1, CDC20, ID3, MKI67, TOP2A, TNFRSF18, and TNFRSF4, and the stromal score used COL1A1, COL1A2, VWF, PECAM1, ACTA2, VWA1, HSPG2, TIE1, and ANGPTL7. The tumor score threshold was selected using the biological expectation that post-treatment tumor cells should be depleted in chemotherapy-sensitive patients, whose tumors were eliminated by therapy, whereas pre-treatment cells and post-treatment resistant cells should retain tumor cells. Cells above the threshold were labeled tumor and all others stromal. This step is circular with respect to the outcome: because the threshold was informed by the expected treatment response, the tumor classification

is not independent of chemotherapy response, and all downstream resistance comparisons inherit this dependency (Section 4.1). The tumor score also contains the proliferation markers MKI67, TOP2A, and CDC20, so the label preferentially selects cycling cells; Section 3.1 shows that this has a large effect on the cell-cycle composition of the labeled cells.

2.6 Resistance classification

Patient-level resistance status was taken directly from the clinical labels provided with GSE205472 (Jia et al., 2024): SC-P04, SC-P05, and SC-P07 were resistant and SC-P01 and SC-P02 were sensitive. These labels were used for all comparisons.

2.7 Differential expression and resistance signature

For each resistant patient, differential expression between post-treatment and pre-treatment tumor cells was computed with the Wilcoxon rank-sum test (Scanpy rank_genes_groups), with post-treatment cells as the test group. Results were filtered to genes with log-fold-change > 0 and adjusted $p < 0.05$. These are cell-level tests, which treat cells as independent and therefore reintroduce the pseudoreplication problem examined in this paper (Section 4.1). Patient SC-P07 yielded no significantly upregulated genes because only six pre-treatment tumor cells were available, so it was excluded from signature derivation but retained as a resistant patient in all hypothesis testing. Rank aggregation across SC-P04 and SC-P05 summed each gene's per-patient ranks, assigning rank 1001 to a gene absent from a patient's 1,000-gene list (one position beyond the end of the list), and the 50 genes with the lowest aggregate rank formed the resistance signature.

2.8 Resistance scoring and hypothesis testing

Per-cell resistance scores were computed across tumor cells with Scanpy's score_genes function, which compares mean expression of the signature genes with an expression-matched control set (Tirosh et al., 2016), using a fixed random seed. Two metrics were computed on pre-treatment tumor cells for each patient: the fraction of cells with a resistance score above 0.3, and the fraction falling within the post-treatment compartment, defined as tumor subclusters that were at least 90% post-treatment cells. The 0.3 cutoff, the 90% composition threshold, and the rank-1001 convention have no independent

justification or sensitivity analysis behind them and should be regarded as arbitrary analysis choices (Section 4.1).

Groups of patients were compared with one-sided Mann-Whitney U tests (Mann & Whitney, 1947) on the per-patient values. With three resistant and two sensitive patients there are only C(5,2) = 10 possible group assignments, so the smallest attainable exact one-sided *p*-value is 1/10 = 0.10. We therefore report exact *p*-values from the full enumeration. Where earlier versions of this analysis reported the normal approximation with tie correction, which is the default in common software when tied values are present, we also give that value and label it as an approximation.

2.9 Robustness checks

Seven checks were applied to the candidate signal. They use the same five patients and tumor labels, so they are complementary and not independent: each probes a different failure mode and none provides an independent replication.

Gene set enrichment. To characterize what the signature measures, over-representation analysis was performed on the original 50-gene signature with the Enrichr web service through the gseapy package (Kuleshov et al., 2016), testing against the MSigDB Hallmark collection (Liberzon et al., 2015; Subramanian et al., 2005), Gene Ontology Biological Process (Ashburner et al., 2000; Gene Ontology Consortium, 2023), KEGG (Kanehisa & Goto, 2000), and Reactome (Milacic et al., 2024) libraries. Enrichr tests whether a fixed gene list contains more members of a pathway than expected by chance; adjusted *p*-values (Benjamini-Hochberg) ranked the terms. Because that is not ranked gene set enrichment analysis of the Subramanian et al. (2005) type, we also ran ranked analysis with gseapy prerank on the full gene matrix. Genes were ranked by the mean, across the three resistant patients, of each patient's pseudobulk change (post minus pre mean log-normalized expression), with 1,000 gene-set permutations, gene sets of 10 to 500 genes, and the same four libraries. The ranked analysis was repeated after cell-cycle regression (below). Ranking by patient-level change describes the data and is not itself a patient-level significance test.

Batch correction. Harmony integration was applied to the tumor-cell principal component embedding (50 components) with patient as the batch variable (Korsunsky et al., 2019). The neighbor graph, Leiden clustering (resolution 1.0), and UMAP were recomputed with the same parameters, the post-treatment compartment was redefined on the corrected clusters with the same 90% threshold, and the hypothesis test was repeated.

Patient-level pseudobulk analysis. Expression was collapsed to one profile per patient per timepoint by averaging log-normalized expression across that patient's tumor cells. A gene was retained in a pseudobulk-derived set only if its expression was higher after treatment than before in every resistant patient. A pseudobulk score was computed per patient per timepoint as the mean expression of a gene set, and pre- versus post-treatment scores were compared across the three resistant patients with an exact paired one-sided signed-rank test ($n = 3$; smallest attainable $p = 0.125$). Under random direction each gene has a 1-in-8 chance of being higher after treatment in all three patients, so the size of the retained set must be judged against chance. We therefore repeated the analysis on the full gene matrix and compared the number of retained genes with a null built by shuffling the timepoint labels of cells within each patient (1,000 permutations).

Cell-cycle regression. Tumor cells were scored for S and G2/M phase with the marker sets of Tirosh et al. (2016) using Scanpy's score_genes_cell_cycle, the scores were regressed out of the expression matrix for every gene with regress_out, and the signature was re-derived with the procedure of Section 2.7. This was done on the full gene matrix (all genes detected in at least 10 tumor cells), so that 42 of 43 S-phase and 52 of 54 G2/M markers were present; an earlier version restricted to the 2,000 highly variable genes had only 3 and 13.

Null simulation. To quantify how much treating cells as independent inflates false positives at this cohort size, we simulated 2,000 cohorts per setting with no true difference between groups. Patients had the observed numbers of pre-treatment tumor cells (sensitive: 477 and 852; resistant: 324, 717, and 6). Each patient's underlying rate was drawn from a Beta distribution with mean 0.10 and concentration κ (1000, nearly identical patients; 20 and 4, increasingly different patients), and cells were drawn from that

rate. A cell-level one-sided two-proportion test and the exact patient-level rank-sum test were applied to each cohort.

Sensitivity analyses. We varied the resistance-score cutoff (0 to 0.5), the post-treatment-compartment cutoff (70% to 95%) crossed with Leiden resolution (0.1 to 1.0), and the missing-gene rank in the signature (501, 1001, 2001, 5001, or requiring a gene in every patient's list), recomputing exact *p*-values each time.

Copy-number-based tumor definition. As an outcome-independent alternative to the marker score, copy-number change was inferred from expression in the manner of Tirosh et al. (2016): full-matrix log-normalized expression of genes detected in at least 5% of cells (9,176 genes) was ordered by genomic position (GENCODE v44), centered on a reference of immune cells, and smoothed with a 101-gene moving average per chromosome, and each cell's score was the mean absolute smoothed value. Cells above the 95th percentile of a held-out half of the reference were called aneuploid-like. This analysis is exploratory because the only non-malignant reference available was immune cells (Section 3.4.6).

3. Results

3.1 Isolation of a tumor cell population

CellTypist assigned the 91,235 cells to fine-grained breast cell types, with luminal hormone receptor and basal epithelial cells forming the largest populations alongside substantial immune infiltrate. Removing immune cells and re-clustering the luminal compartment produced clearer cluster separation than the full-dataset clustering, in which immune and non-immune populations had dominated the structure. Score-based tumor labeling recovered tumor cells across patients and timepoints, with marker expression consistent with malignant epithelium, and post-treatment tumor cells were depleted in the sensitive patients, as the threshold had been chosen to reflect. Figure 1 shows the luminal compartment, marker expression, and tumor versus stromal assignments; Table 1 gives cell counts.

The tumor label has a side effect that matters for everything downstream. Because the tumor score includes proliferation markers, the labeled cells are enriched for cycling cells in some patients: 99.0% and 87.8% of the pre-treatment tumor cells of the sensitive patients SC-P01 and SC-P02 were

assigned to S or G2/M phase, compared with 36.7%, 4.3%, and 16.7% for the resistant patients SC-P04, SC-P05, and SC-P07 (Section 3.4.4). Pre-treatment tumor cells from sensitive patients are therefore far more proliferative than those from resistant patients under this definition.

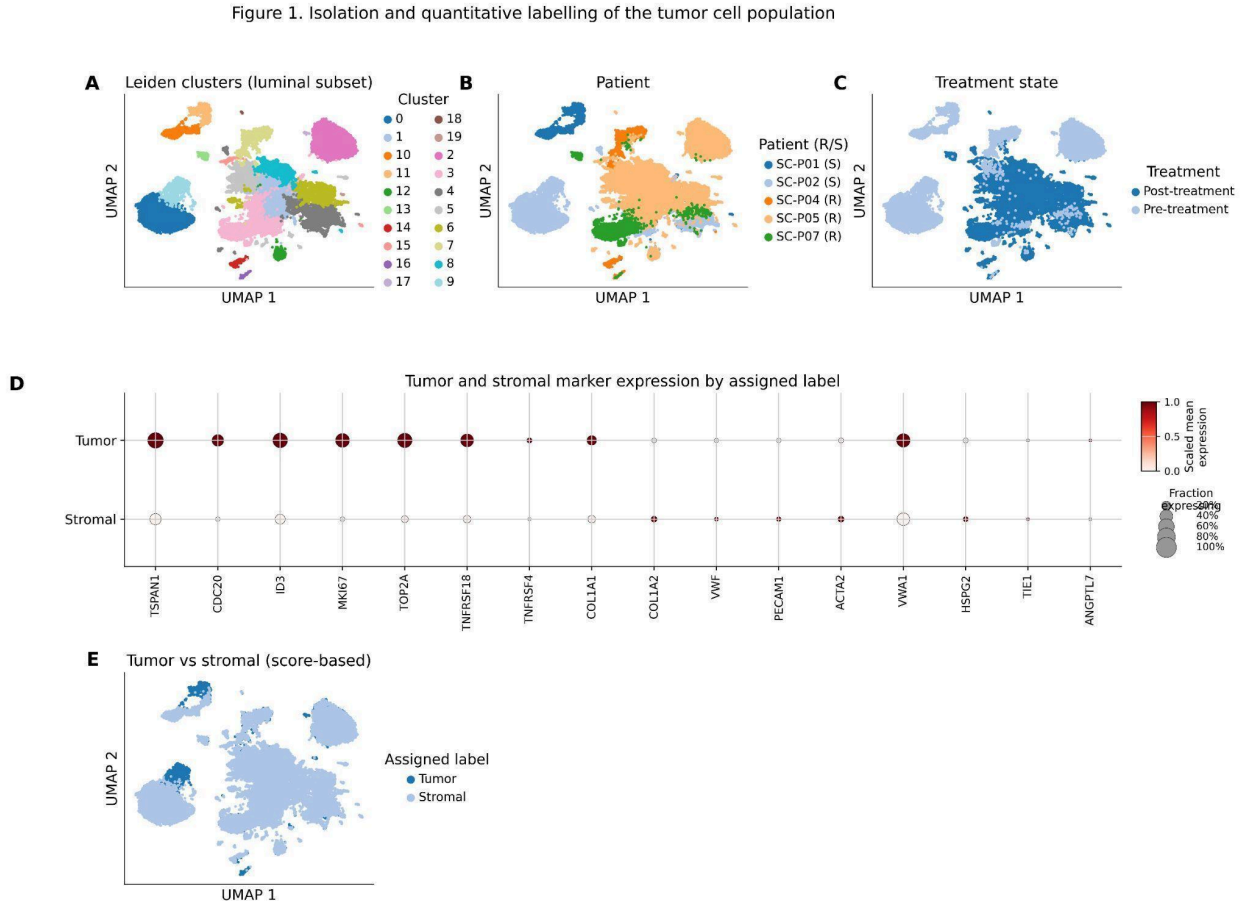


Figure 1. Isolation and quantitative labelling of the tumor cell population. (A) UMAP of the re-clustered luminal epithelial compartment colored by Leiden cluster. (B) The same UMAP colored by patient, with resistant (R) and sensitive (S) status annotated. (C) The same UMAP colored by treatment state. (D) Dot plot of tumor and stromal marker expression by assigned label. (E) The luminal UMAP colored by the score-based tumor versus stromal assignment. Tumor identity was assigned with a tumor gene set score threshold rather than manual cluster annotation.

Table 1. Cell counts at each analysis stage.

Analysis stage	Cells
Cells loaded before quality control	91,235
After quality control (≥ 200 genes per cell; ≤ 20% mitochondrial content)	91,235
After removal of immune-labeled cells	70,510
Luminal epithelial subset	30,544
Tumor cells (score-based label), all patients and timepoints	3,292
of which pre-treatment	2,376
of which post-treatment	916

Quality-control thresholds removed no cells from the deposited matrix. Tumor-cell totals follow from the per-patient counts in Table 2: pre-treatment 477 + 852 + 324 + 717 + 6 = 2,376, and post-treatment 17 + 48 + 76 + 413 + 362 = 916.

3.2 Derivation of a resistance signature

Differential expression between post-treatment and pre-treatment tumor cells identified 378 upregulated genes in SC-P04 and 482 in SC-P05. SC-P07 produced none and was excluded from signature derivation. Rank aggregation across SC-P04 and SC-P05 produced a shared 50-gene signature dominated by proliferation and cell-cycle genes, including TOP2A, MKI67, CDK1, CENPF, PTTG1, BIRC5, ASPM, TPX2, NUSAP1, and UBE2C, together with a smaller set of immune-related genes including TRBC1, CD7, HLA-DQA1, TNFRSF1B, TNFRSF4, and HERC5. The displayed signature genes were more highly expressed after treatment than before in each of the two derivation patients (Figure 2).

Figure 2. Derivation of the resistance signature

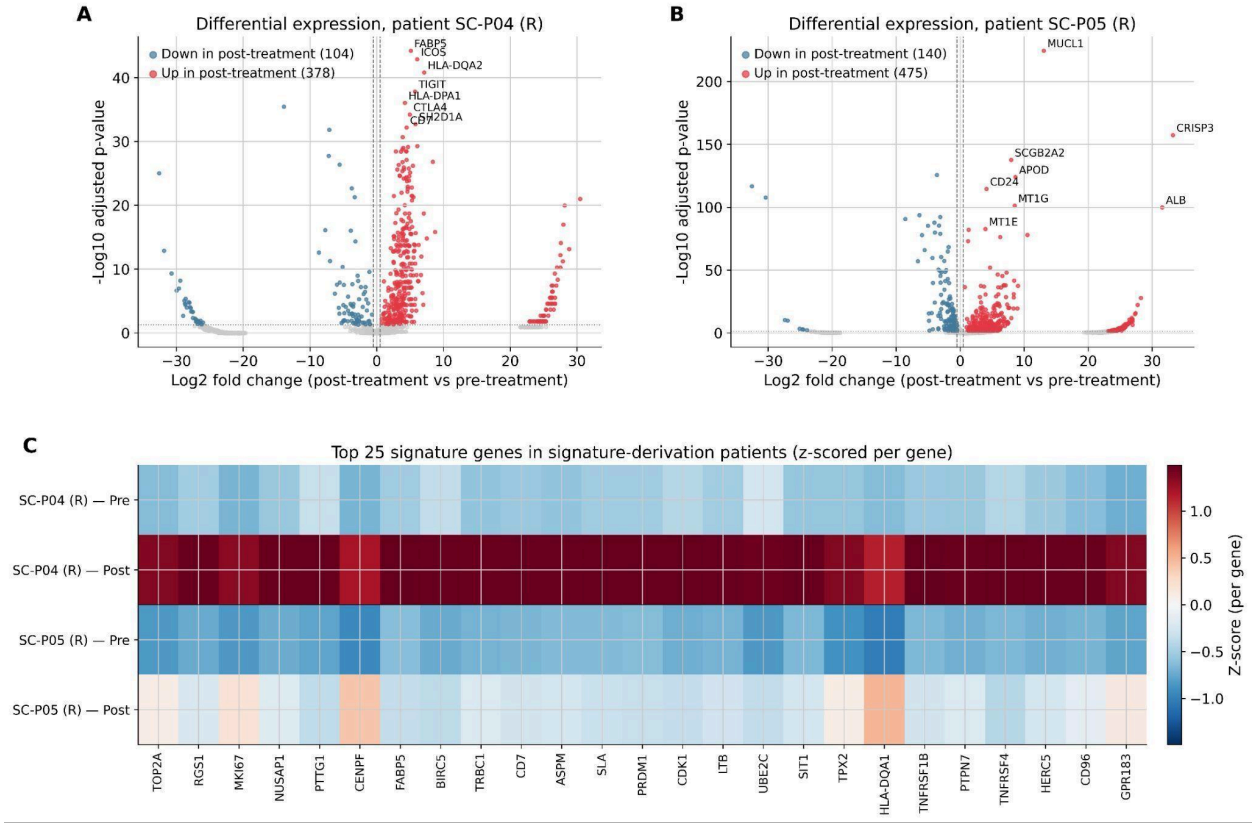


Figure 2. Derivation of the initial candidate resistance signature. (A, B) Volcano plots of post-treatment versus pre-treatment differential expression in tumor cells for SC-P04 and SC-P05, the two resistant patients used to derive the signature; upregulated genes are red and downregulated genes are blue. (C) Heatmap of the top 25 signature genes across the two derivation patients at pre-treatment and post-treatment timepoints, with expression averaged per patient by timepoint and z-scored per gene across the four rows. SC-P07 is not shown because its six pre-treatment tumor cells produced no significantly upregulated genes; it is retained as a resistant patient in the hypothesis test (Figure 3).

3.3 Initial test for pre-existing resistant-like cells

Two metrics were computed per patient on pre-treatment tumor cells (Figure 3, Table 2). The first, the fraction of cells with a resistance score above 0.3, did not support the hypothesis: sensitive patients averaged 15.13% of cells above threshold and resistant patients 1.55%, and every resistant patient fell below both sensitive patients (exact one-sided $p = 1.0$). The second metric, the fraction of

pre-treatment tumor cells within the post-treatment compartment, has to be read patient by patient. All six pre-treatment tumor cells from SC-P07 (100%) fell in the compartment, whereas SC-P04 and SC-P05 had 0.31% and 0.28%, indistinguishable from the 0.00% in both sensitive patients. The resistant-group mean of 33.53% is therefore produced by one patient with six cells.

The one-sided Mann-Whitney U test on the second metric gave $p = 0.069$ by normal approximation with tie correction, the value reported in earlier versions of this analysis. The exact one-sided p-value is 0.10, which is the value annotated in Figure 3C and also the smallest value attainable with three versus two patients (Section 2.8). The clustering test could therefore not have reached $p < 0.05$ in this cohort whatever the data, and 0.069 should not be read as borderline evidence. The first metric's exact p-value is 1.0.

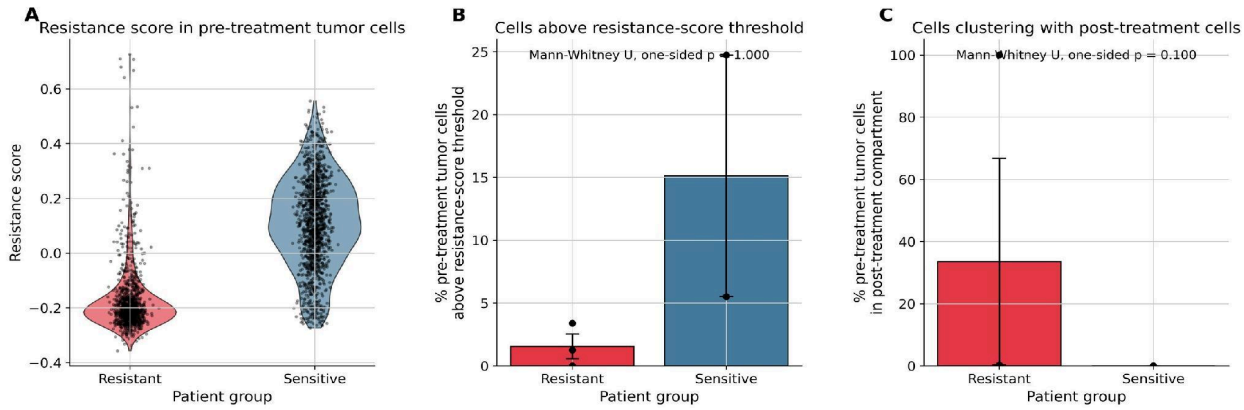


Figure 3. Initial hypothesis test for pre-treatment resistance-like cells, before robustness testing. (A) Distribution of resistance scores across pre-treatment tumor cells by patient group. (B) Fraction of pre-treatment tumor cells with a resistance score above 0.3 by group. (C) Fraction of pre-treatment tumor cells clustering with the post-treatment compartment by group. Points are individual patients and bars show group means with standard error. Annotated p-values are exact one-sided values; the smallest attainable with three versus two patients is 0.10 (Section 3.3).

Patient	Resistance	Pre-treatment tumor cells	% above threshold	% in post-treatment compartment
SC-P01	Sensitive	477	24.74	0.00
SC-P02	Sensitive	852	5.52	0.00
SC-P04	Resistant	324	3.40	0.31
SC-P05	Resistant	717	1.26	0.28
SC-P07	Resistant	6	0.00	100.00

Table 2. Per-patient pre-treatment metrics: the number of pre-treatment tumor cells, the percentage above the resistance score threshold, and the percentage within the post-treatment compartment.

3.4 Robustness checks

Because the initial result came from a small cohort and depended on one patient, seven checks were applied (Section 2.9). Results are reported in order of how directly they address what the signature measures.

3.4.1 Gene set enrichment

Over-representation analysis of the original 50-gene signature showed strong enrichment for cell-cycle and mitotic pathways: Hallmark G2-M Checkpoint (10 of 200 genes, adjusted $p = 1.4 \times 10^{-9}$), Hallmark Mitotic Spindle (7 of 199, adjusted $p = 3.1 \times 10^{-6}$), Hallmark E2F Targets (7 of 200, adjusted $p = 3.1 \times 10^{-6}$), and Reactome Cell Cycle, Mitotic (10 of 523, adjusted $p = 1.3 \times 10^{-4}$). Immune-related terms were also enriched, including Hallmark IL-2/STAT5 Signaling (adjusted $p = 3.1 \times 10^{-6}$) and GO Regulation of T Cell Activation (adjusted $p = 5.5 \times 10^{-6}$), reflecting the smaller immune-associated component of the signature (Figure 4C).

Ranked gene set enrichment analysis, which does not depend on a pre-selected gene list, agreed. Ranking all genes by the mean patient-level change after treatment in the three resistant patients, the Hallmark sets enriched in the post-treatment direction were G2-M Checkpoint (normalized enrichment score [NES] 2.28), E2F Targets (NES 2.16), Allograft Rejection (NES 2.15), Interferon Gamma Response

(NES 2.03), and Mitotic Spindle (NES 1.96), all with false discovery rate below 0.001. Proliferation therefore dominated the genome-wide ranking and not only the 50-gene list.

3.4.2 Harmony batch correction

After Harmony correction for patient identity, the fraction of SC-P07's pre-treatment tumor cells in the post-treatment compartment fell from 100% (6 of 6 cells) to 0%. SC-P04 changed from 0.31% to 3.70% and SC-P05 from 0.28% to 0.14%, and both sensitive patients remained at 0.00% (Figure 4A). The corrected one-sided test gave exact $p = 0.30$, weaker than before correction. The exact corrected values depend on the Harmony and Leiden runs and differed somewhat between earlier runs (for example, 5.56% for SC-P04 in one), but SC-P07 fell to 0% each time.

3.4.3 Patient-level pseudobulk analysis

Collapsing cells to one profile per patient and timepoint on the 2,000-gene matrix, 367 genes were higher after treatment than before in all three resistant patients, including SC-P07 despite its low cell count. A pseudobulk score based on these genes rose from pre- to post-treatment in every resistant patient (SC-P04, 0.037 to 0.221; SC-P05, 0.021 to 0.083; SC-P07, 0.011 to 0.045; Figure 4B). The exact paired one-sided signed-rank test gave $p = 0.125$, the smallest value attainable with three pairs regardless of effect size.

Whether 367 genes is more than expected by chance was tested on the full gene matrix. There, 2,518 of 17,986 genes were higher after treatment in all three resistant patients, but shuffling the timepoint labels within each patient gave a mean of 2,163 such genes (95th percentile 2,723), so the observed count was not unusual (empirical $p = 0.15$; 0.19 when restricted to SC-P04 and SC-P05). The original 50-gene signature, scored on the full matrix, rose from pre- to post-treatment in SC-P04 (0.137 to 0.812) and SC-P05 (0.074 to 0.261) but not in SC-P07 (0.138 to 0.131), giving an exact signed-rank $p = 0.25$. Neither the size of the reproducible set nor its direction therefore provides evidence beyond what random labeling can produce at $n = 3$.

3.4.4 Cell-cycle regression

On the full gene matrix, 42 of 43 canonical S-phase markers and 52 of 54 G2/M markers were present (compared with 3 and 13 in the 2,000-gene matrix), and 509 tumor cells were assigned to S phase, 1,172 to G2/M, and 1,611 to G1. The fraction of cycling (S or G2/M) pre-treatment tumor cells differed sharply between patients (SC-P01, 99.0%; SC-P02, 87.8%; SC-P04, 36.7%; SC-P05, 4.3%; SC-P07, 16.7%). After regressing out the S and G2/M scores and re-deriving the signature from SC-P04 and SC-P05, none of the top 50 genes were canonical cell-cycle genes and none overlapped with the original signature. The new list was not biologically coherent: 20 of its 50 entries were unannotated or antisense non-coding loci (for example AC016027.1, LINC02332, and KTN1-AS1), and it included genes with no evident relevance to breast tumor biology, such as the cell-type-restricted gene CDSN. It is most consistent with noise from low-expression genes and not with a recoverable non-proliferation program.

Ranked enrichment on the regressed data gave a clearer picture. G2-M Checkpoint (NES −0.93, false discovery rate 0.69), E2F Targets (NES −1.09, 0.42), and Mitotic Spindle (NES 1.12, 0.24) were no longer enriched, while immune and interferon sets remained: Allograft Rejection (NES 2.33), Interferon Gamma Response (NES 2.26), and Interferon Alpha Response (NES 2.08), all with false discovery rate 0, and Inflammatory Response (NES 1.60, false discovery rate 0.017). The immune pathways are consistent with an interferon response in tumor cells but are also what contamination by immune transcripts would produce, and the original signature itself contains T-cell genes such as CD7, TRBC1, and ICOS; the data cannot separate the two. Because cell-cycle state and tumor identity are entangled under this tumor definition, regression is also likely to remove some real signal, so the result is best read as showing that the proliferation component is separable from the rest, not that a distinct resistance program has been found.

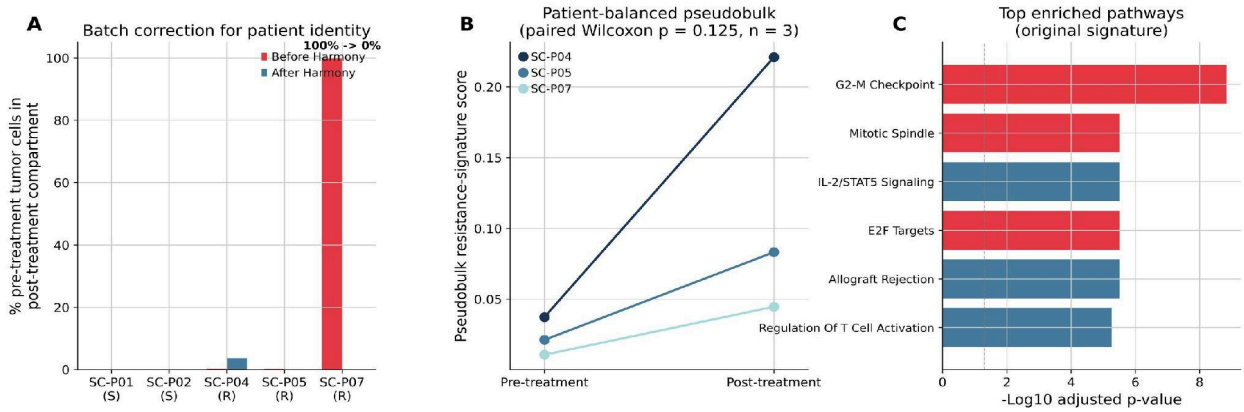


Figure 4. Robustness checks on the candidate signal. (A) Fraction of pre-treatment tumor cells within the post-treatment compartment per patient, before and after Harmony correction for patient identity; the value for SC-P07 fell from 100% to 0% and that for SC-P04 rose from 0.31% to 3.70%. (B) Patient-level pseudobulk score of the 367 genes that rose in all three resistant patients (2,000-gene matrix), pre- versus post-treatment; the exact paired one-sided signed-rank $p = 0.125$ ($n = 3$) is the smallest value attainable with three pairs. (C) Top pathways enriched in the original 50-gene signature by over-representation analysis (Enrichr), ranked by adjusted p-value; red bars are cell-cycle or mitotic pathways, blue bars are immune-associated pathways, and the dashed line marks adjusted $p = 0.05$.

3.4.5 Null simulation

With no true difference between groups, a test that treated every cell as independent produced false positives in 9.7% of simulated cohorts at a nominal 5% when patients were nearly identical (κ = 1000), rising to 37.6% and 45.6% when patients differed (κ = 20 and 4; Table 3). The exact patient-level test rejected in 5.9% to 7.6% of cohorts at its nominal 10%, which is its smallest attainable level at this cohort size, and it cannot reject at 5%. The inflation arises because cells from one patient share that patient's baseline, so thousands of cells carry far less independent information than their number suggests.

Table 3. False-positive rates in 2,000 simulated cohorts per row with no true group difference (three resistant and two sensitive patients, with the cell counts observed here).

Patient heterogeneity (κ)	**Cell-level test, rejection at $p < 0.05$**	**Cell-level test, rejection at $p < 0.10$**	**Patient-level exact test, rejection at $p < 0.10$**
1000 (near identical)	9.7%	15.6%	7.6%
20	37.6%	40.2%	6.8%
4 (very different)	45.6%	46.8%	5.9%

Nominal rates are 5% and 10%. The patient-level test can never reach $p < 0.05$ with three versus two patients.

3.4.6 Sensitivity analyses and an exploratory copy-number tumor definition

Varying the resistance-score cutoff from 0 to 0.5 never favored the resistant group (exact p from 1.0 down to 0.5), because the sensitive patient SC-P01 had the highest fraction of cells above every cutoff. The compartment test depended on arbitrary choices: across Leiden resolutions of 0.1 to 1.0 and compartment cutoffs of 70% to 95%, the exact p-value was 0.10 in 11 of 16 combinations and 0.30 in the other five, and the result was driven by SC-P07 (66.7% to 100% in every setting) except at lower cutoffs, where SC-P04 also contributed (up to 32.4%). The top 50 signature genes were identical under every missing-gene rank convention tested, and 11 of the 50 were canonical cell-cycle genes.

We also attempted an outcome-independent tumor definition based on inferred copy-number change. It did not discriminate: 90% of marker-called tumor cells and 77% of marker-called stromal luminal cells were called aneuploid-like (area under the curve 0.63 for separating the marker labels), and 95% to 100% of pre-treatment luminal cells in every group were called aneuploid-like. The likely reason

is that the reference was immune cells, so any luminal-versus-immune expression difference raises the score; a normal epithelial reference would be needed, and none was available to us. Rerunning the hypothesis test on these cells gave the same direction as before (exact $p = 1.0$; percentage of pre-treatment cells above the score cutoff 28.0% and 5.7% in the sensitive patients and 0.3%, 0.4%, and 0.0% in the resistant patients), but because the cell definition is not validated we do not treat this as a result.

4. Discussion

We asked whether gene set enrichment, supported by patient-level and batch-aware checks, can show what a candidate pre-treatment resistance signature actually measures. In this test case it could. The signature derived from resistant patients was dominated by proliferation genes and strongly enriched for the G2-M Checkpoint, Mitotic Spindle, and E2F Targets pathways, in both over-representation and ranked genome-wide analysis, and the same signature scored higher in sensitive patients' pre-treatment cells, not lower. That pattern is what a proliferation signature is expected to produce. Chemotherapy preferentially eliminates actively cycling cells, and higher proliferation is associated with higher pathological response (Denkert et al., 2013), so a gene list built from post-versus-pre differences in resistant tumors can end up measuring a proliferation axis that points opposite to resistance once more proliferative sensitive tumors are included. Our tumor definition made this more likely: its score contains proliferation markers, and 99.0% and 87.8% of the pre-treatment tumor cells of the two sensitive patients were cycling, compared with 4.3% to 36.7% in the resistant patients. The enrichment result turns an informal impression about gene names into a reproducible, quantified finding, and regressing out cell-cycle scores removed the cell-cycle pathways from the ranking, as expected if they carried that component.

The clustering result tells a different story. The signature-independent test appeared positive, but its smallest attainable exact p-value with three versus two patients is 0.10, so it could never have supported a claim at the conventional threshold, and its group mean came from one patient with six cells. Harmony correction removed that patient's apparent membership in the post-treatment compartment (exact $p = 0.30$ afterward), and the result changed between 0.10 and 0.30 with ordinary changes to the compartment cutoff and clustering resolution. Harmony removes variation associated with patient

identity, which can include real biological differences as well as technical effects, so the drop from 100% to 0% shows that the result was unstable and could not be separated from patient identity. It does not by itself prove that the original pattern was an artifact. Six cells from one patient could not support a claim about a recurring resistant-cell phenotype in either case.

The checks should not be read as independent confirmations. They share the same patients and tumor labels, and several were limited by design: the pseudobulk test cannot go below $p = 0.125$ with three patients, and the copy-number definition was uninformative because of the reference used. What the set supports is modest. It provides no support for a pre-existing resistance signal in this cohort, a clear demonstration that the signature measures proliferation, evidence that the count of genes rising in all three resistant patients is what label shuffling would produce, and a simulation showing why cell-level tests mislead at this scale. The interferon and immune pathways that remained after cell-cycle regression are the one possibly interesting residue, but with three patients, one of whom contributes six pre-treatment cells, and possible immune contamination, they are a hypothesis and not a finding. We treat the combination as showing that the initial result is not evidence of pre-existing resistance, not as grounds to select the most favorable of several discordant tests.

These observations bear on the question of whether chemoresistance is pre-existing or induced by treatment. Under the induced model, therapy reprograms surviving cells into a resistant state, as in the reversible drug-tolerant states described under drug pressure (Sharma et al., 2010); under the pre-existing model, resistant states are present before therapy and selected by it. The source study distinguished innate from acquired programs by copy number and found gene sets consistent with both (Jia et al., 2024). The present analysis does not adjudicate between the models. Given the fragility of the clustering result and the absence of independent validation, these data should not be read as favoring a pre-existing component over a purely induced one.

The approach is not specific to HR+ breast cancer. Any cancer type in which paired pre- and post-treatment samples can be obtained is amenable to the same design, including triple-negative breast cancer treated with chemotherapy, lung adenocarcinoma treated with targeted therapy, or melanoma

treated with checkpoint blockade. In each setting, a larger cohort, a patient-balanced analysis, and an outcome-independent definition of tumor cells would be needed before any pattern could be considered a candidate biomarker.

The practical lesson is that each check was easy to omit and omitting any one would have left a more encouraging but unsupported conclusion in place. Without Harmony correction, the SC-P07 result would have stood as apparent biology; without enrichment analysis, the proliferation confound would have remained an informal observation; without patient-level testing and the exact *p*-value floor, the cell-level statistics would have gone unchallenged. Based on this case, small-cohort single-cell studies that propose pre-treatment resistance or response signatures should at minimum: (1) report hypothesis tests at the patient level, through pseudobulk analysis or by stating how many independent patients support a claimed effect; (2) apply a batch-integration method such as Harmony, scVI, or BBKNN and report whether clustering-based claims survive correction for patient identity, while recognizing that correction can remove real biology; (3) subject any candidate signature to pathway-level enrichment analysis, ideally with a cell-cycle-regressed comparison on the full gene matrix, instead of reporting a gene list alone; (4) report exact *p*-values, or state the approximation used, and the smallest *p*-value attainable at the cohort size; and (5) define tumor cells without outcome information or proliferation markers. These checks are inexpensive relative to the sequencing itself and can change a paper's conclusion.

4.1 Limitations

Statistical power and design. The effective unit of inference is the patient. With three resistant and two sensitive patients, thousands of cells per patient are not thousands of independent observations, and no test here is adequately powered; the exact one-sided Mann-Whitney *p*-value cannot fall below 0.10. The design is retrospective and post hoc: the signature was derived from patients whose outcomes were known, and the same patients' pre-treatment samples were then examined, so the analysis tests retrospective self-consistency and not predictive validity. No independent validation cohort was used. We searched public repositories for an independent pre-treatment single-cell HR+ cohort with known

chemotherapy response and did not find one. Bulk RNA-seq or microarray cohorts with known pathological response exist (for example GSE25066, GSE20194, GSE32646, GSE123845, and GSE163882) and could validate a frozen gene score, but bulk data average over the whole tumor and cannot test a rare-subpopulation hypothesis directly.

Circularity and arbitrary choices. The tumor score threshold was chosen partly from the expected direction of treatment response, and the score contains proliferation markers, so an outcome-informed and proliferation-enriched step feeds every downstream comparison. Our attempt to replace it with an inferred copy-number definition was uninformative because the only reference was immune cells; a normal epithelial reference would be needed. The signature was derived from cell-level differential expression, which repeats the pseudoreplication problem criticized above; pseudobulk differential expression would be more consistent, although with two or three patients its power is also limited. The 0.3 score cutoff, the 90% compartment threshold, and the rank-1001 convention are arbitrary; the sensitivity analyses show that the compartment result varies with them and that the signature is unchanged by the rank convention, but they do not make any choice correct. The tumor-labeled cells also include immune-gene signal, as the T-cell genes in the signature show, so some labeled cells may be immune cells or doublets.

Technical limitations. Cell-cycle regression assumes the marker-based scores capture proliferation without capturing tumor identity; here the two are entangled, so regression may remove real signal, and the regressed top-50 list lacked coherent biological composition. Harmony may also have removed real patient-level biology along with technical variation, and its exact corrected values varied between runs, so its result shows instability and not proof of artifact. The ranked enrichment ordered genes by patient-level change but is not a patient-level significance test, and the permutation null shuffled cell labels within patients, which treats cells as exchangeable. The simulation used one simple model of patient heterogeneity.

Scope of inference. No lineage tracing, clonal, or mutational information was available to show that transcriptionally similar pre- and post-treatment cells are clonally related, so transcriptional

resemblance is not evidence that specific pre-treatment cells gave rise to the surviving population. The study makes only a computational, associative claim. Establishing a causal link between a candidate resistance program and chemotherapy sensitivity would require isolating or generating cells expressing the program, exposing them to the relevant chemotherapy in vitro, and testing whether perturbing a candidate driver alters sensitivity, which is outside the scope of this reanalysis. Taken together, these limitations mean the analysis identifies candidate patterns and concrete methodological pitfalls for future studies and is not evidence of a validated, reproducible pre-existing resistance state.

5. Conclusion

This exploratory, retrospective reanalysis of a paired pre- and post-chemotherapy single-cell dataset from five HR+ breast cancer patients used a candidate resistance signature as a test case for gene set enrichment and related checks. Over-representation and ranked enrichment analysis showed that the signature reflected cell-cycle activity (Hallmark G2-M Checkpoint, adjusted $p = 1.4 \times 10^{-9}$; normalized enrichment score 2.28 in the ranked analysis) and not a resistance-specific program, and regressing out cell-cycle scores on the full gene matrix removed those pathways. The apparently positive clustering result could not have reached significance at this cohort size (smallest exact $p = 0.10$), and it did not survive correction for patient identity (exact $p = 0.30$). The set of genes rising in all three resistant patients was no larger than label shuffling would produce, and simulations showed that treating cells as independent inflates false positives to 38% to 46% when patients differ. None of these results supports pre-existing resistance in this cohort.

The value of the study lies in showing, on real data, how specific and inexpensive checks can separate a reproducible biological signal from an analytical artifact. Small-cohort single-cell studies that propose pre-treatment resistance or response signatures should report patient-level tests, apply and interpret batch integration, test candidate signatures for pathway enrichment, and report exact *p*-values before such signals are described as candidate biomarkers.

References

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics, 25*(1), 25–29. https://doi.org/10.1038/75556

Denkert, C., Loibl, S., Müller, B. M., Eidtmann, H., Schmitt, W. D., Eiermann, W., Gerber, B., Tesch, H., Hilfrich, J., Huober, J., Fehm, T., Barinoff, J., Jackisch, C., Prinzler, J., Rüdiger, T., Erbstößer, E., Blohmer, J. U., Budczies, J., Mehta, K. M., & von Minckwitz, G. (2013). Ki67 levels as predictive and prognostic parameters in pretherapeutic breast cancer core biopsies: A translational investigation in the neoadjuvant GeparTrio trial. *Annals of Oncology, 24*(11), 2786–2793. https://doi.org/10.1093/annonc/mdt350

Domínguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., Suchanek, O., Polanski, K., King, H. W., Mamanova, L., Huang, N., Szabo, P. A., Richardson, L., Bolt, L., Fasouli, E. S., Mahbubani, K. T., Prete, M., Tuck, L., . . . Teichmann, S. A. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science, 376*(6594), Article eabl5197. https://doi.org/10.1126/science.abl5197

Gene Ontology Consortium. (2023). The Gene Ontology knowledgebase in 2023. *Genetics, 224*(1), Article iyad031. https://doi.org/10.1093/genetics/iyad031

Haque, W., Verma, V., Hatch, S., Klimberg, V. S., Butler, E. B., & Teh, B. S. (2018). Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. *Breast Cancer Research and Treatment, 170*(3), 559–567. https://doi.org/10.1007/s10549-018-4801-3

Howlader, N., Altekruse, S. F., Li, C. I., Chen, V. W., Clarke, C. A., Ries, L. A. G., & Cronin, K. A. (2014). US incidence of breast cancer subtypes defined by joint hormone receptor and HER2

status. *Journal of the National Cancer Institute, 106*(5), Article dju055. https://doi.org/10.1093/jnci/dju055

Jia, F., Sun, S., Li, J., Wang, W., Huang, H., Hu, X., Pan, S., Chen, W., Shen, L., Yao, Y., Zheng, S., Chen, H., Xia, W., Yuan, H., Zhou, J., Yu, X., Zhang, T., Zhang, B., Huang, J., & Ni, C. (2024). Neoadjuvant chemotherapy-induced remodeling of human hormonal receptor-positive breast cancer revealed by single-cell RNA sequencing. *Cancer Letters, 585*, Article 216656. https://doi.org/10.1016/j.canlet.2024.216656

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research, 28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., & Navin, N. E. (2018). Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell, 173*(4), 879–893.e13. https://doi.org/10.1016/j.cell.2018.03.041

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods, 16*(12), 1289–1296. https://doi.org/10.1038/s41592-019-0619-0

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research, 44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database hallmark gene set collection. *Cell Systems, 1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., & Theis, F. J. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods, 19*(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform manifold approximation and projection for dimension reduction* (arXiv:1802.03426). arXiv. https://doi.org/10.48550/arXiv.1802.03426

Meacham, C. E., & Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature, 501*(7467), 328–337. https://doi.org/10.1038/nature12624

Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., Haw, R., Jassal, B., Matthews, L., May, B., Petryszak, R., Ragueneau, E., Rothfels, K., Sevilla, C., Shamovsky, V., Stephan, R., Tiwari, K., Varusai, T., Weiser, J., . . . D'Eustachio, P. (2024). The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research, 52*(D1), D672–D678. https://doi.org/10.1093/nar/gkad1025

Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M. A., Wong, K.-K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., & Settleman, J. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell, 141*(1), 69–80. https://doi.org/10.1016/j.cell.2010.02.027

Squair, J. W., Gautier, M., Kathe, C., Anderson, M. A., James, N. D., Hutson, T. H., Hudelle, R., Qaiser, T., Matson, K. J. E., Barraud, Q., Levine, A. J., La Manno, G., Skinnider, M. A., & Courtine, G. (2021). Confronting false discoveries in single-cell differential expression. *Nature Communications, 12*(1), Article 5692. https://doi.org/10.1038/s41467-021-25960-2

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249. https://doi.org/10.3322/caac.21660

Tanay, A., & Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature, 541*(7637), 331–338. https://doi.org/10.1038/nature21350

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J. R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., . . . Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science, 352*(6282), 189–196. https://doi.org/10.1126/science.aad0501

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports, 9*(1), Article 5233. https://doi.org/10.1038/s41598-019-41695-z

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology, 19*(1), Article 15. https://doi.org/10.1186/s13059-017-1382-0

Zimmerman, K. D., Espeland, M. A., & Langefeld, C. D. (2021). A practical solution to pseudoreplication bias in single-cell studies. *Nature Communications, 12*(1), Article 738. https://doi.org/10.1038/s41467-021-21038-1