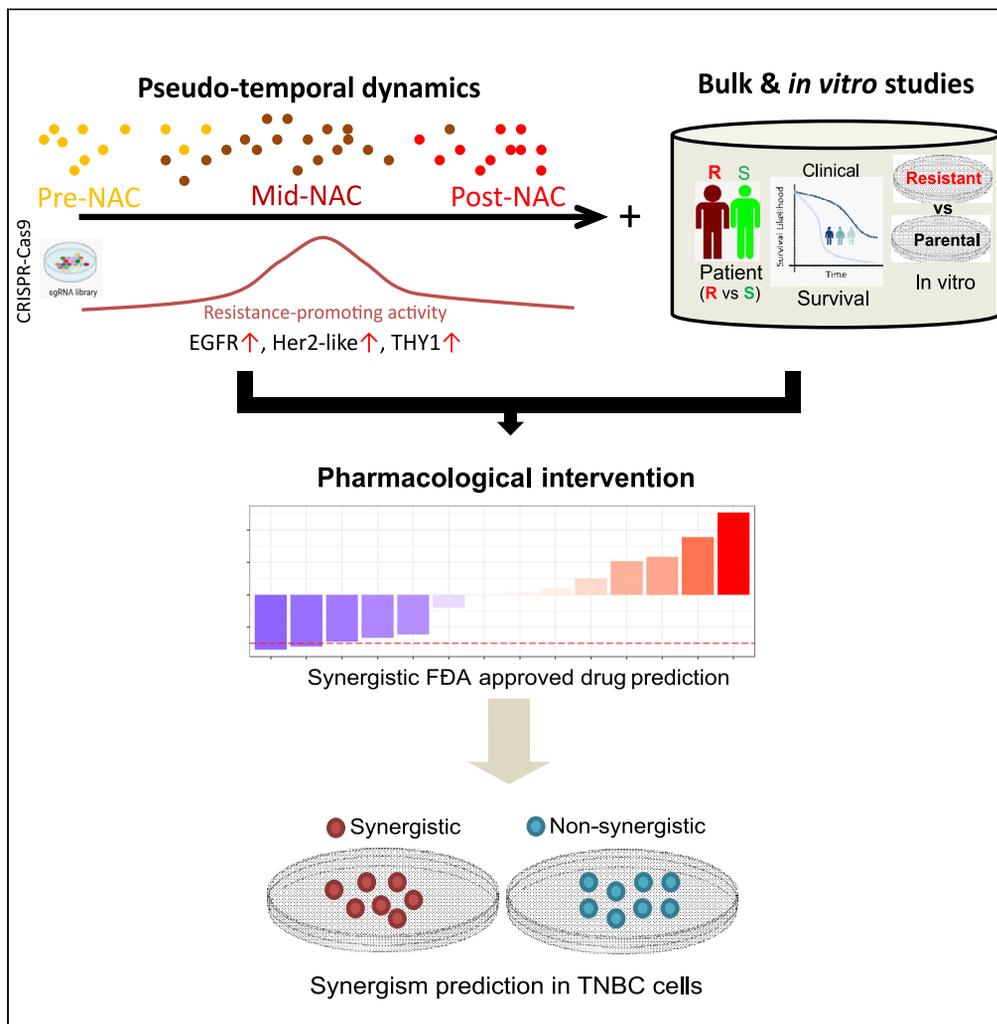**Article**

Pseudo-temporal dynamics of chemoresistant triple negative breast cancer cells reveal EGFR/HER2 inhibition as synthetic lethal during mid-neoadjuvant chemotherapy

Won-Min Song, Pei-Ling Chia, Xianxiao Zhou, Martin Walsh, Jose Silva, Bin Zhang

won-min.song@mssm.edu (W.-M.S.)
jose.silva@mssm.edu (J.S.)
bin.zhang@mssm.edu (B.Z.)

Highlights

Integrative model of pseudo-temporal dynamics in chemo-resistant TNBC is proposed

Chemo-resistant cell signatures are predictive of survival in bulk cohorts

Consensus signature is derived to predict FDA-approved drugs to synergize with paclitaxel

Afatinib efficacy is predicted and validated in TNBC cell lines by suppressing EGFR

Article

Pseudo-temporal dynamics of chemoresistant triple negative breast cancer cells reveal EGFR/HER2 inhibition as synthetic lethal during mid-neoadjuvant chemotherapy

Won-Min Song,[1,2,3,8,9,*] Pei-Ling Chia,[4,6,8] Xianxiao Zhou,[1,2,3] Martin Walsh,[5] Jose Silva,[7,*] and Bin Zhang[1,2,3,5,*]

SUMMARY

In the absence of targetable hormonal axes, chemoresistance for triple-negative breast cancer (TNBC) often compromises patient outcomes. To investigate the underlying tumor dynamics, we performed trajectory analysis on the single-nuclei RNA-seq (snRNA-seq) of chemoresistant tumor clones during neoadjuvant chemotherapy (NAC). It revealed a common tumor trajectory across multiple patients with HER2-like expansions during NAC. Genome-wide CRISPR-Cas9 knock-out on mammary epithelial cells revealed chemosensitivity-promoting knock-outs were up-regulated along the tumor trajectory. Furthermore, we derived a consensus gene signature of TNBC chemoresistance by comparing the trajectory transcriptome with chemoresistant transcriptomes from TNBC cell lines and poor prognosis patient samples to predict FDA-approved drugs, including afatinib (pan-HER inhibitor), targeting the consensus signature. We validated the synergistic efficacy of afatinib and paclitaxel in chemoresistant TNBC cells and confirmed pharmacological suppression of the consensus signature. The study provides a dynamic model of chemoresistant tumor transcriptome, and computational framework for pharmacological intervention.

INTRODUCTION

Triple-negative breast cancer (TNBC) is a type of breast cancer characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER2). TNBC accounts for 10–20% of all breast cancer cases and is associated with a high recurrence/distant metastasis rate and poor survival.[1] Treatment options for TNBC patients remain scarce. The standard-of-care consists of chemotherapy based on anthracycline-taxane regimens as TNBC tumors lack molecular targets for endocrine and HER2-targeted therapies. However, chemoresistance, the insensitivity of cancer cells to therapies, is a key factor of poor prognosis. Only 20% of TNBC patients present a pathological complete response (pCR) after neoadjuvant chemotherapy (NAC), and those with residual disease (RD) have significantly worse overall survival than non-TNBC patients with RD.[2]

The molecular underpinnings of TNBC therapy resistance are heterogeneous. Bai et al. summarized seven mechanistic pillars of TNBC therapy resistance[3] including: (1) Stronger anti-oxidant ability by treatment-modulated mitochondrial damage, (2) increased drug efflux and metabolism to reduce intra-cellular drug accumulation, (3) enhanced tumoral DNA repair capacity to maintain the DNA integrity, (4) anti-apoptosis and autophagy to sustain tumor cell survival, (5) metabolic reprogramming for rapid tumor growth, (6) epithelial-mesenchymal transition pathways giving rise to cancer stem cell and metastatic recurrence, and (7) heterogeneous tumor micro-environment for immune evasion and stem-like phenotypes. These pillars may act independently or crosstalk, and manifest differently for TNBC subtypes showing varying responses to neoadjuvant chemotherapy (NAC).[4]

To tackle the diverse spectrum of TNBC therapy resistance, several therapeutic strategies have shown some promise in clinical and preclinical studies. For instance, platinum-based treatments are now incorporated into neoadjuvant and metastatic treatments.[5] Several targeted therapies were proposed for subsets of TNBC patients, including cyclin-dependent kinases (*CDK1, CDK2, CDK4,* and *CDK6*)[6–8] and several

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[2]Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[3]Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[4]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[5]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[6]Institute of Molecular and Cell Biology, Agency for Science, Technology, and Research (A*STAR), 61 Biopolis Drive, Proteos, Singapore 138673, Singapore

[7]Department of Pathology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[8]These authors contributed equally

[9]Lead contact

*Correspondence:
won-min.song@mssm.edu
(W.-M.S.),
jose.silva@mssm.edu (J.S.),
bin.zhang@mssm.edu (B.Z.)
https://doi.org/10.1016/j.isci.2023.106064

receptor tyrosine kinases (RTKs) including *FGFR, VEGFR, PDFGR, IGFR1, AXL, MET* and *EGFR*.[9] Bromodomain and extra-terminal domain (BET) inhibitors, intervening with epigenetic regulations by BET family proteins, showed some efficacy in TNBC,[10] followed by the rapid emergence of resistance.[11] Overall, the development of TNBC-targeted therapies is well behind that of other breast cancer subtypes, and chemoresistance is still a key contributing factor to the poor outcomes in TNBC.

To this end, the dynamic aspects of TNBC therapy have been a significant contributing factor to TNBC chemoresistance. During the standard-of-care chemotherapy, intrinsic chemoresistant sub-clones may expand (i.e. selection),[12] or tumor cells may acquire resistance through genetic and epigenetic alterations favoring their survival (i.e. adaptation).[13,14]

Single-cell sequencing studies have been instrumental to dissect these cell-level dynamics during TNBC therapies. For instance, the infiltration of immuno-suppressive myeloid cells (IMCs) in tumors arises after HER2-targeted therapy in HER2+ breast cancers and IMC-targeted therapy enhances anti-tumor immunity.[15] Pre-existing *de novo* mutations and acquired mutations during neoadjuvant chemotherapy (NAC) have been found to confer resistance in TNBC.[14] Computational frameworks have also been developed to infer cell-level temporal dynamics, thus 'cell trajectories' capturing cell states stretched across inferred time windows (hence pseudo-time).[16,17] These advances hold the potential to uncover mechanisms underlying the dynamic processes of TNBC chemoresistance. However, systematic efforts to translate the dynamic aspects into potential therapeutic avenues have been lacking.

To address this knowledge gap, we designed a three-step computational workflow (Figure 1). In the first step, we utilize single-cell sequencing data of TNBCs during NAC and inferred the tumor cell trajectory under chemotherapeutic treatment to study activated pathways in chemoresistant cells. Importantly, genome-wide CRISPR loss-of-function studies confirmed that the up-regulated transcriptome along the pseudo-temporal trajectory was enriched in genes whose knock-out promotes chemosensitivity. In the second step, we compared the dysregulated pathways along the tumor cell trajectories to the tumor intrinsic pathways from *in vitro* chemoresistant TNBC models and survival-associated pathways in bulk TNBC cohorts. These comparisons identified recurrent gene signatures across different studies and yielded the *consensus chemoresistance pathways*. In the third step, we investigated the therapeutic potential by predicting FDA-approved drugs to suppress the consensus pathways. Finally, we validated the predicted efficacy of combined treatment of a standard chemo-regimen, paclitaxel, and top predicted drug in the consensus pathway-driven TNBC cells. Overall, the three-step workflow represents a comprehensive framework to model the molecular dynamics of TNBC chemoresistance and effectively predicts its pharmacological intervention strategy.

RESULTS

Single-nuclei transcriptome of TNBC patients undergoing NAC

We analyzed the single-nuclei RNA-sequencing (snRNA-seq) data before (B), in the middle of (M), and after (A) neoadjuvant chemotherapy (NAC) in samples from 6 TNBC patients. Three of the six patients showed NAC-sensitivity (S) and exhibited tumor clonal extinction after NAC whereas the rest were NAC-resistance (R), exhibiting persistent clones after NAC[14] (Table 1). For simplicity, we termed the snRNAseq data as the Kim-snRNA-seq data (Sequencing Reads Archive (SRA): SRP114962).

After data quality control (Figures S1 and S2; see STAR Methods for details), 2,849 cells from all the 6 patients in the Kim-snRNA-seq data[14] were retained for the subsequent analyses. These cells formed 10 clusters (Figure 2A) which corresponded to specific cell types (Figure 2B). T. The cell clusters C3, C4, C9, and C2 highly expressed an epithelial tumor marker, *EPCAM*. Especially, C3, C4, and C9 hihgly expressed an undifferentiated cancer progenitor/stem cell (CSC) marker, *THY1* (also known as CD90),[18] which is a characteristic of cells undergoing epithelial-mesenchymal transition (EMT).[18,19] In contrast, *EPCAM* was not highly expressed in C1, C6, and C10. As these cells were tumors from TNBC patients, the gene expression reflecting receptor status of ER+ (*ESR1*), PR+ (*PGR*), and HER2+ (*ERBB2*) were generally low. This dataset contains primarily the single-nuclei transcriptomes of TNBC epithelial cells (EPCAM+) but was short of immune cells (CD45+), and thus it was limited to elucidating changes in tumor cells.

We identified the genes over-expressed in each cell cluster (i.e., cluster-specific markers) by performing differential expression analysis for each cell cluster against the rest by limma[20] (see STAR Methods for
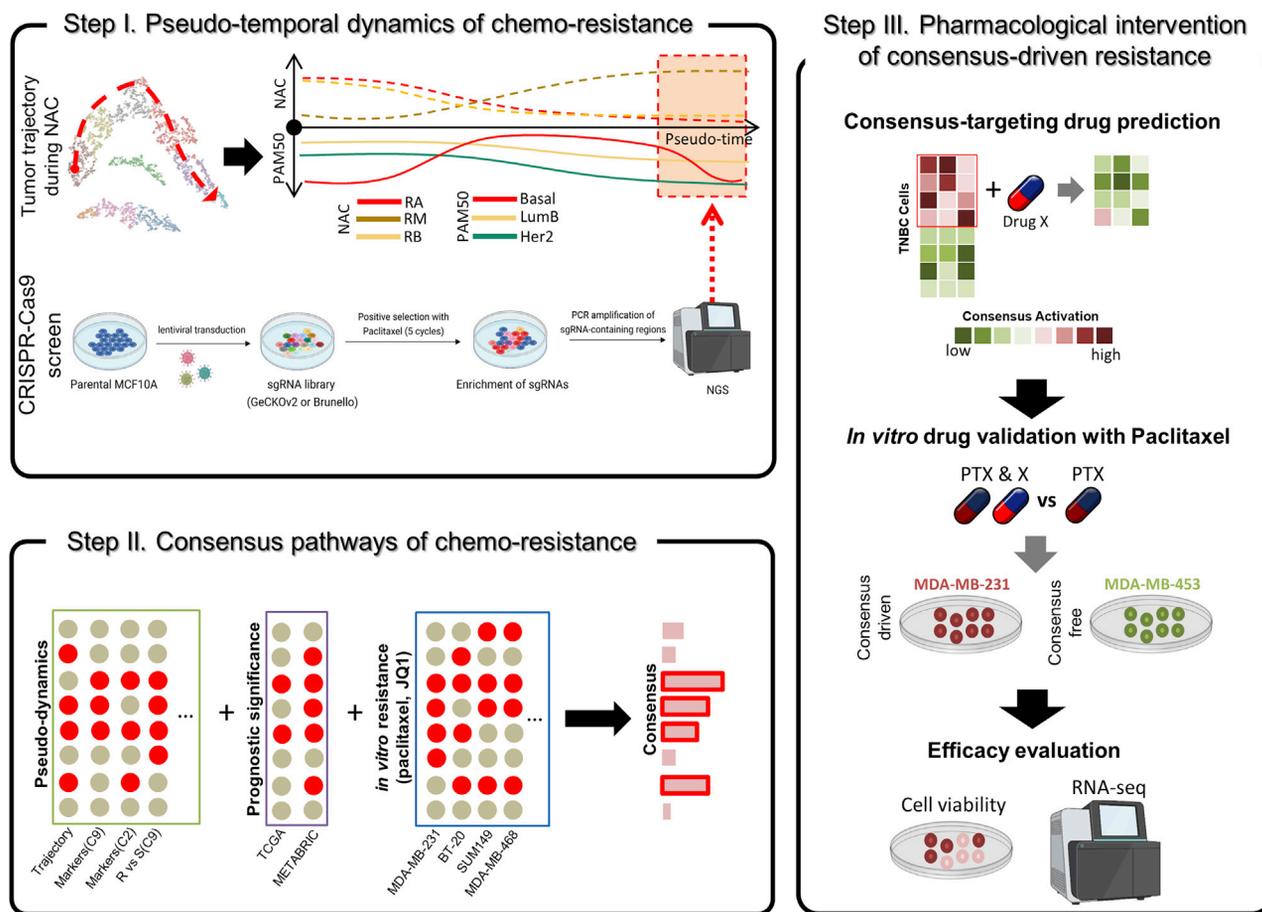
Figure 1. Integrative workflow to functionally characterize and target triple-negative breast tumor (TNBC) chemoresistance

In Step, I, snRNA-seq of neoadjuvant chemotherapy (NAC)-resistant tumors reveal shared cellular trajectory (top left). This yields pseudo-temporal dynamics across treatment stages (RB: treatment-naive, RM: mid-treatment cells, RA: post-treatment cells), and varying molecular subtypes (top right). Functional screening of resistance/sensitivity-promoting genes in epithelial mammary cells (MCF10A, bottom) is incorporated to identify TNBC chemoresistance hotspots in the pseudo-temporal model. InStep II, differentially expressed pathways in the NAC-driven trajectory (green box) are compared with other molecular signatures of TNBC chemoresistance, including TNBC prognostic signatures in bulk cohorts (purple box) and in vitro chemoresistance signatures of TNBC cells (blue box). Recurrent genes across different signatures will be gathered as the consensus pathway of TNBC chemoresistance (pink bars on right). In Step III, we predict FDA-approved drugs targeting the consensus pathways, and validate the prediction by evaluating the efficacy of the combined treatment of paclitaxel and the predicted drug, in comparison with paclitaxel alone. The efficacies in TNBC cells with consensus pathway-driven resistance (MDA-MB-231) were compared to those with consensus-free resistance (MDA-MB-453).

marker analysis; see Table S2). The enriched functions and pathways in these cluster specific signatures suggest aberrant signaling underlying TNBC chemoresistance (Figure S3A). For instance, the chemo-sensitive clusters C1 and C6 are associated with apoptosis-induced DNA fragmentation (FET FDR = 2.46E-6 & 2.11E-5 respectively; Table S3), reflecting their responses to chemotherapy. On the other hand, the chemo-resistant cluster C2 is associated with TNFα signaling while another resistant cluster, C9, is involved in MYC signaling (Figure S3A).

Pseudo-temporal dynamics of adaptive tumor cells during chemotherapy

Then, we performed trajectory analysis to identify pseudo-temporal transcriptional changes in the chemoresistant tumor cells at different stages of NAC. After inferring the cellular trajectories initiated by *slingshot*[21] (see STAR Methods for trajectory inference; Figure S4), two chemoresistant trajectories, namely Trajectory-I and-III (Figure 2A), were identified. Trajectory-I captured the gradual increases of mid-NAC cells from chemoresistant patients (Figures 2A and 2B), with increases in cancer stem cell markers, *EPCAM* and *THY1,* and a genome-wide *de novo* methylation regulator, *DNMT3A* (DNA methyltransferase 3 alpha) (Figure 2D). The top 500 genes (Table S3A) positively correlated with the inferred pseudo-time in

Table 1. The number of cells corresponding to distinct patients (KTN102, KTN126, KTN129, KTN132, KTN302, and KTN615) and NAC resistance/treatment group

Cell group	KTN102	KTN126	KTN129	KTN132	KTN302	KTN615
SB	59	0	0	216	0	266
SM	0	0	0	443	0	96
SA	97	0	0	0	0	0
RB	0	122	406	0	0	0
RM	0	0	0	0	313	0
RA	0	455	376	0	0	0

*S: chemo-sensitive, R: chemoresistance, B: pre-NAC, M: mid-NAC, A: post-NAC.
Patients with persistent tumor clones after NAC were regarded as chemoresistant, and those without any clones were regarded as chemo-sensitive.[14] See the abbreviations described in the table.

trajectory-I were enriched for the hallmark epithelial-mesenchymal transition pathway (corrected FET (cFET) p =3.44E-4, 4.65-fold enrichment (FE); Table S3B; see STAR Methods for the details about trajectory association analysis), which is often over-expressed in invasive tumors.[22] As over-expression of DNA methyltransferases plays causal roles in the generation and maintenance of CSCs,[23] the results suggest that *de novo* epigenetic modifications by *DNMT3A* may be linked to the emergence of CSC-like phenotypes in mid-NAC.

The trajectories also reflected pseudo-temporal changes in breast tumor cell subtypes during NAC. By inferring cellwise molecular subtypes by PAM50[24] (see STAR Methods), trajectory-I showed varying proportions of the molecular subtypes along the path. Her2-like and Luminal B cells gradually increased, whereas basal-like cells gradually decreased toward C4 with a sudden increase in the terminal C9 (Figure S3B). Also, *EGFR* expression gradually increased across trajectory-I, and aligned with the increased Her2-like cells in C9 (Figure S3D). On the other hand, trajectory-III is characterized by high basal-like cell proportion in the cluster C2 which includes cells from a single patient, consistent with the high expression of basal markers (*KRT5* and *KRT14* expression in Figure S3C).

Next, we reasoned that some up-regulated pathways/genes along the trajectories may be involved in the acquisition of chemoresistance. Toward this end, we performed genome-wide CRISPR/Cas9 pooled knockout screening in the immortalized basal non-transformed human mammary epithelial cell line (MCF-10A)[25] with exposure to either Doxorubicin (anthracycline) or Paclitaxel (taxane) (see Step I in Figure 1; see STAR Methods). The high throughput screening identified gene knock-outs that sensitized for chemotherapy or promoted chemoresistance (Table S1).

Remarkably, knock-outs promoting chemo-sensitivity were enriched in trajectory-I. By evaluating the cellwise enrichment of these signatures by Gene Set Variation Analysis (GSVA),[26] we observed gradual enrichment of the chemo-sensitizing knock-outs in the cells along trajectory-I (C5-9; Figure 2C). This enrichment peaked at C9 with most cells enriched for GSVA z-scores >3 (Figure 2D). Many of these knock-outs corresponded to up-regulated genes in C9 (Figure S5), which were the marker genes of C9 (green gene symbols on the right, Figure 2B). On the other hand, the knock-outs promoting chemoresistance were not enriched in the other cell trajectories (trajectory-II: C1 → C6 → C10, trajectory-III: C5 → C2; Figure 2C).

In summary, trajectory-I, characterized by the expansion of Her2-like cells during the mid-NAC phase (Figure 2E) and transformation into adaptive CSC-like tumor cells. Trajectory-I was a shared tumor cell trajectory across multiple TNBC patients, underlying the acquired chemoresistance in TNBC.

Activated genes along trajectory-I associates with poor prognosis in bulk cohorts

Prompted by the findings above, we postulated that the presence of tumor populations along the chemoresistant trajectories (i.e. trajectory-I and –II) would be associated with poor prognosis. To test the hypothesis, we collected publicly available, clinically annotated TNBC bulk cohorts (METABRIC[27]: 288 samples, TCGA[28]: 157 samples). As the snRNA-seq data was short of immune and stromal cells, we inferred immune and stromal cell abundances by ESTIMATE[29] to select the samples with low stroma/immune cell contents, leaving 143 and 77 samples for METABRIC and TCGA respectively (see STAR Methods for details). Then, we
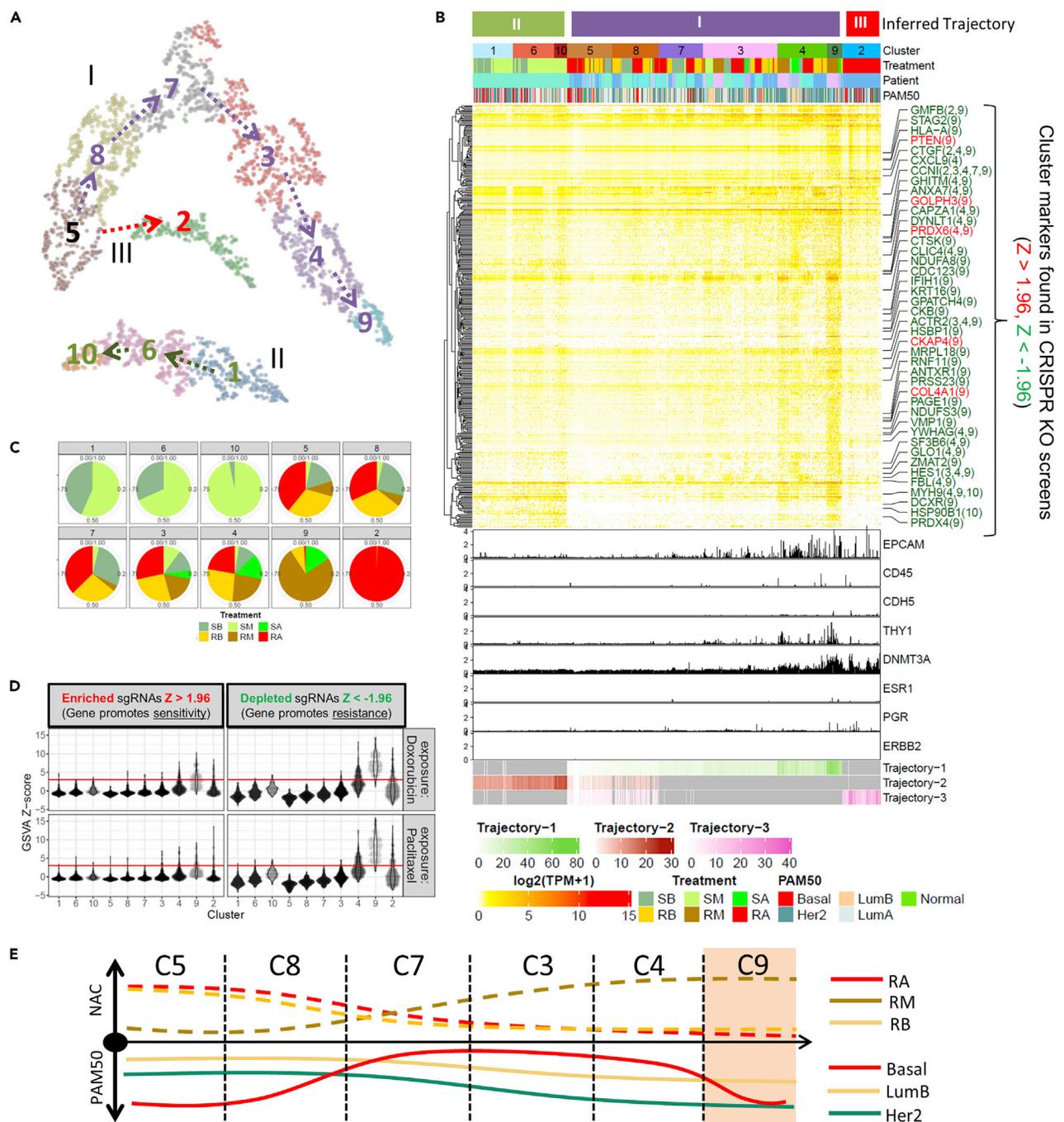
Figure 2. Pseudo-temporal dynamics of single-cell transcriptome during neoadjuvant chemotherapy (NAC)

(A) tSNE plot of cell clusters and associated inferred lineages (I, II and III). The dotted arrows show the inferred tumor trajectories across respective cell clusters labeled by numbers.

(B) Heatmap of cluster markers. The log2(TPM +1) of cluster marker expressions across all nuclei are shown. The top panel shows clinical and molecular annotations for individual nuclei. From the top, the first row depicts cell trajectory assignments, the second row shows cell cluster assignments, the third row shows cell groups as defined by patients' responses (S: chemo-sensitive, R: chemoresistant) and NAC stages (B: pre-NAC, M: mid-MAC, A: post-NAC), the fourth shows the patients from which the nuclei were sourced, and the fifth shows molecular subtypes by PAM50 classifications. On the right, chemo-sensitizing or resistance-promoting gene knockouts from genome-wide CRISPR/Cas9 screening in MCF10A cells are marked (chemo-sensitizing hits: green, resistance-promoting hits: red). The first 5 rows in the bottom panel show key cell type marker genes (epithelial tumor: *EPCAM, CDH5*, immune cells: *CD45*, cancer stem-cell marker: *THY1*, hormone receptors/Her2: *ESR1, PGR,* and *ERBB2*). The bottom three rows show inferred pseudo-time stamps on individual nuclei per trajectory in A.

Figure 2. Continued

(C) Pie charts of cell cluster compositions across different NAC stages and NAC sensitivity. Each pie size represents the proportion of cells coming from chemo-responsive (S) or –resistant (R) patients at different stages of NAC (B: pre-NAC, M: mid-MAC, A: post-NAC). The color codes are shown in the bottom legend.

(D) Enrichments of chemo-sensitizing (left) or resistance-promoting (right) gene knockouts from genome-wide CRISPR/Cas9 screening for Doxorubicin (top) and Paclitaxel (bottom). Each dot shows enrichment of the screening signatures as evaluated by Gene Set Variation Analysis (GSVA). We applied GSVA Z-score >3 or <−3 thresholds (horizontal red lines) to identify significant enrichment or depletion of the knockout signatures in the cellwise transcriptome.

(E) Summary of pseudo-temporal changes in cell populations along trajectory-I. Top panel: Cells by different treatment stages, Bottom panel: Cells by PAM50 molecular subtypes. C9 is marked as the primary cell cluster of interest with therapeutic potentials as indicated by CRISPR-Cas9 screening results in D.

inferred the abundance of the tumor populations by enrichment test for the top 500 most over-expressed genes in the tumor cell cluster marker signature in the bulk samples (see STAR Methods; Figure S6A for overall workflow).

In METABRIC, five distinct clusters emerged (Figure 3A) with significant difference in survival patterns (Figure 3B; logrank pvalue = 4.70E-2). Clusters 1 and 4 were associated with poor prognosis (Figure 3B), where cluster 1 expressed the markers of C3, C4, and C9, and cluster 4 expressed the markers of C2.

Similar patterns were seen in TCGA. Clustering analysis of the tumor cell marker enrichments yielded 7 distinct clusters (Figure 3C). Although TCGA data was limited with a smaller number of samples and shorter periods of patient follow-up than the METABRIC cohort, the clusters still had significant difference in survival (Figure 3D; logrank pvalue = 2.11E-2). Similar to cluster 1 in the METABRIC cohort, cluster 2 from TCGA expressed the markers of C3, C4, and C9, and was associated with poor prognosis (Figure 3C).

In summary, over-expression of C3, C4, and C9 markers, marking the terminal stages of trajectory-I, were consistently associated with poor prognosis in independent bulk cohorts, suggesting the association of C3, C4, and C9 markers in tumors with chemoresistance.

Consensus signature of TNBC chemoresistance compendium

We postulated that trajectory-I captured the adaptive reprogramming of CSCs as a shared molecular feature of TNBC chemoresistance. In other words, some of the core pathways underlying chemoresistance are shared across different models (e.g. EMT,[30] self-renewing cancer stem cells[31]), and should thus be robustly activated.

Hence, we interrogated a broad spectrum of TNBC transcriptome data to extract gene signatures associated with chemoresistance. The gene signatures were derived from the following three studies: (1) The chemoresistant cells against the chemo-responsive cells in the Kim-snRNA-seq data,[14] (2) the therapy-resistant TNBC cells against the treatment naive parental cells underPaclitaxel (GEO: GSE90564) and JQ1 (GEO: GSE63582),[11] (3) and the survival-associated gene signatures in the METABRIC cohort.

Then, we identified the 1,524 frequently up-regulated genes in chemoresistant cases across the gene signatures as the *consensus chemoresistance signaturein TNBC* (hence the consensus chemoresistant pathways; see STAR Methods for details). These genes constitute the pathways robustly activated in TNBC chemoresistance (Figure 4A and Table S5A). Indeed, the consensus signature was also enriched for several hallmark pathways of TNBC therapy-resistance[32] including TNFα signaling via NF-κB (cFET p = 2.57E-16, 5.30 EFC), EMT pathway (cFET p = 1.66E-10, 4.24 EFC) and hypoxia (cFET p = 8.48E-13, 4.71 EFC).

Identification of the consensus signature-driven TNBC models

We identified TNBC models whose chemoresistance was driven by activation of the consensus chemoresistant pathways. We performed GSVA on individual TNBC cell line transcriptomes to evaluate enrichment of the consensus pathways as a proxy for consensus pathway activation. Then, these consensus pathway activation scores were compared with publicly available doxorubicin and paclitaxel efficacy data in GR_{max} index (GR_{max} = −1: cytotoxic to the cells, GR_{max} = 0: cytostatic, i.e., cells are resistant).[33,34] It turned out that MDA-MB-231 and HCC1143 cells were resistant to the regimens and had high consensus activation
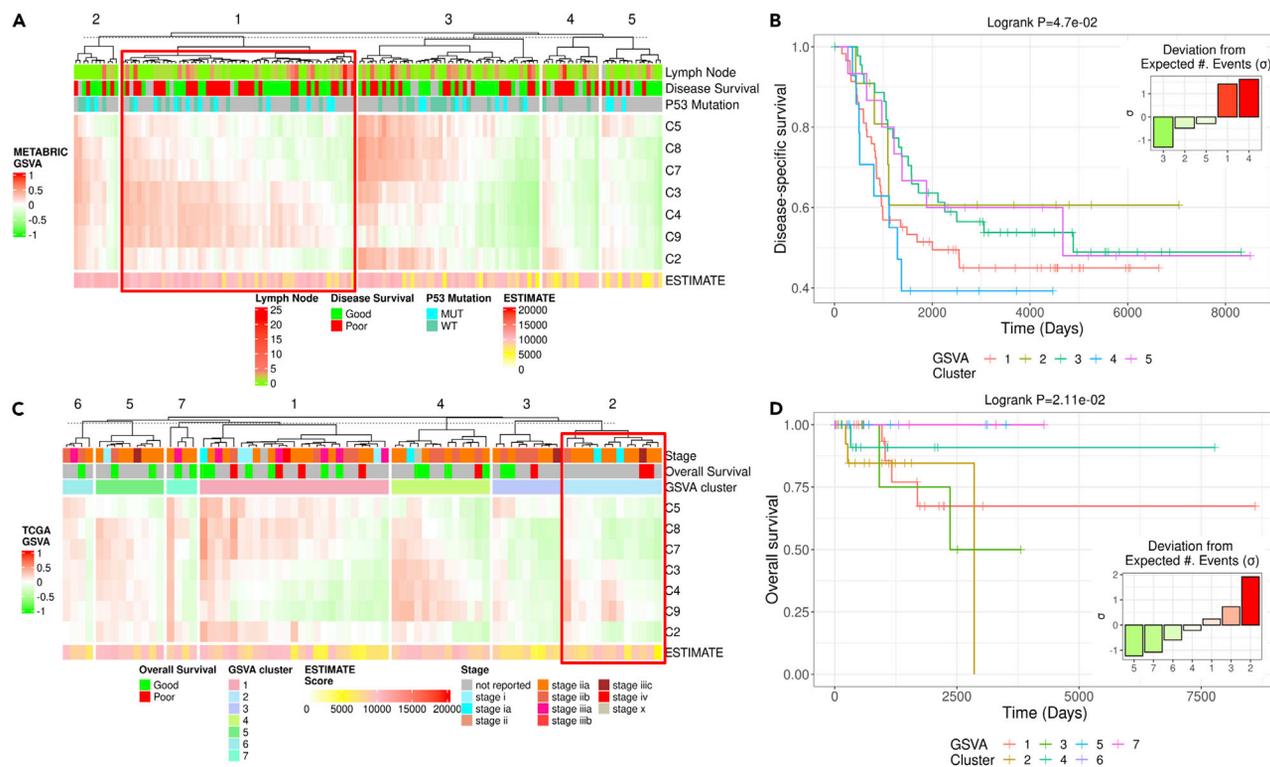
Figure 3. Expressions of the trajectory markers in bulk samples are associated with poor prognosis in METABRIC and TCGA

(A) Heatmap of top 500 tumor cluster marker enrichments in METABRIC TNBC patients. The main heatmap depicts GSVA enrichment score as sample-wise enrichments of the trajectory markers in METABRIC TNBC patients. On top of the heatmap, the annotations depict the number of lymph node metastasis, 5-year disease recurrence (relapse: red, no relapse: green) and TP53 mutation status. On the bottom, the abundance of stromal and immune cells by ESTIMATE score (low: yellow, high: red) is shown in the heatmap bar. The heatmap was clustered by complete linkage hierarchical clustering with 5 clusters as evaluated in Figure S4F. The poor prognosis associated with cluster 1 is highlighted to show enrichments of tumor cell markers from C3, C4, and C9.

(B) Kaplan-Meier plot of METABRIC TNBC patients stratified by hierarchical clustering from the tumor cell marker enrichments in A.

(C) Heatmap of marker GSVA z-scores in TCGA TNBC patients: From the top, the annotation depicts cancer stage, 5-year overall survival (death: red, alive: green), and sample clusters by hierarchical clustering on the GSVA matrix with k = 4. The ESTIMATE scores are shown on the bottom annotation bar. The tumor-enriched clusters with mild C9 and C2 marker enrichments are associated with poor prognosis and being highlighted in a red box.

(D) Kaplan-Meier plot of TCGA TNBC patients stratified by hierarchical clustering (k = 4) from the marker enrichments in C.

scores whereas MDA-MB-453 cells showed a low consensus activation score and were highly resistant to the regimens (Figure 4C).

Such heterogeneity in chemoresistant TNBC cell lines may be explained by several molecular factors. EMT and TNFα pathways, EGFR, and the snRNA-seq markers (C3, C4, and C9) were activated in MDA-MB-231 cells (Figure S7). In contrast, these pathways and genes were lowly expressed in MDA-MB- (Figure S7). These results suggest that MDA-MB-231 is more suitable for modeling chemoresistance for it captures the pseudo-temporal dynamics in the Kim-snRNA-seq data, and MDA-MB-453 can serve as a model with low expression of EGFR.

Data-driven drug prediction identifies EGFR inhibition as an effective strategy to abolish chemoresistance

As the consensus signature represents the shared pathways conferring chemoresistance, these pathways could represent an Achilles heel in the resistant TNBC cells and serve as therapeutic targets to synergize with the established chemo-regimens to improve the overall efficacy.

To address this, we designed a computational framework to predict FDA-approved drugs targeting the consensus pathways (Step III, Figure 1). We first sought to identify an adequate TNBC model whose chemo-resistance is driven by the consensus pathways. To this end, we utilized the GR50 index for evaluating the
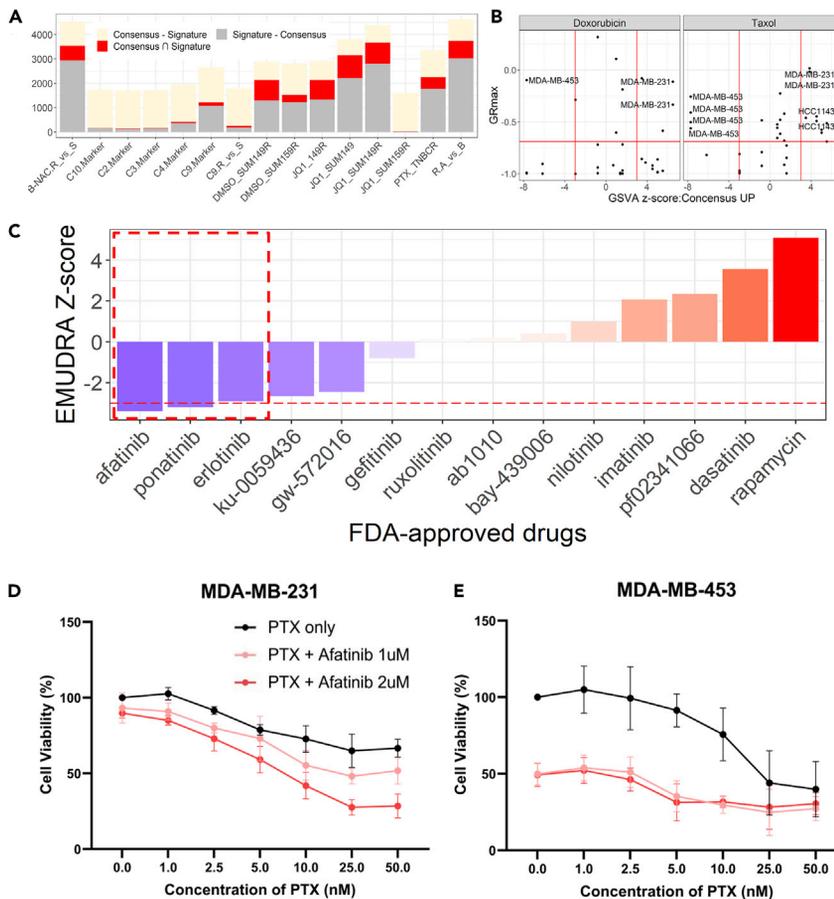
Figure 4. Identification of FDA-approved drugs to target consensus-driven chemoresistant TNBC cells

(A) Overlap between the consensus signature and individual chemoresistance signatures from multiple sources (marked in red). Differences between individual signatures and consensus signatures are also shown in gray and ivory as labeled in the legend.

(B) Scatterplot of consensus signature enrichments in TNBC cell line transcriptome (xaxis: GSVA Zscore of the consensus signature) and efficacy of doxorubicin (on left) and paclitaxel, also known as taxol (on right), by GR_{max} index. The horizontal red line is the median GR_{max} score among TNBC cells per tested drug. Vertical red lines are |GSVA Zscore| = 3,-3 thresholds to identify cells enriched or depleted of the consensus upregulated signature.

(C) Computational prediction of FDA-approved drugs to suppress the consensus chemoresistance pathways by EMUDRA. On the yaxis, the negative EMUDRA Zscore shows the predicted efficacy of the drug to reverse the consensus signature in MDA-MB-231 cells.

(D and E) Dose-dependent TNBC chemoresistant cell viability on Paclitaxel (PTX) only or Paclitaxel + afatinib combination therapy on MDA-MB-231 cells (D) and MDA-MB-453 cells (E). Curves with different afatinib concentrations are highlighted by different colors (0 μM: black, 1 μM: pink, 2 μM: red). Xaxis: paclitaxel concentrations (in nM). Yaxis: cell viability after 72 h of treatment, where 100% cell viability corresponds to untreated cells at 0h time point. The error bars are Mean ± SD.

drug sensitivity of TNBC cells to standard chemo regimens (doxorubicin and paclitaxel).[33] The non-responsive cells with a high degree of consensus pathway activation, hence likely to acquire chemoresistance were inspected. This yielded MDA-MB-231 as the ideal TNBC model, whose chemoresistance is driven by the highest degree of activation of the consensus chemoresistance pathway (Figure 4B).

Then, we leveraged Ensemble of Multiple Drug Repositioning Approaches (EMUDRA)[35] to predict FDA-approved drugs. Briefly, EMUDRA queries cancer cell transcriptome changes by chemical perturbations in LINCS database[36] and examines if each pharmacological perturbation can successfully reverse the transcriptional changes in disease. In our study, we focused on the chemically perturbed MDA-MB-231 transcriptome to search for potential drugs that suppress the consensus chemoresistance signature (Figure 4C and Table S5C).

Of interest, the top three predictions from EMUDRA (afatinib, ponatinib, and erlotinib; Figure 4C) commonly inhibit EGFR. Afatinib is a pan-HER inhibitor that has been approved for Her2+ non-small cell lung cancer and has shown efficacy in Her2+ breast cancers.[37,38] Ponatinib, a multi-targeted tyrosine kinase inhibitor, targets primarily Bcr-Abl tyrosine kinase proteins, but also VEGF receptor, PDGFR, SRC kinase, KIT, and RET.[39] Ponatinib has shown efficacy in chronic myeloid leukemia (CML) and acute lymphoblastic leukemia (ALL).[40] A recent multi-omics analysis of TNBC cell lines showed potential efficacy in the mesenchymal subtype.[41] Erlotinib, an EGFR inhibitor, is an FDA-approved drug for non-small cell lung cancer and pancreatic cancer. In preclinical studies, erlotinib synergized with metformin to show improved efficacy in basal breast cancer cell lines.[42]

We questioned if the top-ranked drug afatinib which targets the EGFR axis can effectively suppress the consensus pathways in TNBC cells (Figure 4D). We compared the afatinib-perturbed transcriptome of TNBC cells in the LINCS database to evaluate the suppression of the consensus chemoresistance signature in TNBC cells. As expected, the consensus chemoresistance signature was significantly suppressed by afatinib in MDA-MB-231 cells (GSEA pvalue = 1.53E-2, normalized enrichment score (NES) = −1.16), and BT-20 cells (GSEA pvalue = 1.29E-5, NES = −1.38).

In smmary, EGFR inhibition targeted by the top predicted drugs is consistent with the pro-tumorigenic roles of EGFR in TNBC, such as modulating CSC phenotypes, recurrence and metastasis.[43,44]

Predicted drugs exhibit improved efficacy when combined with paclitaxel

We performed *in vitro* validation of the predicted treatment in MDA-MB-231 cells as the ideal consensus-driven chemoresistant TNBC model. On the other hand, MDA-MB-453 served as the negative counterpart to MDA-MB-231 because of the absence of the consensus pathways, hence the consensus-free resistant cells (Figure 4B). Then, we compared the efficacies of the combination therapy (afatinib and paclitaxel) in MDA-MB-231 and MDA-MB-453 by cell viability assays (see STAR Methods for details).

As predicted, we observed the synergistic efficacy of paclitaxel and afatinib in MDA-MB-231 (Figure 4E). Notably, the combination therapy with 2 μM of afatinib showed greater efficacy than 1 μM. The higher concentration led to greater inhibition of downstream MAPK/ERK and PI3K/Akt signaling, and this is indicated by a greater reduction in phosphorylation of ERK1/2 and Akt (Figure S8A). On the other hand, the addition of afatinib did not improve response to paclitaxel monotherapy in MDA-MB-453 cells (Figure 4F), as expected by the absence of dependency of EGFR signaling in these cells.

In addition, we evaluated the efficacy of ponatinib and paclitaxel in MDA-MB-231 cells to further validate the drug predictions. As the second-ranked prediction, we argue that the efficacy of ponatinib and paclitaxel would demonstrate the effectiveness of the predictive framework. We tested 250 nM and 400 nM dosages of ponatinib with a range of pacliataxel dosages from 0 to 1000 nM.

We observed additive efficacies of paclitaxel and ponatinib. For both of 250 nM and 400 nM ponatinib dosages, the difference between paclitaxel monotherapy and ponatinib combination curves matched the efficacy of ponatinib monotherapy (Figure S8). Although the observed synergy is less than the afatinib combination, the results demonstrate that ponatinib complements paclitaxel in MDA-MB-231 cells, and targets the paclitaxel-resistant pathways to improve the overall efficacy.

Combined therapy of paclitaxel and afatinib exhibits synergistic efficacy in MDA-MB-231 cells

We systematically evaluated the synergy of afatinib and paclitaxel in the MDA-MB-231 and MDA-MB-453 cells. We utilized the Bliss index, which tests the significant deviation of the observed effects from the expected effects if the two drugs worked independently.[45] Overall, the Bliss index across the range of doses in MDA-MB-231 cells indicated significant synergy with an average of 13.16 across three replicates, and these were significantly higher than MDA-MB-453 cells (Wilcox pvalue = 5E-2, Figure S10).

We investigated the suppression of the chemoresistance pathways by combination therapy. We performed RNA-sequencing on untreated, mono-treated (paclitaxel or afatinib) and the combo-treated MDA-MB-231 cells (see STAR Methods for details), and evaluated enrichments of the chemoresistance signatures in mono- and combo-treated cells by Gene Set Enrichment Analysis (GSEA).[46] In MDA-MB-231 cells, the
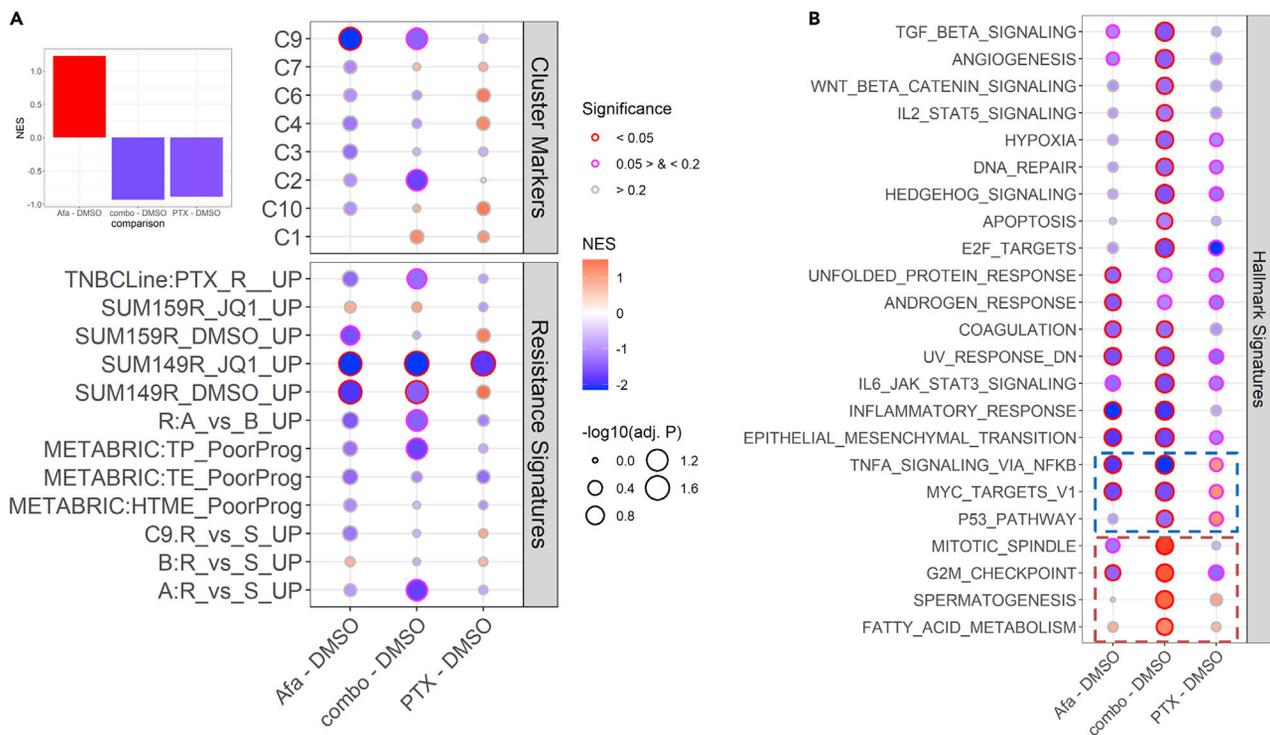
Figure 5. Differentially expressed pathways in Paclitaxel and afatinib treated MDA-MB-231 cells, compared to DMSO treated cells

(A) Enriched chemoresistance signatures in TNBC tumor as evaluated by Gene Set Enrichment Analysis (GSEA). The top left panel shows the normalized enrichment score (NES) of the consensus chemoresistance signature by different treatments in MDA-MB-231 cells. The main panel shows enrichments of various chemoresistance signatures by the drug treatments (Afa: afatinib, PTX: Paclitaxel, combo: afatinib + Paclitaxel). The upper panel includes cell cluster markers from snRNA-seq of chemo-treated patient samples, and the lower panel shows resistance signatures from various independent *in vitro* studies, survival signatures from METABRIC, and differential gene signatures in chemoresistant cells from the snRNA-seq. The colors indicate NES as the degree of the signature activation (in red) or suppression (in blue), point sizes as correlations to the significance of the NES (i.e. –log10(FDR adjusted GSEA pvalue)), and border colors for significance levels as shown in the right side legend.

(B) Enriched hallmark pathways: Similar to A, enrichments of the cancer hallmark signatures in the Molecular Signature Database (MSigDB) in the treated cells are shown. The treatments are shown on the xaxis. The significance of overall activation or suppression is denoted similarly to A. Significantly down-regulated pathways by the combination therapy are highlighted by the blue dashed box. Likewise, upregulated pathways by the combination treatment are highlighted by the red dashed box.

combination treatment most effectively silenced the consensus pathways (top left, Figure 5A) as well as many individual chemoresistant signatures, compared to other mono-treatments (Figure 5A). In the Kim-snRNA-seq, the genes down-regulated by the combination treatment were significantly enriched for the markers of chemoresistant cells, C2 and C9. Furthermore, the combination treatment down-regulated paclitaxel resistance signatures from pan-TNBC cell lines, JQ1 resistance signatures from SUM149 cells, and poor prognosis-associated genes from tumor-purified samples in METABRIC cohort (lower panel, Figure 5A).

On the other hand, the mono-treated cells did not demonstrate as many suppressions of the chemoresistance pathways. Paclitaxel-treated cells did not show any significant enrichments the chemoresistant single-nuclei cluster markers, and afatinib-treated cells down-regulated C9 markers (upper panel, Figure 5A). These results indicate that the combination therapy shows greater efficacy in suppressing the consensus pathways than the respective monotherapies, demonstrating synergistic lethality.

Finally, we investigated the effects of afatinib on other pathways. In comparison with the mono-treatments, the combination treatment most effectively suppressed cancer hallmark pathways (Figure 5B) including TNFα signaling via NF-κB, EMT, and hypoxia, which were enriched in the consensus chemoresistance signature. The combination therapy also significantly down-regulated several pathways upregulated by paclitaxel, including TNFα signaling via NF-κB, MYC, and P53 (highlighted in the red dashed box, Figure 5B).

DISCUSSION

Herein, we developed an integrative computation framework to investigate the clinical problem of chemoresistance in TNBC. Therapy resistance can emerge from the expansion of intrinsic, therapy-resistant sub-clones,[13] or is acquired from genetic or epigenetic events that eventually favor the cancer cells' survival and expansion.[13,14] To address this multi-faceted nature of chemoresistance, we analyzed the single-nucleus transcriptome of TNBC patients undergoing neo-adjuvant chemotherapy (NAC) and identified three tumor cell trajectories. This study revealed that tumor cell trajectory-I was shared across different patients and associated with increased proportions of cells from mid-NAC phase and progression toward expressing markers of several oncogenic pathways including MYC (FET FDR = 2.01E-19, EFC = 5.87), EMT (FET FDR = 2.63E-16, EFC = 5.51) and TNFα signaling (FET FDR = 1.28E-15, 5.41 EFC).

Several marker genes of C9 suggest reprogramming into CSC-like cells during mid-NAC. These include a cancer stem cell marker *THY1*, and a regulator of genome-wide *de novo* epigenetic modifications, *DNMT3A*. In cancer stem cells, *DNMT3A* collaborates with *MYC* to silence the tumor suppressor miR-200b, promoting EMT.[47] The aforementioned activation of MYC and EMT pathways in C9 also supports the DNMT3A-MYC axis in cancer stem cells, suggesting epigenetic reprogramming into cancer stem cells as a potential mechanism of adaptive chemoresistance.

Meanwhile, the inferred trajectory (trajectory-I) reflected the gradual emergence of chemoresistance-promoting genes during NAC. The chemoresistance-promoting genes from our genome-wide loss-of-function CRISPR-Cas9 screening were gradually up-regulated along trajectory-I and peaked at the terminal cluster, C9. As mid-NAC cells also gradually increased along trajectory-I, these imply therapeutic opportunities for TNBC chemoresistance lie in targeting the mid-NAC window.

The robustness of snRNA-seq-based finding was limited by the small number of samples. To mitigate these shortcomings, we interrogated independent TNBC studies including clinically annotated bulk RNA-seq cohorts and *in vitro* studies of TNBC chemoresistance and established the consensus chemoresistance signature by searching for robustly up-regulated genes in the resistant tumor cells.

By querying chemically perturbed transcriptomes of TNBC cells, we predicted afatinib, a pan-HER (EGFR, HER2, and HER3) tyrosine kinase inhibitor, as the most promising FDA-approved drug to suppress the consensus chemoresistance signature genes using EMUDRA.

The consensus signature was also instrumental to identify an adequate TNBC cellular model. The joint analysis of consensus pathway enrichment and drug efficacy data on TNBC cell lines predicted MDA-MB-231 as the responsive paclitaxel-resistant model, in contrast to MDA-MB-453 as the non-responsive counterpart. The predicted synergy between afatinib and paclitaxel was validated *in vitro* by testing cell viability on exposure to mono- or combination therapy of paclitaxel and afatinib in MDA-MB-231 and MDA-MB-453 cells. Indeed, we observed the dosage-dependent synergy in MDA-MB-231 cells, whereas such synergy was not observed in MDA-MB-453 cells. This is in line with the increased presence of Her2-like cells (Figure S3B) and EGFR expression (Figure S3D) in the shared chemoresistant cell trajectory (trajectory-I), and demonstrates the EGFR-axis as a promising therapeutic target.

In human cancers, afatinib, an orally bioavailable irreversible pan-HER inhibitor, has shown some efficacy in metastatic TNBC patients as a monotherapy in a phase 2 clinical trial.[38] Its combination with dasatinib, an SRC-family tyrosine kinase inhibitor, has also shown synergy in several TNBC cell lines through inhibiting ERK and Akt signaling.[48] Afatinib in combination with paclitaxel has shown efficacy in epithelial cancer cells, including ovarian, colon, non-small cell lung carcinoma, and esophageal cancers.[49]

However, EGFR inhibition as a monotherapy or in combination with other treatments has demonstrated only marginal efficacy.[50–56] Although most of these clinical trials were focused on metastatic, advanced TNBC, two recent clinical studies reported that EGFR inhibition is more beneficial in operable primary TNBC cases.[56,57] These imply that post-treatment, metastatic TNBC tumors may not be dependent on EGFR signaling alone,[58] and the emergence of CSC niche should be evaluated during mid-NAC to increase the likelihood of therapeutic efficacy.

As opposed to Her2-like convergence in trajectory-I, the trajectory-III (Figure 2) represents a basal-like tumor cell trajectory with chemoresistant post-NAC cluster C2 as the terminal cluster. The markers of C2 were enriched for TNFα signaling (FET FDR = 4.51E-11, EFC = 17.5) and hypoxia (FET FDR = 1.09E-4, EFC = 10.1). This trajectory was not extensively studied because of its specificity for a single patient, but it is worth an investigation as another distinct chemoresistance axis. Although this trajectory also shows high EGFR expression (Figure S3D), it is distinctively different from Her2-like enriched C9. The CRISPR/Cas9 knockout signatures of chemoresistance were not significantly enriched in C2 (Figure 2C). In the bulk samples, C2 and C9 marker enrichments were mutually exclusive in tumor-enriched, poor prognosis samples (Figures 3A and 3B). Also, the enrichment of hypoxia pathways was unique to C2, indicating the basal-like trajectory captures the hypoxic niche of chemoresistance.[59] Thus, unlike the Her2-like cells expanding in mid-NAC, the basal-like, post-treatment cells in trajectory-III/C2 may not depend on EGFR signaling and require entirely different post-treatment therapeutic approaches.

Luminal A tumors express high levels of the hormone receptors ER+ and PR+, and show a good response to therapies targeting ER-dependent axis such as anti-estrogen and aromatase inhibitors which directly inhibit ER production.[60] On the other hand, luminal B tumors are express lower amount of these receptors and are characterized by higher cellular proliferation and poorer responses to the anti-hormonal therapies. During ER-axis targeted therapy, it has been observed that luminal A tumors are converted into the more resistant subtype, luminal B evolving into ER-independent tumors, which confers resistance to these therapies.[60]

Trajectory-I revealed the expansion of luminal B-like cells as well as the regression of luminal A cells (Figures 2C and 2E), suggesting the chemoresistant mechanisms in TNBC cells may be shared with the resistance in ER+ breast cancers. Further studies are necessary to determine the clinical consequences of this finding.

In summary, we have developed an integrative computational framework to dissect the pseudo-temporal dynamics of TNBC undergoing neoadjuvant chemotherapy (NAC). This study has provided an innovative molecular landscape of the adaptive dynamics of TNBCs and identified shared tumor cell trajectories across multiple patients leading to chemotherapy resistance. Our study presents a more holistic understanding of tumor evolution during treatment and provides promising windows for therapeutics against TNBC.

Limitations of the study

Although our study has led to several findings, it is worth discussing its limitations. First, the number of single-cell transcriptome of chemo-treated TNBC patient samples that are publicly available and that we could analyze is low. Thus, the cellular dynamics portrayed in this study is likely incomplete. For the drug prediction framework, we were restricted to the cell lines with GR50 index (used to evaluate chemoresistance) and the list of cell lines with chemically perturbed transcriptome by FDA-approved drugs in the LINCS database. These limitations significantly reduced the overall list of TNBC cells for adequate validation experiments and led to the nomination of MDA-MB-231 as the only adequate model to test suppression of the consensus-driven chemoresistance.

In this regard, BT-20 showed a moderate degree of activation of the consensus chemoresistance signature (GSVA Zscore = 1.12) and a high expression of EGFR (log2-transformed counts per million (CPM) = 10.67). Thus, BT-20 could potentially serve as another model for consensus-driven chemoresistant TNBC to examine the synergistic efficacy of afatinib andpacliataxel. As *in vitro* chemotherapy efficacy on BT-20 was missing in the GR50 database, so it was omitted in the formal TNBC model nomination.

Further study is needed to to identify and characterize more chemoresistant trajectories and test novel drug combinations to restore chemosensitivity.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106064.

AUTHOR CONTRIBUTIONS

Conceptualization, W.M.S. and B.Z, and J.S.; Methodology, W.M.S. and P.L.C.; Writing – Original draft, W.M.S. and P.L.C.; Investigation, W.M.S., P.L.C., and X.Z.; Resources, J.S. and P.L.C.; Supervision, B.Z. and J.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Dent, R., Trudeau, M., Pritchard, K.I., Hanna, W.M., Kahn, H.K., Sawka, C.A., Lickley, L.A., Rawlinson, E., Sun, P., and Narod, S.A. (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. Clin. Cancer Res. *13*, 4429–4434.

2. Liedtke, C., Mazouni, C., Hess, K.R., André, F., Tordai, A., Mejia, J.A., Symmans, W.F., Gonzalez-Angulo, A.M., Hennessy, B., Green, M., et al. (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J. Clin. Oncol. *26*, 1275–1281.

3. Bai, X., Ni, J., Beretov, J., Graham, P., and Li, Y. (2021). Triple-negative breast cancer therapeutic resistance: where is the Achilles' heel? Cancer Lett. *497*, 100–111.

4. Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M.V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., and Pietenpol, J.A. (2016). Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. PLoS One *11*, e0157368.

5. Lebert, J.M., Lester, R., Powell, E., Seal, M., and McCarthy, J. (2018). Advances in the systemic treatment of triple-negative breast cancer. Curr.Oncol. *25*, S142–S150.

6. Horiuchi, D., Kusdra, L., Huskey, N.E., Chandriani, S., Lenburg, M.E., Gonzalez-Angulo, A.M., Creasman, K.J., Bazarov, A.V., Smyth, J.W., Davis, S.E., et al. (2012). MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition. J. Exp. Med. *209*, 679–696.

7. Knudsen, E.S., and Witkiewicz, A.K. (2017). The strange case of CDK4/6 inhibitors: mechanisms, resistance, and combination strategies. Trends Cancer *3*, 39–55.

8. Witkiewicz, A.K., Chung, S., Brough, R., Vail, P., Franco, J., Lord, C.J., and Knudsen, E.S. (2018). Targeting the vulnerability of RB tumor suppressor loss in triple-negative breast cancer. Cell Rep. *22*, 1185–1199.

9. Hwang, S.Y., Park, S., and Kwon, Y. (2019). Recent therapeutic trends and promising targets in triple negative breast cancer. Pharmacol.Ther. *199*, 30–57.

10. Sahni, J.M., Gayle, S.S., Bonk, K.L.W., Vite, L.C., Yori, J.L., Webb, B., Ramos, E.K., Seachrist, D.D., Landis, M.D., Chang, J.C., et al. (2016). Bromodomain and extraterminal protein inhibition blocks growth of triple-negative breast cancers through the suppression of aurora kinases. J. Biol. Chem. *291*, 23756–23768.

11. Shu, S., Lin, C.Y., He, H.H., Witwicki, R.M., Tabassum, D.P., Roberts, J.M., Janiszewska, M., Huh, S.J., Liang, Y., Ryan, J., et al. (2016).

Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer. Nature *529*, 413–417.

12. Polyak, K. (2011). Heterogeneity in breast cancer. J. Clin. Invest. *121*, 3786–3788.

13. Zong, Y., and Goldstein, A.S. (2013). Adaptation or selection–mechanisms of castration-resistant prostate cancer. Nat. Rev. Urol. *10*, 90–98.

14. Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N.E. (2018). Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. Cell *173*, 879–893.e13.

15. Wang, Q., Guldner, I.H., Golomb, S.M., Sun, L., Harris, J.A., Lu, X., and Zhang, S. (2019). Single-cell profiling guided combinatorial immunotherapy for fast-evolving CDK4/6 inhibitor-resistant HER2-positive breast cancer. Nat. Commun. *10*, 3817.

16. Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. *37*, 547–554.

17. Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., and Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. Nat. Commun. *11*, 1201.

18. Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*, R68.

19. Wang, X., Liu, Y., Zhou, K., Zhang, G., Wang, F., and Ren, J. (2015). Isolation and characterization of CD105+/CD90+ subpopulation in breast cancer MDA-MB-231 cell line. Int. J. Clin. Exp. Pathol. *8*, 5105–5112.

20. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

21. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genom. *19*, 477.

22. Hall, A. (2009). The cytoskeleton and cancer. Cancer Metastasis Rev. *28*, 5–14.

23. Subramaniam, D., Thombre, R., Dhar, A., and Anant, S. (2014). DNA methyltransferases: a novel target for prevention and therapy. Front. Oncol. *4*, 80.

24. Gendoo, D.M.A., Ratanasirigulchai, N., Schröder, M.S., Paré, L., Parker, J.S., Prat, A., and Haibe-Kains, B. (2016). Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics *32*, 1097–1099.

25. Soule, H.D., Maloney, T.M., Wolman, S.R., Peterson, W.D., Jr., Brenz, R., McGrath, C.M., Russo, J., Pauley, R.J., Jones, R.F., and Brooks, S.C. (1990). Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res. *50*, 6075–6086.

26. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7.

27. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

28. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

29. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

30. De Las Rivas, J., Brozovic, A., Izraely, S., Casas-Pais, A., Witz, I.P., and Figueroa, A. (2021). Cancer drug resistance induced by EMT: novel therapeutic strategies. Arch. Toxicol. *95*, 2279–2297.

31. Li, Y., Wang, Z., Ajani, J.A., and Song, S. (2021). Drug resistance and Cancer stem cells. Cell Commun. Signal. *19*, 19.

32. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

33. Hafner, M., Niepel, M., Chung, M., and Sorger, P.K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods *13*, 521–527.

34. Hafner, M., Heiser, L.M., Williams, E.H., Niepel, M., Wang, N.J., Korkola, J.E., Gray, J.W., and Sorger, P.K. (2017). Quantification of sensitivity and resistance of breast cancer cell lines to anti-cancer drugs using GR metrics. Sci. Data *4*, 170166.

35. Zhou, X., Wang, M., Katsyv, I., Irie, H., and Zhang, B. (2018). EMUDRA: Ensemble of multiple drug repositioning approaches to improve prediction accuracy. Bioinformatics *34*, 3151–3159.

36. Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D.J., Turner, J.P., Vidovic, D., Forlin, M., Kelley, T.T., D'Urso, A., et al. (2018). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res. *46*, D558–D566.

37. Hurvitz, S.A., Shatsky, R., and Harbeck, N. (2014). Afatinib in the treatment of breast cancer. Expert Opin.Investig. Drugs *23*, 1039–1047.

38. Schuler, M., Awada, A., Harter, P., Canon, J.L., Possinger, K., Schmidt, M., De Grève, J., Neven, P., Dirix, L., Jonat, W., et al. (2012). A phase II trial to assess efficacy and safety of afatinib in extensively pretreated patients with HER2-negative metastatic breast cancer. Breast Cancer Res. Treat. *134*, 1149–1159.

39. Frankfurt, O., and Licht, J.D. (2013). Ponatinib–a step forward in overcoming resistance in chronic myeloid leukemia. Clin. Cancer Res. *19*, 5828–5834.

40. Tan, F.H., Putoczki, T.L., Stylli, S.S., and Luwor, R.B. (2019). Ponatinib: a novel multi-tyrosine kinase inhibitor against human malignancies. OncoTargets Ther. *12*, 635–645.

41. Lehmann, B.D., Colaprico, A., Silva, T.C., Chen, J., An, H., Ban, Y., Huang, H., Wang, L., James, J.L., Balko, J.M., et al. (2021). Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. Nat. Commun. *12*, 6276.

42. Lau, Y.K.I., Du, X., Rayannavar, V., Hopkins, B., Shaw, J., Bessler, E., Thomas, T., Pires, M.M., Keniry, M., Parsons, R.E., et al. (2014). Metformin and erlotinib synergize to inhibit basal breast cancer. Oncotarget 5, 10503–10517.

43. Liu, X., Adorno-Cruz, V., Chang, Y.F., Jia, Y., Kawaguchi, M., Dashzeveg, N.K., Taftaf, R., Ramos, E.K., Schuster, E.J., El-Shennawy, L., et al. (2021). EGFR inhibition blocks cancer stem cell clustering and lung metastasis of triple negative breast cancer. Theranostics *11*, 6632–6643.

44. Lev, S. (2020). Targeted therapy and drug resistance in triple-negative breast cancer: the EGFR axis. Biochem. Soc. Trans. *48*, 657–665.

45. Bliss, C.I. (1939). The toxicity of poisons applied jointly. Ann. Appl. Biol. *26*, 585–615.

46. 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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

47. Pang, Y., Liu, J., Li, X., Xiao, G., Wang, H., Yang, G., Li, Y., Tang, S.C., Qin, S., Du, N., et al. (2018). MYC and DNMT3A-mediated DNA methylation represses microRNA-200b in triple negative breast cancer. J. Cell Mol. Med. *22*, 6262–6274.

48. Canonici, A., Browne, A.L., Ibrahim, M.F.K., Fanning, K.P., Roche, S., Conlon, N.T., O'Neill, F., Meiller, J., Cremona, M., Morgan, C., et al. (2020). Combined targeting EGFR and SRC as a potential novel therapeutic approach for the treatment of triple negative breast cancer. Ther. Adv. Med. Oncol. *12*. 1758835919897546.

49. Suder, A., Ang, J.E., Kyle, F., Harris, D., Rudman, S., Kristeleit, R., Solca, F., Uttenreuther-Fischer, M., Pemberton, K., Pelling, K., et al. (2015). A phase I study of daily afatinib, an irreversible ErbB family

blocker, in combination with weekly paclitaxel in patients with advanced solid tumours. Eur. J. Cancer *51*, 2275–2284.

50. Ferraro, D.A., Gaborit, N., Maron, R., Cohen-Dvashi, H., Porat, Z., Pareja, F., Lavi, S., Lindzen, M., Ben-Chetrit, N., Sela, M., and Yarden, Y. (2013). Inhibition of triple-negative breast cancer models by combinations of antibodies to EGFR. Proc. Natl. Acad. Sci. USA *110*, 1815–1820.

51. Finn, R.S., Press, M.F., Dering, J., Arbushites, M., Koehler, M., Oliva, C., Williams, L.S., and Di Leo, A. (2009). Estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 (HER2), and epidermal growth factor receptor expression and benefit from lapatinib in a randomized trial of paclitaxel with lapatinib or placebo as first-line treatment in HER2-negative or unknown metastatic breast cancer. J. Clin. Oncol. *27*, 3908–3915.

52. Harbeck, N., Schmidt, M., Harter, P., Possinger, K., Jonat, W., and Lück, H. (2009). BIBW 2992, a Novel Irreversible EGFR/HER1 and HER2 Tyrosine Kinase Inhibitor for the Treatment of Patients with HER2-Negative Metastatic Breast Cancer after Failure of No More than Two Prior Chemotherapies (AACR).

53. Carey, L.A., Rugo, H.S., Marcom, P.K., Mayer, E.L., Esteva, F.J., Ma, C.X., Liu, M.C., Storniolo, A.M., Rimawi, M.F., Forero-Torres, A., et al. (2012). TBCRC 001: randomized phase II study of cetuximab in combination with carboplatin in stage IV triple-negative breast cancer. J. Clin. Oncol. *30*, 2615–2623.

54. Layman, R.M., Ruppert, A.S., Lynn, M., Mrozek, E., Ramaswamy, B., Lustberg, M.B., Wesolowski, R., Ottman, S., Carothers, S., Bingman, A., et al. (2013). Severe and prolonged lymphopenia observed in patients treated with bendamustine and erlotinib for metastatic triple negative breast cancer. Cancer Chemother.Pharmacol. *71*, 1183–1190.

55. Baselga, J., Gómez, P., Greil, R., Braga, S., Climent, M.A., Wardley, A.M., Kaufman, B., Stemmer, S.M., Pêgo, A., Chan, A., et al. (2013). Randomized phase II study of the anti-epidermal growth factor receptor monoclonal antibody cetuximab with cisplatin versus cisplatin alone in patients with metastatic triple-negative breast cancer. J. Clin. Oncol. *31*, 2586–2592.

56. Nabholtz, J.M., Abrial, C., Mouret-Reynier, M.A., Dauplat, M.M., Weber, B., Gligorov, J., Forest, A.M., Tredan, O., Vanlemmens, L., Petit, T., et al. (2014). Multicentric neoadjuvant phase II study of panitumumab combined with an anthracycline/taxane-based chemotherapy in operable triple-negative breast cancer: identification of biologically defined signatures predicting treatment impact. Ann. Oncol. *25*, 1570–1577.

57. Matsuda, N., Wang, X., Lim, B., Krishnamurthy, S., Alvarez, R.H., Willey, J.S., Parker, C.A., Song, J., Shen, Y., Hu, J., et al. (2018). Safety and efficacy of panitumumab Plus neoadjuvant chemotherapy in patients with primary HER2-negative inflammatory breast cancer. JAMA Oncol. *4*, 1207–1213.

58. Nedeljkovic, M., and Damjanovic, A. (2019). Mechanisms of chemotherapy resistance in triple-negative breast cancer-how we can rise to the challenge. Cells *8*, 957.

59. Cosse, J.P., and Michiels, C. (2008). Tumour hypoxia affects the responsiveness of cancer cells to chemotherapy and promotes cancer progression. Anti Cancer Agents Med. Chem. *8*, 790–797.

60. Wong, C., and Chen, S. (2012). The development, application and limitations of breast cancer cell lines to study tamoxifen and aromatase inhibitor resistance. J. Steroid Biochem. Mol. Biol. *131*, 83–92.

61. Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq reads with STAR. Curr.Protoc. Bioinformatics *51*, 11.14.1.

62. Chen, Y., Lun, A.T.L., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res. *5*, 1438.

63. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

64. Joung, J., Konermann, S., Gootenberg, J., Abudayyeh, O., Platt, R., Brigham, M., Sanjana, N., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.

65. Pons, P., and Latapy, M. (2005). Computing communities in large networks using random walks. Lect. Notes Comput.Sci. *3733*, 284–293.

66. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

67. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

68. Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat. Commun. *10*, 380.

69. Jain, A.K., Murty, M.N., and Flynn, P.J. (1999). Data clustering: a review. ACM Comput.Surv. *31*, 264–323.

70. Harrington, D.P., and Fleming, T.R. (1982). A class of rank test procedures for censored survival data. Biometrika *69*, 553–566.

71. Therneau, T.M. (1997). Extending the Cox model. In Proceedings of the First Seattle Symposium in Biostatistics (Springer), pp. 51–84.

72. Park, J.H., Ahn, J.H., and Kim, S.B. (2018). How shall we treat early triple-negative breast cancer (TNBC): from the current standard to upcoming immuno-molecular strategies. ESMO Open *3*, e000357.

73. Cox, D.R. (1972). Regression models and life-tables. J. Roy. Stat. Soc. B *34*, 187–202.

74. Dobin, A., and Gingeras, T.R. (2016). Optimizing RNA-seq mapping with STAR. Methods Mol. Biol. *1415*, 245–262.

75. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

76. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
phospho-Tyr	CST	Cat #9411; RRID: RRID:AB_331228
phospho-EGFR	CST	Cat #4407; RRID:AB_331795
EGFR	CST	Cat #2232; RRID:AB_331707
phospho-ERK1/2	CST	Cat #4370; RRID:AB_2315112
ERK1/2	Santa Cruz	sc-94; RRID:AB_2140110
phospho-Akt	CST	Cat #3787; RRID:AB_331170
Akt	CST	Cat #4691; RRID:AB_915783
β-actin	Santa Cruz	sc-47778; RRID:AB_2714189
HRP-conjugated anti-rabbit or anti-mouse antibodies	GE healthcare	Cat#NA934V, #NA931V
Chemicals, peptides, and recombinant proteins		
DMEM	Corning	Cat #10-013-CM
Fetal bovine serum	Corning	Cat #35-016-CV
penicillin/streptomycin	Gibco	Cat #15140-122
trypsin-EDTA	Gibco	Cat #15400-054
Paclitaxel	Mount Sinai Hospital	N/A
Doxorubicin	Mount Sinai Hospital	N/A
Afatinib (BIBW2992)	Selleckchem	Cat #S1011
Ponatinib	MedChemExpress	Cat # HY-12047
polystyrene microplates	Corning	Cat #3903
EGF	Sigma	Cat #E9644
protease inhibitor cocktail	Roche	Cat #04693132001
Laemmli buffer	Bio-rad	Cat #1610747
chemiluminescent substrate SuperSignal West Pico Plus	Thermo Scientific	Cat #34580
Critical commercial assays		
GeCKOv2	Addgene	Cat #1000000048
Brunello	Addgene	Cat #73179
CellTiter-Glo® Luminescent Cell Viability Assay	Promega	Cat #G7571
Bio-Rad Protein Assay Dye	Bio-rad	Cat #5000001
PCR	Qiagen	Cat #28104
Deposited data		
JQ1 resistant TNBC cell line transcriptome	Shu et al.[11]	GEO: GSE63582
FASTQ files for published snRNA-seq	Kim et al.[14]	Sequence read archive (SRA): SRP114962
RNA-sequencing data of MDA-MB-231 and MDA-MB-468 cells treated with Afatinib, Paclitaxel, and the combined drugs	This paper	Gene Expression Omnibus (GEO): GSE199779
Processed data and codes	This paper	Zenodo DOI: https://zenodo.org/badge/latestdoi/475991227

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
transcriptome of resistant TNBC cell lines after prolonged exposure to Paclitaxel	N/A	GEO: GSE90564
JQ1 resistant TNBC cell line transcriptome	Shu et al.[11]	GEO: GSE63582
Experimental models: Cell lines		
MDA-MB-231	AmericanType Culture Collection (ATCC; Manassas, VA, USA)	N/A
MDA-MB-468	AmericanType Culture Collection (ATCC; Manassas, VA, USA)	N/A
Software and algorithms		
STAR aligner	Dobin et al.[61]	2.5.0a
Scran	Lun et al.[62]	1.10.1
R	Comprehensive R Archive Network (CRAN)	3.6.1
Slingshot	Street et al.[21]	1.2.0
ESTIMATE	Yoshihara et al.[29]	1.0.13
DESeq2	Love et al.[63]	1.20.0
Survival	N/A	2.3-7
Limma	Ritchie et al.[20]	3.46.0
GSVA	Hanzelmann et al.[26]	1.32
Other		
License for BioRender contents used in Figure 1	BioRender	Agreement number: OM24SF3YYA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Won-Min Song (won-min.song@mssm.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: The RNA-sequencing data of MDA-MB-231 and MDA-MB-468 cells treated with Afatinib, Paclitaxel, and the combined drugs are deposited at GEO and publicly available as of the date of publication. Accession numbers are listed in the key resources table.

- Code: All original code and associated processed data have been deposited at Zenodo, and is publicly available as of the date of publication. DOI is listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

MDA-MB-231 and MDA-MB-453 cells were obtained from AmericanType Culture Collection (ATCC; Manassas, VA, USA) and cultured in DMEM (Corning, Cat #10-013-CM) supplemented with 10% fetal bovine serum (Corning, Cat #35-016-CV) and 1% penicillin/streptomycin (Gibco Cat #15140-122). Cells were cultured in a humidified incubator at 37°C with 5% CO_2 and passaged every 3 – 4 days with 0.05% trypsin-EDTA (Gibco, Cat #15400-054).

METHOD DETAILS

Drugs

Paclitaxel and Doxorubicin were obtained from Mount Sinai Hospital. Afatinib (BIBW2992) was obtained from Selleckchem (Cat #S1011). Ponatinib was obtained from MedChemExpress (Cat # HY-12047).

CRISPR/Cas9 knock-out screening

CRISPR genome-wide loss-of-function (LOF) screens were performed using human CRISPR knockout pooled libraries GeCKOv2 (Addgene Cat #1000000048) and Brunello (Addgene Cat #73179) using a well-established protocol.[64] The one-vector system cloned into a lentiCRISPRv2 backbone was selected for both libraries. MCF10A cells were lentivirally transduced with each library at a multiplicity of infection (MOI) of 0.3. 35 million cells were transduced for Brunello and 60 million cells for GeCKOv2. To maintain adequate library representation, a sufficient number of cells were transduced to ensure at least 100X representation assuming ~30% of the cells were successfully transduced. Cells were then selected with 1 μg/ml puromycin for 72h, and passaged for 14 days to ensure an efficient CRISPR knockout was achieved before screen selection was done. For the screens, 2 million cells were plated in 8 x 15cm dishes for a total of 16 million cells. sgRNA library expressing cells were positively selected with 5 cycles of Paclitaxel (PTX) at 7.5nM or Doxorubicin (DOX) at 30nM. Concentrations used reduced cell viability by 85% (PTX) and by 50% (DOX), as determined by CellTiter-Glo® assays. Cells were treated with paclitaxel for 24h before washout, allowed to recover before passaging at least once, and propagated to have enough cells for the next cycle. Each cycle lasted around 7 – 10 days. Genomic DNA was extracted from cells before and after 5 cycles of selection (Agilent, Cat #200600). A 2-step PCR was performed to amplify the sgRNA-containing region of interest and then to add adaptors and barcodes to the amplicons for multiplex deep sequencing on the Illumina HiSeq 2500 platform. For the 1st step, multiple tubes of DNA amounting to a total of 45μg (GeCKOv2) and 30μg (Brunello) were amplified. PCR clean-up was performed (Qiagen, Cat #28104) before gel extraction of the expected band. 60ng of amplified DNA was subjected to 2nd step PCR to add P5 and P7 primers. Both PCR steps were performed using PfuUltra II Hotstart PCR Master Mix following the manufacturer's protocol. The PCR cycling parameters for both steps were as follows:

Temperature	Time	No. of cycles
95°C	2 min	1
95°C	30s	30
53°C	30s	30
72°C	30s	30
72°C	10min	1
4°C	Hold	N/A

The FASTQ sequences of both libraries were separately counted and annotated, and then merged by homemade R scripts. Quality control (QC) and the identification of gene-level hits were conducted using the meta-analysis algorithm ScreenBEAM. sgRNAs from both GeCKOv2 and Brunello libraries were combined for hit identification using the benchmark of Z-score >1.96.

1 replicate was done for each library (positive screening process). 2 independent rounds of sequencing were done. The 1st round had 4 replicates for the parental libraries and 2 replicates for the PTX/DOX experimental groups, where gDNA extraction and subsequent steps were performed independently. The 2nd round of sequencing had 1 replicate for each experimental group (parental, PTX, DOX).

Cell viability

Cell viability was assayed using CellTiter-Glo® Luminescent Cell Viability Assay (Cat #G7571). Cells were seeded at comparable confluency at 10,000 cells/well for MDA-MB-231 and 25,000 cells/well for MDA-MB-453 in 96-well clear bottom, white polystyrene microplates (Corning, Cat #3903) and treated with respective drugs the following day. Cells were treated for 72h before drug-containing media was removed and washed once with 1X PBS. Equal volumes of fresh media and CellTiter-Glo® reagent were added.

Plates were incubated on a rocker for 2 minutes, after which endpoint readings at luminescent settings were performed using a SpectraMax M5 microplate reader (Molecular Devices).

Western blotting

MDA-MB-453 or MDA-MB-231 cells were seeded in 6 cm dishes and grown to near-confluency. Cells were washed with 1X PBS and treated with increasing concentrations of Afatinib (25nM to 2μM) for 24h in 0.5% FBS DMEM and then stimulated with 50nM of EGF (Sigma, Cat #E9644) for 15min before washing with 1X PBS and lysis in 1X RIPA buffer containing a protease inhibitor cocktail (Roche, Cat #04693132001). Protein quantification was performed using Bio-Rad Protein Assay Dye (Bio-rad, Cat #5000001), and samples were prepared in 4X Laemmli buffer (Bio-rad, Cat #1610747) and heated at 95°C for 5min. Proteins were resolved by 8% SDS-PAGE gels, transferred to nitrocellulose membranes, and blocked with 5% BSA in PBS-T for 1h. The membranes were probed with the following primary antibodies according to the manufacturer's recommended dilutions: phospho-Tyr (CST, #9411), phospho-EGFR (Y1173) (CST, #4407), EGFR (CST #2232), phospho-ERK1/2 (CST, #4370S), ERK1/2 (Santa Cruz, sc-94), phospho-Akt (S473) (CST, #3787), Akt (CST, #4691), β-actin (Santa Cruz, sc-47778). Membranes were washed 3 times for 5min each with 1X PBS-T and incubated with HRP-conjugated anti-rabbit or anti-mouse antibodies (GE healthcare, Cat#NA934V, #NA931V) at 1:5000 dilution in 5% milk in PBS-T for 1h. After washing with 1X PBS-T, signals were detected using chemiluminescent substrate SuperSignal West Pico Plus (Thermo Scientific, Cat #34580).

QUANTIFICATION AND STATISTICAL ANALYSIS

snRNA-seq analysis

Data download and alignment to hg38

Following the initial study of single-nuclei transcriptome of 6,862 cells across 8 patients who underwent pre-, mid-, and post-neoadjuvant chemotherapy (NAC)[14], we were able to access a subset of data including paired-end.fastq files for 3,546 cells from Sequencing Reads Archive (SRA) under study id 'SRP114962'. These reads were mapped to hg38 using STAR aligner[61] (–outFilterScoreMinOverLread=0.3,–outFilter-MatchNminOverLread=0.3). As a first-pass filter for quality control, cells with alignment rate <30%, and a number<100,000 reads were filtered out, leaving 3,167 cells for further analysis.

Expression quantification

The mapped transcript counts were called by hg38 Ensembl transcript id, and processed according to the recommended workflow from R package *scran*.[62] Cells were further screened with multiple criteria within the workflow (Figure S1A). Firstly, outlier cells with low library size and endogenous reads with MAD <-3 were filtered out. To determine the mitochondrial rate threshold to filter out apoptotic cells, we tested the enrichment of TNBC chemoresistance signatures across a range of the thresholds (Figure S2), and identified >50% as a reasonable threshold that balances the signature enrichments and the number of remaining cells (Figures S1A and S1C). Overall, 2,849 cells eventually remained, and Transcripts Per Million (TPM) values were calculated for further analyses (see Table 1 for a breakdown of cell numbers). The normalized, log-transformed counts were further adjusted for loading date with systematic batch and mitochondrial rate as biological confounders using *removeBatchEffect()* from limma.[20]

Unsupervised cell clustering

We utilized *scran* (1.10.1) R package for unsupervised clustering of the single-cell RNA-seq data. Unsupervised clustering based on random walk community detection, namely walktrap algorithm,[65] was applied by constructing k-nearest neighbor (kNN) graph in first 10 PCs. With kNN=$\sqrt{N_c}$ (N_c = number of cells), it yielded 10 cell clusters (see Figure 2A). We utilized "*findMarkers()*" function in *scran* R package to identify the cell cluster markers with FDR <0.05 and |log2(overall fold change)| > log2(1.2).

Molecular subtyping of individual cells

We also performed PAM50 subtyping on the single-nuclei transcriptome to classify their molecular characteristics. Using normalized, log-transformed single-cell expression with '*molecular.subyping()*' function from *genefu* R package (v2.16.0),[24] the molecular subtypes of individual cells were called with subtype probability >50%.

Trajectory inference

We utilized R package *slingshot* (v 1.2.0) to infer the trajectories of cancer cells depicting pseudo-temporal transcriptomic changes during chemotherapy. s*lingshot* requires cell cluster assignments and the initial cell cluster as the inputs. To identify the initial cell cluster, we performed slingshot for each patient to track patient-wise temporal dynamics and merged the individually inferred pseudo-times. For each patient, unsupervised cell clustering was performed and clusters with the most cells from pre-NAC were selected as the initial cluster for the lineage inference (Figures S4A andS4B). These pseudo-times were then summarized per cell cluster in the merged data, and the cell cluster with the overall minimum pseudo-time was selected as the initial time point to infer the overall cell lineages. The gene expression correlating to the lineages was identified by Spearman;s correlation analysis using *cor.test()* function in stat R package. Among the significantly upregulated genes with FDR <0.05, the top 500 genes were chosen for pathway/function enrichment analysis by Fisher's Exact Test (FET). The results from Spearman correlation and FET are reported in Table S3.

Chemoresistance signature collection

We collected TNBC chemoresistance signatures from single nuclei RNA-seq, bulk data (METABRIC), and *in vitro* models of TNBC. A detailed description of statistics and signature summaries can be found in Table S4.

Chemoresistant signatures from scRNA-seq

We analyzed differentially expressed genes (DEGs) by comparing cells from chemoresistant (R) or-sensitive (S) patients, and post- or pre-neoadjuvant chemo-treatments (post-/pre-NAC). Firstly, NAC response signatures within the chemoresistant patients were first identified within each patient using *DESeq2* (version 1.20.0)[63] by contrasting cells from post-NAC with pre-NAC (R: Post-vs- Pre). After obtaining up- or downregulated signatures |fold change| (|FC|) > 3 and FDR <0.05 within each patient, we intersected the signatures across the patients to finalize the signature, yielding 446 and 1,150 up-and down-regulated genes respectively. Next, we identified time point-specific chemoresistance signatures by contrasting cells from chemoresistant patients with those from chemo-sensitive patients within post-NAC or pre-NAC cells (Post:R-vs- S or Pre:R-vs- S). For Post:R-vs- S comparison, post-NAC cells across chemoresistant patients were aggregated and compared against post-NAC cells across chemosensitive patients by *DESeq2* pipeline with |FC| > 3 and FDR <0.05 (8,611 and 528 up- and down-regulated genes respectively). Similarly, for Pre:R-vs- S comparison, there were 3,528 and 583 up- and down-regulated genes respectively. The stringent fold change threshold was due to the inflated DEGs with lower fold changes. For instance, |FC| > 2 with FDR <0.05 yielded over 6,000 up-regulated genes in chemoresistant cells compared to sensitive cells. As single-cell RNA sequencing data are known to be noisy and dependent on distinct cell populations, we applied |FC| > 3 to further eliminate false positives emerging from such noisy data.

Further, we identified chemoresistance signatures within each cell cluster by contrasting cells from the chemoresistant patients with those from chemosensitive patients with |FC| > 2 and FDR <0.05.

Chemoresistant signatures from in vitro experiments

We searched Gene Expression Omnibus (GEO) and identified a Paclitaxel chemoresistance signature by contrasting acquired chemoresistant TNBC cell lines against the respective parental cells across multiple TNBC cell lines (BT20, SUM149, MDA-MB-231, MDA-MB-436, and MDA-MB-468) under accession number GSE90564. The DEGs were called with (FC > 1.2 or FC < 1/1.2) and FDR <0.05 using limma.[20] Similarly, the resistance signatures of BET bromodomain inhibitor, JQ1, were obtained by contrasting TNBC cells (SUM149 and SUM159) with acquired resistance against the parental cells (GEO accession: GSE63582).[11] Two types of exposure, DMSO and JQ1, were considered separately per cell line, and this led to cell line and exposure-specific JQ1 resistance signatures.

Sample clustering by cell cluster markers and immune/stromal cell abundances

We sought to calculate sample-wise enrichment of the tumor cell cluster markers from chemoresistance promoting trajectories (trajectory-I &-III), in the 288 and 157 TNBC samples in METABRIC[27] and TCGA,[28] respectively. Overall, computational workflow consists of three major steps i) tumor cell cluster marker identification as the gene signature, ii) identification of stromal/immune contents low bulk samples to infer

the tumor cell signature enrichments, then iii) Gene Set Variation Analysis (GSVA)[26] to evaluate the tumor cell signature enrichments. The overall computational workflow is illustrated in Figure S6A.

We chose GSVA over current state-of-art cell type deconvolution methods such as CIBERSORT,[66] xCell,[67] or MuSiC[68] to infer the tumor cell abundances in the bulk transcriptomes. These cell deconvolution methods assume gene signatures for all cell types are available to infer the cell type compositions in the bulk data, but Kim-snRNA-seq data lacked major immune and stromal cell populations (Figure 2B), and were insufficient to apply the cell type deconvolution analysis. Alternatively, we postulated that tumor samples with low stromal/immune contents are more suitable to leverage the tumor cell signatures from the Kim-snRNA-seq data, encoding aberrantly activated tumoral pathways in acquired chemoresistance. Thus, we utilized Gene Set Variation Analysis (GSVA) to evaluate the sample-wise enrichments of the tumor cell signatures and inferthe degree of chemoresistance driving pathway activation.

To extract the tumor cell signatures from the Kim-snRNA-seq data, we performed the marker detection analysis in scran with "*findMarkers()*" function[62] and extracted the top 500 marker genes showing the largest expression fold changes with FDR <0.05 per tumor cluster. We note that the top 500 genes were chosen uniformly to all tumor cell clusters so that the enrichment of the equally sized gene signatures allow for comparisons across different tumor cell clusters, in contrast to the marker signatures with different sizes by numerical thresholds (e.g. expression fold change). We also compared the GSVA results by using top 250 markers to evaluate the robustness of the findings (Figure S5). Overall, we repetitively observed that findings from using the top 500 markers in the results from using the top 250 markers. These include poor prognosis of patients showing enrichments of C3, C4, and C9 signatures (cluster 4 and 6, Figure S10A), analogous to cluster 1 in Figure 3A, and good prognosis of patients showing the absence of C9 signatures (cluster 2, Figure S10A), analogous to cluster 4 in Figure 3A. With the robustness of the results in different numbers of top cluster markers, we chose the top 500 marker genes for the subsequent analysis of METABRIC and TCGA bulk transcriptome cohorts.

To choose the bulk samples with low immune/stromal contents, we utilized 'Estimation of STromal and Immune cells in MAlignant Tumours using Expression data' (ESTIMATE) method[29] to infer the stromal and immune cell abundance, where a lower ESTIMATE score indicates higher tumor purity. Then, we selected the samples below the mean ESTIMATE scores in METABRIC (Figures S6B andS6C), and TCGA (Figures S6D andS6E). For METABRIC cohort, the mean was calculated by modeling the ESTIMATE score distribution with bimodal normal distribution as the distribution was not fully explained by the unimodal model (Figure S6B) by Shapiro-Wilk test (p=1.34E-4). The mean value from the dominant peak was used as the threshold to identify low stromal/immune content samples (Figure S6B). For TCGA cohort, the distribution was normally distributed (Shapiro-Wilk p=8.11E-1), hence the mean of the unimodal model was used as the threshold (Figure S6E).

To calculate the enrichment scores, we performed GSVA by running "gsva()" command with method="gsva" parameter in GSVA R package[26] on the stroma/immune-low samples. The "gsva" method calculates the normalized enrichment score within the range of [-1,1], where the enrichments are captured in (0,1], depletion in [-1,0), and 0 corresponds to no patterns of enrichment or depletion in individual samples.

Then, we performed sample clustering on stroma/immune-low samples by grouping samples with similar enrichment patterns of the tumor cell marker genes. The sample-to-sample distance was calculated by correlation distance, $d_{ij} = \sqrt{2(1 - \rho_{ij})}$, ρ_{ij} = Pearson's correlation, a metric distance with d_{ij}=0 if ρ_{ij}=1, d_{ij}=2 if ρ_{ij}=1. This was followed by hierarchical clustering with complete linkage distance.[69] The adequate number of clusters (k) was inspected by checking for common elbows in k-vs- average silhouette width curve, and k-vs- adjusted Dunn's index[69] in k ranging in [2,10] (Figures S4F andS4G).

Finally, the prognostic significance of the clustering results was evaluated by testing if the survival curves in different sample clusters (hence patient clusters) are significantly different via G^ρ family of tests[70] implemented in "survdiff()" function (from the survival R package).[71] Then, the individual clusters associated with poor or good prognoses were identified by Mantel-Haenszel chi-square statistic,[71] measuring the deviation of the observed number of events (i.e. death or disease recurrence) from the expected number of events.

Survival signatures from METABRIC

Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) included the largest TNBC cohort of 286 bulk samples with disease-free progression follow-up, and we sought to identify prognostic gene signatures from METABRIC transcriptome data. Although disease-free progression itself is not a direct indicator of chemoresistance, it may serve as a sub-optimal indicator as chemotherapy is the standard-of-care option for diagnosed patients, and relapse within 3-5 years is the major risk factor leading to poor prognosis.[72] Within the chemoresistant transcriptome, we postulated that both intra-tumoral pathways and tumor micro-environment with varying stromal or immune cell abundance are the confounding factors.

We hypothesized that tumor-purified (TP) samples would show distinctively low ESTIMATE scores, in contrast to samples with high immune/stromal micro-environment contents (HTME) showing high ESTIMATE scores. Thus, we modeled the ESTIMATE scores from METABRIC cohort with bimodal distributions. The fitted bimodal distribution yielded two peaks, namely: a high tumor purity peak (μ_{TP}, red peak, Figure S12B), and a high stromal/immune contents peak (μ_{HTME}, green peak, Figure S12B). We then stratified the samples according to the peak locations. TP samples were defined as low ESTIMATE score samples with $< \mu_{TP} + 3\sigma_{TP}$ (σ_{TP}: standard deviation of the red peak in Figure S12B). Similarly, HTME samples were defined as high ESTIMATE score samples with $> \mu_{HTME}$. We classified the intermediate samples with $\mu_{TP} + 3\sigma_{TP} <$ ESTIMATE score $< \mu_{HTME}$ as tumor-enriched (TE) samples. Overall, we classified tumors into tumor purified (TP) with low ESTIMATE scores, tumor enriched (TE) with intermediate ESTIMATE scores, and high tumor micro-environment (HTME) with high ESTIMATE scores to further guide the sample clustering analysis.

We classified the samples into TP, TE, and HTME groups, and Cox proportional hazard (Cox PH) model[73] was used within each group to evaluate the correlation between individual gene expression and disease-free prognosis. Furthermore, samples were stratified by median values of individual gene expression and tested for the significant difference in disease-free prognosis between expression-high and –low groups. Overall, gene expression associated with good prognosis was defined as Hazard ratio <1 with Cox p-value <0.05 by Cox PH and log-rank p < 0.05, and gene expression associated with poor prognosis was defined as Hazard ratio >1 with Cox p-value <0.05 by Cox PH and log-rank p < 0.05.

Deriving consensus signature of TNBC chemoresistance

We interrogated NAC-treated single-nuclei transcriptome,[14] the transcriptome of resistant TNBC cell lines after prolonged exposure to Paclitaxel (GEO accession: GSE90564) or JQ1 (GEO accession: GSE63582)[11] and survival-associated gene signatures in METABRIC cohort (see chemoresistance signature collection in STAR Methods).

Then, we identified robustly activated pathways in chemoresistant tumors (i.e., consensus signatures) which a synthetically lethal drug could then target with a conventional chemo-regimen, Paclitaxel. We reasoned that these pathways should overlap across multiple chemoresistance signatures due to their robustness, and yet manifest slightly differently from study to study due to noise. To account for this, we evaluated the overlapping genes co-occurring across the pool of collected chemoresistance signatures to detect the adequate co-occurrence threshold balancing the signature robustness over the noise.

In the co-occurrence-vs- overlap size curve (Figure S13A), the co-occurrence of 4 emerged as the elbow point, indicating this as the optimal co-occurrence threshold to derive the overall consensus signature (red line, Figure S13A), and yielded 1,524 consensus genes. The overlap between the consensus and individual seed signatures is shown in Figure 4A.

RNA-sequencing data analysis of treated and untreated MDA-MB-231 cells

The 150bp paired-end RNA-sequencing data were generated from Illumina NovaSeq platform. The reads were mapped to hg38 by STAR aligner (v2.7.5b),[74] and the gene counts for fragments mapped into exonic regions were called for hg38 ensembl genes by featureCounts in subread package (v2.0.1, parameters:-t exon-g gene_id-p).[75] Overall, uniquely mapped reads ranged from 34 to 54 million reads (Table S6A). The gene counts were subsequently normalized by Counts Per Million (CPM), followed by Trimmed Mean of M (TMM) scaling normalization.[76] t-distributed stochastic neighbor embedding (t-SNE) of the normalized expressions revealed a distinct separation between two groups of replicates (rep 1 & 2 against

rep 3 & 4), regardless of drug treatments (Figure S14B), and these groups coincided with sample RNA concentrations. To evaluate the impact of RNA concentration, we adjusted each gene expression for RNA concentration by utilizing 'lm()' in R (v3.6.1), and the adjustment gathered each treatment group into distinct clusters in the adjusted t-SNE plot (Figure S14C). Thus, we treated RNA concentration as the confounding variable in the subsequent differential expression analysis between each drug treatment group (Paclitaxel, afatinib, and a combination of both) and DMSO-treated samples by performing limma (v3.46.0).[20]