Transcriptional Characterization of Resistance in Early Drug Response

Cole Ruoff[1], Allison Mitchell[2], Priya Mondal[2], Vishaka Gopalan[1], Michael Gottesman[2], Sridhar Hannenhalli[1*]

[1]Cancer Data Science Lab, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

[2]Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

[*]Corresponding author: Sridhar.hannenhalli@nih.gov

Abstract

Therapeutic resistance is the leading cause of treatment failure and death from cancer. While resistance can be driven by genetic mutations, mounting evidence also points to an epigenetic basis of resistance. Much of this epigenetic, or non-genetic, resistance has been attributed to drug-resistant transcriptional cell states that are either induced by drug treatment or pre-exist in a fraction of cells selected by treatment. However, the extent to which long-term resistance is manifested in the early inherent cellular response to drugs is poorly understood, and which aspects, if any, of this early response to drug-induced transcriptional response are evolutionarily conserved properties of cells. To address these questions, we integrate datasets of long-term drug resistance and early drug response data across multiple cell lines with drug response and resistance data from bacteria and yeast. Our findings suggest that cancer cell states in both the drug-naive populations as well as in populations shortly after treatment share transcriptional properties with fully established resistant cell populations, and CRISPR-cas9 knockout of

transcription factors predicted to regulate the resistant transcriptional programs result in inceased drug sensitivity. Furthermore, the resistance states manifested as early drug response are evolutionarily conserved. Finally, we show that early resistant states discriminate responders from non-responders across multiple human cancer trials.

Introduction

While numerous chemotherapeutic, targeted, and monoclonal antibody-based drugs are approved for various cancers, resistance to these drugs invariably emerges resulting in relapse and eventual death of the patient. Indeed, therapeutic resistance is the leading cause of treatment failure and cancer mortality (Ramos et al., 2021). Several mechanisms driving resistance have been identified (Lei et al., 2020). While mutation-driven resistance has been the predominant paradigm thus far (Nussinov et al., 2021), more recent evidence points to an epigenetic, or transcriptional basis of resistance, whereby resistance results from a subpopulation of malignant cells occupying drug resistant transcriptional states that allow for their survival through the treatment (Frank and Rosner, 2012; Bell et al., 2019). Pre-existing, reversible resistant cells are referred to as drug tolerant persister (DTP) cells (Sharma et al., 2010) distinguishing them from the emergent, stable, resistant states (Marine et al., 2020).

Recent literature has identified and characterized these resistant cell states (Goyal et al., 2023; Oren et al., 2021) but much is unknown about the timeline in which these states develop. For instance, it is not clear whether, and to what extent, long-term resistance is evident soon after drug treatment. Furthermore, these immediate drug responses leading

up to long-term drug resistance may be an inherent property of cells shared with unicellular ancestors (Bigger, 1944). Here, we investigate whether (i) there is a similarity between early transcriptional response to drugs and long-term resistance, (ii) these responses are conserved in evolution, and (iii) this information has clinical implications.

Toward these aims, we obtained a large scRNA-seq dataset consisting of three well characterized cancer cell lines (A549, K562, MCF7) treated with 188 different compounds at 24 hours post-treatment (Srivatsan et al., 2019). These data represent the early transcriptional response to drugs. We then compared the single cell clusters in this early response dataset with transcriptional signatures of resistant cells derived from published studies of long-term resistance across multiple cell lines and drugs (Goyal et al., 2023; Shaffer et al., 2017; Oren et al., 2021). We additionally assessed the evolutionary conservation of early response to drugs based on previously published transcriptional response datasets in the bacterium *E. coli* (Bie et al., 2023) and the fungus *C. auris* (Zhou et al., 2021).

Our findings suggest that transcriptional states associated with long-term resistance are manifested early on after the drug treatment; across all three cell lines investigated here, we observed transcriptional states similar to resistant states. Furthermore, our analyses suggest that these early harbingers of long-term resistance may exist prior to drug treatment as well as being induced by drug treatment. These pre-existing and early drug response states share characteristics across the cell lines, including higher ribosome activity, EMT, and MYC activity. The gene signatures of early resistant states were also

found to have prognostic value in multiple cancer types and are up-regulated in patients that did not respond to therapies. Interestingly, we also found that similar states emerge in *E. coli* and *C. auris,* after multiple drug treatments, reflecting that this mode of resistance is an evolutionarily conserved mechanism.

Results

Multiple cell states in pre-treatment and early response data express long-term resistance programs

To identify and characterize early post-treatment manifestation of long-term drug resistance, we first created a robust gene signature of long-term resistance. To this end, we compiled a drug resistance gene signature composed of 882 common upregulated genes (Table S1) in four different drug-resistant samples drawn from previously published studies devised specifically to characterize cancer samples with fully established drug resistance (Goyal et al., 2023; Shaffer et al., 2017) (Methods). The long-term resistance experiments were done in breast cancer and melanoma cell lines, and thus the derived gene set represents a general drug resistance program. We validated the 882 gene signature in an additional dataset comprising multiple timepoints throughout resistance development in yet another cancer type -- lung adenocarcinoma (Oren et al., 2021). Using gene set enrichment analysis, we calculated the enrichment of the resistance signature among the upregulated genes at each time point (Methods). Encouragingly, we saw a strong enrichment at day 14 (Figure 1a), thus supporting the broader validity of the resistance signature. We thus further refined the 882 gene resistance signature by

retaining the upregulated genes from day 14 in the lung adenocarcinoma study resulting in a high-confidence, broadly applicable, 97-gene resistance signature (Methods; Table S1) that we utilized in downstream analyses.

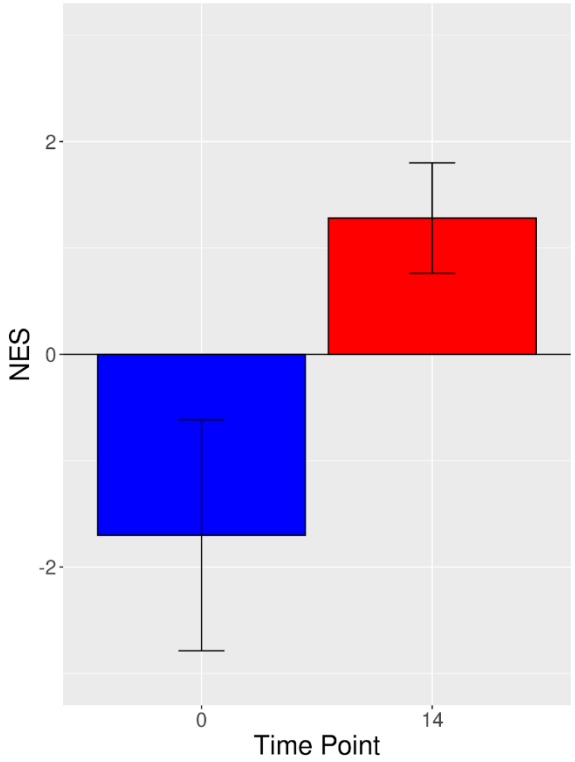

Figure 1a: Enrichment of resistance signature along time points. Barplot showing the initial drug resistance signature gene set enrichment score at day 0 and day 14 time points in the Oren et al. (2021) dataset. NES: Normalized Enrichment Score.

Next, we turned to a large scRNA-seq dataset containing drug-naive cells as well as cells 24 hours post-treatment, in three well-characterized cancer cell lines (A549, K562, MCF7) treated with 188 different compounds (Srivatsan et al., 2019). We downloaded the dataset of three single cell RNA-seq samples grouped into 49 transcriptional clusters. Independently, in each of the three cell lines, we assessed the relative expression of the

resistance signature in each cluster, using AUCell to get the raw signature score in each cell (Aibar et al., 2017) (see Methods). We observed that in each cell line, some of the transcriptional clusters had significantly higher resistance signature score distributions (Figure 1b). This suggests either pre-existence of the resistance signature expression or emergence within 24 hours of drug treatment.

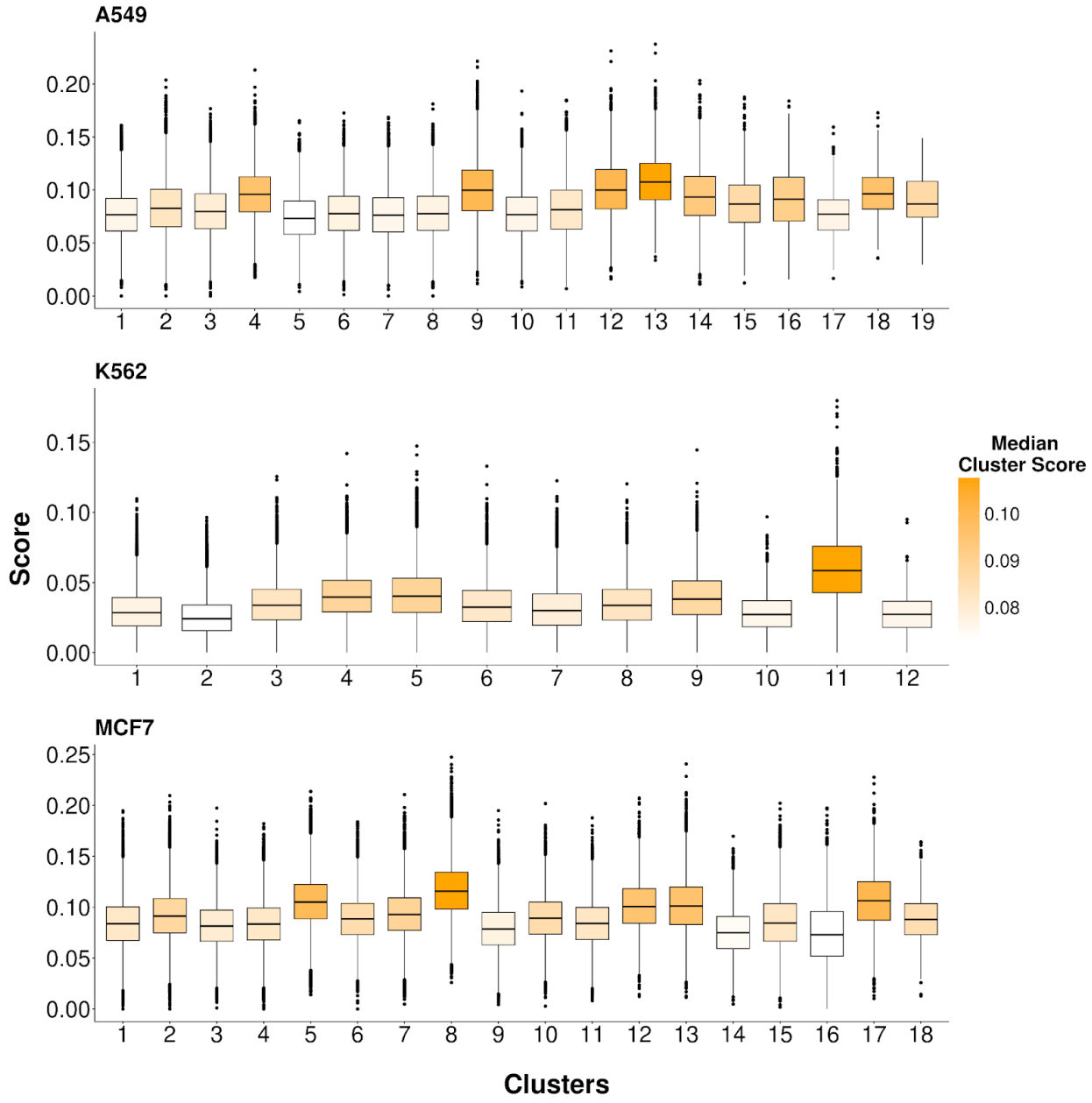

Figure 1b: Resistance signature AUCell scores in cell line clusters. Raw AUCell score distributions for each cluster in each cell line. Each cell was scored with the refined drug resistance signature and the resulting distributions are shown as boxplots. The y-axis represents the raw AUCell score and each boxplot represents a transcriptional cluster. The color of a box plot represents the median raw AUCell score of that cluster.

Next, to identify the cell states that may represent the early stages of resistance, based on the AUCell score threshold (see Methods) to distinguish active from inactive cells for the resistance gene signature, we calculated the relative enrichment of active cells in each cluster relative to all other clusters of the same cell line as an odds ratio (see Methods; Figure 1c). At an odds ratio threshold of 1.5 and a p-value threshold of 0.05, we identified multiple clusters in each cell line with markedly higher fractions of active cells compared to the baseline across the entire cell line. We termed them Resistance Activated Clusters (RAC).

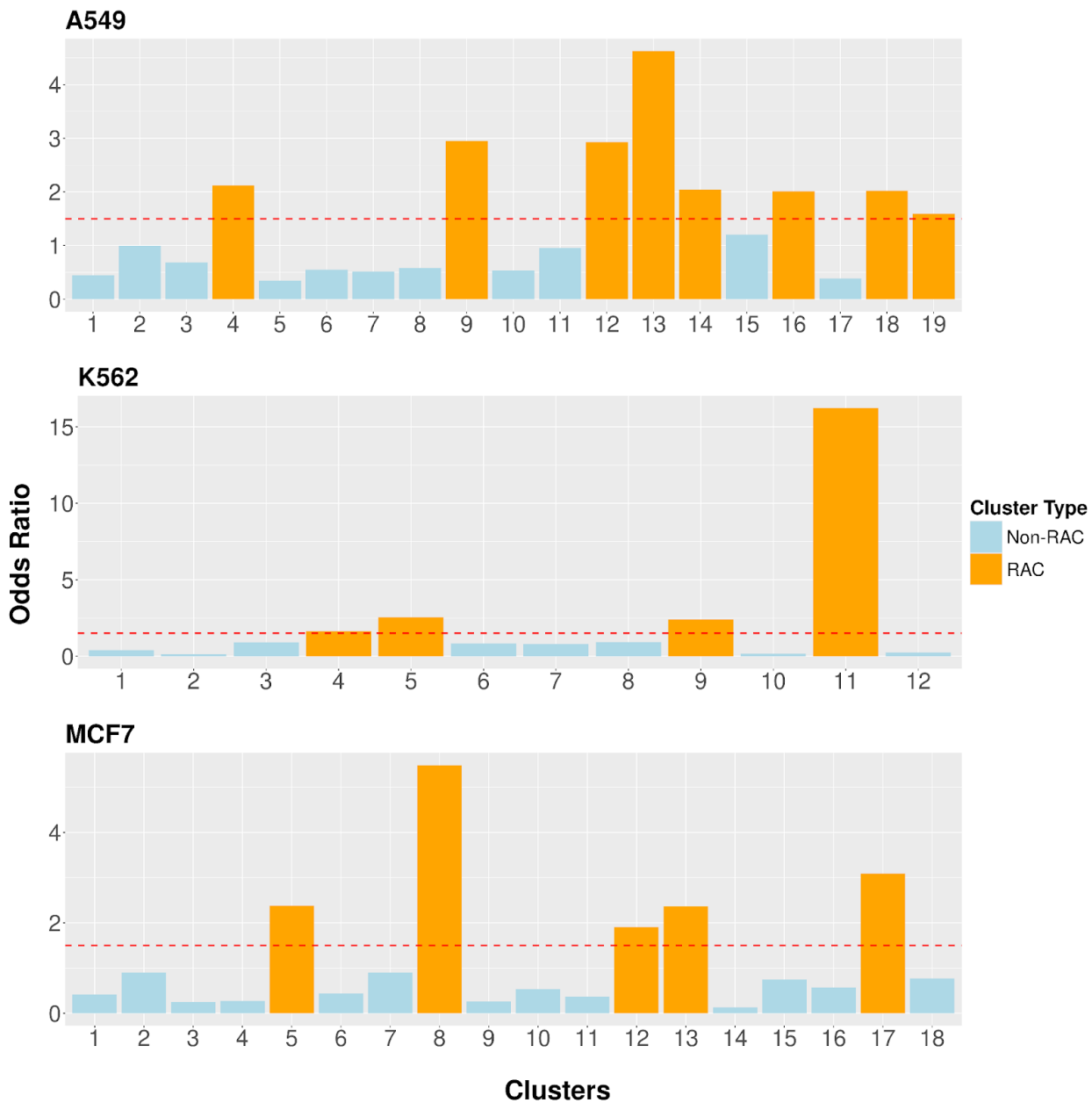

Figure 1c: Resistance-active to inactive cell odds ratio of each cluster in each cell line. The y-axis represents the odds ratio value and each bar represents a transcriptional cluster. Any cluster with an odds ratio greater than 1.5 and a p-value less than 0.05 is considered a Resistance Activated Cluster (RAC) and is colored orange, while the remaining clusters are non-RACs and are colored blue.

We further assessed whether the relative fraction of resistant cells (quantified using the odds ratio as above) changed with the drug dosage. Indeed, in many RACs, we observed an increase in the odds ratio as the drug dose increased (Figure 1d), suggesting a dosage effect on the relative fitness of resistant cells, consistent with previous reports (França et al., 2022).

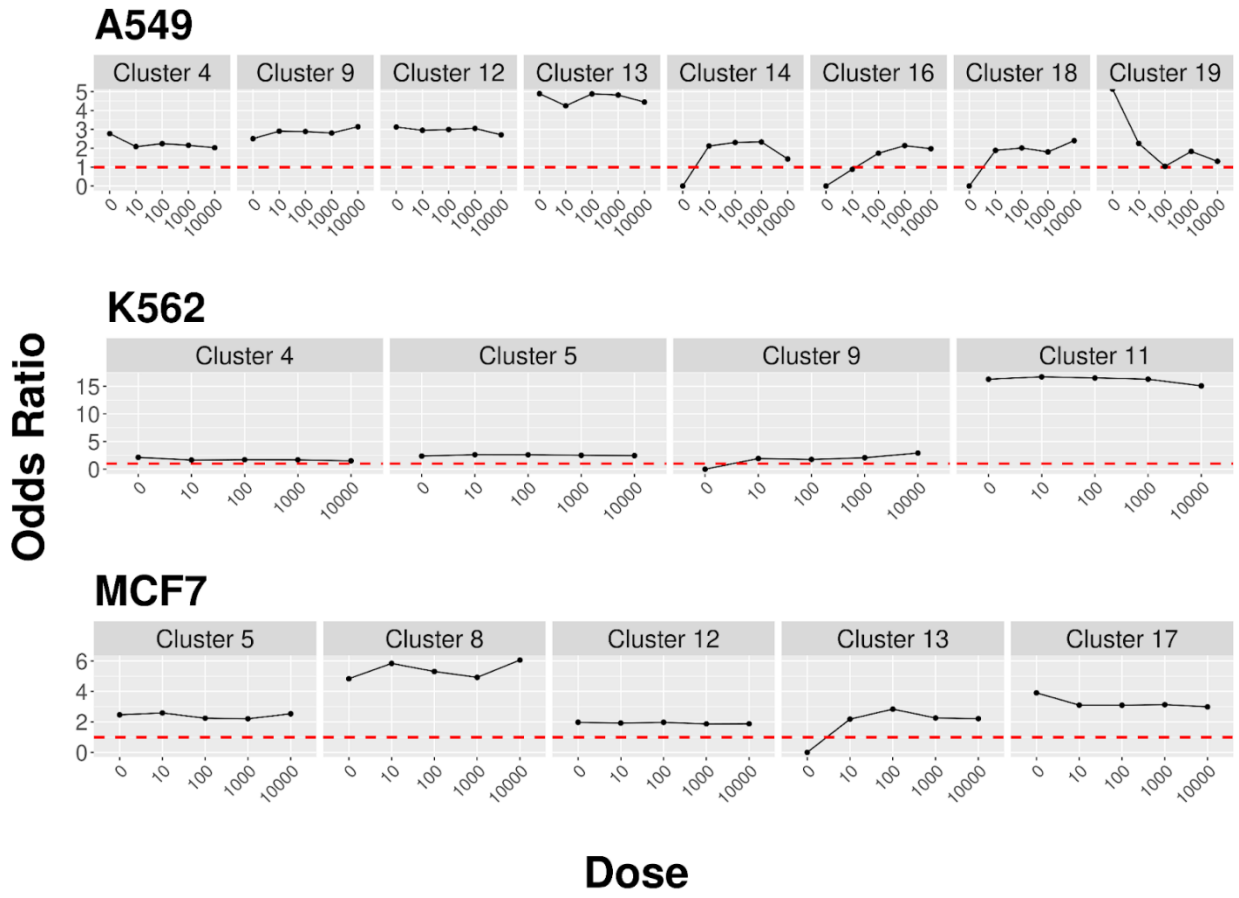

Figure 1d: Resistant active to inactive odds ratio of each RAC in each cell line across the four doses of drugs. The y axis represents the odds ratio. Each panel represents a RAC and within each panel the four points in the line represent the increasing doses of drug treatment.

Furthermore, we noticed that in each cell line some of the transcriptional clusters already had high levels of resistance signature expression even before the treatment, consistent with pre-existing resistance (Figure S1). On the other hand, many states are emergent, meaning they only exist after the drug treatment suggesting acquisition of resistance. These data clearly indicate that multiple cell states in pre-treatment and early drug response stages express resistance programs similar to those observed in long-term resistant samples.

Resistance activated clusters are enriched for resistance pathways conserved across cell lines

Having identified RACs in each cell line, we aimed to characterize them functionally. We first scored all cells in each cell line with cancer hallmark signatures (from MsigDB; Liberzon, 2015) using AUCell and plotted the mean score of each hallmark for each cluster (Methods; Figure 2a).

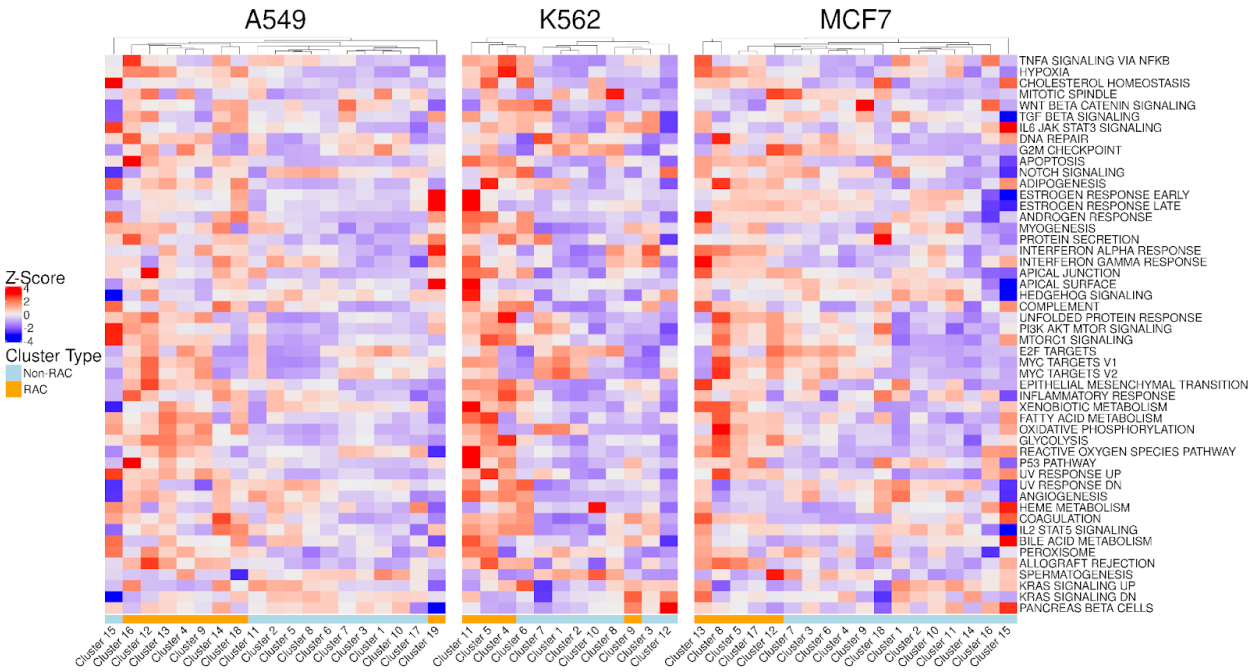

Figure 2a: Cancer Hallmarks mean AUCell score for all clusters in each cell line. The rows represent the different hallmark genesets and the columns represent each transcriptional cluster, separately for each cell line. The colors in the bottom row signify which cluster is a Resistance Activated Cluster (RAC) (orange = RAC, blue = Non-RAC).

Overall, the hierarchical clustering of clusters based on their hallmark scores separated RACs from non-RAC clusters (Figure 2a). However, we also observed subtle differences across RACs in the same cell line, consistent with previously reported distinct, functionally diverged, resistant states (Goyal et al., 2023). Notably, in each cell line, there are both RACs with high enrichment and low enrichment of cell cycling signatures, consistent with previously noted cycling variability among resistance states (Oren et al., 2021). Interestingly, across all three cell lines, virtually all RACs have high activation of genes associated with hypoxia and epithelial to mesenchymal transition (EMT). We repeated this analysis with additional genesets, known as intratumor heterogeneity (ITH) meta-

programs -- transcriptional programs that are commonly expressed across malignant cells in patient tumors (Gavish et al., 2023). The results of this analysis showed similar enrichment as the cancer hallmark analysis (Figure S2a). These characteristics of the RACs further suggest their role in the early stages of drug resistance; hypoxia is known to directly cause drug resistance in cancer by decreasing drugs' cytotoxic ability (Jing et al., 2019), and both hypoxia and EMT indicate increased cellular plasticity enabling these cells to survive drug treatment through adaptive responses (Song and Faber, 2019; De Las Revis at al., 2021).

As a complement to the above analysis, in each cell line, we determined the differentially expressed genes between all RAC cells and all other cells using the Seurat FindAllMarkers function (see Methods), and in the marker genes, assessed for enrichment of Cancer Hallmarks, ITH meta-programs, and various GO biological processes and pathways (Methods; Figure S2b). Consistent with the above analysis, across all three cell lines, we observed common enrichment for EMT. We specifically observed enrichment for the meta-program EMT-III which includes both mesenchymal and epithelial markers, consistent with a hybrid cellular state. These characteristics all have previously been shown to contribute to the development of drug resistance (Celià-Terrassa and Kang et al., 2024).

Other drug resistance-associated pathways enriched in at least two of the cell lines include MYC signaling, which is known to promote plasticity (Torborg et al., 2022). Also, MTORC1 signaling and oxidative phosphorylation are both associated with drug

resistance in cancer (Donati and Amati, 2022; Pan et al., 2023; Zhao et al., 2023). The enrichment of these pathways further suggests that RAC cells may represent the early stages of drug resistant cell states.

RACs Shared Across Cell Lines.

Encouraged by the functional consistencies among RACs within each cell line, we next aimed to identify pan-cell line resistance states. We hierarchically clustered the RACs across all three cell lines based on a gene expression similarity metric (see Methods) which yielded two highly correlated groups spanning RACs from all three cell lines; we refer to the two pan-cell line RAC groups as superclusters (Figure 2b). To further probe the pan-cell line conserved function of the superclusters, we constructed consensus gene sets representing each supercluster by selecting the genes that are upregulated in all three component cluster signatures for the supercluster (Table S2). Functional enrichment analysis of these genesets showed both distinct and shared enriched transcriptional programs (see Methods; Figure 2c). Supercluster 1 is enriched for androgen response, which has been implicated in therapy resistance in some cancer types (De Amicis et al., 2010). Cancer hallmarks enriched in supercluster 2 included MYC targets, E2F targets, and MTORC1 signaling. While both superclusters are enriched for EMT, a notable difference among the superclusters is the varying proportions of cell cycle phases (Figure 2d). This suggests variable proliferation rates across resistant states, previously attributed to differences in metabolism (Oren et al., 2021). Specifically, supercluster 1, exhibiting the largest fraction of cells in G0 phase, is composed entirely of emergent cell states, i.e., with no representation in pre-treatment samples. This may

suggest that cell state transitions induced by drug treatment lead cells to enter a more quiescent state (Kluska et al., 2023).

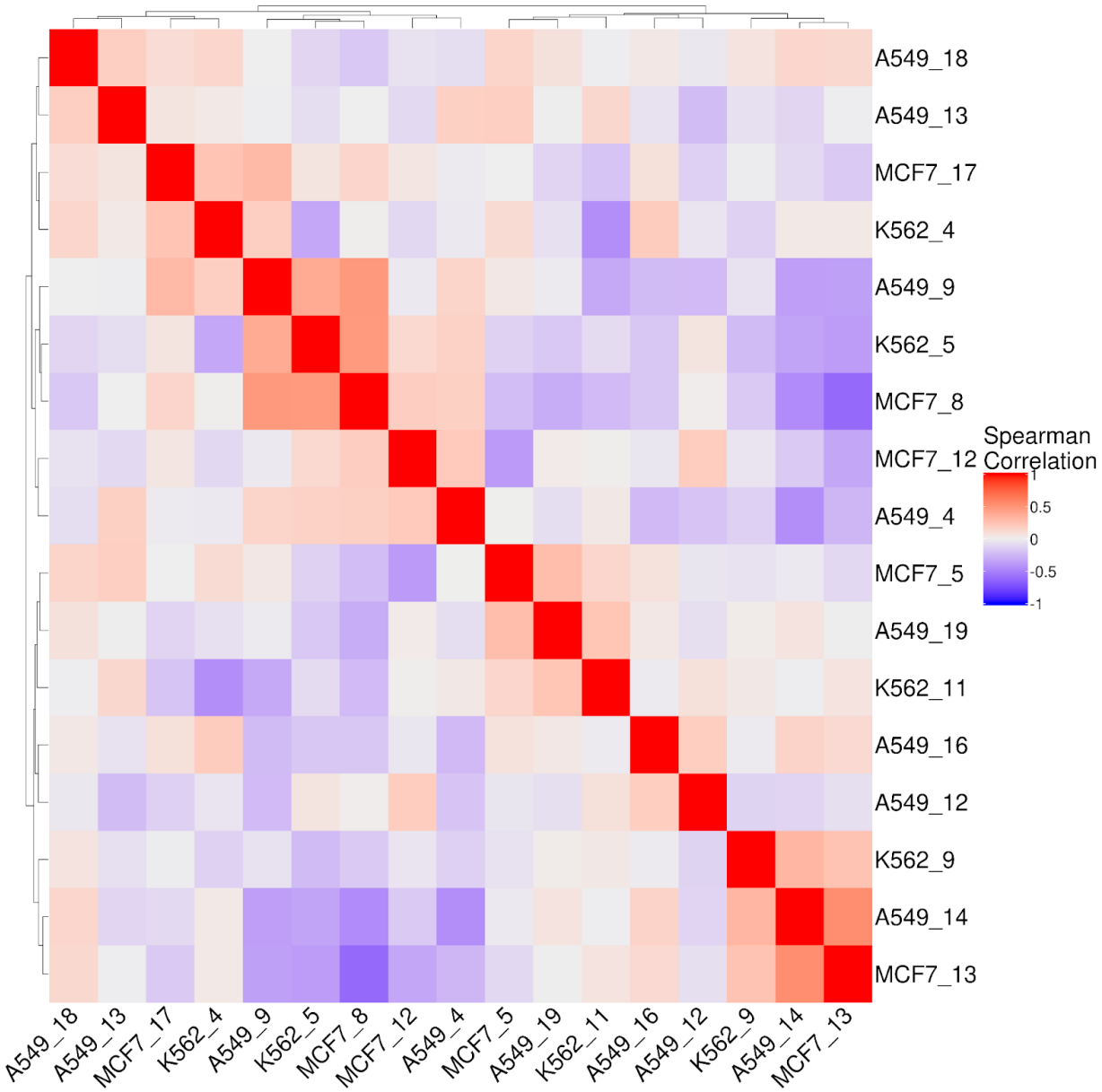

Figure 2b: Correlations of RACs Differential Mean Expression of Most Variable Genes. Heatmap displaying the hierarchical clusters of an all by all correlation matrix based on the Spearman correlation of the cluster differential mean expression vectors (see Methods).

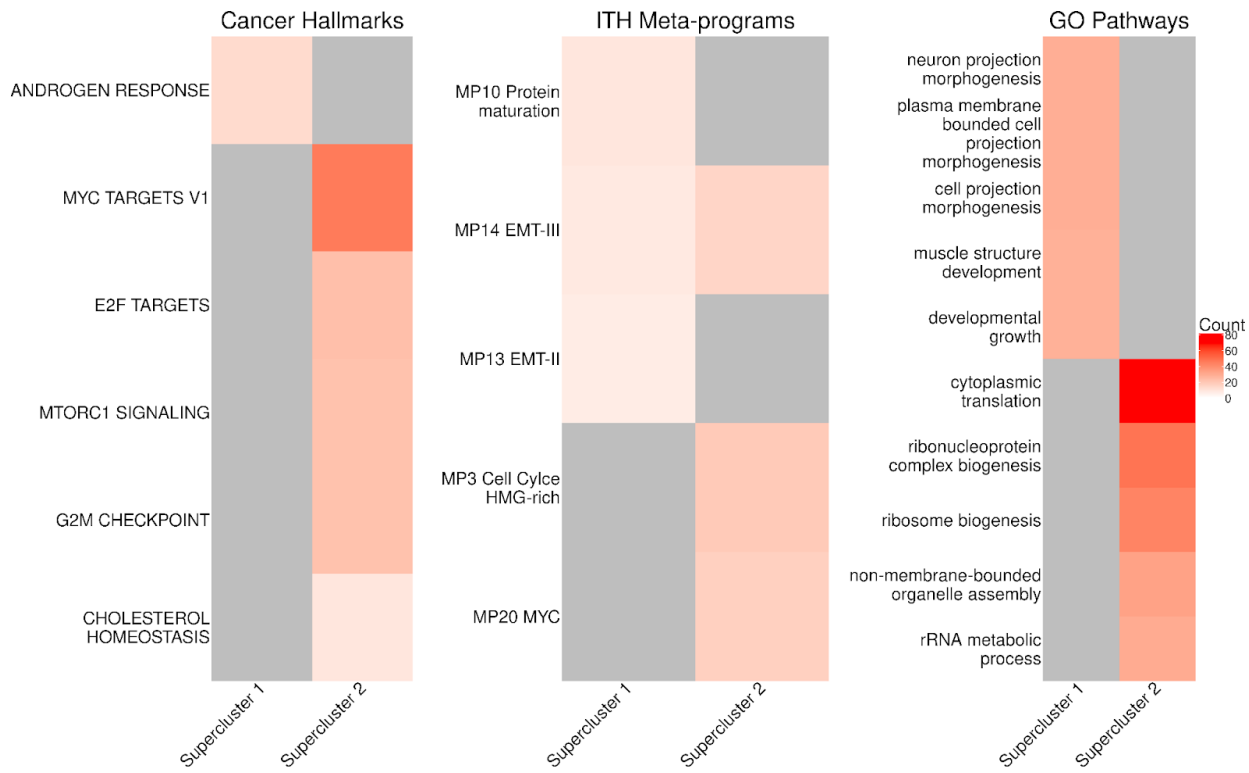

Figure 2c: Supercluster Signatures Enrichment. The heatmaps from left to right represent cancer hallmarks, ITH meta-programs, and GO pathways respectively. Within each heatmap, the columns represent the two superclusters

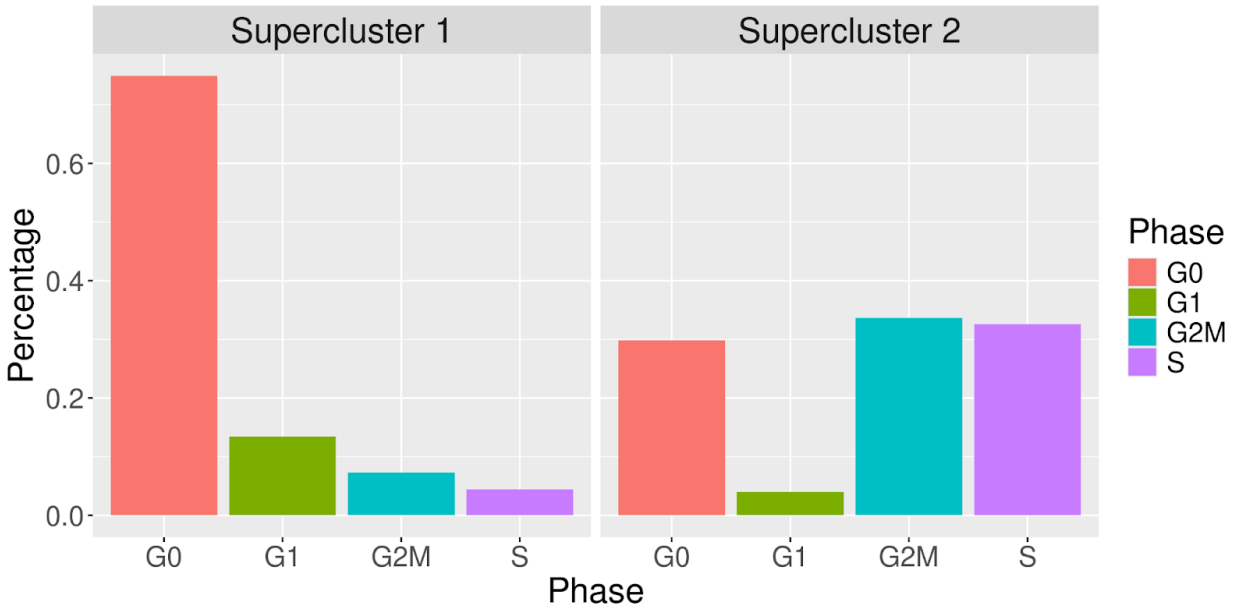

Figure 2d: Superclusters Cell Cycle Phase Percentages. Each panel represents the percentage of total cells in each cell cycle phase for each supercluster. (G0 = red; G1 = green; G2M = blue; S = purple)

Cell Line Specific Properties of Superclusters.

Although our primary interest is in the conserved pan-cell line properties of resistance, we also investigated the cell line-specific aspects of the two superclusters (Figure S2c). In terms of supercluster 1, there were few cell line-specific properties. Among the few differences were an enrichment for hypoxia and EMT in MCF7 and slight differences in GO pathway enrichments relating to cell mobility and wound healing response. On the other hand supercluster 2 had more cell line distinct properties. Although all three cell lines were enriched for hallmarks relating to proliferation, each cell line was enriched for different cell cycling pathways. For example, K562 was enriched for G2M pathways while MCF7 displayed enrichment for G1/S pathways. Overall, this reveals that the

superclusters have few differences among them in each cell line and mostly share the same properties.

Drug Class Specific Superclusters.

Recall that our primary dataset includes responses to 188 drugs from 15 classes of drugs. We therefore assessed the associations of the superclusters with specific drug classes. To do this, in each cell line, we calculated the enrichment of cells treated with each drug class in the RACs of each supercluster relative to the respective cell line as a whole. We observed that supercluster 1 is largely composed of cells treated with drugs targeting epigenetic regulation, while supercluster 2 contains cells that are treated with various drugs (Figure 2e), suggesting that drugs may be associated with specific resistance states. This led us to determine if these drug class associated resistance states increase in proportion across dose. We completed a similar analysis to detect the relative growth of each supercluster's component RACs as the treatment intensity increased. We found that, across all three cell lines, the proportion of cells in supercluster 1 does in fact increase with the treatment intensity by drugs targeting epigenetic regulation pathways. As for supercluster 2, not one singular drug class results in a clear increase in proportion. The relative proportion among the drugs is about the same, indicating that supercluster 2 represents resistance mechanisms induced by various drugs (Figure S2d). These findings lead us to believe that some drugs may induce the same resistance state regardless of the cancer type meaning that the transcriptional response is largely dictated by the treatment rather than the cancer type.

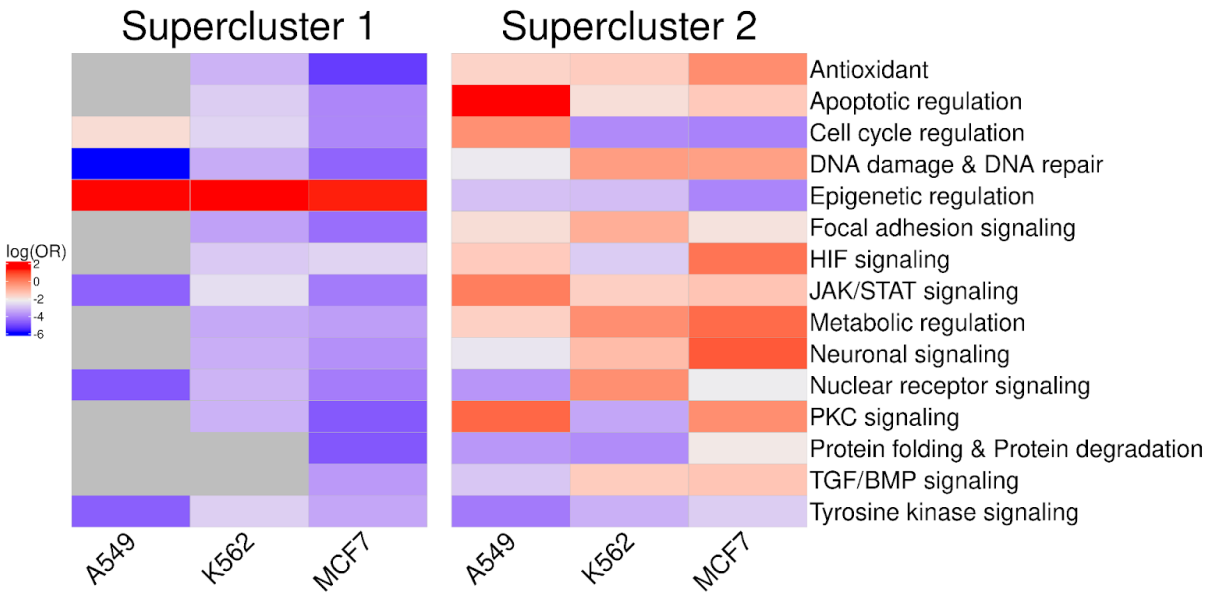

Figure 2e: Drug Class Treated Cells Overrepresentation in Superclusters. Within each heatmap, the rows represent the various pathways targeted by the drug treatments and the columns represent the RACs of each supercluster. The heatmap cell is gray if there are no cells treated with that drug class in a given supercluster component.

CRISPR KO of transcription factors driving resistance states sensitize the cells to the drug

Here we aimed to identify transcription factors (TF) potentially driving the resistance states. Starting with 1186 TFs along with their regulons (CollecTRI; Müller-Dott et al., 2023), we scored all cells for the activity of each regulon using AUCell and identified the differential activity of each regulon in every cluster, relative to all non-RAC clusters in the cell line (see Methods). Then for each of the two superclusters, we created a consensus list of top (most differentially active) and bottom (most differentially inactive) regulons. This top list represents the TFs that are potential regulators of the conserved supercluster

states (Table S3), while the bottom list serves as a control. Some notable TFs that are most differentially active in supercluster 1 include SMARCD3 and DMRTA1, both of which have been shown to be associated with drug resistance (Ferguson et al., 2023; Zhang et al., 2023). For supercluster 2, notable TFs include SPZ1 which has also been shown to be important for developing resistance (Liu, 2018).

Next, we experimentally assessed whether knocking out the TFs activated in, and potentially driving, the resistant superclusters increased sensitivity to cytotoxic drugs. Toward this, we performed a genome-wide CRISPR knockout screen and quantified the effect of each gene's KO on the sensitivity of an ovarian cell line (OVCAR8) to a CHK1,2 checkpoint kinase inhibitor, prexasertib, for 10 generations (see Methods). We scored the genes based on the increase in drug sensitivity upon KO; lower scoring genes in this list are likely to mediate drug resistance. Encouragingly, we found that in supercluster 1, the top TFs have significantly lower CRISPR-based scores than the bottom TFs (p-value = 0.01; Figure 3a). Top ranking TFs in supercluster 1 that were validated in the CRISPR screen, includes TRIM5 which is known to regulate EMT and lead to drug resistance (Huang et al., 2022). Even though we did not see a statistical significance in supercluster 2, among the top ranking validated TFs in supercluster 2 were SIN3B and MESP2, known mediators of resistance (Morales-Valencia et al., 2023; Smith and Bhowmick, 2016). The finding that the transcriptional regulators of supercluster 1 score better in the prexasertib screen is consistent with the suspected mechanisms of action -- replication catastrophe (King et al., 2015). Due to low cell cycling of supercluster 1, cells occupying this state would therefore be more resistant to this treatment. Supercluster 2 however, is more

sensitive to prexasertib treatment due to its higher expression of G2M checkpoint leading to susceptibility to replication catastrophe and cell death. This analysis suggests that disruption of some potential regulators of the resistance-associated transcriptional states in early response could contribute to drug tolerance.

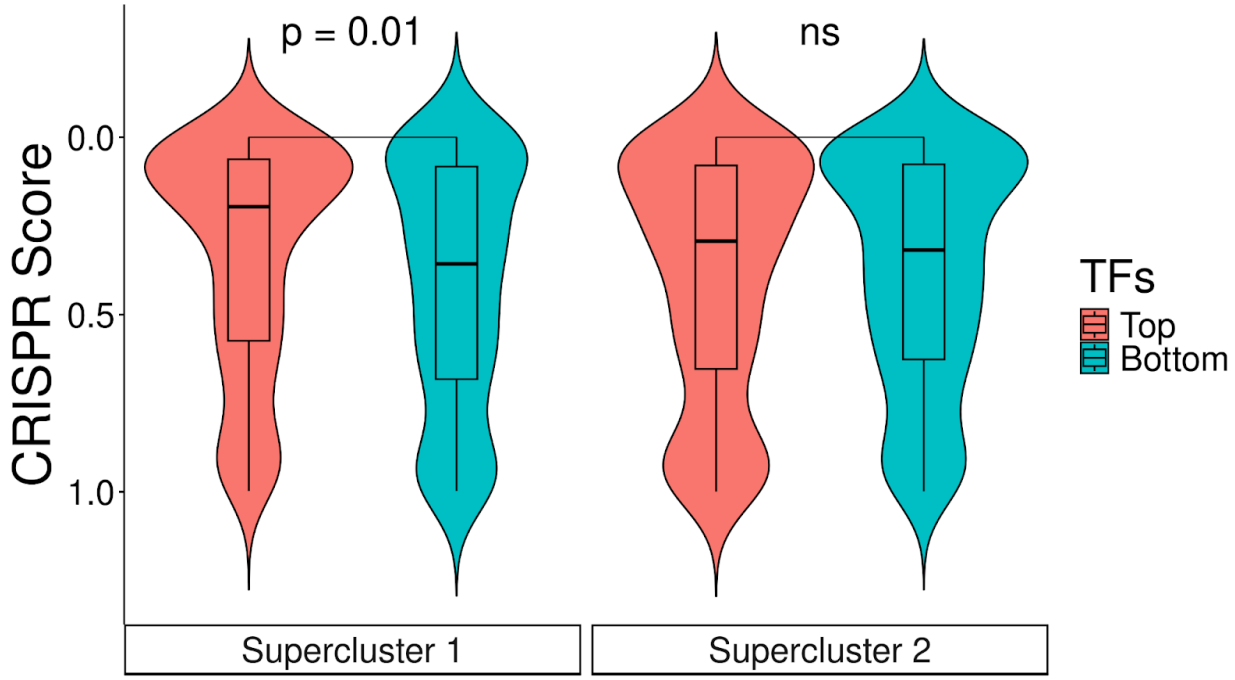

Figure 3a: Distributions for CRISPR scores for each supercluster top and bottom 25% differential activity transcription factors (red = top; blue = bottom)

Clinical Relevance of Resistant States

Having identified, and validated to some extent, gene signatures associated with drug resistance, we next assessed the clinical relevance of each cell line's global RAC signatures as well as the signatures of both superclusters. First, for each cell line, we assessed the association of the global RAC signature with patient survival in the matching

cancer type in TCGA -- A549 signatures in lung cancer, K562 signatures in blood cancer, and MCF7 signatures in breast cancer. We scored the samples with each global RAC signature using ssGSEA (Hänzelmann et al., 2013) and performed a Cox regression analysis controlling for age, sex, and tumor purity (Methods). In TCGA-LUAD and TCGA-LAML, high expression of the global RAC signature is associated with worse survival with hazard ratios of 76.2 and 2.2 respectively (Figure 4a). These results indicate that the signatures associated with resistance derived from early drug response are prognostic of overall survival in two of the three cancer types. Since superclusters are representative of multiple cancer types, we repeated the above analysis with the supercluster signatures across all TCGA cohorts (Figure S3). We found that supercluster 1 is associated with worse overall survival in 24 of the 33 TCGA cohorts, with 5 of them being significant. For supercluster 2, we found 25 of the 33 cohorts are associated with worse overall survival with 11 of them being significant (Figure 4b).

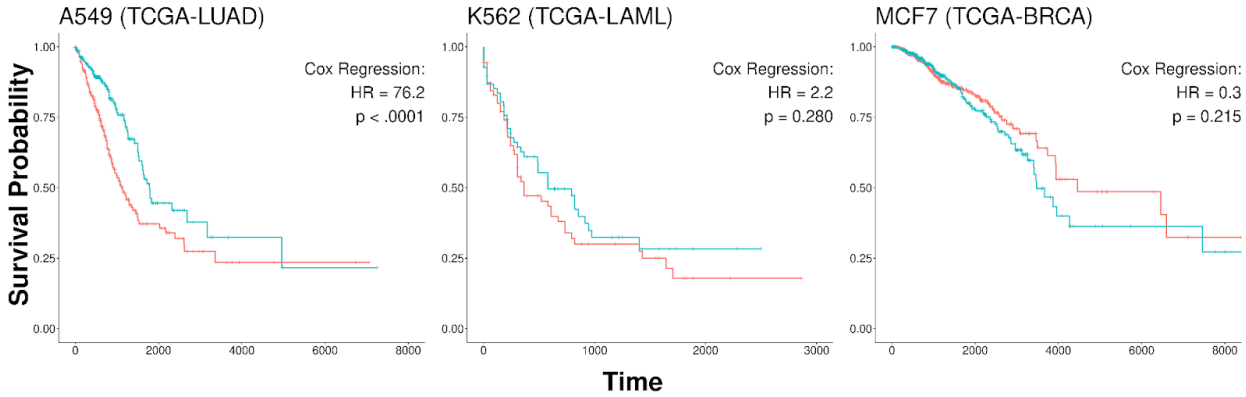

Figure 4a: Global RAC Signatures are associated with overall survival in the corresponding cancer. Kaplan-Meier plots for the survival analysis of each RAC Type signature from each cell line. The patients are stratified into 2 groups representing their

level of expression for the given cell line signature (red = high, blue = low). The HR and p-values are obtained from Cox regression.

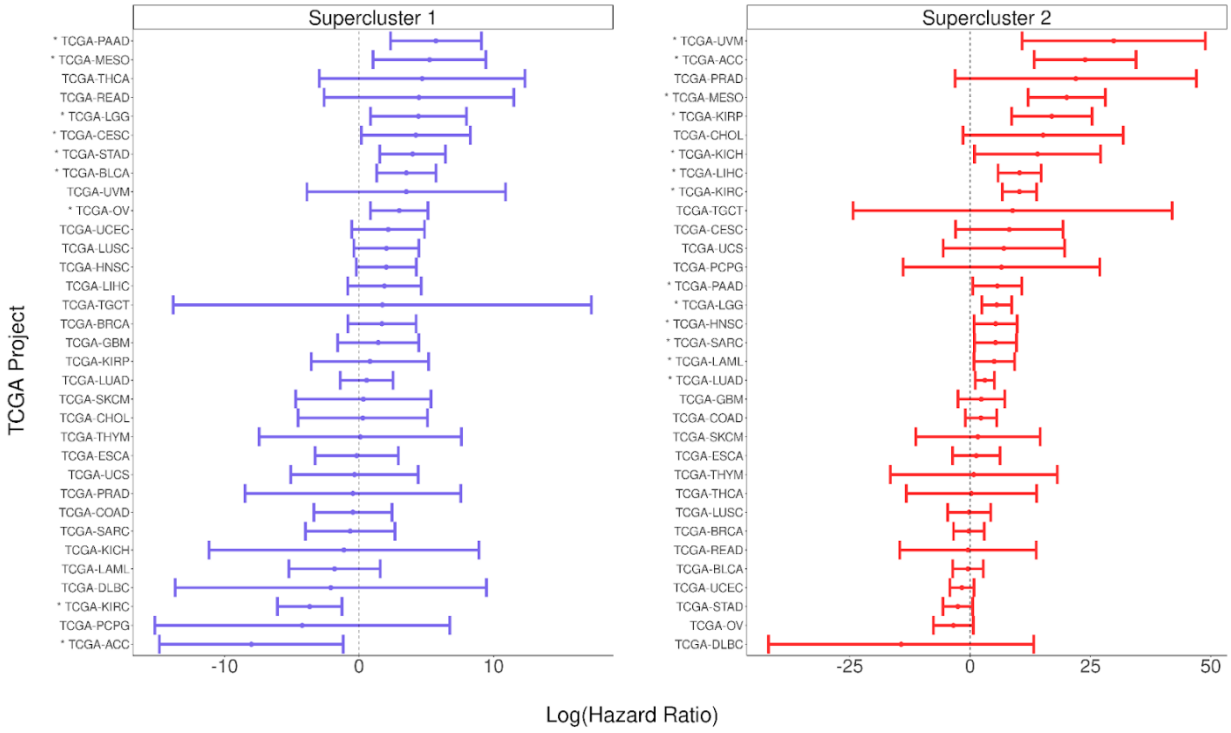

Figure 4b: Hazard Ratios for Supercluster Signatures Across TCGA Cancer Types with 95% confidence interval of log hazard ratios for both supercluster signatures in all 33 TCGA projects. Significant hazard ratios are indicated by a star next to the TCGA project name.

Due to the RAC and supercluster signatures' association with patient survival, we next assessed their ability to predict therapy response. Toward this, first, we acquired a dataset that contains pre-treatment tumor RNA-seq samples from 8 patients with EGFR-mutant lung cancer treated with the tyrosine kinase inhibitor, osimertinib. All patients had

relapsed after initial response and were classified into those with short relapse time (4 patients, < 8.6 months) and those with long relapse time (4 patients; >= 12 months; Gurule et al., 2021). We scored each patient's pre-treatment sample with the A549 global RAC signature (cancer type matched) as well as with the two supercluster signatures using ssGSEA. We observed that the patients with short relapse times had higher resistance scores for all three signatures; p-values are not significant due to the small sample size. This suggests pre-existing resistance mechanisms in short-term responder patients.

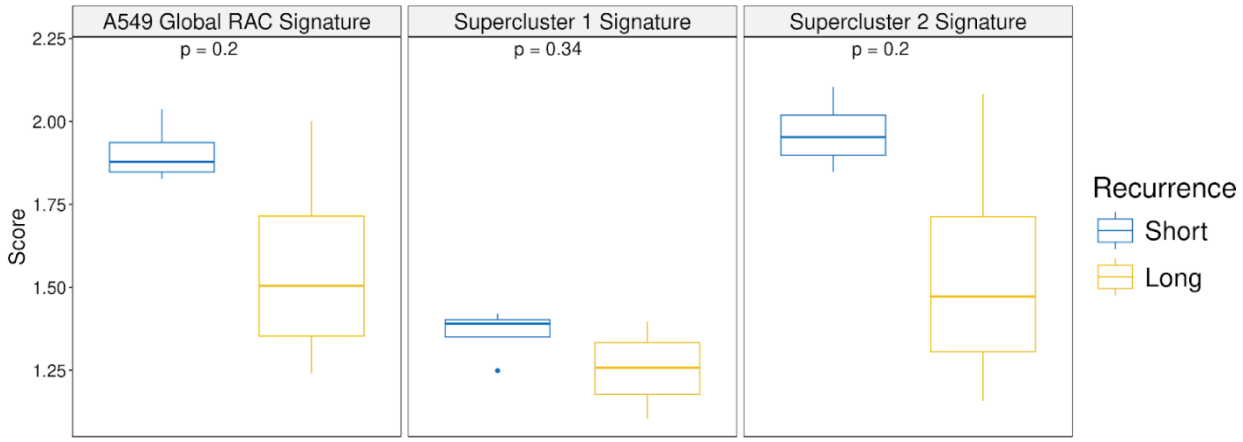

Figure 4c: Pre-treatment RAC signatures scores in lung cancer patients before tyrosine kinase inhibitor treatment. The y-axis represents the ssGSEA score and the x-axis stratifies patients by recurrence time (blue = short term recurrence, yellow = long term recurrence).

We next extended this analysis to other cancer and drug types, using the two supercluster signatures. The data we used consisted of bulk RNA sequencing and microarray data spanning 6 cancer types and 7 drug treatments, where in each study the patients were classified into responders and non-responders (Dinstag et al., 2023) (Table S4). We

compared scaled supercluster signature scores between the responders and non-responders and in all cohorts we analyzed, at least one of the supercluster signatures was associated with non-responders.

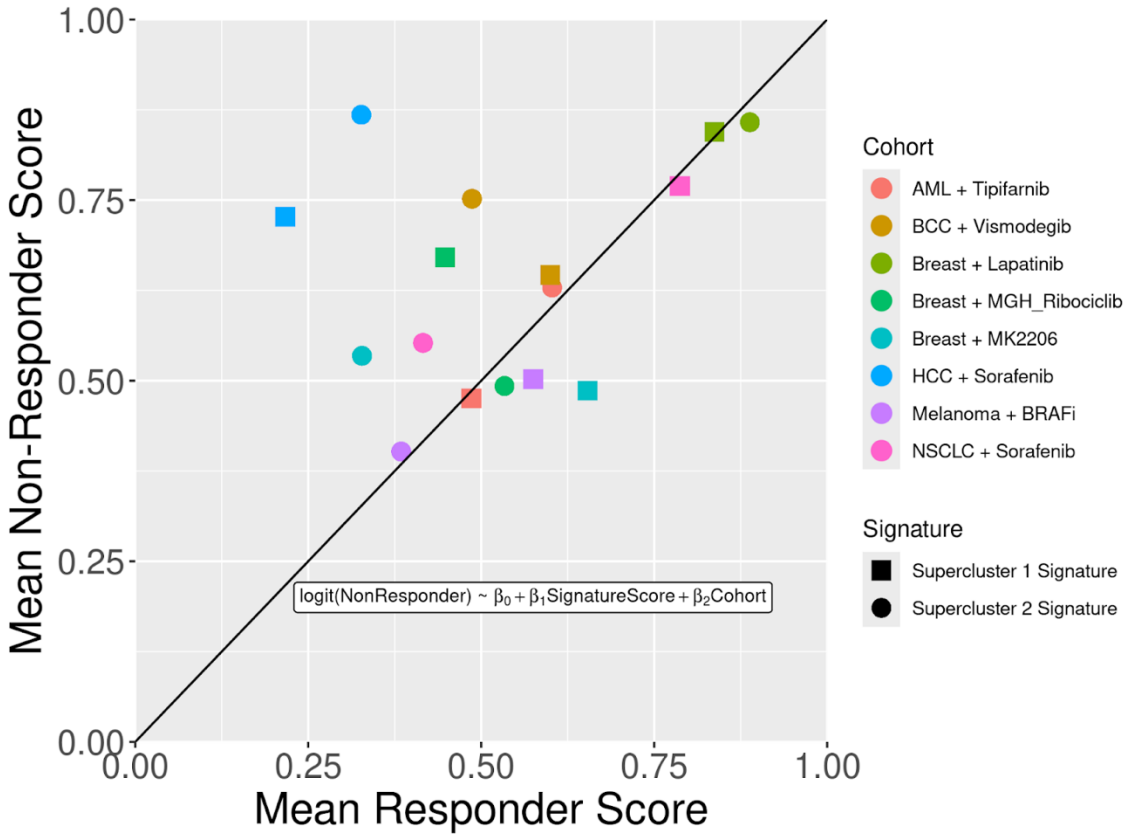

Figure 4d: Mean supercluster signature scores across cancer treatment cohorts (cohort = drug x cancer combo). All 8 cohorts have a higher mean non-responder score than mean responder score for at least 1 supercluster signature.

To determine whether the supercluster signature score of a patient within a given cohort could be used to classify the patient as a responder or non-responder to therapy, we fit a

logistic regression model to the data. Controlling for cohort ID, we found that the signature score does in fact significantly contribute to whether a patient is a responder or non-responder (Supercluster 1 score p-value: 0.0025; Supercluster 2 score p-value: 1e^-6). Overall, our results strongly suggest that our cancer-specific RAC as well as pan-cancer supercluster resistance signatures are prognostic of patient survival and therapy response across cancers and therapy types.

Evolutionary Conservation of Cellular Resistance Mechanisms

Having identified common mechanisms of resistance across the three cancer cell lines, and observing their prognostic utility across cancer types, we next assessed the extent to which these resistance-associated transcriptional responses represent inherent cellular response to chemical-induced stressors and therefore may be evolutionarily conserved across different species. To test this hypothesis, we first obtained a gene signature comprising the genes that are upregulated in the yeast *Candida auris* resistant to antifungal drugs (Zhou et al., 2021) and mapped them to their human orthologs using information from OMA Browser (Altenhoff et al., 2020) (Table S5). Next, we scored all cells in the early-response dataset using AUCell and compared the score distributions across the various supercluster cells and the non-RAC cells. In supercluster 2, we saw significantly higher signature scores for the ortholog yeast resistance geneset compared to all non-RAC cells, suggesting that supercluster 2 represents conserved mechanisms of drug response between yeast and human (Figure 5a).

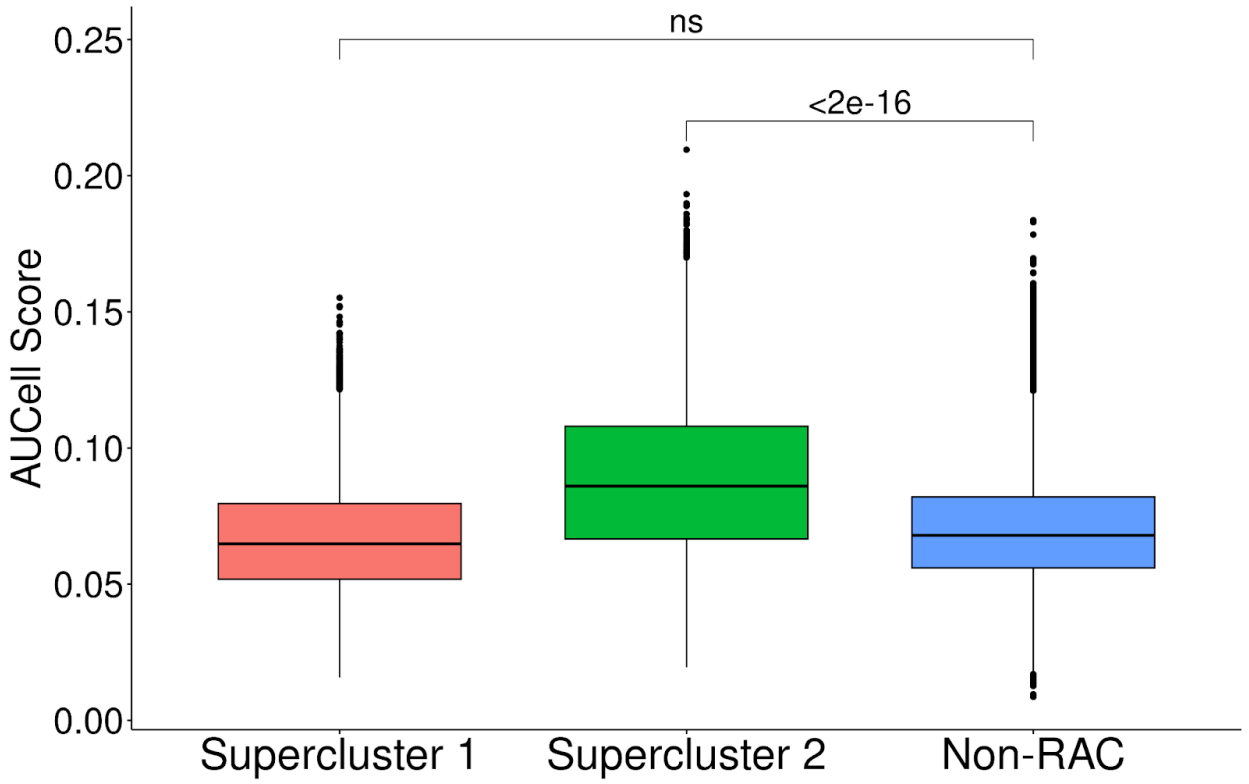

Figure 5a: Yeast antifungal resistance gene signature scores in human resistance and non-resistance clusters. Each distribution contains the scores for every cell in each supercluster or non-RAC clusters across the three cell lines.

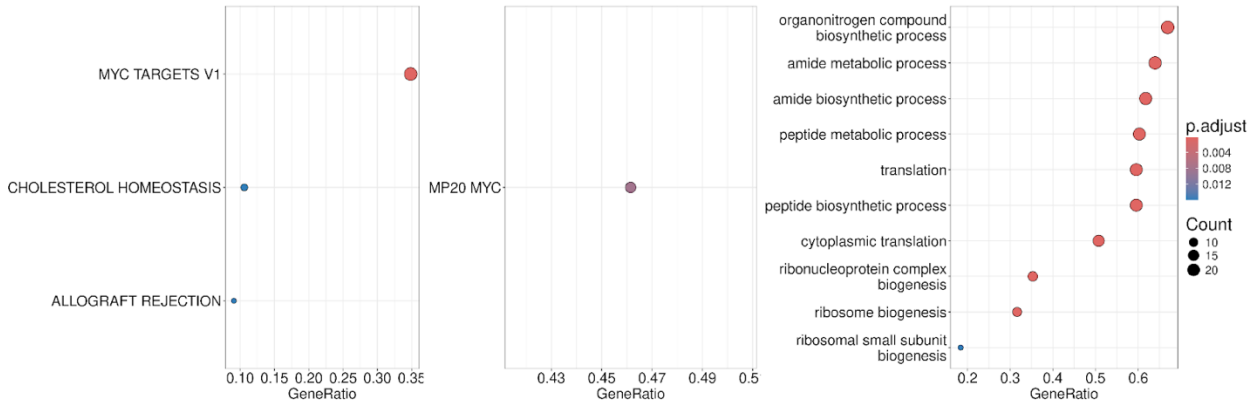

Figure 5b: Pathway enrichment of shared genes between yeast resistance orthologs and supercluster 2.

To further probe into the underlying shared mechanisms, we performed functional enrichment on the genes that are common in both the antifungal resistant ortholog geneset and genes upregulated in supercluster 2 (Figure 5b). Similar to the pathways enriched in the global RAC signatures in respective cell lines, this analysis shows enrichment for translation and MYC activity, suggesting that these processes are conserved properties of cellular response to chemicals and contribute to resistance.

Next, we compiled experimental data of drug-treated *E. coli* strains and obtained differentially upregulated genes in resistant relative to drug-naive samples for 9 antibacterial drugs (Table S5). With the 9 drug resistant upregulated genesets, we created a consensus antimicrobial resistance gene set and mapped these genes to their human orthologs, as above. As in the case of yeast, we scored all cells for the gene signature using AUCell. Again, in supercluster 2, we saw significantly higher resistance scores compared to all non-RAC cells. (Figure 5c). We next performed functional enrichment on the genes that are shared in both the antimicrobial resistant ortholog genesets and genes upregulated in supercluster 2 (Figure 5d). Though this analysis was underpowered due to the small number (30) of genes, we found among the shared genes multiple ribosomal proteins including RPL11, which is shown to be associated with lung cancer progression (Chen et al., 2023) and RPL23, which is associated with multidrug resistance and worse prognosis (Shi et al., 2004; Kang et al., 2021). Additionally, APEX1 and DDX5 have been shown to promote therapy resistance (Cao et al., 2020; Li, 2023). The mechanisms that are evolutionarily conserved between *E. coli* and human cancer

seem to stem from dysregulated ribosomal proteins which have been noted to be associated with tumor progression and therapeutic resistance (Elhamamsy at al., 2022). Overall, these analyses suggest that supercluster 2 represents an evolutionarily conserved mechanism underlying drug resistance between human, yeast, and bacteria.

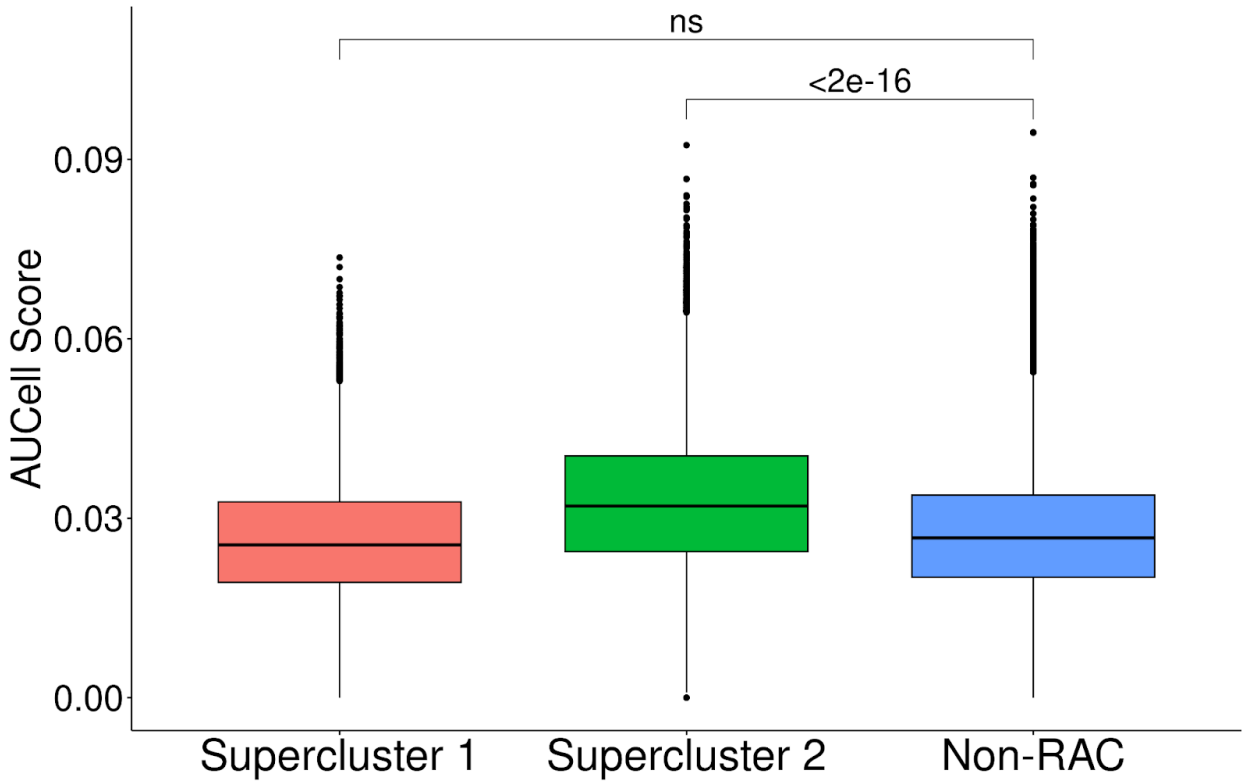

Figure 5c: E. coli antimicrobial resistance gene signature scores in human resistance and non-resistance clusters. Each distribution contains the scores for every cell in each supercluster or non-RAC clusters across the three cell lines.

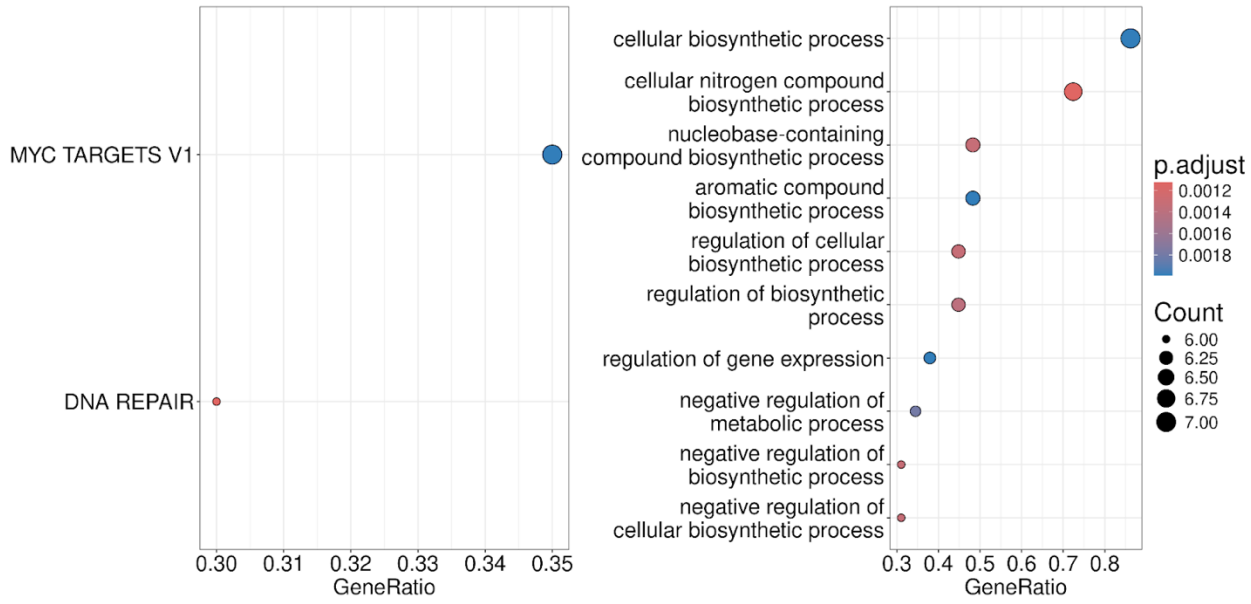

Figure 5d: Pathway enrichment of shared genes between E. coli resistance orthologs and supercluster 2.

Discussion

In this work, we have comprehensively assessed the extent to which the transcriptional programs that characterize drug-resistant cancers are manifested before drug treatment or early in the course of drug treatment, and thus may represent inherent cellular response to cytotoxic stressors, some of which form the basis for long-term resistance. We find that many of the transcriptional programs that are observed in established resistance such as epithelial to mesenchymal transition (Song and Faber, 2019) and MYC signaling (Donati and Amati, 2022; Doha and Sears, 2023) are indeed present just 24 hours after drug treatment. Furthermore, not only are these early manifestations of resistance programs evident in multiple cancer cell lines, but aspects of them exhibit deep evolutionary conservation in bacteria and yeast, suggesting a conserved inherent cellular

response to cytotoxic stressors, ultimately linked to long-term survival of the cell. Using a whole genome CRISPR knockout screen, we validated that the transcriptional regulators of our identified early drug resistance states, when knocked out, render the cell less tolerant to the drug.

Consistent with previous reports (Goyal et al., 2023), our analysis reveals distinct modes of resistance. The resistant transcriptional states conserved across cancer types, namely supercluster 1 and supercluster 2, while sharing common characteristics such as EMT, also exhibit varying cell cycle states and association with distinct drug classes. These aspects of resistance are frequently outlined in the literature as responses to drug treatment (Oren et al., 2021; Song and Faber, 2019; Kluska et al., 2023; De Las Rivas et al., 2021). Interestingly, the identified early response signatures associated with long-term resistance are prognostic of survival and predictive of therapy response across multiple cancer and treatment types.

Since this prediction is based on cancers obtained at the time of diagnosis, and not after cancer chemotherapy, expression of the identified resistant states may be present before treatment and subsequently selected for upon treatment. This is consistent with our observation that the gene signatures of these states can be prognostic even in pre-treated samples (Figure 4c). On the other hand, multiple transcriptional states appear strictly post-treatment and may represent adaptive responses (Figure S2a). These findings corroborate recent discussion of drug-tolerant persister (DTP) cells, which are rare, slow-cycling persister cells that transiently occur post-treatment and can revert to cycling upon

release from treatment and become sensitive to the initial therapy. However, it is still unclear whether DTPs are pre-existing states or the result of therapeutically induced transcriptional reprogramming (Marine et al., 2020). Furthermore, an additional confounder to consider is that rare populations of pre-existing resistant cells may go undetected in single cell profiling and may appear as emergent states post-treatment.

Notably, supercluster 2 gene signature include many genes encoding ribosomal proteins. These ribosomal protein genes frequently show up in drug resistance selection experiments (Li, 2023) and the dysregulation of ribosomal proteins appear to play a role in cancer resistance (Elhamamsy et al., 2022; El Khoury and Nasr, 2021). Many of these ribosomal proteins have extrarobosomal functions including chemoresistance (Zhou et al., 2015).

The drug response data used in our study includes responses to a large array of drugs and drug classes and therefore the identified resistance-associated signatures are relevant across multiple drug classes (Srivatsan et al., 2019). One of the prominent mechanisms underlying multidrug resistance (MDR) involves ATP-binding cassette (ABC) transporters. The gene signatures of some of the RACs characterized in this work include ABC transporter genes (cluster 16 in A549, cluster 11 in K562, and cluster 13 and 17 in MCF7), yet they were not present in either of the supercluster signatures, suggesting variability in their expression or functional role in resistance across cell lines. While it has been shown that the expression of one ABC transporter, P-glycoprotein 1 (Pgp), can be

induced by treatment (Robinson and Tiriveedhi, 2020), it is unclear to what extent Pgp plays a role in the resistance characterized in this work.

Novel treatment strategies to modulate resistance programs in cancer are currently being investigated (Lei et al., 2020). Among them are targeting biomarkers associated with drug resistant cancer cells, and the supercluster signatures identified in this work may represent potential avenues of further investigation (Song and Faber, 2019). Although much more research is needed to fully understand the mechanisms of resistance and how to inhibit them, this work contributes to the growing knowledge of the molecular characteristics of resistant transcriptional states shortly after treatment which may inform novel adjuvant treatment strategies to overcome resistance.

Methods

Creating resistance signature

We created a drug resistance gene signature (882 genes) that consists of the intersection of upregulated genes in drug resistance samples across 4 experiments (Table S1). The upregulated genes from Goyal et al. (2023) and Shaffer et al. (2017) were identified with two different approaches. For Goyal et al. (2023), we obtained data for the cell lines, WM989 A6-G3 and MDA-MB-231-D4 treated with vemurafenib and paclitaxel, respectively. We then repeated the single cell clustering approach as detailed in the publication's methods section then pseudo-bulked the counts data based on Seurat cluster classification (Hao et al., 2021). This was followed by computing the differentially expressed genes using DESeq2 (Love et al., 2014) between the pseudo-bulked resistant

samples to a control sample of the same cell line acquired from another publication (Shaffer et al., 2020). We then selected the upregulated (log2FC > 0 and adjusted p value < 0.05) genes for these experiments.

For Shaffer et al. (2017), we acquired data for the cell lines WM989 and WM983B both treated with vemurafenib. Utilizing the differential expression results directly from the publication data, we selected the genes that are upregulated (log2FC > 0 and adjusted p value < 0.05) in the resistant samples compared to the untreated samples. We subsequently filtered these genes further by removing the genes that are strongly upregulated (log2FC > 2). This left us with four genesets (two from each publication) of upregulated genes across multiple experiments and we used the intersection of these genesets as the initial drug resistance signature.

We validated the 882 gene signature in time-course data of drug resistance development - PC9 lung cancer cells were treated with osimertinib and single cell transcriptomic profiling was done at days 0, 3, 7, and 14. For each timepoint in the Oren et al. data, we calculated fold change values for each gene using the Seurat FoldChange() function with cell identities being set to the time points. With these ranked gene lists, at each timepoint, we calculated the enrichment of the 882 gene resistance signature at day 0 and day 14 using GSEA from the fgsea R package (Korotkevich, 2019). We then plotted the normalized enrichment score for the two time points as a line plot.

After validating the initial resistance signature from Shaffer et al and Goyal et al., in an additional dataset acquired from Oren et al., we filtered the signature to only retain the genes that are also upregulated (log2FC > 0 and p value < 0.05) at day 14 cells from Oren et al., resulting in 97 genes (Table S1).

AUCell Threshold

For all applications of AUCell, to estimate a threshold for the raw AUCell score to designate a cell 'active' or 'inactive' for the gene signature, we created a background distribution of AUCell scores based on 100 control geneset with similar gene expression levels as the given geneset. To generate these control gene sets, we follow a procedure similar to that outlined in the AddModuleScore function in Seurat. We first assigned all expressed genes into ten bins based on their mean normalized expression, and assigned each signature gene to a bin based on its normalized expression level.

To generate a single control gene set, we picked a single gene at random from the same expression bin corresponding to each signature gene. The activity of the control gene set was then scored using AUCell, where the 95th percentile of the AUCell scores was stored as a putative activity threshold. This entire process was repeated 100 times, where the largest threshold computed is the final activity threshold. Any cell with an AUCell score higher than this threshold is considered to have an active resistance signature.

Odds ratio for resistance signature activity in a cell cluster calculation

The odds ratio for each cluster is calculated as (# of active cells in cluster X/# of inactive cells in cluster X)/(total # of active cells/total # of inactive cells). We use an odds ratio threshold of 1.5 and p-value threshold of 0.05 as the cutoff for resistance-activated clusters.

Cancer hallmark heatmap creation

We scored every cell for each cancer hallmark geneset (Liberzon, 2015) with AUCell then plotted the mean cluster score for each hallmark and the Z-scored ((0,1)-normalization) values for each hallmark across clusters within a cell line.

Global RAC gene signatures

To acquire the gene signatures that define the RAC cells in each cell line we used the Seurat FindAllMarkers function to compare all RAC cells to all non-RAC cells. We further filtered the resulting differentially expressed genes for adjusted p value < 0.05 and average log2FC > 0. We sorted the genes by descending fold change and selected the top 200 genes.

Global RAC signatures functional enrichment

To identify which relevant pathways are enriched in the global RACs, we utilized the clusterProfiler R package. The pathways we investigated for enrichment included mSigDB hallmarks (Liberzon, 2015), intratumoral heterogeneity meta-programs (MPs) (Gavish et al., 2023), and GO biological process terms. We performed GSEA on the ranked gene lists via the GSEA (Subramanian et al., 2005; Mootha et al., 2003) function

for hallmarks and MPs and the gseGO() function for GO terms. The background gene universe included all genes that are expressed (count > 0) in at least 1% of cells in each respective cell line.

Grouping the RACs across cell lines into superclusters

We first created a gene list that consists of the intersection of the top 2000 most variable genes in each cell line. With the resulting union of 3011 genes, for each gene, we calculated its differential mean cluster expression in a RAC relative to all other cells in each respective cell line. We then performed an all-by-all (for all clusters across the three cell lines) Spearman correlation of these differential gene expression vectors and hierarchically grouped the cell line-specific clusters. We defined superclusters as highly correlated (correlation among top 90th percentile) groups of RACs including components from all three cell lines.

Supercluster Consensus Signature creation

To create the signature to represent the supercluster spanning all three cell lines, we first selected all genes that are upregulated (log2FC > 0 and p-value < 0.05) in the component RACs in the respective cell line. We then selected the genes that are upregulated in all three component cluster signatures for the supercluster.

Supercluster Consensus Signature Functional enrichment

To identify which relevant pathways are enriched in the supercluster consensus signatures, we utilized the clusterProfiler R package. The pathways we investigated for

enrichment included mSigDB hallmarks, intratumoral heterogeneity meta-programs (MPs), and GO biological process terms. We performed a hypergeometric test via the enricher() function for hallmarks and MPs and the enrichGO() function for GO terms. The background gene universe used included all genes that are expressed (count > 0) in at least 1% of cells in each cell line.

Supercluster Cell Cycle phase plots

To calculate the percentage of cells in a supercluster that are in each of the cell cycle phases, we first scored all cells in each supercluster for gene signatures associated with each phase using the Seurat CellCycleScoring (Hao et al., 2021) function. This function was modified to include an additional G0 signature derived from Wiecek et al. (2023). Each cell was assigned a phase based on the highest scoring cell cycle phase signature in that cell. With each cell assigned to a phase, we plotted the percentage of each phase for each supercluster

Cell line specific supercluster enrichment

To determine which pathways are enriched in just one of three clusters composing each supercluster we first selected the genes that are upregulated only in one of the cell line-specific component clusters. With this gene list we repeated the enrichment analysis described above.

Drug class specific RACs

To determine if cells treated with a specific drug class are overrepresented in some RACs, we computed an odds ratio comparing the proportion of cells treated with a given drug class in a single RAC to the proportion of cells treated with the same drug overall.

Inferring transcription factors potentially regulating the supercluster programs

We obtained target genesets (regulons) of 1186 TFs from CollecTRI (Müller-Dott et al., 2023). For each regulon we calculated an odds ratio comparing the mean AUCell regulon expression score in a given RAC compared to the mean regulon expression score in all non RACs in the same cell line. After this, we selected the top 50% and bottom 50% scoring regulons for each RAC that make up the two superclusters. To get the final top and bottom TF lists for each supercluster, we selected the regulons that appear in all three clusters that make up a given supercluster.

Kaplan-Meier plots

To create the Kaplan-Meier plots in this work we utilized the survminer R package. Samples were split into high and low expression groups based on top and bottom 50th percentiles.

CRISPR Screening

A genome-wide CRISPR Brunello knockout (KO) lentiviral pooled library was used to assess genes required for resistance to the CHEK1 inhibitor, Prexasertib. The Brunello library contains 76,441 gRNAs targeting 19,114 genes and 1,000 unique non-targeting sgRNA controls (Addgene # 73178). First, OVCAR8 ovarian cancer cells were selected

for stable Cas9 expression (Addgene # 52962-LV). Then, the Brunello sgRNA library was transduced at ~0.3 MOI to ensure unique sgRNA distribution, and cells were selected with 5 µg/ml of puromycin for three days. The initial pooled cell population was split into separate study arms with an estimated 500x library coverage per condition. The treatment (30ng Prexasertib) and control (DMSO) arms were maintained in culture for 10 generations at a density required for library representation. Genomic DNA was extracted using the QIAmp DNA blood Cell Maxi Kit (Qiagen) according to the manufacturer's protocol. The sgRNA barcodes were amplified using Taq polymerase and adapted for sequencing. The desired DNA product was purified with 6% TBE gel (Invitrogen) and samples were sequenced on an Illumina HiSeq2000. The sgRNA read count and hit calling were analyzed using MAGeCK (v0.5.7). Gene scoring and ranking were determined by Robust Rank Aggregation (RRA) with normalization relative to the distribution of non-targeting gRNAs. Genes that were negatively selected in the treatment arm compared to control were selected for downstream analysis.

References

1. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14, 1083–1086 (2017).

2. Altenhoff, A. M. et al. OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. Nucleic Acids Research 49, D373–D379 (2021).

3. Bell, C. C. et al. Targeting enhancer switching overcomes non-genetic drug resistance in acute myeloid leukaemia. Nat Commun 10, 2723 (2019).

4. Bie, L. et al. Comparative Analysis of Transcriptomic Response of Escherichia coli K-12 MG1655 to Nine Representative Classes of Antibiotics. Microbiol Spectr 11, e0031723 (2023).

5. Bigger, JosephW. TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION. The Lancet 244, 497–500 (1944).

6. Cao, L., Cheng, H., Jiang, Q., Li, H. & Wu, Z. APEX1 is a novel diagnostic and prognostic biomarker for hepatocellular carcinoma. Aging (Albany NY) 12, 4573–4591 (2020).

7. Celià-Terrassa, T. & Kang, Y. How important is EMT for cancer metastasis? PLoS Biol 22, e3002487 (2024).

8. Chen, J. et al. RPL11 promotes non-small cell lung cancer cell proliferation by regulating endoplasmic reticulum stress and cell autophagy. BMC Mol Cell Biol 24, 7 (2023).

9. De Amicis, F. et al. Androgen receptor overexpression induces tamoxifen resistance in human breast cancer cells. Breast Cancer Res Treat 121, 1–11 (2010).

10. De Las Rivas, J. et al. Cancer drug resistance induced by EMT: novel therapeutic strategies. Arch Toxicol 95, 2279–2297 (2021).

11. Dinstag, G. et al. Clinically oriented prediction of patient response to targeted and immunotherapies from the tumor transcriptome. Med 4, 15-30.e8 (2023).

12. Doha, Z. O. & Sears, R. C. Unraveling MYC's Role in Orchestrating Tumor Intrinsic and Tumor Microenvironment Interactions Driving Tumorigenesis and Drug Resistance. Pathophysiology 30, 400–419 (2023).

13. Donati, G. & Amati, B. MYC and therapy resistance in cancer: risks and opportunities. Mol Oncol 16, 3828–3854 (2022).

14. El Khoury, W. & Nasr, Z. Deregulation of ribosomal proteins in human cancers. Biosci Rep 41, BSR20211577 (2021).

15. Elhamamsy, A. R., Metge, B. J., Alsheikh, H. A., Shevde, L. A. & Samant, R. S. Ribosome Biogenesis: A Central Player in Cancer Metastasis and Therapeutic Resistance. Cancer Res 82, 2344–2353 (2022).

16. Ferguson, L. P. et al. Smarcd3 is an epigenetic modulator of the metabolic landscape in pancreatic ductal adenocarcinoma. Nat Commun 14, 292 (2023).

17. França, G. S. et al. Drug-Induced Adaptation along a Resistance Continuum in Cancer Cells. http://biorxiv.org/lookup/doi/10.1101/2022.06.21.496830 (2022) doi:10.1101/2022.06.21.496830.

18. Frank, S. A. & Rosner, M. R. Nonheritable cellular variability accelerates the evolutionary processes of cancer. PLoS Biol 10, e1001296 (2012).

19. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. Nature 618, 598–606 (2023).

20. Goyal, Y. et al. Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. Nature 620, 651–659 (2023).

21. Gurule, N. J. et al. A tyrosine kinase inhibitor-induced interferon response positively associates with clinical response in EGFR-mutant lung cancer. npj Precis. Onc. 5, 41 (2021).

22. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics 14, 7 (2013).

23. Hao, Y. et al. Integrated analysis of multimodal single-cell data. Cell 184, 3573-3587.e29 (2021).

24. Huang, N. et al. TRIM family contribute to tumorigenesis, cancer development, and drug resistance. Exp Hematol Oncol 11, 75 (2022).

25. Jing, X. et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Mol Cancer 18, 157 (2019).

26. Kang, H. et al. USP19 and RPL23 as Candidate Prognostic Markers for Advanced-Stage High-Grade Serous Ovarian Carcinoma. Cancers 13, 3976 (2021).

27. King, C. et al. LY2606368 Causes Replication Catastrophe and Antitumor Effects through CHK1-Dependent Mechanisms. Mol Cancer Ther 14, 2004–2013 (2015).

28. Kluska, M., Piastowska-Ciesielska, A. W. & Tokarz, P. Cell Cycle Status Influences Resistance to Apoptosis Induced by Oxidative Stress in Human Breast Cancer Cells, Which Is Accompanied by Modulation of Autophagy. CIMB 45, 6325–6338 (2023).

29. Lei, Z.-N. et al. Understanding and targeting resistance mechanisms in cancer. MedComm (2020) 4, e265 (2023).

30. Li, F. et al. Role of the DEAD-box RNA helicase DDX5 (p68) in cancer DNA repair, immune suppression, cancer metabolic control, virus infection promotion, and human microbiome (microbiota) negative influence. J Exp Clin Cancer Res 42, 213 (2023).

31. Li, M., Sun, J. & Shi, G. Application of CRISPR screen in mechanistic studies of tumor development, tumor drug resistance, and tumor immunotherapy. Front Cell Dev Biol 11, 1220376 (2023).

32. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417–425 (2015).

33. Liu, X. et al. SPZ1 is critical for chemoresistance and aggressiveness in drug-resistant breast cancer cells. Biochem Pharmacol 156, 43–51 (2018).

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

35. Marine, J.-C., Dawson, S.-J. & Dawson, M. A. Non-genetic mechanisms of therapeutic resistance in cancer. Nat Rev Cancer 20, 743–756 (2020).

36. Mootha, V. K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34, 267–273 (2003).

37. Morales-Valencia, J., Petit, C., Calderon, A., Saini, S. & David, G. Chromatin-Associated SIN3B Protects Cancer Cells from Genotoxic Stress-Induced Apoptosis and Dictates DNA Damage Repair Pathway Choice. Mol Cancer Res 21, 947–957 (2023).

38. Müller-Dott, S. et al. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. Nucleic Acids Research 51, 10934–10949 (2023).

39. Nussinov, R., Tsai, C.-J. & Jang, H. Anticancer drug resistance: An update and perspective. Drug Resist Updat 59, 100796 (2021).

40. Oren, Y. et al. Cycling cancer persister cells arise from lineages with distinct programs. Nature 596, 576–582 (2021).

41. Pan, Z., Zhang, H. & Dokudovskaya, S. The Role of mTORC1 Pathway and Autophagy in Resistance to Platinum-Based Chemotherapeutics. Int J Mol Sci 24, 10651 (2023).

42. Ramos, A., Sadeghi, S. & Tabatabaeian, H. Battling Chemoresistance in Cancer: Root Causes and Strategies to Uproot Them. Int J Mol Sci 22, 9451 (2021).

43. Robinson, K. & Tiriveedhi, V. Perplexing Role of P-Glycoprotein in Tumor Microenvironment. Front Oncol 10, 265 (2020).

44. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature 546, 431–435 (2017).

45. Shaffer, S. M. et al. Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. Cell 182, 947-959.e17 (2020).

46. Sharma, S. V. et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141, 69–80 (2010).

47. Shi, Y. et al. Ribosomal proteins S13 and L23 promote multidrug resistance in gastric cancer cells by suppressing drug-induced apoptosis. Exp Cell Res 296, 337–346 (2004).

48. Smith, B. N. & Bhowmick, N. A. Role of EMT in Metastasis and Therapy Resistance. J Clin Med 5, 17 (2016).

49. Song, K.-A. & Faber, A. C. Epithelial-to-mesenchymal transition and drug resistance: transitioning away from death. J Thorac Dis 11, E82–E85 (2019).

50. Srivatsan, S. R. et al. Massively multiplex chemical transcriptomics at single-cell resolution. Science 367, 45–51 (2020).

51. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550 (2005).

52. Torborg, S. R., Li, Z., Chan, J. E. & Tammela, T. Cellular and molecular mechanisms of plasticity in cancer. Trends Cancer 8, 735–746 (2022).

53. Wiecek, A. J. et al. Genomic hallmarks and therapeutic implications of G0 cell cycle arrest in cancer. Genome Biol 24, 128 (2023).

54. Zhang, R. et al. The DMRTA1-SOX2 positive feedback loop promotes progression and chemotherapy resistance of esophageal squamous cell carcinoma. Oncol Res 31, 887–897 (2023).

55. Zhao, Z., Mei, Y., Wang, Z. & He, W. The Effect of Oxidative Phosphorylation on Cancer Drug Resistance. Cancers 15, 62 (2022).

56. Zhou, W. et al. A Comparative Transcriptome Between Anti-drug Sensitive and Resistant Candida auris in China. Front Microbiol 12, 708009 (2021).

57. Zhou, X., Liao, W.-J., Liao, J.-M., Liao, P. & Lu, H. Ribosomal proteins: functions beyond the ribosome. J Mol Cell Biol 7, 92–104 (2015).