A heterogeneous pharmaco-transcriptomic landscape induced by targeting a single oncogenic kinase

Ross M. Giglio[1], Nicholas Hou[2], Adeya Wyatt[2], Justin Hong[3], Lingting Shi[10], Mathini Vaikunthan[2], Henry Fuchs[2], Jose Pomarino Nima[2], Seth W. Malinowski[4], Keith L. Ligon[4,5], José R. McFaline-Figueroa[6], Nir Yosef[7,8,9], Elham Azizi[2,3,10,11,12], José L. McFaline-Figueroa[2,10,11,#]

[1]Department of Molecular Pharmacology and Therapeutics, Columbia University Medical Center, New York, NY 10032, USA.
[2]Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA.
[3]Department of Computer Science, Columbia University, New York, NY 10027, USA.
[4]Department of Oncologic Pathology, Brigham and Women's Hospital, Boston Children's Hospital, Dana-Farber Cancer Institute, Boston, MA 02215, USA.
[5]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.
[6]Center for Neuro-Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.
[7]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720, USA.
[8]Center for Computational Biology, University of California, Berkeley, Berkeley, CA 94720, USA.
[9]Department of Systems Immunology, Weizmann Institute of Science, Rehovot, Israel.
[10]Irving Institute for Cancer Dynamics, Columbia University, New York, NY 10027, USA.
[11]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA
[12]Data Science Institute, Columbia University, New York, NY 10027, USA.

[#]Correspondence: jm5200@columbia.edu

Abstract

Over-activation of the epidermal growth factor receptor (EGFR) is a hallmark of glioblastoma. However, EGFR-targeted therapies have led to minimal clinical response. While delivery of EGFR inhibitors (EGFRis) to the brain constitutes a major challenge, how additional drug-specific features alter efficacy remains poorly understood. We apply highly multiplex single-cell chemical genomics to define the molecular response of glioblastoma to EGFRis. Using a deep generative framework, we identify shared and drug-specific transcriptional programs that group EGFRis into distinct molecular classes. We identify programs that differ by the chemical properties of EGFRis, including induction of adaptive transcription and modulation of immunogenic gene expression. Finally, we demonstrate that pro-immunogenic expression changes associated with a subset of tyrphostin family EGFRis increase the ability of cytotoxic T-cells to eradicate tumor cells. Our study provides a framework that considers each agent's unique and often unknown poly-pharmacology to prioritize compounds that induce clinically favorable molecular responses.

Introduction

The response of cancer to therapy is driven by a complex interplay between a tumor's genetic background and the mode of anti-tumor therapy. The development of bulk shRNA and CRISPR-based genome-wide screens has drastically increased our ability to identify the genetic requirements of cancer cell response to therapy. These screens use gross phenotypic assays (e.g., viability), specific molecular measurements (reporter activity), or, more recently, single-cell profiling as readouts[1–4]. Variability across the drug's mechanisms of action and polypharmacology can also result in substantial heterogeneity in a tumor's response to therapy. Additionally, the classification of a drug by its predicted or previously annotated target has not always aligned with functional outcomes[5,6], and in some cases, has led to inefficacy or unwanted off-target toxicity in clinical settings[7,8]. Altogether, we lack a comprehensive understanding of how different approaches to targeting oncogenic activity lead to heterogeneity in the molecular response to therapy.

A prime example of variable response to therapy is observed in IDH-wild type glioblastoma (GBM) - the most common and aggressive primary brain cancer. Despite a multimodal treatment regimen composed of surgical resection, radiotherapy, chemotherapy, and tumor-treating fields, most patients succumb to their disease within 2 years of diagnosis[9,10]. Large-scale efforts have defined the genomic landscape of GBM. These studies identified near-ubiquitous overactivation of receptor tyrosine kinase (RTK) signaling, loss of the negative regulation of the RAS/MAPK and PI3K effector pathways, and inactivation of the Rb and p53 tumor suppressor pathways as core alterations in oncogenic signaling[11,12]. Unfortunately, this genetic characterization of the disease has yet to translate into a therapeutic gain despite numerous clinical evaluations of targeted therapies in GBM.

Amplification and over-activating mutations in the epidermal growth factor receptor (EGFR) tyrosine kinase are found in approximately half of all GBM patients[11], with the most prevalent mutation being truncation of exons 2-7 of EGFRs extracellular domain (EGFRvIII) which render the kinase constitutively active. Although targeting of EGFR has demonstrated clinical efficacy in EGFR-mutated non-small cell lung cancer, no such benefit has been identified in GBM despite strong selection for EGFR activity in the disease. This inefficacy is consequent to a myriad of factors, including poor brain penetrance of small molecule inhibitors across the blood-brain barrier, though few studies have focused on the mechanisms of innate and adaptive resistance to brain-penetrant agents. A case study detailing the clinical response with a brain-penetrant EGFR inhibitor has renewed interest in the use of these inhibitors in GBM[13]. More recently, a retrospective study found that a subset of patients treated with a combination of an EGFR inhibitor (osimertinib) and an anti-angiogenic (bevacizumab) experienced a long-term benefit[14]. In addition, resistance to the treatment was associated with secondary mutations in RTK-associated pathway components (*MET*, *IGF1R*, *PTEN*, *PDGFR*). These resistance mechanisms highlight the central role of EGFR and RTK signaling in the disease, the importance of understanding how GBM responds to the loss of EGFR activity, and the need to identify therapies that block RTK activity while leveraging their polypharmacology to minimize the emergence of resistant subclones.

Here, we apply ultra-high throughput single-cell chemical transcriptomic profiling to map variability in tumor cell response to targeting a single oncogenic event in one tumor type. We focus on the single-cell molecular profiling of patient-derived neurosphere GBM models (patient-derived cell lines or PDCLs) treated with 72 compounds and biomolecules with anti-EGFR activity and controls across ~1,800 unique conditions (PDCL x EGFRi x dose). We first examine the changes in gene expression networks associated with RTK signaling, tumor cell proliferation, and adaptive transcriptional resistance to identify EGFR inhibitors that decrease viability with minimal induction of a resistance program in surviving cells. We then use Multi-resolution Variational Inference (MrVI)[15], a hierarchical probabilistic model that learns cell-specific sample stratifications, to identify drug-dose combinations with similar effects on distinct molecular programs. We leverage the information on how each compound alters these programs in each genetic background to arrive at a molecular classification of EGFR inhibitors. Amongst the transcriptional programs modulated by distinct EGFR inhibitors, we focus on expression changes of clinical interest, primarily the upregulation of the antigen processing and presentation machinery (APM), identifying a subset of tyrphostin family inhibitors that regulate MHC class I signaling and demonstrating a functional outcome associated with upregulation.

Results

Single-cell chemical transcriptomics identifies shared and distinct expression changes driven by a drug's polypharmacology.

Transcriptomic profiling of chemically exposed cells can identify commonalities in molecular response across compounds with similar mechanisms of action[5]. Their recent application at single-cell resolution allowed for a detailed description of how chemical perturbation alters cellular states and provided insight into the mechanisms of action not apparent across population-averaged data[16]. However, studies have yet to examine whether this approach can detect subtle differences in response within a compound class that inform on a drug's off-target effects and polypharmacology.

To determine whether single-cell chemical transcriptomics profiling can identify differences in response across compounds targeting the same molecular activity, we devised a pilot screen using six small molecules that inhibit EGFR activity (afatinib, brigatinib, CUDC-101, EAI045, neratinib, osimertinib). Four of the six inhibitors target EGFR irreversibly (afatinib, EAI045, neratinib, osimertinib), one binds an allosteric site (EAI045), and two have been designed with activity against another enzyme (brigatinib[17]: an ALK inhibitor with anti-EGFR activity; CUDC-101[18]: a bifunctional EGFR & HDAC class I/II inhibitor). We exposed 3 established glioblastoma cell lines (A172, T98G, U87MG) and 1 PDCL (BT333) to 7 doses (0.005 - 10 μM) of each compound or DMSO vehicle control for 24 hours. To determine the effect of each compound on transcription, we processed conditions across two replicates via sci-Plex multiplexing and subjected cells to single-cell combinatorial indexing RNA-seq[19] (**Fig. 1A**). After filtering, we collected 188,746 single-cell transcriptomes across our 4 GBM models and 43 exposures for a total of 172 unique conditions (**Supp. Fig. 1A-B**).

Differential expression (DE) analysis identified robust changes in gene expression in response to exposure to 5 of the 6 EGFR inhibitors as a function of dose (median of 733 DE genes per EGFRi, FDR < 0.1%, normalized beta > 0.05, **Supp. Fig. 1B-C**). We used integration[20], dimensionality reduction[19,21], and Leiden community detection [22] to visualize the similarities and differences in response across our models, which varied as a function of the precise inhibitor used (**Fig. 1B-D)**. For example, exposure to high doses of either afatinib, brigatinib, neratinib or osimertinib led to the induction of a cell state associated with increased MTOR and decreased KRAS signaling (cluster 7, **Fig. 1D**). Conversely, cells exposed to high doses of the dual EGFR/HDAC inhibitor CUDC-101 gave rise to a distinct cell state associated with changes in TNF-α signaling and an IFN response (cluster 6, **Fig. 1D**). To further explore whether the response to each inhibitor was similar across our GBM models, we performed a joint analysis to summarize the covariance of gene-level effects across the union of dose-dependent DE genes, which we previously showed can recover commonalities in drug-induced expression programs[16] (**Fig. 1F-H, Supp. Fig. 1D**). The results of our analysis also suggest that transcriptional response to a compound is similar across our models.

We next used hierarchical clustering to group modules of dose-dependent DE genes that covary across our experiment, identifying 17 and 15 modules of genes (each composed of > 30 genes) that are similarly and uniquely upregulated or downregulated, respectively, by one or more EGFRis (**Fig I-L**, **Supp. Fig. 2A-C**, **Supp. Table S1**). Amongst these gene modules, we identified a 48 gene signature that was upregulated in a dose-dependent manner broadly across all EGFR inhibitors, leading to a measurable effect on transcription (shared EGFR module, **Fig. 1I-K**). Interestingly, this signature had a significant and negative correlation with overall survival across two GBM cohorts from Brennan et al.[11] and The Cancer Genome Atlas (TCGA PanCancer)[23] (**Fig. 1M** and **Supp. Fig. 2D**). Given that these cohorts are unlikely to contain patients treated with EGFR targeted therapy, this gene module may report on a general feature of populations of cells that survive cytotoxic or genotoxic therapy. Consistent with this hypothesis, this gene module correlates with increasing tumor grade and inversely correlates with survival across all gliomas (TCGA PanCancer, **Supp. Fig. 2D-F**).

In addition to identifying broadly shared programs, our analysis identified transcriptional changes specific to one or a subset of compounds. We identified modules of genes unique to CUDC-101 exposure, which likely report on transcriptional effects downstream of HDAC class I/II inhibition (**Fig. 1K**, **Supp. Fig. 2A-C**). We also identified a shared 70 gene signature between CUDC-101 and osimertinib-exposed cells (**Fig. 1L, Supp. Fig. 2A-C**). Interestingly, the EGFR targeting quinazoline functional group of CUDC-101[18] is most structurally similar to that of afatinib[24], not osimertinib, among the 6 compounds tested. Moreover, this module is weakly correlated with proliferation index (**Supp. Fig. 2G-H**), suggesting this shared program is not simply a consequence of loss of viability. Overall, and towards our goal of classifying compounds by their polypharmacology, our results show that a single-cell chemical transcriptomics approach captures differences between compounds targeting the same oncogenic alteration, including unexpected associations not readily explained by a drug's molecular structure.

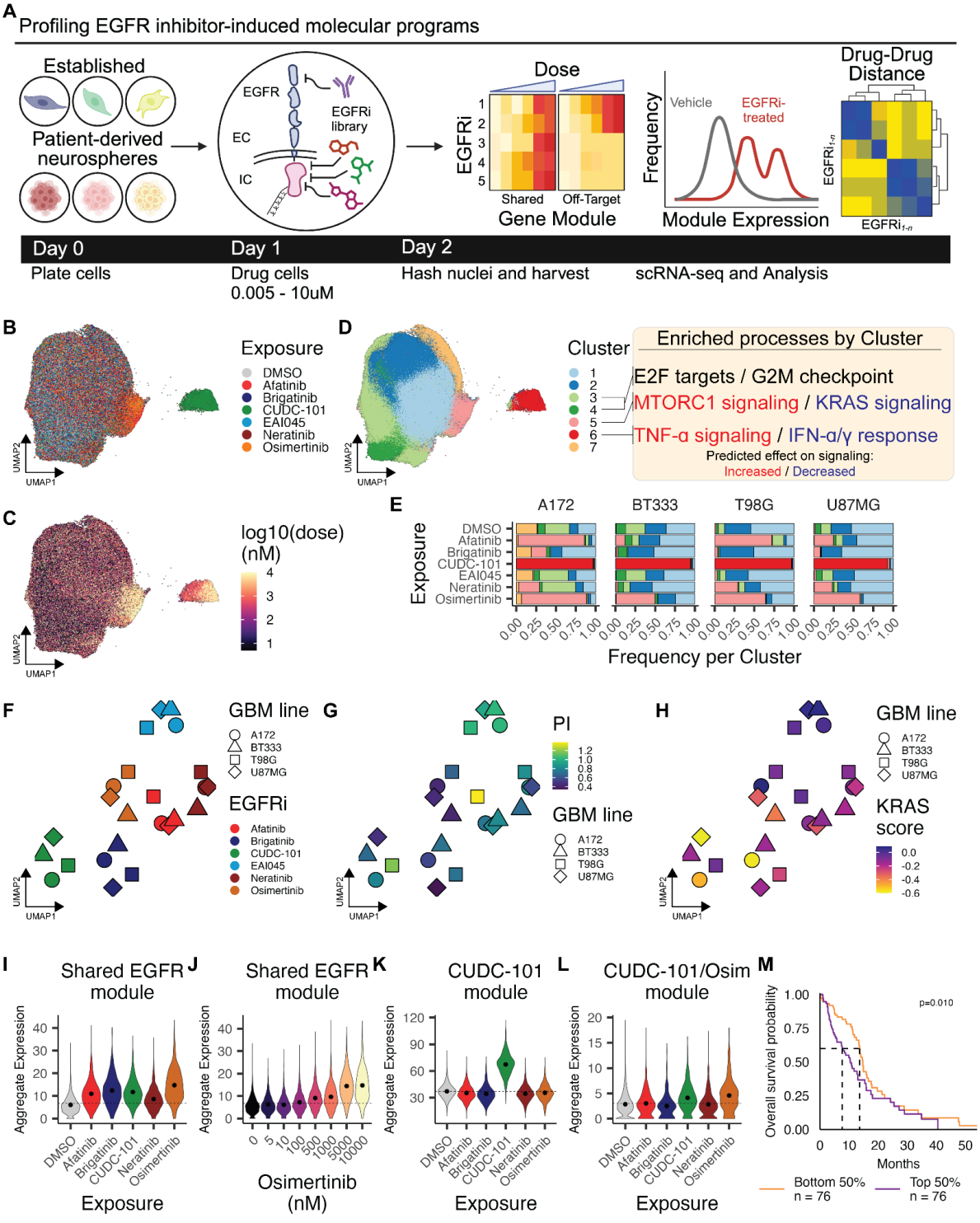


Figure 1. Single-cell chemical transcriptomics recovers drug-induced expression signatures that vary due to a compound's polypharmacology. A. Schematic depicting comprehensive screening of EGFR inhibitors in GBM. **B-C**. Integrated UMAPs of A172, T98G, U87MG, and BT333 GBM models exposed to one of 6 EGFR inhibitors or DMSO control. Color denotes the EGFR inhibitor (**B**) or the dose of inhibitor (**C**) that a cell was exposed to. **D**. Left: Integrated UMAP as in **B-C** colored by the cluster assignment from Leiden-based community

detection of in sample-integrated aligned PCA space. Right: Results of gene set enrichment analysis for MSigDB Hallmarks gene sets of biological processes that are significantly enriched (FDR < 5%, hypergeometric test) amongst differentially expressed genes (FDR < 0.1%, see Methods) for clusters associated with proliferation (clusters 3 & 4) and those that are dependent on EGFR inhibitor exposure (clusters 5 and 6). **E**. Distribution of GBM cells across clusters from **D**. **F-H**. UMAP projections summarizing the correlation (Pearson's coefficient of effects size estimates across the union of differentially expressed genes) between cell line-drug effects on transcription, annotated by cell line and exposure (**F**), relative proliferation index (PI) (**G**), and relative aggregate expression of genes associated with active KRAS signaling (**H**). **I-J.** Violin plots of the aggregate expression of a shared EGFR inhibitor transcriptional module across BT333 cells exposed to a 10 μM dose of the specified EGFRi or DMSO control (**I**) or increasing doses of osimertinib (**J**). Dashed line denotes the median aggregate expression of DMSO control. **K-L**. Violin plots of the aggregate expression of a CUDC-101 specific (**K**) or shared CUDC-101/osimertinib (**L**) transcriptional module across BT333 cells exposed to a 10 μM dose of the specified EGFRi or DMSO control. Dashed line denotes the median aggregate expression of DMSO control. **M**. Overall survival of GBM patients from Brennan *et al*. [11] grouped by the relative expression of the shared EGFR inhibitor transcriptional module (n = 152 patients, p: Wilcoxin-Gehan test).

Natively heterogeneous patient-derived GBM models vary in the expression of key RTK pathway components and evolution of cell states in response to EGFR inhibition.

To comprehensively define pharmacological variability in the transcriptional response to EGFR inhibition, we profiled three PDCL cultures. In addition to being amenable to screening a large number of exposure conditions, these models maintain the genetic and molecular diversity observed in GBM tumors[25–27]. The three models chosen are representative of the diversity of EGFR status in the disease. All lines display amplification of the receptor, and two of three carry extracellular mutations that render EGFR constitutively active (**Fig. 2A**). In addition, the tumor suppressor landscape of these models is also representative of GBM, including mutations in negative regulators of the RTK pathway (*NF1*, *PTEN*) and the p53 and Rb pathways (*TP53*, *CDKN2A*, *CDKN2B*) (**Fig. 2A**).

Multiplex single-cell RNA-seq profiling of PDCL cultures identified substantial heterogeneity in the expression of genes associated with EGFR and RTK signaling. For example, while all three cell lines expressed EGFR, the expression of EGFR dimerization partners varied, with *ERBB2*, *ERBB3*, and *MET* highest in BT112 cells (**Supp. Fig. 3A**). The expression of other RTKs and RTK pathway components were similarly variable across models, including the expression of FGFR and RAF family members (**Supp. Fig. 3A**). We also identified heterogeneity in the expression of genes associated with TP53, RB, and CDK signaling (**Supp. Fig. 3B-C**).

Our analysis identified subpopulations within two of the three PDCL models profiled (**Fig. 2B, 2G, Supp. Fig. 3D**), which could not be explained solely by differences in proliferation or transcripts captured per cell (**Fig. 2C-D**, **2H-I**, **Supp. Fig. 3E-H**). Examining the genes that vary across subpopulations (**Supp. Table S2**) revealed RTK pathway components whose expression differed,

including *BRAF* expression in BT228 cells and *FGFR1* expression in BT333 cells (FDR < 1%, **Fig. 2E-F**, **2J-K**). Previous studies have demonstrated that transcriptional plasticity renders a subset of cancer cells resistant to the effect of BRAF kinase inhibition in $BRAF^{V600E}$-driven melanoma cells[28]. Similarly, this pre-existing heterogeneity in RTK pathway-associated gene expression may contribute to their ability to adapt and recur after targeting oncogenic driver mutations such as EGFR.

We exposed PDCL models to each of 71 EGFR targeting small molecule inhibitors (**Supp. Table S3**), an EGFR neutralizing antibody (Panitumumab), a small molecule activator of EGFR (NSC228155), and puromycin (viability control) each at 4 doses, or PBS, Media, or DMSO controls for 24 hours and in replicate for a total of ~1,800 unique conditions. These EGFR targeting compounds serve as a comprehensive panel of EGFR inhibitors varying by generation, FDA approval, binding mechanism, activity against EGFR mutants, and known additional targets (**Fig. 2L**, **Supp. Table S3**). We subjected EGFR inhibitor-exposed samples to a modified version of our multiplex single-cell RNA-seq workflow (sci-Plex-v2) that incorporates advances in transcript capture[29,30] and captures isoform information[30] while remaining compatible with sci-Plex multiplexing, capturing a total of 181,485 cells with a median of 165 cells per unique combination of PDCL, drug, and dose (**Supp. Fig. 4A-D**), similar to the complexity per condition of prior large-scale single-cell chemical genomics screens [16]. In our large screen, we again identified subpopulations associated with increased *BRAF* and *FGFR1* expression. In BT228 cells, the proportion of cluster 3 (**Fig. 2M**) associated with *BRAF* high cells (**Fig. 2N**) appeared to expand moderately as a function of EGFR inhibition (**Fig. 2O**). This effect was more pronounced when examining the proportion of *BRAF* high cells exposed to osimertinib (**Fig. 2P**), the strongest inducer of the shared EGFR signature in our pilot screen (**Fig. 1I**). In contrast, BT333 *FGFR1* high cells did not appear to expand in response to EGFR inhibition (**Fig. 2Q-R**). These results are consistent with studies demonstrating that heterogeneity in the expression of factors associated with RTK signaling can be selected for upon treatment with inhibitors targeting RTK pathway overactivation[28].

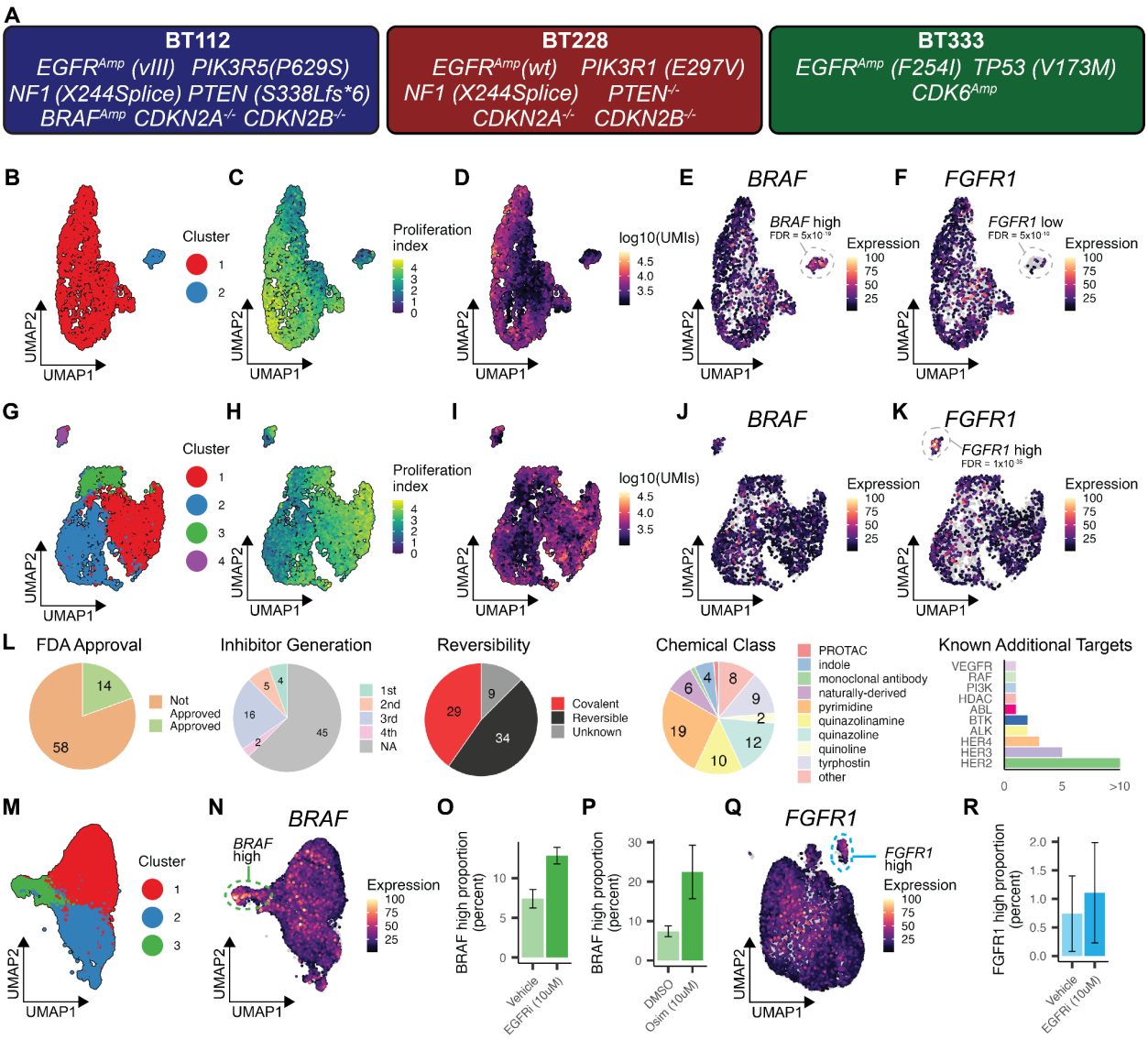


Figure 2. Inter- and intra-model variation in RTK pathway expression across patient-derived glioblastoma models and selection upon EGFR inhibition. A. Genetic status of EGFR and key GBM-associated genes. **B-K**. UMAP embeddings of untreated BT228 (**B-F**) and BT333 (**G-K**) PDCLs colored by cluster (**B, G**), the aggregate expression of genes associated with proliferation (**C, H**), and total number of transcripts (UMIs) captured per cell (**D, I**), the expression of *BRAF* (**E, J**) and *FGRF1* (**F, K**) as percent of maximal expression. Dotted circles in E-F & J-K denote subpopulations across BT228 and BT333 with a significant difference in *BRAF* and *FGFR1* expression (quasi-Poisson regression, Wald test, FDR < 1%). **L**. Characteristics of the panel of 72 EGFR modulating compounds and biomolecules in our screen. **M-N**. UMAP embedding of BT228 cells exposed to 72 EGFR modulating compounds and controls with cells colored by cluster (**M**) and BRAF expression as percent of maximal expression (**N**). **O**. Percentage of *BRAF* high BT228 cells (cluster 3 in panel **M**) for cells treated with vehicle controls or the highest dose of EGFRi's. Error bars denote the standard deviation from the mean across replicates. **P**. Percentage of *BRAF* high BT228 cells (cluster 3 in panel **M**) for cells treated with DMSO or 10μM osimertinib. Error bars as in **O**. **Q**. UMAP embedding of BT333 cells exposed to 72 EGFR

modulating compounds and controls with cells colored by *FGFR1* expression as percent of maximal expression. **R**. Percentage of *FGFR1* high BT333 cells (cluster 4 in **G**) for cells treated with vehicle controls or the highest dose of EGFRi's. Error bars as in **O**.

The global transcriptional response to EGFR inhibition is only partially shaped by a drug's effect on proliferative- and resistance-associated gene expression.

Exposure to our library of EGFR-targeting compounds had profound effects on transcription, with a mean of 6761 DE genes per PDCL and 662 DE genes per compound (FDR < 1%, **Supp. Fig. 4E**, **Supp. Table S4**). We first investigated how a compound's impact on global gene expression is linked to its potency in altering expression programs associated with proliferation. Across all PDCLs, we identified subsets of EGFRis that led to decreases in the expression of MKI67 and an aggregate score of proliferative gene expression that also displayed substantial effects on global transcription (**Fig. 3A, Supp. Fig. 5A-C & Supp. Fig. 6A**). For example, CUDC-101 and tyrphostin9 led to pronounced dose-dependent decreases in MKI67 and proliferative gene expression and had some of our screen's largest effects on global transcription (as # of DEGs). However, although MKI67 and total gene expression changes were strongly negatively correlated, the relationship appeared weaker for some inhibitors such as EGFRInhibitor in BT112, afatinib in BT228, and PD153035 in BT333 (**Fig. 3A, Supp. Fig. 6A**). Therefore, the results suggest that the mode of EGFR targeting differentially impacts the expression of diverse gene modules, with or without effects on proliferation.

Given the minimal efficacy of EGFR inhibitors in inducing a clinical response in GBM patients in the clinic, we next examined the expression of programs associated with acquired resistance to therapy. Recently, we identified an adaptive resistance signature across established GBM cell lines and glioma stem cell models in response to the inhibition of RTK signaling, most prominently by inhibition of MEK via trametinib[31]. Across the EGFRi signatures from our pilot screen, we found that the shared EGFR signature (**Fig. 1I-J**), associated with decreased patient survival (**Fig. 1M**, **Supp. Fig. 2D**) and tumor grade (**Supp. Fig. 2E**), had the highest overlap with our adaptive program (**Supp. Fig. 2C**), suggesting it perhaps reports on tyrosine kinase activity (at baseline or after adaptation). To determine whether EGFR inhibitors in our screen differentially activate this program, we scored cells by its aggregate expression (**Fig. 3B**). Amongst the compounds that led to the strongest induction of the adaptive signature were several compounds with pronounced effects on proliferative gene expression, such as osimertinib, AZ5104 and WZ3146 (**Fig 3A**, **Supp. Fig. 6B**), members of the pyrimidine chemical class of EGFR inhibitors, a class that is enriched amongst compounds with the strongest effect on adaptive gene expression (**Fig. 3B**). Importantly, this induction was not observed for our puromycin viability control, suggesting that induction of the adaptive program is not merely a consequence of loss of viability. This analysis also identified several compounds that led to decreased proliferative gene expression with minimal induction of adaptive gene expression. For example, MTX-211, a dual EGFR and PI3K inhibitor, displayed the strongest ability to affect proliferative gene expression while minimizing adaptive gene expression in all PDCL models (blue highlight in **Fig. 3A-B**, **Supp. Fig. 6B)**.

We next investigated the association between the induction of adaptive transcription and the ability of cells to persist upon repeated exposure to EGFR inhibitors. We exposed PDCLs to an initial dose of the drug at day 0, a rechallenge at day 3, and profiled cells 7 days post-initial exposure using sci-Plex-v2. Examining the dynamics of adaptive transcription between cells exposed to acute or repeated exposure revealed chemical and PDCL-dependent changes in the acquisition and maintenance of the adaptive resistance program. For instance, adaptive program scores decreased for BT228 and BT333 cells after repeated exposure to osimertinib but not BT112 (**Fig. 3C**, *top*). Examining the proportion of cells at each condition as a proxy for viability[16] revealed that decreases in the adaptive program were associated with decreases in cell viability to repeated osimertinib exposure in BT228 and BT333 but not BT112 (**Fig. 3C**, *bottom*). In contrast, cell viability decreased across all PDCLs in response to repeated MTX-211 exposure, consistent with minimal acute or long-term induction of the adaptive transcription program (**Fig. 3D**). More broadly, we observed that chemicals that increase adaptive transcription upon repeated exposures were less effective in decreasing viability compared to compounds where the program is not induced, maintained or subsided (**Fig. 3E**). Taken together, our results provide evidence that a survey of the differences in the induction of molecular programs across related inhibitors can prioritize chemicals based on maximizing cytotoxicity while minimizing the induction of resistance programs (**Fig. 3F**).

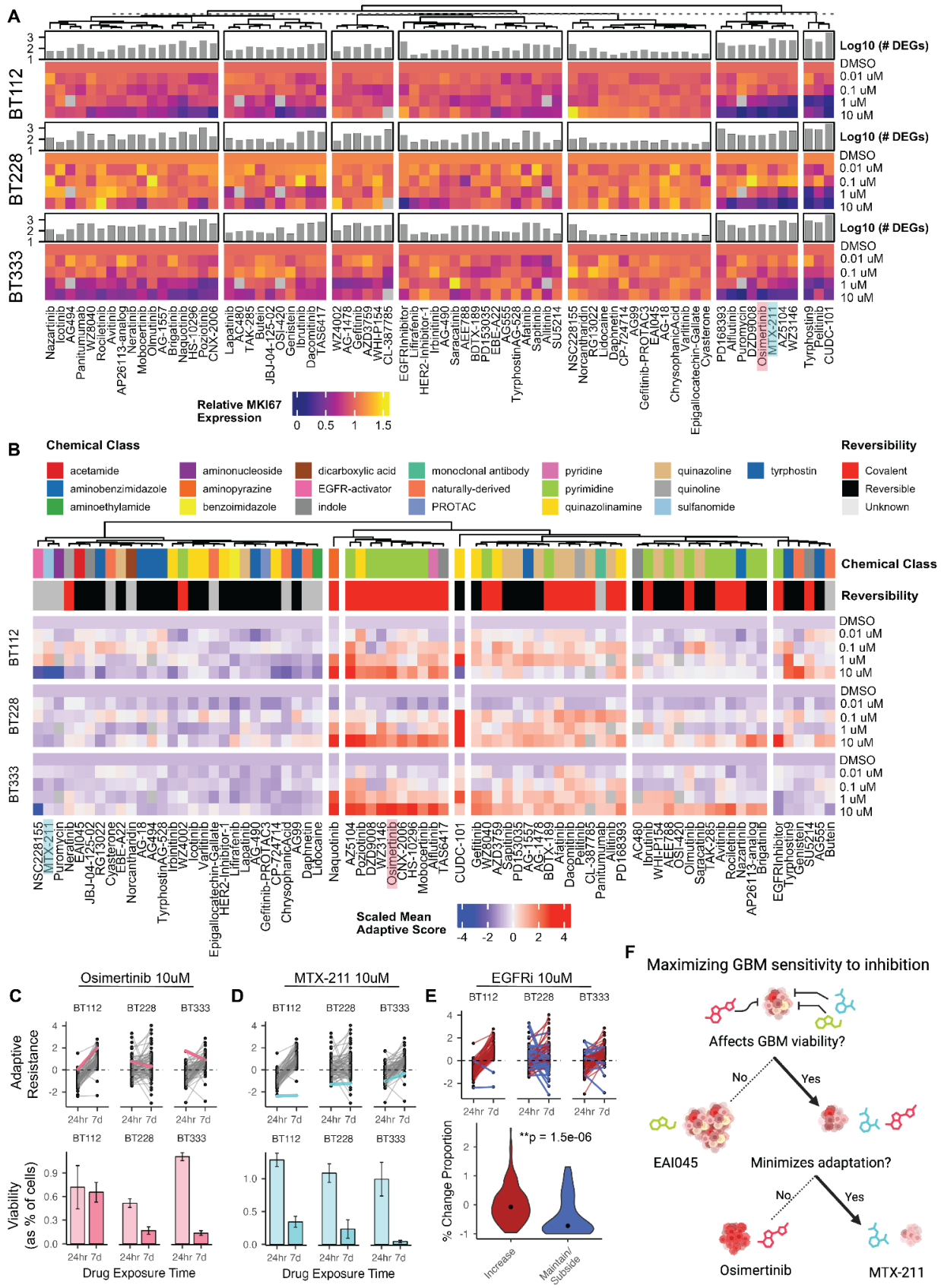


Figure 3. Response to EGFR inhibition is imperfectly correlated to the effect on proliferative transcription. A. Relationship between the total number of differentially expressed genes as a function of EGFR inhibitor dose (quasi-poisson regression, Wald test FDR < 1%) for the specified EGFR inhibitor (top annotations) with the relative expression of the proliferation marker MKI67 across BT112, BT228, and BT333 PDCLs. The chemicals osimertinib and MTX-211 are highlighted pink and blue, respectively. **B.** Heatmap of z-scored mean aggregate expression of genes previously associated with adaptation[31] to inhibition of the RTK pathway for BT112, BT228, and BT333, annotated by chemical classification and EGFR binding mechanism. The chemicals osimertinib and MTX-211 are highlighted pink and blue, respectively. **C-D**. Adaptive resistance changes between 24hr and 7d time points for 10uM EGFRi, with osimertinib (**C**) and MTX-211 (**D**) highlighted (top). Viability, as the proportion of cells treated with osimertinib (**C**) and MTX-211 (**D**) compared to all 10uM treated cells with respect to time point (bottom). Osimertinib decreases viability for BT228 and BT333 but not BT112, while MTX-211 decreases viability across PDCLs, consistent with adaptive resistance changes. **E.** Adaptive resistance changes between 24hr and 7d time points for 10uM EGFRi (top), with inhibitors highlighted and binned by adaptive resistance dynamics, specifically as increasing vs maintain/subside. The group in which there is minimized induction of the program is characterized by a significantly larger drop in viability (bottom, Wilcoxon test). **F.** A model demonstrating the prioritization of inhibitors for use in GBM that decrease viability and minimize the induction of adaptation.

A classification of anti-EGFR agents by their induction of molecular programs.

We next sought to classify our panel of EGFRis and summarize their relative effect on transcription as well as their unique and shared polypharmacology. We used MrVI[15], a deep generative model that performs sample stratification at single-cell resolution, to annotate EGFRi-induced transcriptional responses and classify EGFRis by their ability to induce them (**Fig. 4A**). MrVI's hierarchical probabilistic framework presumes cells to be generated from nested experimental designs, such as our study, in which each sample, defined as a drug-dose treatment condition, is drawn from one of several experimental batches. MrVI learns two distinct latent feature spaces from a given scRNA-seq dataset, a sample-unaware *U*-space that is decoupled from the sample-of-origin, and a sample-aware *Z*-space that incorporates sample-of-origin effects while accounting for technical factors across both spaces. These latent spaces can be used to dissect sample-specific heterogeneity in cell states across populations of interest in downstream analyses (**Supp. Fig. 7**).

We trained separate models for each of our 3 PDCLs with unique drug-dose combinations defined as the sample-of-origin and batch replicate as the technical factor. We used the union of the top 100 highly variable genes for each unique PDCL-drug subset as features (total of 2,767 genes). MrVI defines the distance between samples for a given cell as the Euclidean distance between sample-specific counterfactual cell states, also known as the sample-sample distance. The single-cell resolution of the sample-sample distances reveals heterogeneous transcriptional responses to drug-dose treatment, even within individual drug-dose conditions (**Supp. Fig. 10, Supp. Fig. 11 & Supp. Fig. 12**). Hierarchical clustering of the mean sample-sample distances across cells revealed 22, 21, and 23 distinct groups of drug-dose combinations that resulted in similar

transcriptional responses in BT112, BT228, and BT333 PDCLs (termed **R**esponse **M**odules or RMs) (**Fig. 4B** & **Supp. Fig. 8A**). For example, in BT228, control treatments (DMSO, PBS, and media) and cells exposed to several EGFRis at low doses comprised RM14, which altogether represent exposures with minimal effects on transcription. The other response groups vary in their composition of chemicals and doses; for instance, RM1 and RM3 are comprised of CUDC-101 at high (1uM and 10uM, respectively) doses, while RM8 is represented by 29 unique chemicals ranging from 100nM to 10uM (**Supp. Fig. 9**).

We used covariate-specific differential expression analysis to identify differentially expressed (DE) genes in cells of each RM compared to DMSO as a function of the counterfactual-based sample-sample distances identifying 563, 929, and 1021 DE genes for BT112, BT228, and BT333 PDCLs (LFC > 0.1 and FDR < 5%) (**Supp. Fig. 13, Supp. Fig. 14 & Supp. Fig. 15**, see *Methods*). Gene set enrichment analysis across DE genes using the MSigDB Hallmark, MSigDB Oncogenic Signatures, Kinase Perturbations from GEO up and Kinase Perturbations from GEO down gene set collections revealed that RMs are enriched for genes associated with diverse biological processes, including the cell cycle, PI3K/mTOR signaling, kinase perturbation, induction of epithelial to mesenchymal transition (EMT) processes, increased KRAS or BRAF signaling, and TNF-alpha signaling through the NFkB pathway (**Fig. 4C**, **Supp. Fig. 16**).

We next sought to use the ability of specific EGFRis to populate shared and distinct RMs to incorporate response modules across PDCLs. We hierarchically clustered RMs across PDCLs based on the normalized covariate-specific log-fold changes (LFC) in gene expression for the union of RM DE genes (**Fig. 4A, Supp. Fig. 8B**, **Supp. Table S6**). Although these RM clusters are largely occupied by one PDCL, they allow us to further summarize drug effects by collapsing drugs with similar RM cluster membership profiles (**Fig. 4D-E**, **Supp. Fig. 17**). This process resulted in 14 **T**ranscriptional **D**rug **C**lasses (TDCs) across 24 RM clusters with similar PDCL- and drug-specific responses (**Supp. Table S5**).

Seven TDCs consisted of multiple agents, while seven TDCs consisted of single-agents, some of which were expected based on known protein target profiles. For example, CUDC-101 (EGFRi/HDACi) and NSC228155 (EGFR activator, dimerization-domain binder[32]) were classified into distinct groups, TDC1 and TDC3, respectively. Control chemicals DMSO, PBS, and media were grouped in TDC6 along with compounds like EAI045, EBE-A22 (the inactive metabolite of PD153035), and cyasterone all of whom had little effect on transcription and viability (**Fig. 3**). Similarly, closely related compounds with strong effects on gene expression and similar EGFR off-target profiles grouped together; for instance, the ALK inhibitor brigatinib and its analog-AP26113 were both members of TDC11, and osimertinib and its active metabolite AZ5104 were both members of TDC9.

Interestingly, several drugs fell into or were excluded from TDCs in a manner that is not readily explained by their specificity (wild type vs mutant EGFR), drug generation, or chemical class. For instance, saracatinib, a SRC family TKI[33] with an order of magnitude higher activity against wild type vs. mutant EGFR[34], clustered within TDC11 with WZ3146 and nazartinib, both mutant EGFR (T790M) selective inhibitors[35,36]. Recently, WZ3146 was also found to target the SRC family

kinases LYN and FYN[37], suggesting that nazartinib may have a similar off-target profile. TDC11 also contains the pan-TKI inhibitors brigatinib, its analog-AP26113, and the molecular degrader gefitinib-PROTAC3. Interestingly, structural predictions and in vitro studies suggest that gefitinib can target SRC and the Src family kinase LCK as off-targets[38]. The differential TDC grouping for gefitinib-PROTAC3 and gefitinib and the former's inclusion in TDC11 may highlight an unknown effect specific to the molecular degrader with regard to decreased activity of Src-family kinases.

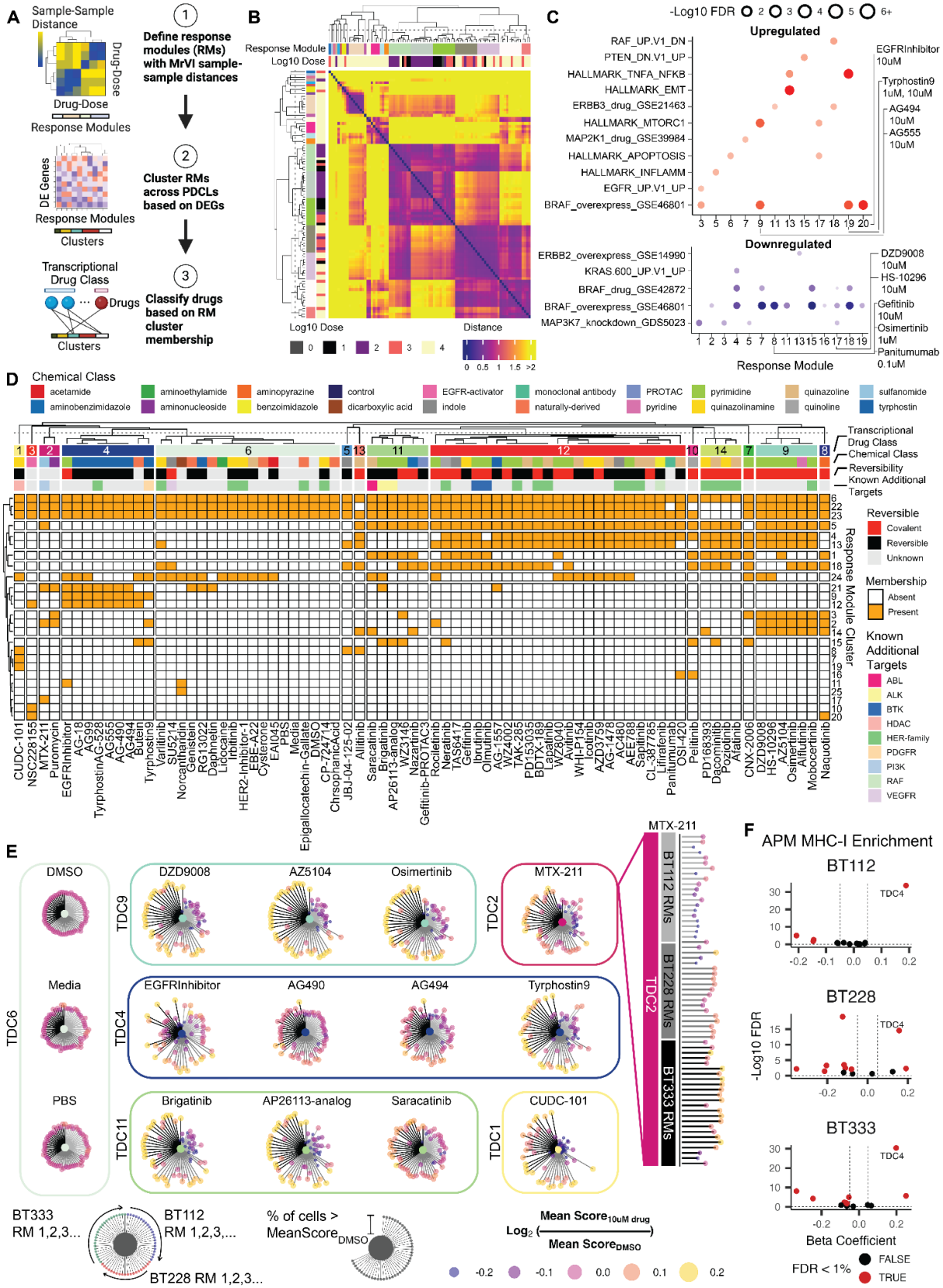


Figure 4. EGFR inhibitor classification with MrVI. **A.** Schematic depicting utilization of MrVI and covariate-specific differential expression analysis to define transcriptional drug classes of EGFRi **B.** Heatmap of the mean normalized distances between drug-dose combinations across BT228 cells, estimated with MrVI. Hierarchical clusters were identified and labeled as Response Modules (RMs) **C.** Gene set enrichment analysis results for the up-regulated (top) and down-regulated (bottom) genes of BT228 RM transcriptional signatures (FDR < 5%). Representative gene sets were chosen for visualization, and full GSEA results can be found in **Supp. Fig. 16**. **D.** Binary heatmap of each drugs' membership across RM clusters. Hierarchical clustering of the drugs reveals generalized transcriptional drug classes (TDCs) of EGFRi. **E.** Circularized lollipop plots for individual drugs. An example plot before circularization is depicted for MTX-211. The color of the center indicates the TDC from **(D)**. Each individual lollipop extending from the center represents the chemical's 10uM mean aggregate expression of a RM's upregulated genes detected by covariate-specific DE analysis, ordered by PDCL. The length of the line is proportional to the percent of cells within the drug's high dose population expressing the RM signature score greater than the mean score amongst DMSO-treated cells respective of PDCL. The color of the line's point is the log2-fold-change in mean RM signature score between the drug's high dose population and DMSO-treated cells. **F.** Volcano plots displaying enrichment for APM MHC-I score across all TDCs for each PDCL. Beta coefficients and FDR values were obtained from a general linear regression model, in which APM MHC-I score was fitted to a singular term TDC. TDC4, represented by tyrphostin-family compounds, is highlighted as enriched for APM MHC-I in each PDCL.

A subset of EGFR inhibitors modifies GBM tumor cell immunogenicity.

GBM is one of the least immunogenic tumor types, displaying low expression of the antigen processing and presentation machinery (APM) and low infiltration of cytotoxic immune cells[39,40]. Several studies have identified changes in tumor cell immunogenicity as a function of inhibiting the kinase activity of EGFR[41,42] or MEK[43,44]. Furthering prior studies, our profiling across a large number of compounds can provide insight into chemical classes of EGFR inhibitors that drive pro-immunogenic response and report on whether an off-target effect contributes to this clinically beneficial outcome. We therefore evaluated whether any TDCs were enriched for genes associated with the MHC class I (MHC-CI) or MHC class II (MHC-CII) antigen processing and presentation machinery[45,46]. Our analysis identified TDC4, induced by tyrphostin compounds such as AG18, AG99, AG490, AG494, AG555, and tyrphostin9 as significantly enriched for MHC-CI APM genes across all three PDCLs (**Fig. 4E-F**).

To more directly determine whether EGFR inhibition leads to increased APM expression in GBM PDCLs, we scored EGFRi-exposed PDCL cells for the expression of MHC-CI and MHC-CII genes. We used linear regression to identify compounds with a dose-dependent effect on APM expression. We identified 42 compounds significantly affecting MHC-CI APM expression and 46 that significantly affected MHC-CII APM expression across one or more PDCLs (**Fig. 5A-B**). Consistent with our TDC analysis, there was a marked enrichment for tyrphostin class compounds amongst MHC-CI APM-modifying EGFR inhibitors (**Fig. 5A**). Of note, several of these tyrphostin compounds had minimal effects on proliferative and moderate effects on global gene expression

(AG18, AG490, AG494, AG555; **Fig. 3A**) yet led to consistent increases in APM scores. Examining changes at the level of individual genes revealed that changes induced by tyrphostins are distinct from other classes. For example, across conditions that increased MHC-CI APM expression in 2 or more PDCLs, all exposures were associated with increased *B2M*, *CALR*, *CANX*, *HLA-A*, *HLA-B*, and *HLA-C* expression. In contrast to tyrphostins, however, exposure to CUDC-101 either did not alter or was associated with the downregulation of factors such as *ERAP2*, *PSMB5*, *PSMB6*, *PSMB7*, and *TAPBP* (**Fig. 5C**, dashed boxes).

Surprisingly, our puromycin viability control induced APM expression in 2 of 3 PDCLs. Previous studies have reported that low-dose puromycin has anti-tyrosine kinase activity independent of its effect on translation[47]. However, given that only a narrow subset of EGFRis modulate APM, it is more likely that other factors downstream of puromycin exposure (puromycilation, inhibition of translation) alter APM expression independent of modulation of tyrosine kinase, or at least EGFR, activity. We also identified broader effects of EGFRis on signaling pathways associated with immunogenicity, such as significant increases in *JAK1*, *STAT1*, and *STAT3* expression by the dual EGFR/HDAC targeting agent CUDC-101 (**Fig. 5E**). However, while changes to MHC-CI APM expression were observed across all models, changes in MHC-CII APM were largely confined to BT333 (**Fig. 5B, 5D**), such as increases in the MHC-CII antigen-presenting subunits *HLA-DMA* and *HLA-DMB*[47,48] (**Fig. 5F-G**).

Lastly, we sought to determine whether changes in gene expression associated with tumor immunogenicity lead to functional changes in the interactions between GBM and immune cells, specifically, the ability of tyrphostin EGFRi's to alter cytotoxic T-cell-mediated tumor cell killing. We used an allogeneic GBM:T-cell co-culture system consisting of GBM cells overexpressing the tumor antigen NY-ESO-1[49] and T-cells expressing anti-NY-ESO-1 T-cell receptors (TCR)[50], similar to recent studies modeling GBM:T-cell interactions in vitro [51]. We pre-treated NY-ESO-1 GBM cells with a range of doses to tyrphostin family EGFRi's (AG18, AG490, AG494, AG555, tyrphostin9) that increased APM expression across our PDCLs, EGFRi's that do not alter APM expression but have a range of effects on transcription (CL-387785, OSI-420, osimertinib, PD153035) and DMSO vehicle control. We used CFSE labeling of GBM cells, measuring the fluorescent area of each well to quantify how EGFRi exposure alters tumor cell killing by cytotoxic T-cells (**Fig. 5H**). We then co-cultured pre-treated NY-ESO-1 GBM and terminally differentiated anti-NY-ESO-1 $CD8^+$ T-cells at GBM:T-cell ratios of 1:0, 1:0.25, and 1:0.5, the latter ratios resulting in approximately 5 to 20% decreases in GBM cell viability 48 hours post-co-culture, respectively. To account for differences in viability across inhibitors, we normalized the fluorescent area to that of the no T-cell condition (**Fig. 5I**). Exposure of cells to tyrphostin9, the strongest inducer of MHC-CI APM machinery in the screen, led to a significant increase in T-cell mediated killing (**Fig. 5I, 5J & Supp. Fig. 20**). In addition, exposure of cells to AG18, AG490, AG494, and AG555 led to moderate increases in T-cell mediated killing at the lower GBM:T-cell ratio of 1:0.25 (**Supp. Fig. 20**), consistent with their effect on APM expression (**Fig. 5A, 5C**).

To the best of our knowledge, the known polypharmacology of these inhibitors does not provide a straightforward explanation of which molecular target is responsible for these changes in immunogenicity. However, our results suggest that EGFR is not the primary effector, given the

small subset of compounds that induce the response (**Fig. 5A**), their differing potency in modifying proliferative gene expression (**Fig. 3A**), and the lack of compensatory feedback that characterizes the most potent EGFR inhibitors in our screen (**Fig. 3B**). Our findings imply that the choice of EGFR inhibitor can result in complex molecular state changes with the potential to alter the immunogenicity of GBM and warrant further investigation for their implications for GBM immunotherapy.

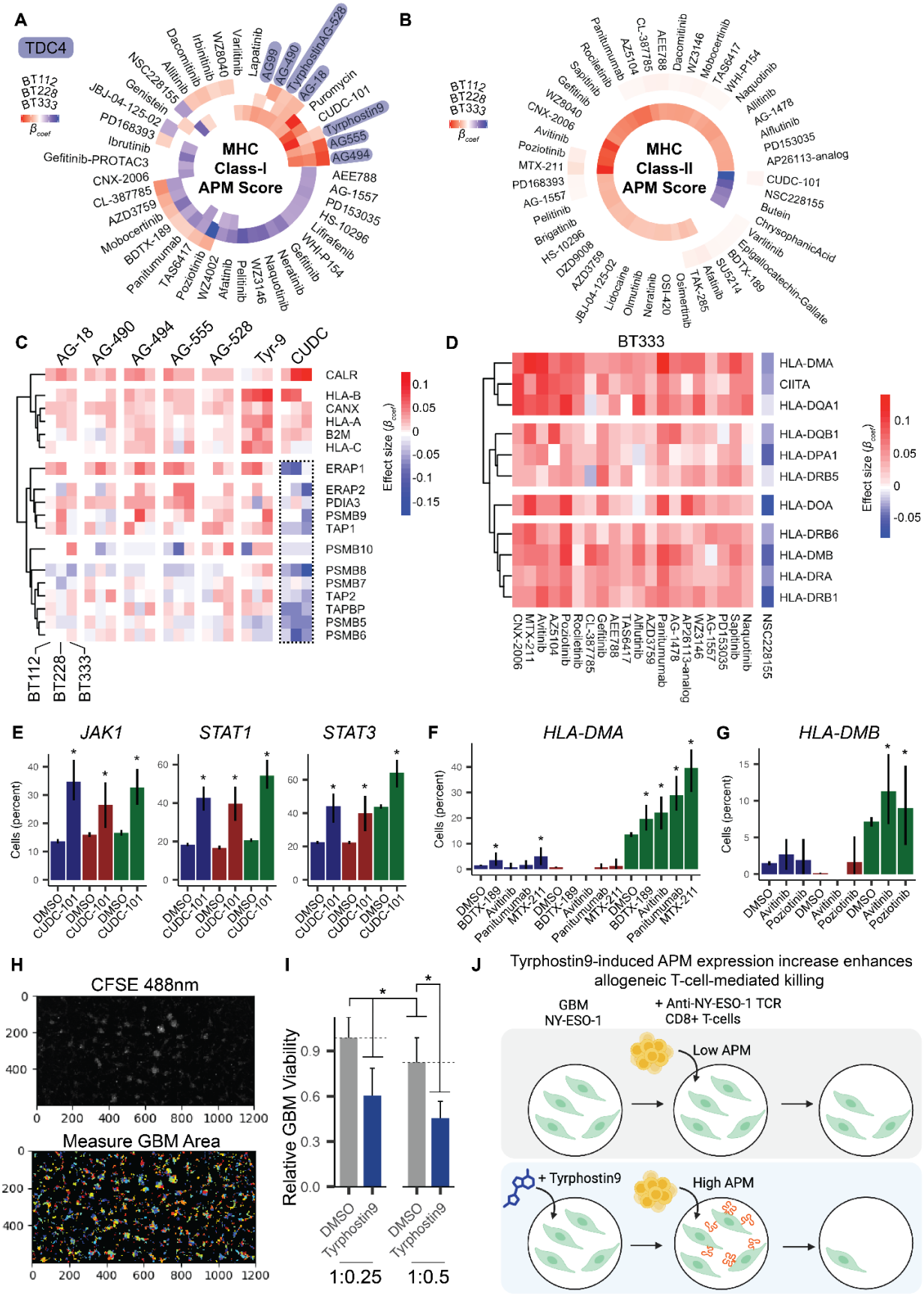


Figure 5. A subset of EGFRi's modulate immunogenicity. A-B. Circos heatmaps of beta coefficients for the dose term of a linear regression model of the effect of EGFR inhibition on the expression of genes associated with the MHC class I (**A**) or II (**B**) antigen processing and presentation machinery (APM) in one or more PDCLs. Only significant coefficients (FDR < 5%, Wald test) are shown. **C**. Heatmap of the effect of EGFR inhibition on genes associated with the MHC-CI APM for EGFRi's that significantly increase aggregated MHC-CI APM expression in two or more PDCLs from (**A**). **D**. Heatmap of the effect of EGFR inhibition on genes associated with the MHC-CII APM for a subset of EGFRi's that significantly increase aggregated MHC-CI APM expression in BT333 from (**B**). Note that EGFR activation using NSC228155 leads to decreases in MHC-CII APM. **E**. Percent of cells expressing *JAK1*, *STAT1* or *STAT3* after exposure to 10 μM CUDC-101 or DMSO vehicle control. Colors denote PDCL models: navy = BT112, dark red = BT228, dark green = BT333. Asterisks denote exposures that lead to a significant change in exposure relative to control (FDR < 5%). **F-G**. Percent of cells expressing *HLA-DMA* (**F**) or *HLA-DMB* (**G**) after exposure to the top dose of the specified agent. Asterisks denote exposures that lead to a significant change in exposure relative to control (FDR < 5%). Colors as in (**E**). **H**. Example 488nM fluorescent CFSE image tile and identification of GBM-covered area with CellProfiler software. **I**. U87MG NY-ESO-1 cell viability post-10 μM tyrphostin9 or DMSO control in combination with increasing T-cell exposure (1:0.25, 1:0.5) quantified as CFSE pixel area of GBM:T-cell conditions 1:0.25 and 1:0.5, relative to 1:0 (drug alone) ($n_{DMSO,T\text{-}cell}$ = 18, $n_{Tyrphostin9,T\text{-}cell}$ = 9). Wilcoxon test, * denotes p < 0.01. **J.** Proposed mechanism of tyrphostin9-induced APM machinery increasing GBM susceptibility to T-cell-mediated killing.

Discussion

Biochemical and chemical proteomic studies have established the fact that kinase inhibitors have varying degrees of selectivity for their intended targets[52–55]. In some cases, a drug's higher degree of polypharmacology may enhance clinical efficacy[56], while for others, off-target effects may result in unacceptable side-effects that lead to its failure in the clinic. In the context of glioblastoma, clinical trials for EGFR inhibitors have yielded little success despite the high prevalence of activating EGFR mutations in the patient population and despite the success in targeting EGFR across other tumor types. It is apparent that the molecular response in targeting this oncogenic activity is not fully understood. Here, in targeting EGFR with a diverse inhibitor library composed of small molecules and a biologic -- each with different chemical structure and binding properties -- we define a landscape of responses to an entire inhibitor class and further group the chemical agents into transcriptional drug classes based on the shared induction of molecular responses.

Our high-throughput targeted screen allowed for the comparison of inhibitor efficacy by standard metrics such as proliferation (e.g. MKI67 or aggregate proliferation score) obtained with traditional viability-based assays. However, we show that an inhibitor's impact on viability is not always correlated with global transcriptional response, indicating that inhibitor responses extend beyond viability, consistent with previous chemical genomic profiling efforts[5,16]. Moreover, we show that EGFRi differentially induce an adaptive resistance molecular response recently associated with MEK inhibition[31] and that accounting for this adaptive response will help lead to identifying clinically effective compounds.

The generation of single-cell chemical genomics datasets necessitates the development of novel computational approaches that efficiently summarize the large-scale high-dimensional data to gain novel biological insight into the effects of chemical perturbation on cellular and molecular states. Recent studies have developed techniques that study transcriptional differences in patients, cell types/states, and molecular perturbation screens[57–59]. In addition to the development of large data resources that can help refine these efforts, we demonstrate the suitability of MrVI[15], a recently developed deep generative model designed to estimate sample (drug-dose) heterogeneity at the cellular level while jointly correcting for batch effects, for summarizing transcriptional responses in large-scale genomics screens and grouping chemical agents by their shared induction of cellular states. We show that our experimental and analytical framework to refine the grouping of chemical agents targeting the same oncogene on the basis of their induction of distinct molecular programs is highly sensitive, revealing associations that reflect a drug's polypharmacology. This framework enabled the discovery of off-target effects related to increased immunogenicity due to the induction of the antigen processing and presentation machinery. Future rational design of inhibitors could incorporate similar approaches to identify compounds that simultaneously block oncogenic kinase activity and have desirable off-target effect profiles that induce responses synergistic with combination immunotherapy.

Our results provide motivation for the single-cell profiling of chemical inhibitors against a single target toward the multiplex identification and annotation of designed and off-target molecular programs. Thanks to advances in combinatorial indexing single-cell RNA-seq library generation methods[29,30], hundreds of thousands to millions of cellular transcriptomes across thousands of unique conditions can be obtained robustly and at reasonable cost[16,31,60]. Previous studies have included large-scale efforts to establish atlases of molecular responses to a massive amount of perturbations across a spectrum of cancer cell lines, utilizing probe-based fluorescence for targeted gene expression measurements[5], with limitations in drug repurposing applications [61]. While these efforts can be beneficial for applications like annotating broad inhibitor mechanisms of action, there is a need for high-resolution cellular resolved characterization of responses to chemicals, limited to a chemical class and targeting a specific cancer type. Future endeavors that would further strengthen our approach include novel ways to incorporate methods like lineage tracing[62], beneficial for answering questions of clonal expansion under inhibition[63], across large-scale experiments with limited cell, and therefore clonal, coverage per condition. Additionally, while we obtain a global view of EGFRi responses by studying three genetically-diverse models, future screens will need to incorporate EGFR genotype as a variable across a larger number of models to discover the effects of GBM-specific mutations or through the study of isogenic alterations in wildtype backgrounds.

Materials and Methods

Cell culture

A172, T98G, and U87MG glioblastoma cell lines were purchased from ATCC. Cells were cultured in DMEM media (ThermoScientific) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin (P/S, ThermoScientific) according to ATCC instructions. Glioma neurosphere cell lines BT112, BT228, and BT333 were obtained from the DFCI Center for Patient Derived Models (CPDM) under a material transfer agreement and maintained as described previously (Touat et al 2020 Nature). Briefly, cells were grown in Neurocult NS-A Proliferation Media (StemCell) supplemented with 0.0002% heparin (StemCell Technologies), EGF (20 ng/ml), and FGF (10 ng/ml; Miltenyi) in a humidified atmosphere of 5% CO2 at 37 °C on low-attachment plates [64] and were dissociated with Accutase (StemCell Technologies) for passaging and plating.

Pilot screen chemical perturbation procedure and materials

Cells were plated at a density of 2.5E4 cells/well onto 96-well flat-bottom plates for adherent cell lines and V-bottom plates for spheroid cell lines and were allowed to acclimate overnight. Chemical agents (SelleckChem) were purchased as a powder and resuspended in DMSO as 10mM (afatinib, CUDC-101, EAI045, neratinib, osimertinib) or 5mM (brigatinib) stocks according to the manufacturer's instructions. 10-fold dose concentrations (5nM, 10nM, 50nM, 100nM, 500nM, 1uM, 5uM, 10uM) were prepared in media as 10X stocks and were added to cells with all wells were normalized to 0.2%v/v DMSO (final conc. in well). Nuclei were hashed, fixed, and harvested 24 hours post-exposure.

Large screen chemical perturbation procedure and materials

Cells were plated at a density of 2.5E4 cells/well on 96-well V-bottom plates as in the pilot screen. Chemical agents (SelleckChem) were purchased as frozen 10mM stocks in respective solvents (DMSO, 1X PBS). 10-fold dose concentrations (0.01uM, 0.1uM, 1uM, 10uM) were prepared in media at 10X stocks and added to cells. Wells were DMSO or PBS normalized according to drug solvent, and vehicle-treated wells were included. Nuclei were hashed, fixed, and harvested at 24 hours post-exposure. For 7d time point exposures, cells were drugged initially in the same manner as in the 24hr condition. At 3 days post-exposure, wells were rechallenged with 10uL of drug in media equimolar to the well condition. Finally, nuclei hashed, fixed, and harvested at 7 days post-initial exposure.

Nuclei hashing and fixation

Nuclei hashing and fixation procedures were adapted from Srivatsan et al. and Sziraki et al. Briefly, adherent cells were trypsinized and moved to V-bottom plates, and spheroid cells were dissociated in place. Upon washing with ice-cold 1X PBS, cells were lysed with EZ Lysis Buffer (Sigma) supplemented with 1% diethyl pyrocarbonate (Sigma), 0.1% SuperaseIn RNase Inhibitor (Thermo), and 500 fmol of hashing oligo. After lysis, nuclei were fixed with the addition of 1.25% formaldehyde in 1.25X PBS (final well conc. 1% and 1X, respectively) and incubated on ice for 10 minutes. Nuclei were pooled into a plastic reservoir and moved into a 50mL conical for centrifugation at 650xg for 5 minutes at 4°C. Supernatant was removed from the nuclei pellet, and nuclei were washed once with nuclei suspension buffer (NSB; 10 mM Tris-HCl, pH 7.4, 10 mM

NaCl, 3 mM MgCl2, 1% Superase RNA Inhibitor (Thermo Fisher), 1% 0.2mg/mL Ultrapure BSA (New England Biosciences)). Nuclei were resuspended in NSB, slow-frozen in 10% DMSO, and stored at -80°C until sci-RNA-seq processing.

Library preparation and sequencing

Hashed nuclei were thawed and subjected to 3-level combinatorial indexing protocols adapted from previous methods (Cao et al, Martin et al, Sziraki et al). Nuclei were spun, resuspended in NSB, and sonicated at low power for 12s (Bioruptor). Upon counting, 21uL nuclei were moved to 96-well low adhesion PCR plates with 2uL 10mM dNTP, 2uL 100uM indexed oligo-shortdT primers, 2uL 100uM indexed random hexamer primers, and 14uL of a reverse transcription master mix consisting of 14.29% 100mM DTT, 14.29% 100mM RNaseOUT Ribonuclease Inhibitor, 57.14% 5X SuperScript IV First-Strand Buffer, and 14.29% SuperScript IV Reverse Transcriptase. Reverse transcription was carried out with an increasing temperature gradient. Post-reverse transcription, nuclei were pooled and distributed as 10uL into a 96-well plate(s) for ligation steps. Briefly, 8uL of indexed ligation primers were added to each well, along with a 4.8uL 3:2 master mix of T4 ligase buffer:T4 ligase (New England Biosciences, NEB) and 9.4uL of nuclei buffer with BSA (NBB; 10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 1% 0.2mg/mL Ultrapure BSA). Ligation was carried out at 25°C for 1 hour. Resulting nuclei were pooled, washed with NBB, and distributed as 1500 nuclei in 5uL NBB per well, where some plates were stored for future processing. Next, 5uL of a second strand synthesis mix consisting of 60% elution buffer (Qiagen), 27% second strand synthesis buffer (NEB), and 13% second strand synthesis enzyme mix (NEB) was added, and second strand synthesis was carried out at 16°C for 3 hours. Post-second strand synthesis, tagmentation was performed at 55°C for 5 min after the addition of 1/50uL of N7-adaptor loaded Tn5 and subsequent quenching with DNA binding buffer (Zymo) for 5 min at room temperature. Resulting dsDNA was purified using a 1X SPRIbead clean-up within the 96-well plate, and the resulting product was subjected to USER digestion (80% ddH2O, 10% 10X rCutsmart, 10% USER enzyme (NEB)). The dsDNA was eluted in buffer EB then moved to a clean 96-well plate. In addition to 16uL of eluted product, 2uL P5 PCR primer and 2uL P7 PCR primer were added to wells in an indexed well-specific combination. Further, 20uL 2X NEBnext PCR master mix (NEB) was added, and PCR to add the adaptors was carried out. The final PCR product was pooled and subjected to a 0.7X SPRIbead cleanup for library cDNA purification and 1X cleanup for hash fraction purification. Library concentrations were determined by Qubit (Invitrogen) and were visualized by TapeStation DNA D1000. The resulting libraries were sequenced on the Illumina NextSeq550 for the pilot screen and on the Element Biosciences AVITI for the large EGFRi screen according to the manufacturer's instructions.

Data preprocessing and generation of count matrix

Raw base call files were obtained from Illumina BaseSpace or AVITI storage and were used to generate fastq files using bcl2fastq v2.20.0.422 or bases2fastq version 1.5.0.962525890, respective of sequencing platform. A custom data processing pipeline, adapted from Srivatsan et al., was used to process fastq data into a single-cell count matrix. First, reverse transcription and ligation barcodes were assigned to reads with a mismatch allowance of 1bp, and reads assigned to oligo-shortdT primers were separated from those assigned to random hexamer primers. After index assignment, polyA sequences were trimmed using TrimGalore version 0.6.10 and CutAdapt

version 2.6. Upon polyA trimming, reads were aligned to human GRCh38 using the STAR aligner version 2.7.9a. Aligned reads were filtered for quality and duplicates and were assigned to genes using bedtools version 2.26.0, as described previously. The resulting unique read assignments from both primers were combined and collapsed by cell and gene, and a *celldataset* (CDS) object was generated using the raw sparse count matrix, cell annotations, and gene annotations with the R package *monocle3*. Cell barcodes were determined to be cells upon filtering the CDS by a UMI cutoff determined visually with the kneeplot of cell rank by UMI count. Lastly, doublets were detected with *scrublet* and were filtered based on the doublet score distribution.

In parallel, hash assignments were determined from demultiplexed untrimmed oligo-shortdT reads as described previously (Srivatsan et al, McFaline-Figueroa et al). Briefly, hash barcodes were assigned to reads with a mismatch allowance of 1bp and if the read was adjacent to repeated A sequences, corresponding with hash sequence design. Duplicate hash reads were filtered by UMI and were collapsed into hash assignment counts by cell. Hashes were assigned to cells by two criteria: (1) a cell having ≥20 (pilot) and ≥5 (large screen) hash UMIS and (2) a ratio of the cell's top hash UMI to second best hash UMI of 3 (pilot) or 2.5 (large screen). The monocle3 package was used to manipulate 19,65,66, batch align 20, and visualize 21 the resulting data.

Defining shared inhibitor signature

For the pilot EGFRi screen, differentially expressed genes (DEGs) were calculated by fitting expression of a gene in each cell line/drug condition to a generalized linear model (GLM) modeled as a function of dose and replicate using the R package *monocle3*'s *fit_models* function. The tests were limited to genes expressed in at least 1% of all cells in the experiment. P-values for each DEG test were FDR corrected (Benjamini-Hochberg), and significant DEGs were defined as FDR < 0.01 and a normalized effect magnitude (beta coefficient) of > 0.05. The DEGs associated with each drug were joined for each cell line, and signatures were defined as the unique intersection of DEGs between all combinations of agents used. Of note, the inhibitor EAI045 was excluded from signature analysis given its low number of DEGs indicating its overall low impact on transcription. Only DEG intersections with at least 30 genes were analyzed. For each cell, genes for each signature were size-factor normalized, log transformed, and aggregated to amount to a signature score.

Patient survival analysis

GBM TCGA data (Cell 2013) was accessed and downloaded through the cBioPortal for Cancer Genomics, specifically clinical survival information and paired RNASeqV2 mRNA expression values. Normalized bulk expression values were log10(RSEM) transformed, and signature scores for each donor were calculated as the sum of expression across signature genes. Donors were categorized as high or low signature expression of the signature based on top and bottom 50% of the cohort. Using *survival, ggsurvfit,* and *condsurv* R packages, Kaplan-Meier curves were generated for each group and compared for significant differences in survival using the Gehan-Wilcoxon test or Mantel-Haenszel as denoted. For signature comparison of other cancer types and glioma grade, TCGA PanCancer Atlas data was downloaded and processed in the same way.

Large-screen differential gene expression analysis and calculation of aggregated gene scores for proliferation, resistance, and APM gene programs

As before, differentially expressed genes (DEGs) were calculated by fitting expression of a gene in each cell line/drug condition to a generalized linear model (GLM) modeled as a function of dose and replicate using the R package *monocle3*'s *fit_models* function. The tests were limited to the union of genes expressed in at least 1% of cells within each PDCL. P-values for each DEG test were FDR corrected (Benjamini-Hochberg), and significant DEGs were defined as FDR < 0.01 and a normalized effect magnitude (beta coefficient) of > 0.05.

Proliferation, resistance, and antigen-presentation machinery scores were calculated as in Srivatsan et al[16]. Briefly, for each cell, raw gene expression for genes within a program were retrieved from the count matrix and were size-factor normalized. The normalized expression values were aggregated by group and log normalized with a pseudocount of 1. Downstream scores were visualized as group medians or z-scored group means using R packages *Pheatmap* and *ComplexHeatmap*. GLMs of aggregate scores as a function of dose were fit using the R package *speedglm*. For resistance program scores that include both 24hr and 7d exposures, mean aggregate expression was z-scored jointly to visualize changes over time.

U87MG NY-ESO-1 generation and induction of cytotoxic T-cell killing

U87MG were engineered to express the antigen NY-ESO-1 constitutively. Briefly, a plasmid containing the sequences for NY-ESO-1 (CTAG1B) was ordered from OriGene Technologies. qPCR were performed to obtain the NY-ESO-1 fragments with forward and reverse primers. Overhangs were added to NY-ESO1 fragments by gibson ligation to insert into a vector backbone containing a blasticidin resistance sequence. Virus was produced by transfecting HEK293T cells with the plasmid and the Lipofectamine 3000 kit (Thermo Fisher Scientific). U87MG were transduced with the supernatant containing the NY-ESO-1 virus. After 72 hours of viral exposure, U87MG NY-ESO-1 cells were selected with 10ug/mL blasticidin in culture media for 3 days. Cells were regularly exposed to 1ug/mL blasticidin throughout expansion to ensure blasticidin resistance and NY-ESO-1 expression.

For the GBM cytotoxic T-cell killing experiment, terminally differentiated human $CD8^+$ anti-NY-ESO-1 T-cells, derived by activation of $CD8^+$ T cells via antigen presenting cells with NY-ESO-1 peptide, were ordered from Charles River Lab. T cells were thawed according to the manufacturer's protocol and used for the experiment immediately. U87MG NY-ESO-1 cells were cultured and were subjected to TrypLE (Gibco) to be made single-cell in suspension. Cells were stained with CellTrace CFSE Cell Proliferation Kit (Invitrogen) according to instructions (5uM CFSE in 1X PBS, cell density 1E6/mL) to emit fluorescence at 488nm and were subsequently plated in a 96-well flat bottom plate at a density of 10E3 cells/well, in media consisting of DMEM, 10% FBS, and 10mM penicillin-streptomycin, as well as in standard culture conditions of 37°C, humidified 5% CO2/balance air environment. After adhering and acclimating overnight, the cells treated with 10uM EGFRi (AG18, AG490, AG494, AG555, Tyrphostin9, CL-387785, OSI-420, Osimertinib, or PD153035) or DMSO control — the inhibitors span transcriptional responses of proliferation, adaptive resistance, and antigen presentation. After 24 hours of EGFRi pre-treatment, EGFRi-supplemented media was removed, and the GBM cells were exposed to T-cells

at various GBM:T-cell ratio (1:0 i.e no T-cells, 1:0.25, and 1:0.5). In total, there were 18 well replicates per DMSO + T-cell condition, and 11 well replicates for each EGFRi (excluding Tyrphostin9) + T-cell condition, and 9 well replicates for Tyrphostin9 + T-cell condition. T-cells and U87MG NY-ESO-1 cells were co-cultured in complete T-cell culture media consisting of RPMI 1640 supplemented with 10mM HEPES, 10mM L-glutamine, 10% FBS, 0.34% β-mercaptoethanol, and 10mM penicillin-streptomycin, as in [67]. After 48 hours post-T-cell exposure, plates were imaged with the Zeiss confocal microscope and software, capturing 4x4 tiles of 50x total magnification images of wells with both brightfield and 488nm fluorescence. Tile images were stitched, exported as .tif files, and analyzed using CellProfiler [68]. First, each raw 488nm image was cropped to each tile, excluding well edges, and fluorescent GBM cells were identified with the *IdentifyPrimaryObjects* function. The area occupied (pixels) was calculated with the *MeasureImageAreaOccupied* function, and the resulting data was exported for downstream analysis and visualization in R. The well area for each EGFRi + T-cell condition was aggregated and normalized to the respective plate-specific drug-alone condition CFSE area.

MrVI model training

MrVI[15] is a deep generative model that performs sample stratification at single-cell resolution. As mentioned in the results, the model's hierarchical probabilistic framework presumes cells to be generated from nested experimental designs, in which each sample is drawn from one of several experimental batches. MrVI learns two distinct latent feature spaces from a given scRNA-seq dataset, a sample-unaware *U*-space that is decoupled from the sample-of-origin and a sample-aware *Z*-space that incorporates sample-of-origin effects while accounting for technical factors across both spaces. These latent spaces can be used to dissect sample-specific heterogeneity in cell states across populations of interest in downstream analyses. The updated version of MrVI we applied features multi-head cross-attention-based decoders from the U-space to the Z-space and from the Z-space to the observed space for improved integration. Additionally, the U-space can be constrained to a lower dimensional space relative to the Z-space to further improve the mixing of different samples in the U-space.

We trained a MrVI model for each PDCL (i.e. BT112, BT228 and BT333) with the sample key defined as unique drug-dose combination and the batch key defined as replicate. For all trained models, we used the recommended default model arguments: *n_latent=30; n_latent_u=10; qz_nn_flavor="attention"; px_nn_flavor="attention"; use_map (qz_kwargs)=True; stop_gradients (qz_kwargs)=False; stop_gradients_mlp (qz_kwargs)=True; dropout_rate (qz_kwargs)=0.03; stop_gradients (px_kwargs)=False; stop_gradients_mlp (px_kwargs)=True; h_activation (px_kwargs)="nn.softmax"; low_dim_batch (px_kwargs)=True; dropout_rate (px_kwargs)=0.03; learn_z_u_prior_scale=False; z_u_prior=True;* and *u_prior_mixture=False.* We used the following training arguments in tandem with the recommended default model arguments: *max_epochs=100; batch_size=256; early_stopping=True; early_stopping_patience=15; check_val_every_n_epoch=1; train_size=0.9; lr (pl_kwargs)=2e-3; n_epochs_kl_warmup (plan_kwargs)=20; max_norm (plan_kwargs)=40; eps (plan_kwargs)=1e-8;* and *weight_decay (plan_kwargs)=1e-8*. To select for gene features inputted into the MrVI models, we obtained the union of the 100 highly variable genes (determined in Scanpy) for each unique PDCL-drug data-subset then filtered for expression in at least 5% of cells included in the total dataset. To visualize

the output of the MrVI models, we generated UMAPs from each the *U* and *Z* latent spaces. Specifically, we sought a single cluster with minimal substructure in the sample-unaware *U*-space that mapped to multiple clusters or blatant substructure in the sample-aware *Z*-space.

Defining response modules

The functional relationship between the sample-unaware *U-space* and the sample-aware *Z-space* in MrVI can be used to directly estimate single-cell resolution sample-sample distance matrices. The sample-sample distances are obtained by computing the Euclidean distances between sample-specific counterfactual cell states. To avoid technical factors, counterfactual cell state predictions are made at the level of the Z latent space. In the context of our study, MrVI enabled us to relate the transcriptional effects between unique drug-dose combinations with respect to each PDCL. As we expect each PDCL to be relatively homogeneous in their response to the treatments, we take the average of the normalized distance matrices across all of the cells of each PDCL so our downstream analysis is robust to noise in individual distance matrix estimates. Hierarchical clustering of the mean sample-sample distances across cells resulted in distinct response modules (RMs) containing drug-dose combinations that induced similar transcriptional responses in a given PDCL. The hierarchical clustering depth was evaluated and determined by observation of silhouette plots.

Covariate-specific DE analysis

To obtain a transcriptional signature for each RM we performed covariate-specific DE analysis for each PDCL, a procedure that leverages a fitted MrVI model. Along with updates to the model architecture, the updated version of MrVI includes local measures of differential expression and differential abundance at a single-cell resolution. In particular, the differential expression procedure takes in sample-specific covariates and compares the counterfactual Z latent vectors for a given U latent vector. By decoding the result of a covariate's average impact in the Z latent space, MrVI can report log fold change (LFC) values for each gene. In simpler terms, the LFC values provide a gene-level characterization of each RM. In our analysis, we set DMSO-treated cells as the vehicle when measuring differential expression for each RM. The parameters we used to run the covariate-specific DE analysis function follow: *batch_size=32; use_vmap=True; normalize_design_matrix=True; add_batch_specific_offsets=False; mc_samples=100; store_lfc=True; store_lfc_metadata_subset=None; store_baseline=False; eps_lfc=1e-3; filter_donors=False; lambd=0.0;* and *delta=0.3*. To filter for genes included in each RM signature, we set a cutoff of absolute LFC value > 0.1 and Benjamin-Hochberg adjusted p-value < 0.05. Gene-level adjusted p-values are unobtainable from MrVI covariate-specific DE analysis. Hence, we utilized the p-values associated with general linear models (GLMs) fitted across all genes inputted into MrVI for each PDCL [69], performed with the R package *monocle3*'s *fit_model* function. In each GLM, RM and replicate are set as terms. As in the MrVI covariate-specific DE analysis, we set DMSO-treated cells to be vehicles. Resulting p-values were adjusted for multiple hypothesis correction with the Benjamini-Hochberg method.

Defining response module clusters across PDCLs

We merged the MrVI covariate-specific DE analysis LFCs for RMs across all PDCLs. The associated genes were filtered to include strictly the union of significant DEGs each RM signature

(FDR < 5% & LFC > 0.1). Hierarchical clustering of the normalized LFCs (z-scored for each gene) revealed RM clusters. The aim of hierarchical clustering here was to find drugs with similar transcriptional responses and normalizing LFCs helped relate RMs across PDCLs despite variation in the magnitude of their response to drug treatment. While the resulting RM clusters were largely representative of one PDCL, RMs with similar dynamics at the gene level were collapsed together indicating the RMs are representative of broader transcriptional programs.

Defining Transcriptional Drug Classes (TDCs)

We plotted binary membership of each drug across the RM clusters as a heatmap (i.e. whether a drug is present in an RM cluster across any of its doses). Hierarchical clustering of the drugs based on their RM clusters membership enabled the formation of generalized transcriptional drug classes (TDCs). TDCs represent groups of EGFRi's that induce a distinct transcriptional response across the PDCLs.

Gene set enrichment analysis

For each response module, we filtered differentially expressed genes by FDR < 0.05 and an absolute LFC > 0.1. We performed gene set enrichment analysis (GSEA) separately for increasing and decreasing DEGs using the R package *piano*. Briefly, we utilized the function *runGSAhyper* to perform a Fisher's exact test of the lists of genes against the gene sets "MSigDB Hallmark 2020", "MSigDB Oncogenic Signatures", "Kinase Perturbations from GEO up" and "Kinase Perturbations from GEO down" against a background of expressed genes. Resulting p-values were adjusted for multiple hypothesis correction with the Benjamini- Hochberg method. Representative results are displayed in Figure 5 for conciseness, and results for RMs of each cell line can be found in Supp. Fig. 16.

Visualization of similarity of EGFRi transcriptional effects

Based on these TDCs, we generated lollipop plots representing the magnitude and heterogeneity of a drug's expression of each response module (Fig. 4E, Supp. Fig. 17, Supp. Fig. 19). Briefly for a given drug, each response module's respective up-regulated DE genes (identified by covariate-specific DE analysis) were aggregated, averaged, and log2 normalized to the mean of its correspondent DMSO control. In the lollipop plots, this quantity is represented by the size and color of each lollipop. Additionally, because aggregate score is calculated at the single-cell level, heterogeneity in response module expression could be characterized as percent cells within a drug's high dose population with greater expression than DMSO control. This quantity is represented by the length of each lollipop. As a result, control condition lollipop plots for DMSO, media, and PBS appear largely monotone with even length. In these plots, the differences between TDCs is made more apparent. For instance, TDC5 contains only AZ5104, a demethylated metabolite of osimertinib which itself is part of TDC4. Although chemically the two compounds are almost identical, their drug class separation can be attributed to differences in response modules like BT333 RM9 and BT112 RM12 (Supp. Fig. 18).

Data and code availability

Raw and processed data can be accessed and downloaded from NCBI GEO under accession number GSE261618. The code necessary to reproduce the analyses in this study can be found

at github https://github.com/mcfaline-figueroa-lab/sci-Plex-EGFRi. The code to run MrVI can be found at https://github.com/YosefLab/mrvi under the “paper_reproducibility” tag.

Acknowledgments

The authors would like to thank all members of the McFaline-Figueroa and Azizi labs for helpful suggestions and critical discussion on the study. The authors would like to thank Erin Bush and the JP Sulzberger Columbia Genome Center for next-generation sequencing support. **Funding**: J.L.M.-F acknowledges support from grants from the NIH (R35HG011941) and the NSF (2146007). E.A. acknowledges support by grant number 2022-253560 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. These studies used the resources of the Cancer Center Flow Core Facility at Columbia University funded in part through Center Grant P30CA013696. **Author contributions**: J.L.M.-F conceived the project and provided overall supervision of the study; R.M.G. designed experiments; R.M.G. and N.H. analyzed and interpreted the data; R.M.G., N.H., A.W., J.H., L.S., M.V., H.F., J.P.N. and S.W.M. performed experiments or aided in analysis. K.L.L. provided PDCL models. K.L.L., J.R.M.-F, N.Y., and E.A. provided additional supervision. R.M.G., N.H., and J.L.M-F wrote the manuscript. All authors reviewed and contributed to the manuscript. **Declaration of Interests:** K.L.L. reports research support to DFCI from Bristol Meyers Squibb (BMS) and Eli Lilly, consulting and advisory roles to BMS, Travera, Blaze Bioscience, and Servier, and equity in Travera.

References

1. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

2. Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301.

3. Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell *167*, 1883–1896.e15.

4. Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1882.e21.

5. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell *171*, 1437–1452.e17.

6. Mitchell, D.C., Kuljanin, M., Li, J., Van Vranken, J.G., Bulloch, N., Schweppe, D.K., Huttlin, E.L., and Gygi, S.P. (2023). A proteome-wide atlas of drug mechanism of action. Nat. Biotechnol. *41*, 845–857.

7. Lin, A., Giuliano, C.J., Palladino, A., John, K.M., Abramowicz, C., Yuan, M.L., Sausville, E.L., Lukow, D.A., Liu, L., Chait, A.R., et al. (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Sci. Transl. Med. *11*. 10.1126/scitranslmed.aaw8412.

8. Lin, A., Giuliano, C.J., Sayles, N.M., and Sheltzer, J.M. (2017). CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. Elife *6*. 10.7554/eLife.24179.

9. Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. Preprint, 10.1056/nejmoa043330 10.1056/nejmoa043330.

10. Stupp, R., Taillibert, S., Kanner, A.A., Kesari, S., Steinberg, D.M., Toms, S.A., Taylor, L.P., Lieberman, F., Silvani, A., Fink, K.L., et al. (2015). Maintenance Therapy With Tumor-Treating Fields Plus Temozolomide vs Temozolomide Alone for Glioblastoma: A Randomized Clinical Trial. JAMA *314*, 2535–2543.

11. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477.

12. The Cancer Genome Atlas Research Network (2008). Comprehensive genomic

characterization defines human glioblastoma genes and core pathways. Preprint, 10.1038/nature07385 10.1038/nature07385.

13. Makhlin, I., Salinas, R.D., Zhang, D., Jacob, F., Ming, G.-L., Song, H., Saxena, D., Dorsey, J.F., Nasrallah, M.P., Morrissette, J.J., et al. (2019). Clinical activity of the tyrosine kinase inhibitor osimertinib in -mutant glioblastoma. CNS Oncol *8*, CNS43.

14. Cardona, A.F., Jaramillo-Velásquez, D., Ruiz-Patiño, A., Polo, C., Jiménez, E., Hakim, F., Gómez, D., Ramón, J.F., Cifuentes, H., Mejía, J.A., et al. (2021). Efficacy of osimertinib plus bevacizumab in glioblastoma patients with simultaneous EGFR amplification and EGFRvIII mutation. J. Neurooncol. *154*, 353–364.

15. Boyeau, P., Hong, J., Gayoso, A., Kim, M., McFaline-Figueroa, J.L., Jordan, M.I., Azizi, E., Ergen, C., and Yosef, N. (2024). Deep generative modeling of sample-level heterogeneity in single-cell genomics. bioRxiv. 10.1101/2022.10.04.510898.

16. Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J., Packer, J., Pliner, H.A., Jackson, D.L., Daza, R.M., Christiansen, L., et al. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. Science *367*, 45–51.

17. Huang, W.-S., Liu, S., Zou, D., Thomas, M., Wang, Y., Zhou, T., Romero, J., Kohlmann, A., Li, F., Qi, J., et al. (2016). Discovery of Brigatinib (AP26113), a Phosphine Oxide-Containing, Potent, Orally Active Inhibitor of Anaplastic Lymphoma Kinase. J. Med. Chem. *59*, 4948–4964.

18. Lai, C.-J., Bao, R., Tao, X., Wang, J., Atoyan, R., Qu, H., Wang, D.-G., Yin, L., Samson, M., Forrester, J., et al. (2010). CUDC-101, a multitargeted inhibitor of histone deacetylase, epidermal growth factor receptor, and human epidermal growth factor receptor 2, exerts potent anticancer activity. Cancer Res. *70*, 3647–3656.

19. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502.

20. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

21. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 10.48550/ARXIV.1802.03426.

22. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

23. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291–304.e6.

24. Li, D., Ambrogio, L., Shimamura, T., Kubo, S., Takahashi, M., Chirieac, L.R., Padera, R.F., Shapiro, G.I., Baum, A., Himmelsbach, F., et al. (2008). BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. Oncogene *27*, 4702–4711.

25. Brown, K.E., Chagoya, G., Kwatra, S.G., Yen, T., Keir, S.T., Cooter, M., Hoadley, K.A., Rasheed, A., Lipp, E.S., Mclendon, R., et al. (2015). Proteomic profiling of patient-derived glioblastoma xenografts identifies a subset with activated EGFR: implications for drug development. J. Neurochem. *133*, 730–738.

26. Lee, J., Kotliarova, S., Kotliarov, Y., Li, A., Su, Q., Donin, N.M., Pastorino, S., Purow, B.W., Christopher, N., Zhang, W., et al. (2006). Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. Cancer Cell *9*, 391–403.

27. Touat, M., Li, Y.Y., Boynton, A.N., Spurr, L.F., Iorgulescu, J.B., Bohrson, C.L., Cortes-Ciriano, I., Birzu, C., Geduldig, J.E., Pelton, K., et al. (2020). Mechanisms and therapeutic implications of hypermutation in gliomas. Nature *580*, 517–523.

28. Emert, B.L., Cote, C.J., Torre, E.A., Dardani, I.P., Jiang, C.L., Jain, N., Shaffer, S.M., and Raj, A. (2021). Variability within rare cell states enables multiple paths toward drug resistance. Nat. Biotechnol. *39*, 865–876.

29. Martin, B.K., Qiu, C., Nichols, E., Phung, M., Green-Gladden, R., Srivatsan, S., Blecher-Gonen, R., Beliveau, B.J., Trapnell, C., Cao, J., et al. (2023). Optimized single-nucleus transcriptional profiling by combinatorial indexing. Nat. Protoc. *18*, 188–207.

30. Sziraki, A., Lu, Z., Lee, J., Banyai, G., Anderson, S., Abdulraouf, A., Metzner, E., Liao, A., Banfelder, J., Epstein, A., et al. (2022). A global view of aging and Alzheimer's pathogenesis-associated cell population dynamics and molecular signatures in the human and mouse brains. bioRxiv. 10.1101/2022.09.28.509825.

31. McFaline-Figueroa, J.L., Srivatsan, S., Hill, A.J., Gasperini, M., Jackson, D.L., Saunders, L., Domcke, S., Regalado, S.G., Lazarchuck, P., Alvarez, S., et al. (2023). Multiplex single-cell chemical genomics reveals the kinase dependence of the response to targeted therapy. bioRxiv. 10.1101/2023.03.10.531983.

32. Sakanyan, V., Angelini, M., Le Béchec, M., Lecocq, M.F., Benaiteau, F., Rousseau, B., Gyulkhandanyan, A., Gyulkhandanyan, L., Logé, C., Reiter, E., et al. (2014). Screening and discovery of nitro-benzoxadiazole compounds activating epidermal growth factor receptor (EGFR) in cancer cells. Sci. Rep. *4*, 3977.

33. Hennequin, L.F., Allen, J., Breed, J., Curwen, J., Fennell, M., Green, T.P., Lambert-van der Brempt, C., Morgentin, R., Norman, R.A., Olivier, A., et al. (2006). N-(5-chloro-1,3-benzodioxol-4-yl)-7-[2-(4-methylpiperazin-1-yl)ethoxy]-5- (tetrahydro-2H-pyran-4-yloxy)quinazolin-4-amine, a novel, highly selective, orally available, dual-specific c-Src/Abl kinase inhibitor. J. Med. Chem. *49*, 6465–6488.

34. Formisano, L., D'Amato, V., Servetto, A., Brillante, S., Raimondo, L., Di Mauro, C., Marciano, R., Orsini, R.C., Cosconati, S., Randazzo, A., et al. (2015). Src inhibitors act through different mechanisms in Non-Small Cell Lung Cancer models depending on EGFR and RAS mutational status. Oncotarget *6*, 26090–26103.

35. Zhou, W., Ercan, D., Chen, L., Yun, C.-H., Li, D., Capelletti, M., Cortot, A.B., Chirieac, L., Iacob, R.E., Padera, R., et al. (2009). Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. Nature *462*, 1070–1074.

36. Jia, Y., Juarez, J., Li, J., Manuia, M., Niederst, M.J., Tompkins, C., Timple, N., Vaillancourt, M.-T., Pferdekamper, A.C., Lockerman, E.L., et al. (2016). EGF816 Exerts Anticancer Effects in Non-Small Cell Lung Cancer by Irreversibly and Selectively Targeting Primary and Acquired Activating Mutations in the EGF Receptor. Cancer Res. *76*, 1591–1602.

37. Park, Y.H., Kim, D.K., Kim, H.S., Lee, D., Lee, M.B., Min, K.Y., Jo, M.G., Lee, J.E., Kim, Y.M., and Choi, W.S. (2019). WZ3146 inhibits mast cell Lyn and Fyn to reduce IgE-mediated allergic responses in vitro and in vivo. Toxicol. Appl. Pharmacol. *383*, 114763.

38. Verma, N., Rai, A.K., Kaushik, V., Brünnert, D., Chahar, K.R., Pandey, J., and Goyal, P. (2016). Identification of gefitinib off-targets using a structure-based systems biology approach; their validation with reverse docking and retrospective data mining. Sci. Rep. *6*, 33949.

39. Wang, S., He, Z., Wang, X., Li, H., and Liu, X.-S. (2019). Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. Elife *8*. 10.7554/eLife.49020.

40. Sener, U., Ruff, M.W., and Campian, J.L. (2022). Immunotherapy in Glioblastoma: Current Approaches and Future Perspectives. Int. J. Mol. Sci. *23*. 10.3390/ijms23137046.

41. Pollack, B.P., Sapkota, B., and Cartee, T.V. (2011). Epidermal growth factor receptor inhibition augments the expression of MHC class I and II genes. Clin. Cancer Res. *17*, 4400–4413.

42. Sadagopan, A., Michelakos, T., Boyiadzis, G., Ferrone, C., and Ferrone, S. (2022). Human Leukocyte Antigen Class I Antigen-Processing Machinery Upregulation by Anticancer Therapies in the Era of Checkpoint Inhibitors: A Review. JAMA Oncol *8*, 462–473.

43. Stopfer, L.E., Rettko, N.J., Leddy, O., Mesfin, J.M., Brown, E., Winski, S., Bryson, B., Wells, J.A., and White, F.M. (2022). MEK inhibition enhances presentation of targetable MHC-I tumor antigens in mutant melanomas. Proc. Natl. Acad. Sci. U. S. A. *119*, e2208900119.

44. Prasad, M., Zorea, J., Jagadeeshan, S., Shnerb, A.B., Mathukkada, S., Bouaoud, J., Michon, L., Novoplansky, O., Badarni, M., Cohen, L., et al. (2022). MEK1/2 inhibition transiently alters the tumor immune microenvironment to enhance immunotherapy efficacy against head and neck cancer. J Immunother Cancer *10*. 10.1136/jitc-2021-003917.

45. Wang, S., He, Z., Wang, X., Li, H., and Liu, X.-S. (2019). Author response: Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. Preprint at eLife Sciences Publications, Ltd, 10.7554/elife.49020.036 10.7554/elife.49020.036.

46. Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell *39*, 779–792.e11.

47. Davidoff, A.N., and Mendelow, B.V. (1992). Puromycin is a potent and specific inhibitor of tyrosine kinase activity in HL-60 cells. Anticancer Res. *12*, 1761–1766.

48. Fling, S.P., Arp, B., and Pious, D. (1994). HLA-DMA and -DMB genes are both required for

MHC class II/peptide complex formation in antigen-presenting cells. Nature *368*, 554–558.

49. Chen, Y.T., Scanlan, M.J., Sahin, U., Türeci, O., Gure, A.O., Tsang, S., Williamson, B., Stockert, E., Pfreundschuh, M., and Old, L.J. (1997). A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. Proc. Natl. Acad. Sci. U. S. A. *94*, 1914–1918.

50. Bethune, M.T., Li, X.-H., Yu, J., McLaughlin, J., Cheng, D., Mathis, C., Moreno, B.H., Woods, K., Knights, A.J., Garcia-Diaz, A., et al. (2018). Isolation and characterization of NY-ESO-1-specific T cell receptors restricted on various MHC molecules. Proc. Natl. Acad. Sci. U. S. A. *115*, E10702–E10711.

51. Mathewson, N.D., Ashenberg, O., Tirosh, I., Gritsch, S., Perez, E.M., Marx, S., Jerby-Arnon, L., Chanoch-Myers, R., Hara, T., Richman, A.R., et al. (2021). Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis. Cell *184*, 1281–1298.e26.

52. Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., and Zarrinkar, P.P. (2011). Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol. *29*, 1046–1051.

53. Elkins, J.M., Fedele, V., Szklarz, M., Abdul Azeez, K.R., Salah, E., Mikolajczyk, J., Romanov, S., Sepetov, N., Huang, X.-P., Roth, B.L., et al. (2016). Comprehensive characterization of the Published Kinase Inhibitor Set. Nat. Biotechnol. *34*, 95–103.

54. Reinecke, M., Brear, P., Vornholz, L., Berger, B.-T., Seefried, F., Wilhelm, S., Samaras, P., Gyenis, L., Litchfield, D.W., Médard, G., et al. (2023). Chemical proteomics reveals the target landscape of 1,000 kinase inhibitors. Nat. Chem. Biol. 10.1038/s41589-023-01459-3.

55. Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P.-A., Reinecke, M., Ruprecht, B., Petzoldt, S., Meng, C., et al. (2017). The target landscape of clinical kinase drugs. Science *358*. 10.1126/science.aan4368.

56. Morphy, R. (2010). Selectively nonselective kinase inhibition: striking the right balance. J. Med. Chem. *53*, 1413–1437.

57. Levitin, H.M., Yuan, J., Cheng, Y.L., Ruiz, F., Jr, Bush, E.C., Bruce, J.N., Canoll, P., Iavarone, A., Lasorella, A., Blei, D.M., et al. (2019). gene signature identification from single-cell RNA-seq with hierarchical Poisson factorization. Mol. Syst. Biol. *15*, e8557.

58. Zhao, W., Dovas, A., Spinazzi, E.F., Levitin, H.M., Banu, M.A., Upadhyayula, P., Sudhakar, T., Marie, T., Otten, M.L., Sisti, M.B., et al. (2021). Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. Genome Med. *13*, 82.

59. Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., Hetzel, L., Ji, Y., Ibarra, I.L., Srivatsan, S.R., Naghipourfar, M., Daza, R.M., Martin, B., et al. (2023). Predicting cellular responses to complex perturbations in high-throughput screens. Mol. Syst. Biol. *19*, e11517.

60. Jiang, L., Dalgarno, C., Papalexi, E., Mascio, I., Wessels, H.-H., Yun, H., Iremadze, N., Lithwick-Yanai, G., Lipson, D., and Satija, R. (2024). Systematic reconstruction of molecular pathway signatures using scalable single-cell perturbation screens. bioRxiv.

10.1101/2024.01.29.576933.

61. Lim, N., and Pavlidis, P. (2021). Evaluation of connectivity map shows limited reproducibility in drug repositioning. Sci. Rep. *11*, 17624.

62. Kong, W., Biddy, B.A., Kamimoto, K., Amrute, J.M., Butka, E.G., and Morris, S.A. (2020). CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. Nat. Protoc. *15*, 750–772.

63. Goyal, Y., Busch, G.T., Pillai, M., Li, J., Boe, R.H., Grody, E.I., Chelvanambi, M., Dardani, I.P., Emert, B., Bodkin, N., et al. (2023). Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. Nature *620*, 651–659.

64. Kuroda, Y., Wakao, S., Kitada, M., Murakami, T., Nojima, M., and Dezawa, M. (2013). Isolation, culture and evaluation of multilineage-differentiating stress-enduring (Muse) cells. Nat. Protoc. *8*, 1391–1415.

65. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

66. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

67. Shi, L., Lim, J.Y., and Kam, L.C. (2023). Substrate stiffness enhances human regulatory T cell induction and metabolism. Biomaterials *292*, 121928.

68. Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A., and Goodman, A. (2021). CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics *22*, 433.

69. Wells, S.B., Rainbow, D.B., Mark, M., Szabo, P.A., Ergen, C., Maceiras, A.R., Caron, D.P., Rahmani, E., Benuck, E., Amiri, V.V.P., et al. (2024). Multimodal profiling reveals tissue-directed signatures of human immune cells altered with age. bioRxiv. 10.1101/2024.01.03.573877.

Supplementary Figures

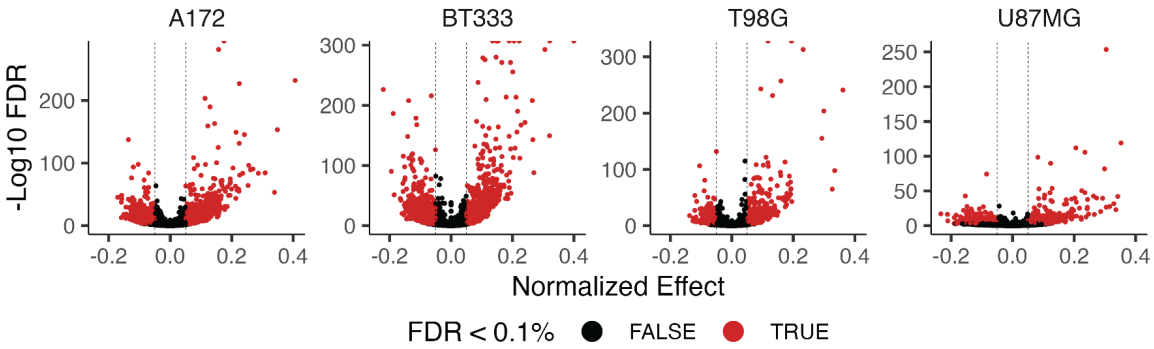
A

Cell Line	Cell Count	Median UMI	Median Hash UMI
A172	40562	533	128
BT333	76654	275	109
T98G	45062	487	119
U87MG	26468	216	110
Total	188746	358	116

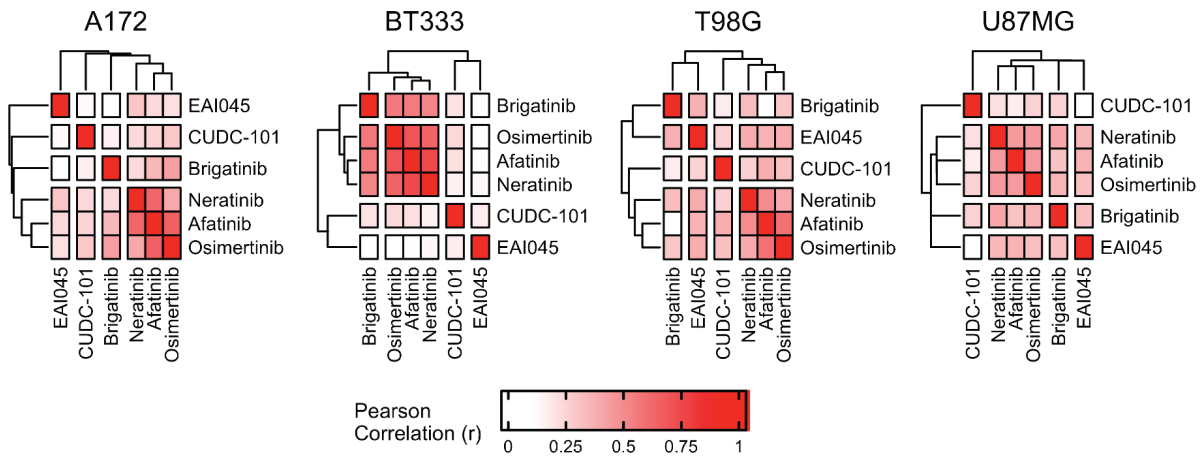
B

Drug (all doses)	Cell Count	Total DEGs	DEGs upreg	DEGs downreg
DMSO	25594	—	—	—
Afatinib	26409	1650	905	745
Brigatinib	25171	1153	546	607
CUDC-101	24701	6423	2997	3426
EAI045	31408	178	77	101
Neratinib	30124	1341	707	634
Osimertinib	25339	1494	759	735

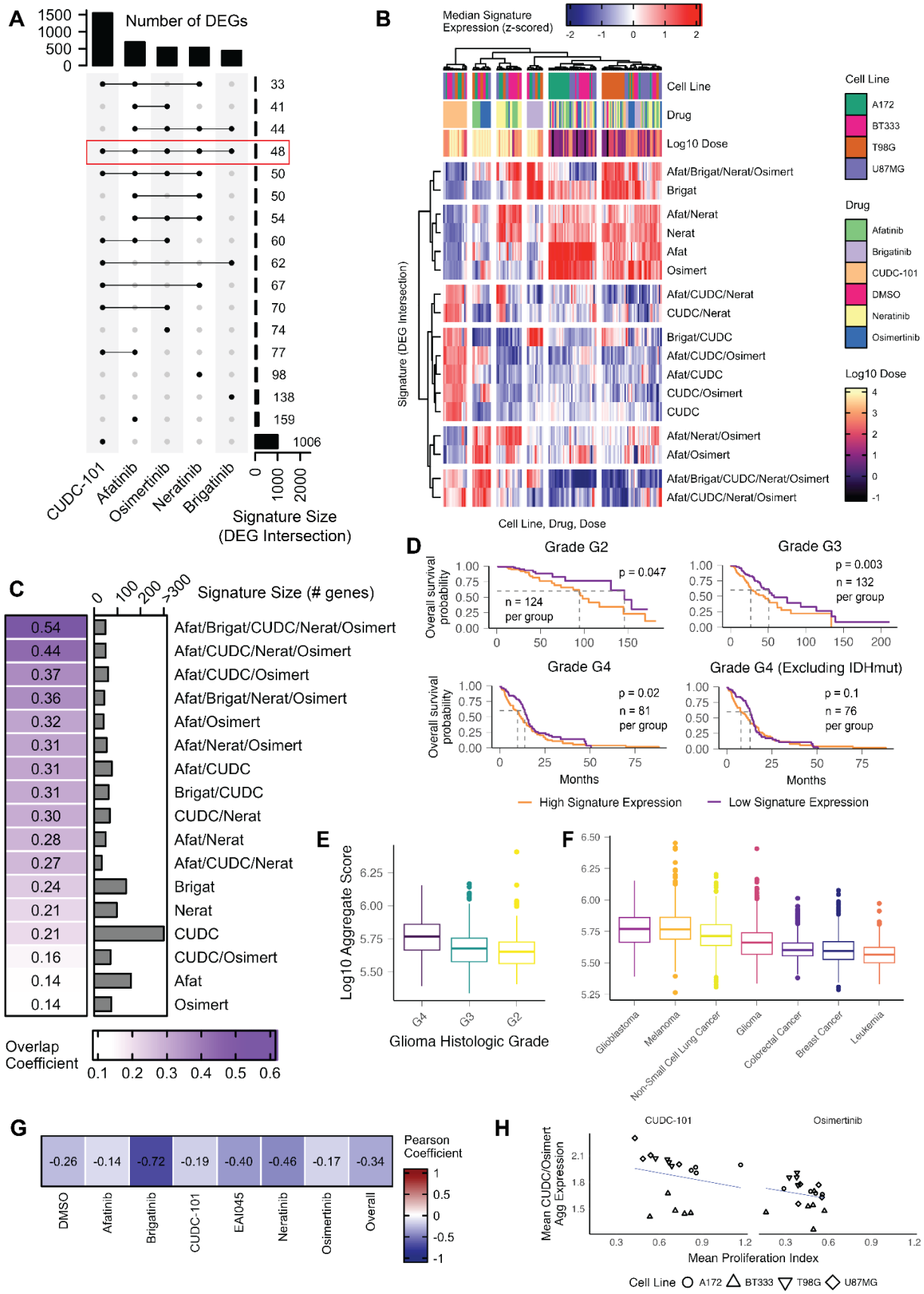
C



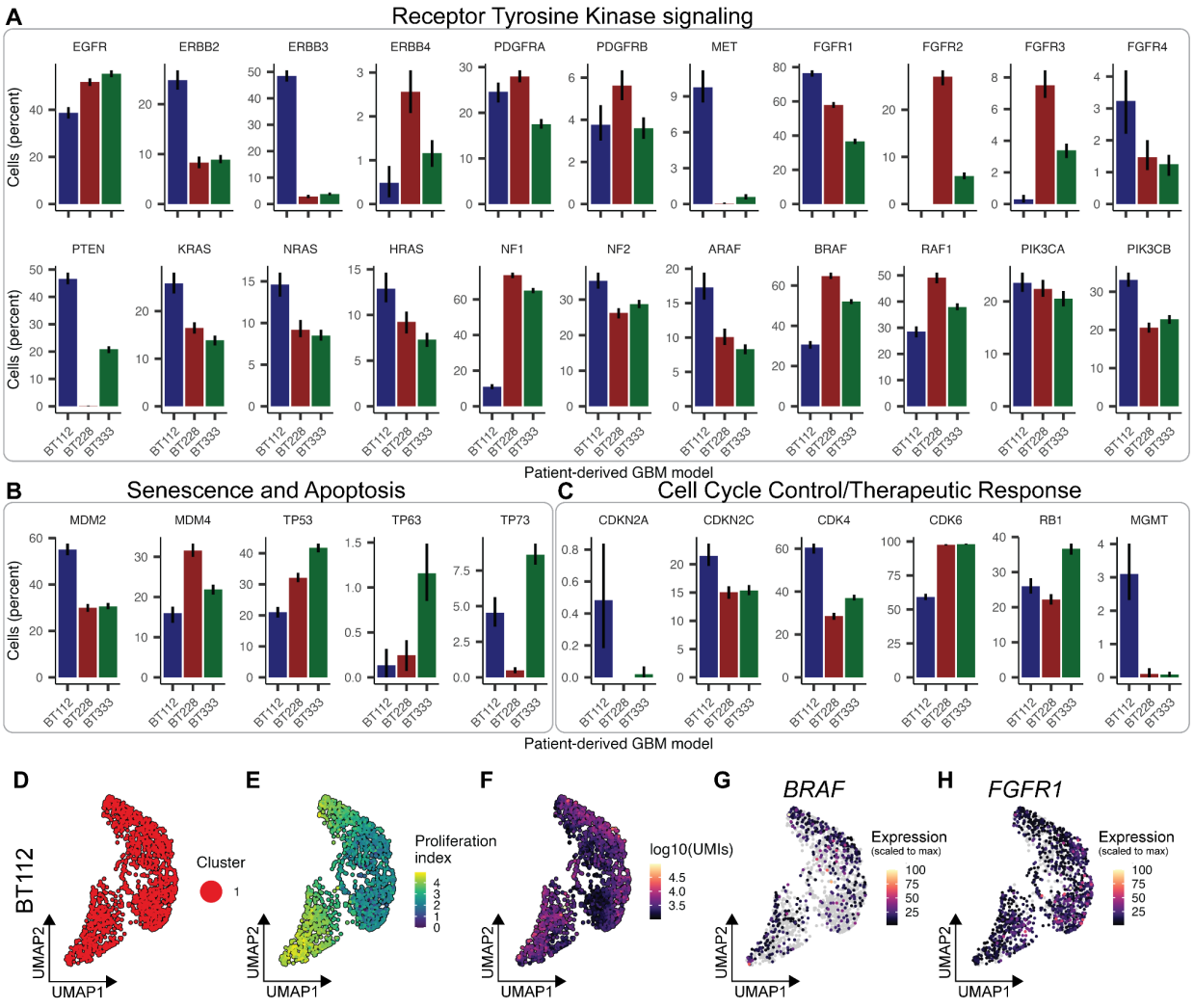
D



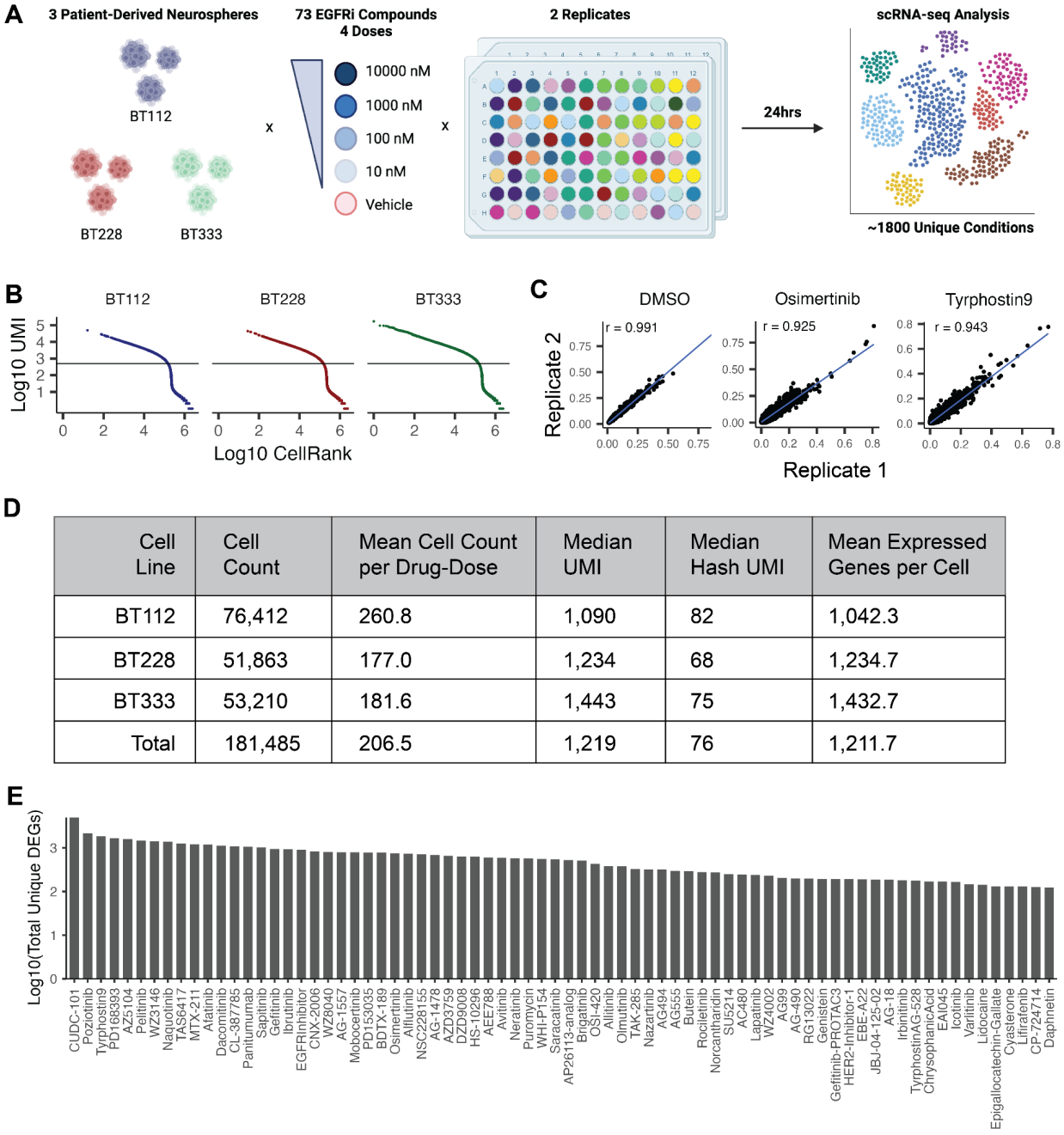
Supplemental Figure 1. Pilot EGFRi screen QC and differential gene expression analysis **A.** Table of experimental summary metrics by cell line. **B.** Table of experimental summary metrics by chemical agent. **C.** Volcano plots displaying the differential gene expression results for all drugs within a cell line. Only the terms from the contribution of dose as log10(dose + 0.01) from the generalized linear model fit to expression is shown. **D.** Pairwise Pearson correlation coefficients between drug's normalized beta coefficients for significant genes (FDR < 0.1%) for each cell line.



Supplemental Figure 2. Defining chemical signatures from DEGs in the pilot EGFRi screen. **A.** Upset plot displaying the unique intersections (non-overlapping) of DEGs between the five drugs that induced substantial transcriptional changes. DEGs were limited to those upregulated in at least one cell line (normalized beta coefficient > 0.05, FDR < 0.1%). Highlighted by a red box is the 48 gene signature shared by all five inhibitors. **B.** Heatmap of z-scored aggregate signature scores by cell line, drug, dose conditions. Hierarchical clustering of the columns shows that the defined signatures are representative shared drug effects across cell lines. Hierarchical clustering of the rows illustrates that the shared five inhibitor signature is transcriptionally distinct from the other defined signatures. **C.** Overlap of pilot EGFRi signature genes with a recently defined adaptive resistance signature across established GBM cell lines and glioma stem cell models in response to the inhibition of RTK signaling [16]. Note that overlap does not correlate with signature size. **D.** Overall survival of glioma patients (by tumor grade) from TCGA [23] grouped by the relative expression (top and bottom 50%) of the shared EGFR inhibitor transcriptional module (p: Gehan-Wilcoxon test). **E.** Aggregate expression distribution of the five inhibitor shared signature by glioma histologic grade from TCGA patient samples, arranged by highest median aggregate expression *(left to right)*. **F.** Aggregate expression distribution of the five inhibitor shared signature by major cancer types from PanCancer TCGA patient samples, arranged by highest median aggregate expression *(left to right)*. **G.** Heatmap of Pearson correlation coefficients between mean CUDC-101/Osimertinib signature aggregate score and mean proliferation index score, for EGFRi (10uM) pseudobulked by cell line and well replicate. The module is weakly anti-correlated with viability as measured by proliferation. **H.** Example of correlation between CUDC-101/osimertinib signature mean aggregate score and mean proliferation index score, for cells treated with 10uM CUDC-101 (*left*) and osimertinib (*right*).

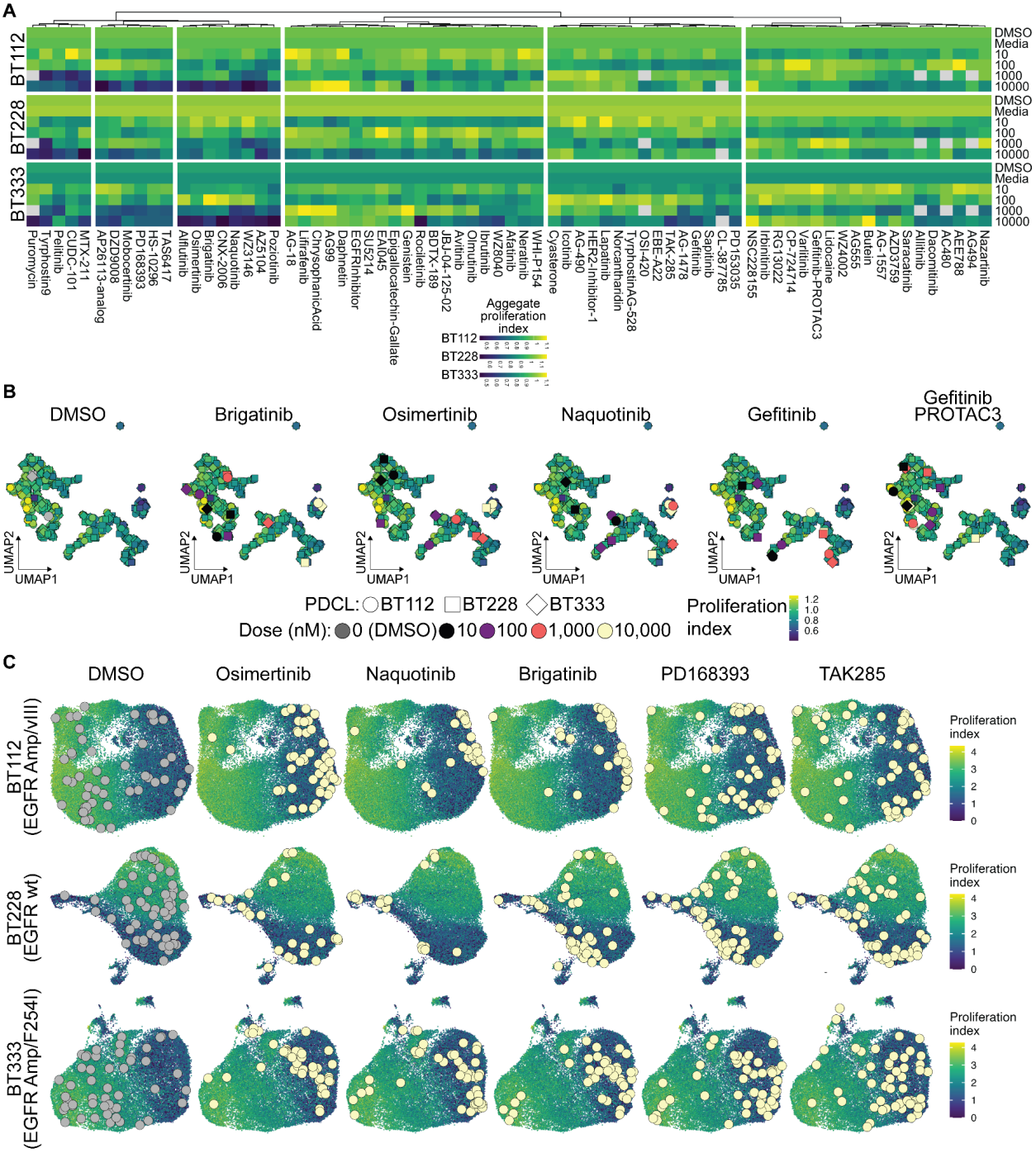


Supplemental Figure 3. **Inter- and intra-model variation in RTK pathway expression across patient-derived glioblastoma models. A**. Expression of genes associated with receptor tyrosine kinase and EGFR-specific signaling across BT112, BT228, and BT333 patient-derived models (PDCLs). **B**. Expression of genes associated with senescence and apoptosis across PDCLs. **C**. Expression of genes associated with cell cycle control and therapeutic response across PDCLs. **D**. UMAP embeddings of BT112 colored by PCA cluster (**D**), the aggregate expression of genes associated with proliferation (**E**)**,** and total number of transcripts (UMIs) captured per cell (**F**). **G-H**. UMAP embedding as in C-E colored by the expression of *BRAF* (**G**) and *FGRF1* (**H**).



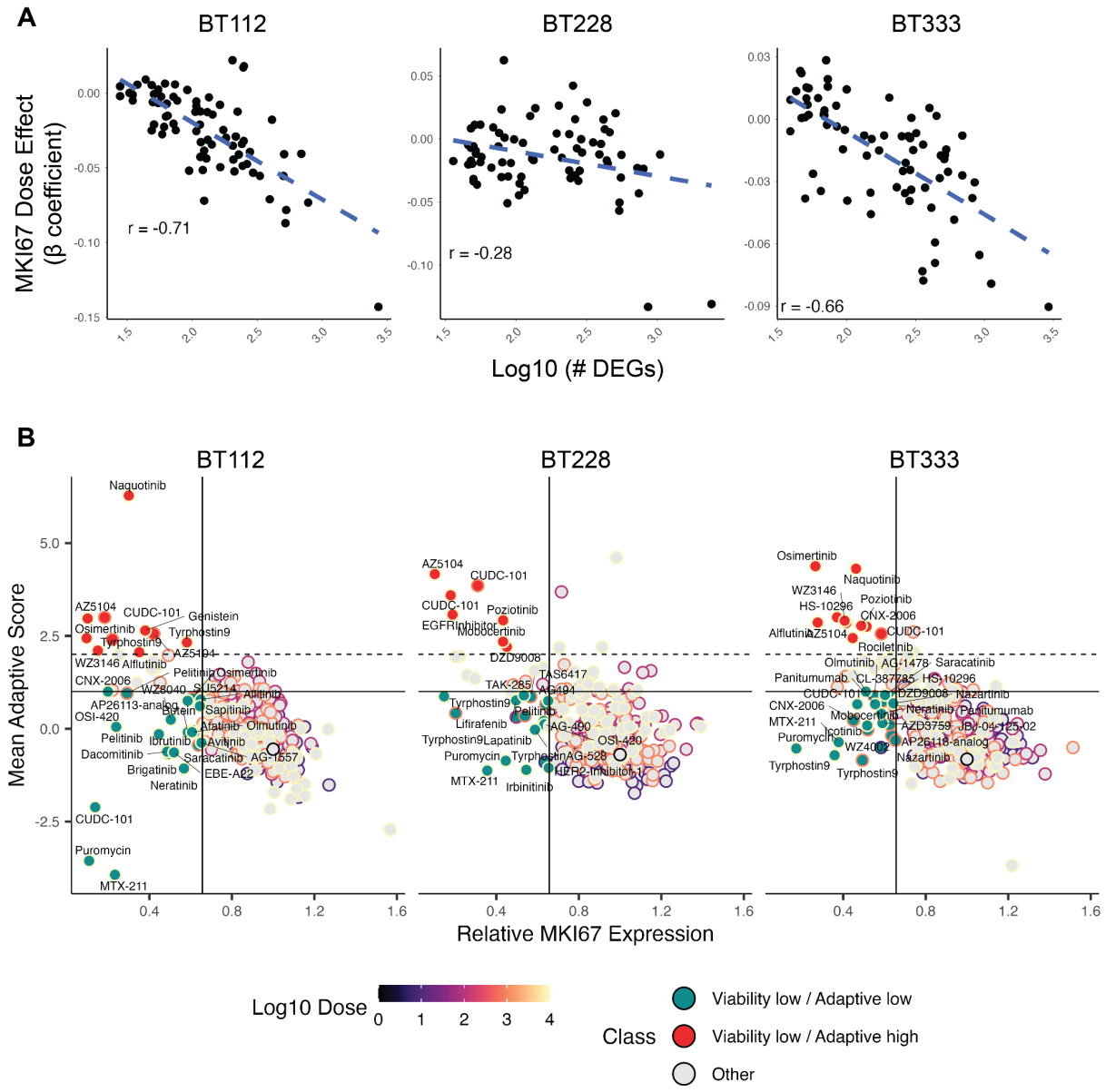
Cell Line	Cell Count	Mean Cell Count per Drug-Dose	Median UMI	Median Hash UMI	Mean Expressed Genes per Cell
BT112	76,412	260.8	1,090	82	1,042.3
BT228	51,863	177.0	1,234	68	1,234.7
BT333	53,210	181.6	1,443	75	1,432.7
Total	181,485	206.5	1,219	76	1,211.7

Supplemental Figure 4. Comprehensive screen of EGFRi in patient-derived models of GBM using sci-Plex for a single-cell transcriptional readout. A. Outline of 24-hour large EGFRi library screen. **B.** Kneeplots for each PDCL as UMI vs cell rank with a cut-off at 500 UMI. **C.** The normalized gene expression of replicates of the screen is highly correlated across chemical agents (DMSO, 10uM Osimertinib, 10uM Tyrphostin9 from left to right). **D.** Table of experimental summary metrics by cell line. **E.** Differentially expressed genes per compound ordered by decreasing count.

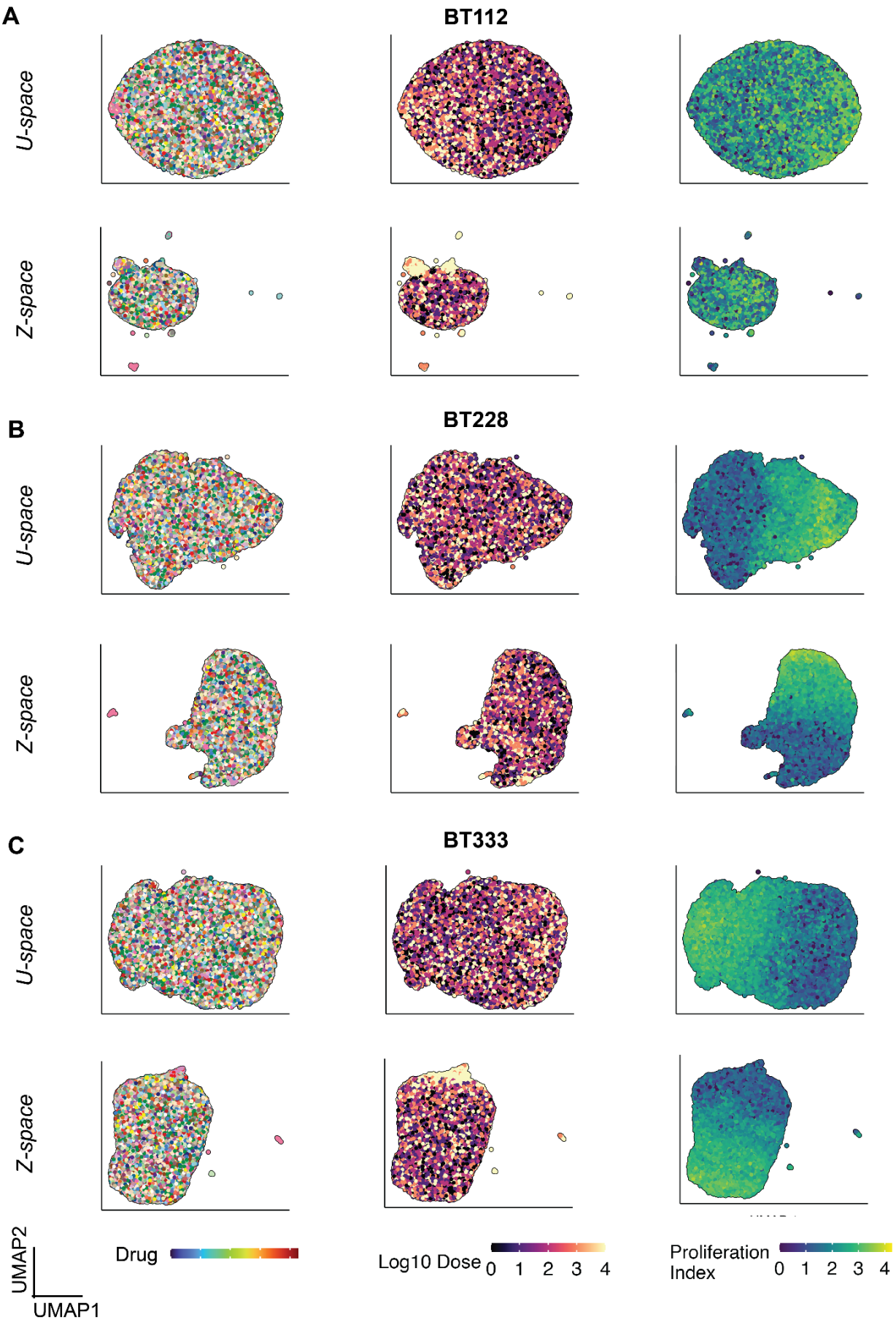


Supplemental Figure 5. Visualizing EGFRi effects on proliferative gene expression in PDCLs. A. Heatmap of mean aggregate proliferation score as for each PDCL, drug, dose, as in Fig. 3A. **B.** UMAP projection of EGFRi PDCL, drug, dose points represented by features as log2 normalized expression to DMSO control. Mean aggregate proliferation index is represented by the underlying viridis color scale, and select drugs are faceted and colored to highlight the differential effects of each drug-dose combination. **C.** UMAP projection of single-cells treated with EGFRi embedded individually for each PDCL. Mean aggregate proliferation index is represented

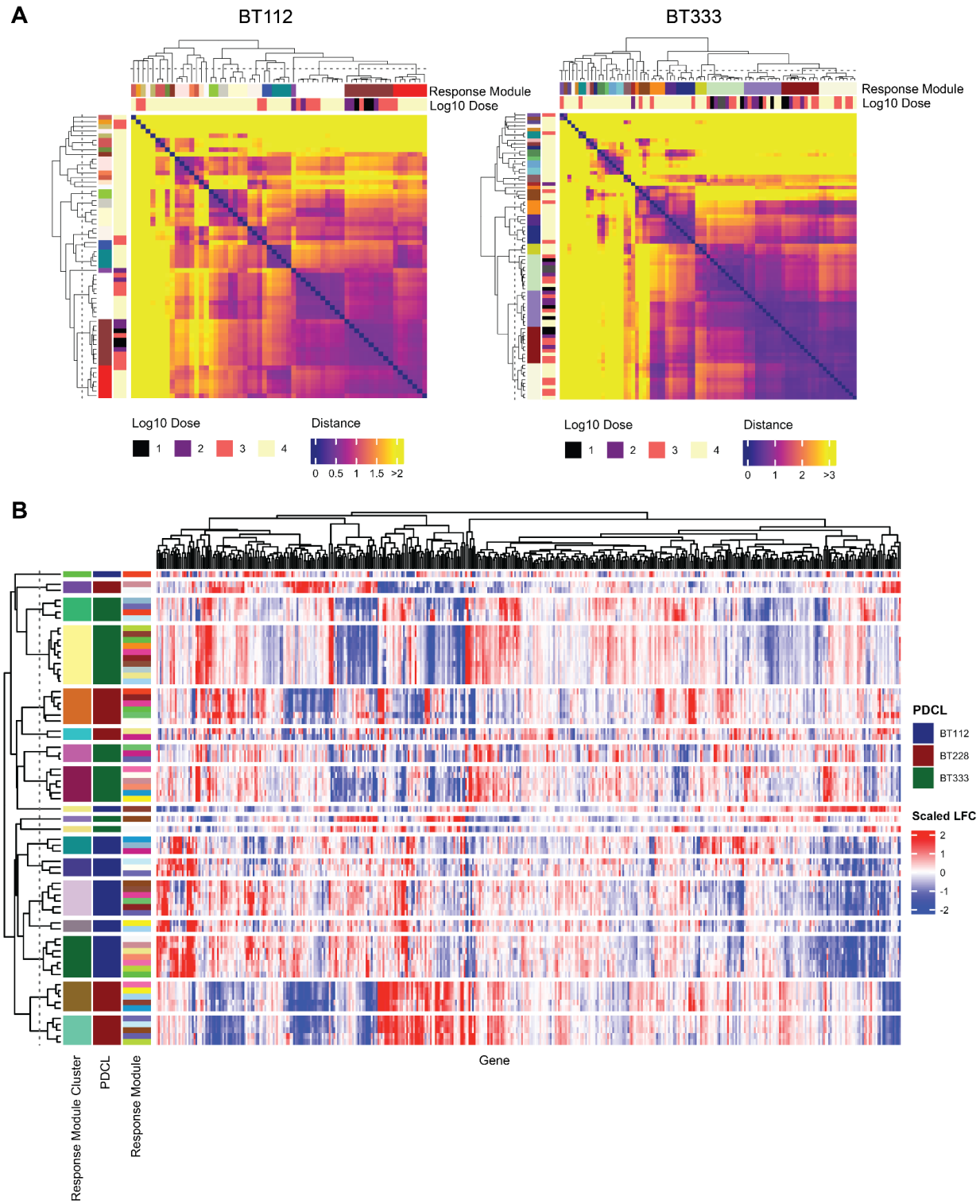
by the underlying viridis color scale, and a sample of cells that were exposed to 10uM EGFRi are faceted and colored to highlight differential effects.



Supplemental Figure 6. Correlation between MKI67 expression and global transcriptional impact or adaptive resistance signaling. **A.** MKI67 beta coefficient has a varying correlation with global transcriptional impact for BT112, BT228, and BT333 (left to right). The scatter plots represent the Pearson correlation within a cell line of inhibitor impact on MKI67 (normalized beta coefficient from quasi-Poisson regression of MKI67 as a function of dose) versus overall transcriptional changes (as total number of DEGs). **B.** Drug-dose group mean adaptive score versus mean MKI67 expression for BT112, BT228, and BT333 (left to right) reveals MTX-211 as an agent that alters viability while minimizing an adaptive response. Outside color of points denotes the dose while the fill color of points designates the ability to alter viability and adaptive signatures above or below the mean + sd or mean + 2sd of all drug-dose groups.

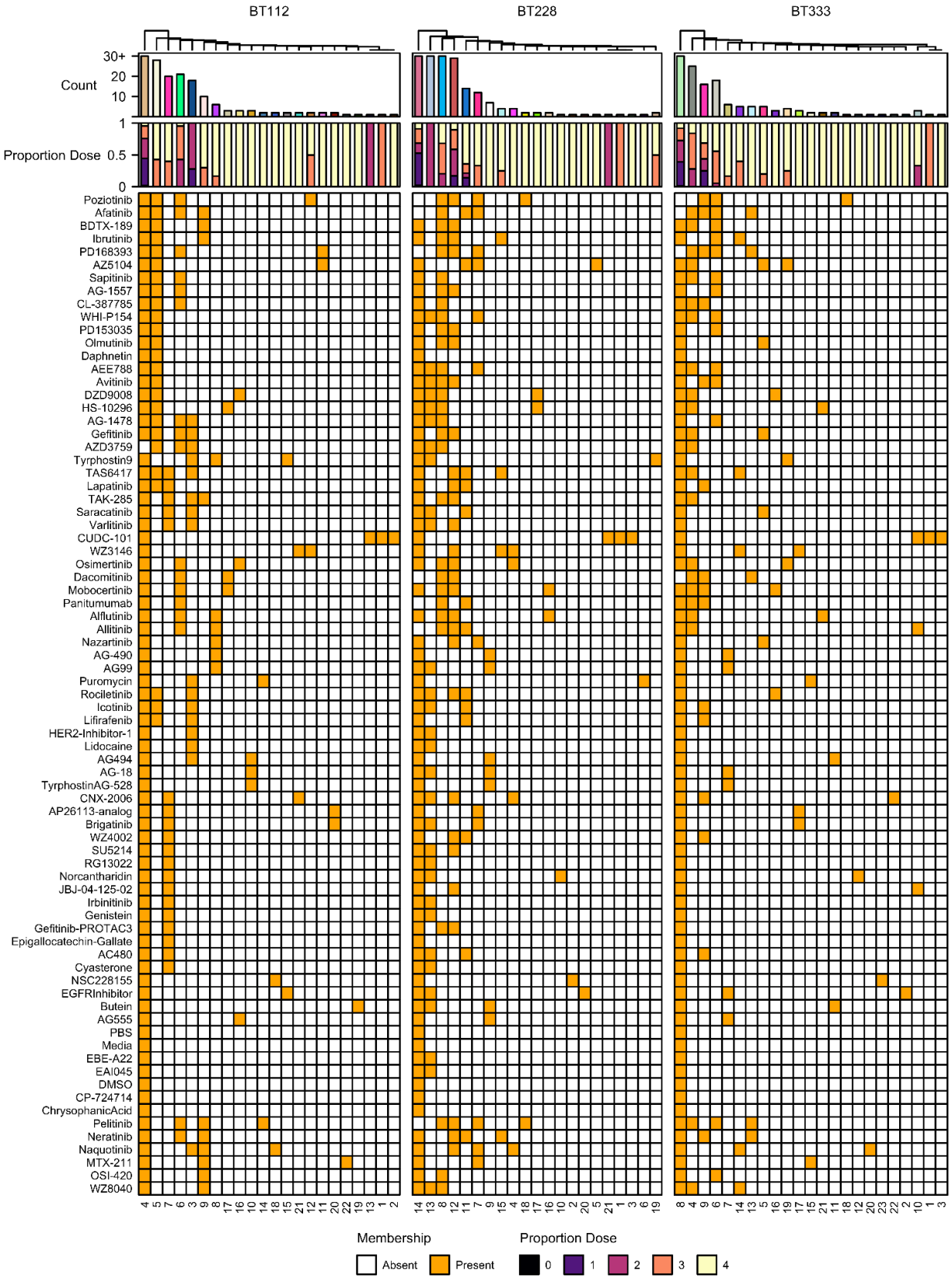


Supplemental Figure 7. MrVI learns a sample-unaware latent U-space embedding as well as a sample-aware latent Z-space embedding. UMAP dimensionality reduction is performed on *U-space* (top) and *Z-space* (bottom) latent embeddings, resulting in two-dimensional representations of **(A)** BT112, **(B)** BT228, and **(C)** BT333. Note that in the UMAPs of the Z-space, the largest clusters consist largely of cells treated with low doses or compounds that do not have a substantial molecular effect.

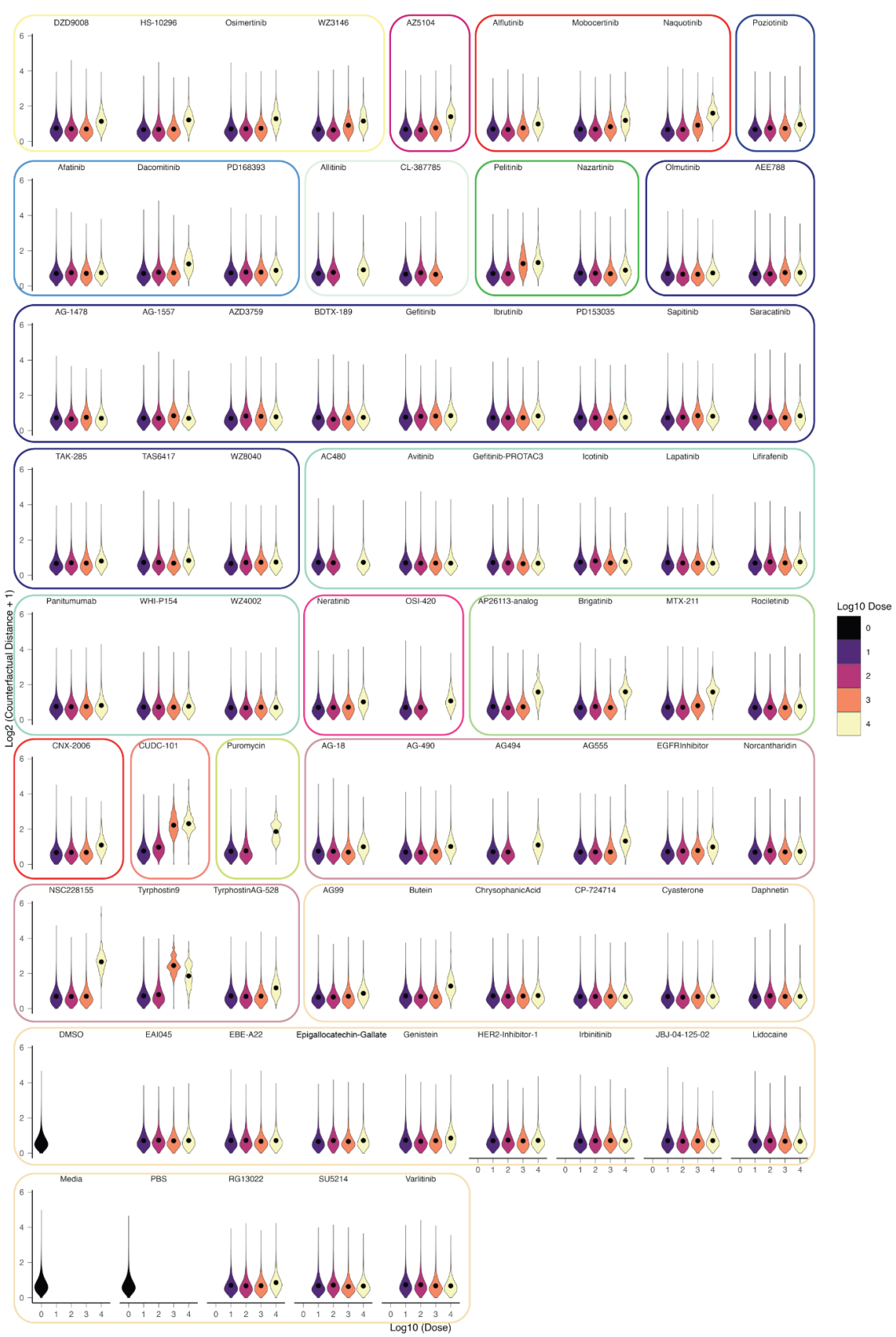


Supplemental Figure 8. Sample-sample distances represent transcriptional similarity between drug-dose conditions. Sample-sample distances for a given cell are more specifically computed by calculating the Euclidean distances between sample-specific counterfactual cell states. Counterfactual predictions are estimated by leveraging the functional relationship between the sample-unaware U-space and sample-aware Z-space. **A.** Mean sample-sample distances

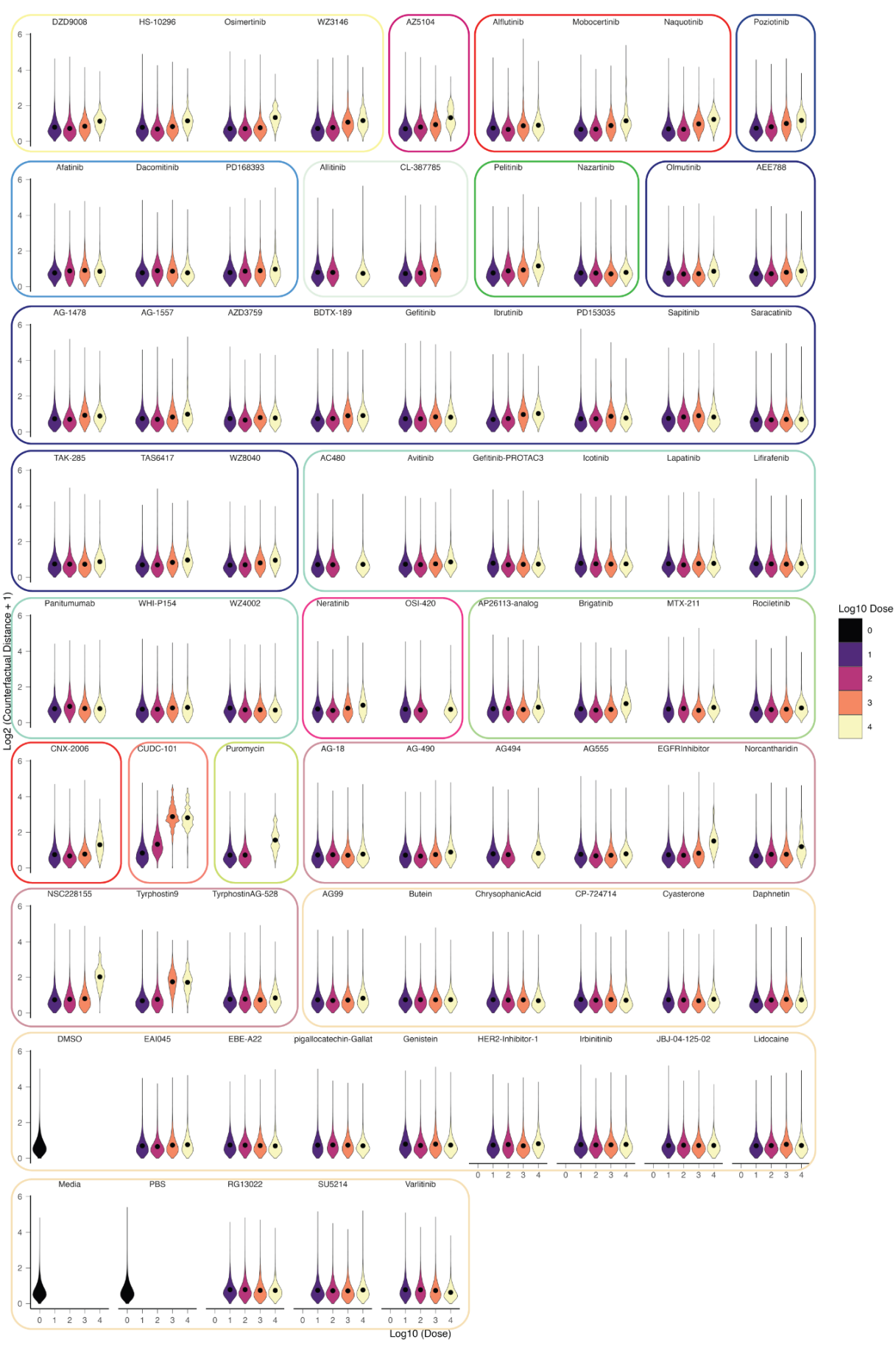
across cells shown for BT112 and BT333, with the number of drug-dose conditions sampled for display. To sample, at most 10 random drug-dose conditions were selected from each response module. **B.** Log fold-changes (LFCs) across differentially expressed genes for each response module, as determined by covariate-specific differential expression analysis. Genes were considered differentially expressed if they had a LFC > 0.1 and an FDR < 5%. FDR values were obtained with pseudo-bulk general linear models.



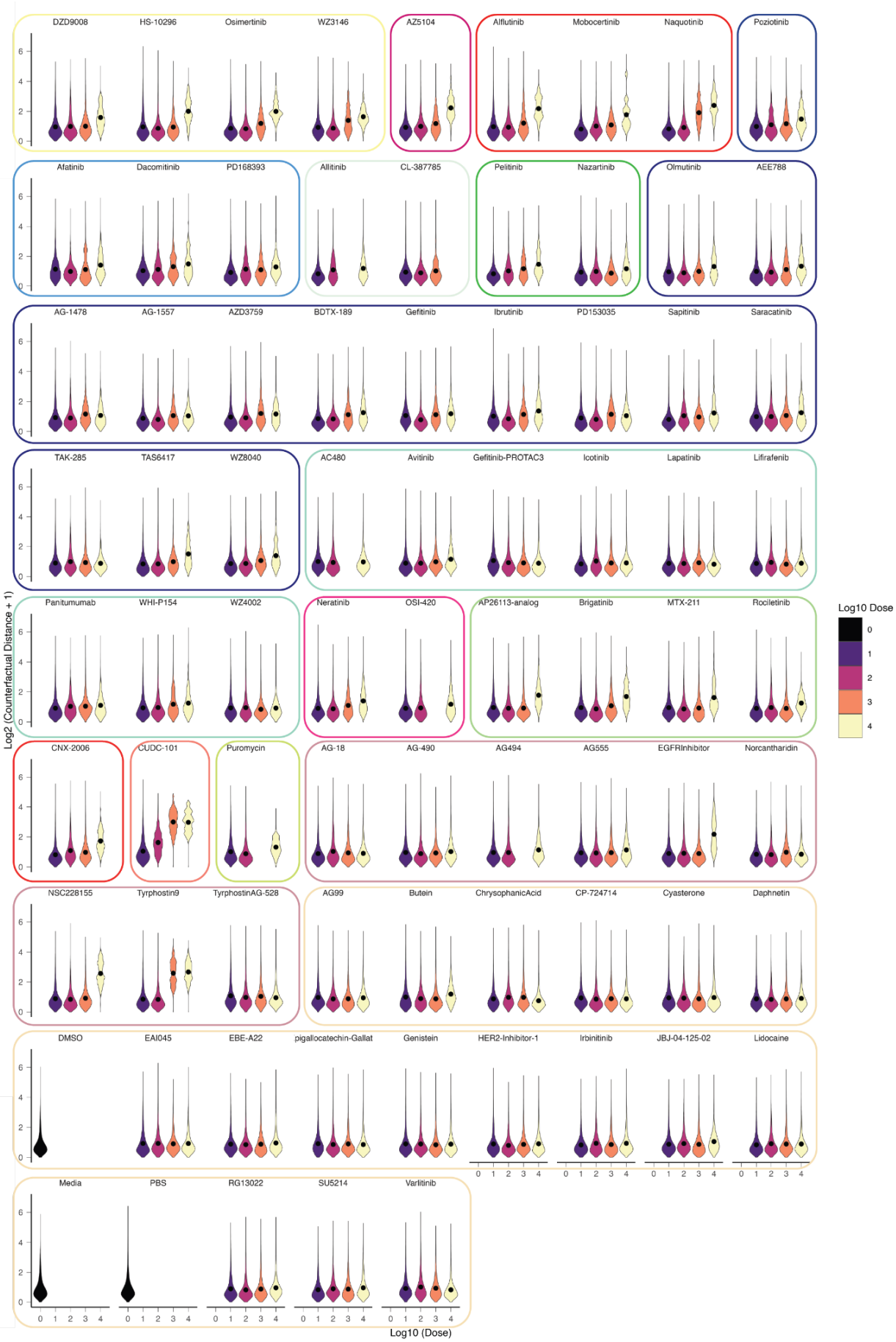
Supplemental Figure 9. **Drug-dose membership across response modules for each PDCL.** The response module axes belonging to each PDCL are hierarchically clustered individually, and the drugs axis for each PDCL is clustered according to the hierarchical clustering of the BT112 drug axis. Here, drugs are shown to be similarly grouped across cell lines in accordance to their ability to induce specific transcriptional responses.



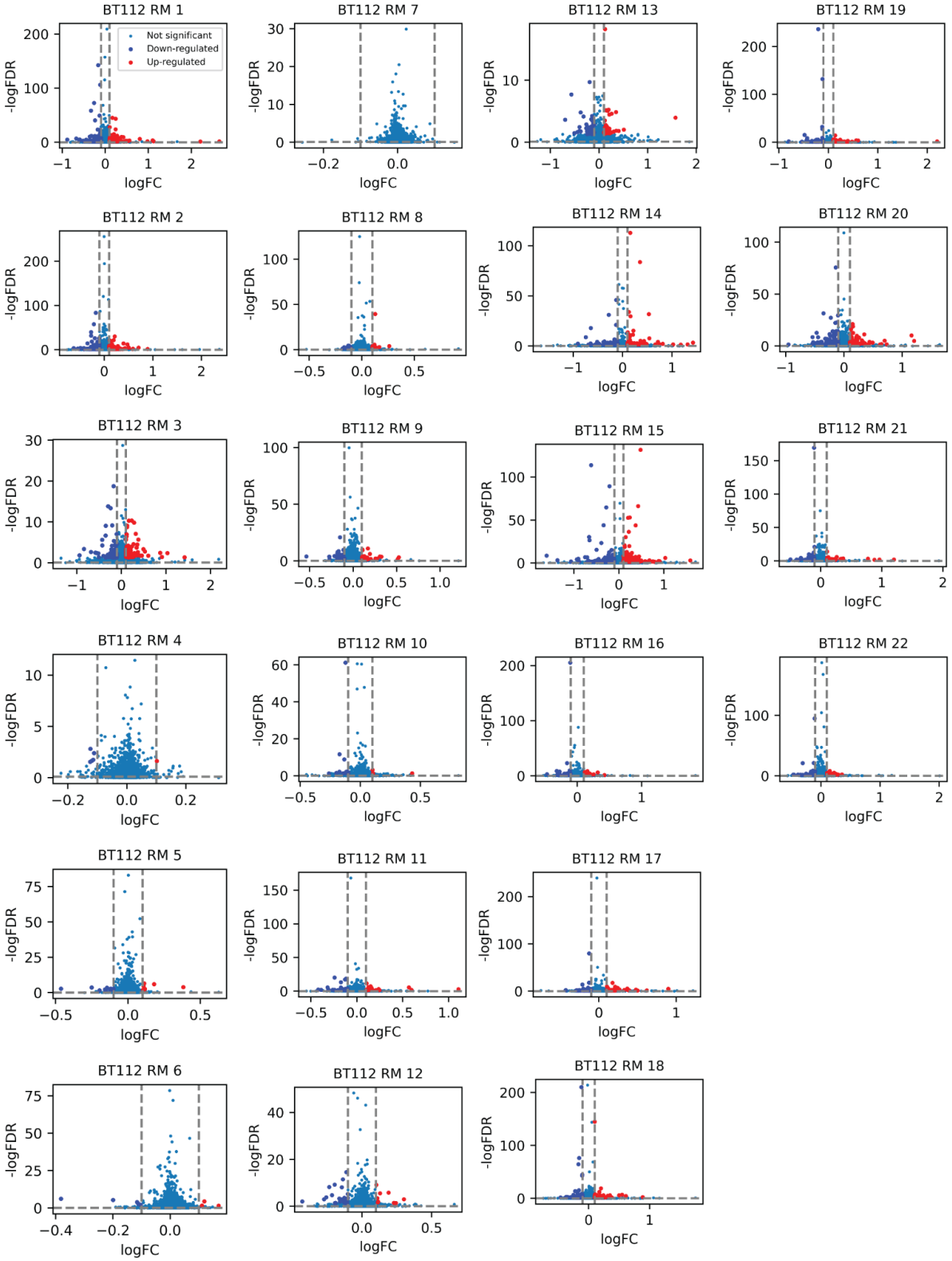
Supplemental Figure 10. Violin plots displaying the distribution of counterfactual sample-sample distances across drug-dose conditions for BT112. The sample-sample distances from a given cell's drug-dose condition to all other drug-dose conditions, including its own, are included as individual points in the violin plots. Hence, the number of cells x the number of drug-dose conditions data points are utilized. These plots exhibit the single-cell heterogeneity of drug-dose-specific induced transcriptional responses uncovered by MrVI.



Supplemental Figure 11. Violin plots displaying the distribution of counterfactual sample-sample distances across drug-dose conditions for BT228. The sample-sample distances from a given cell's drug-dose condition to all other drug-dose conditions, including its own, are included as individual points in the violin plots. Hence, the number of cells x the number of drug-dose conditions data points are utilized. These plots exhibit the single-cell heterogeneity of drug-dose-specific induced transcriptional responses uncovered by MrVI.

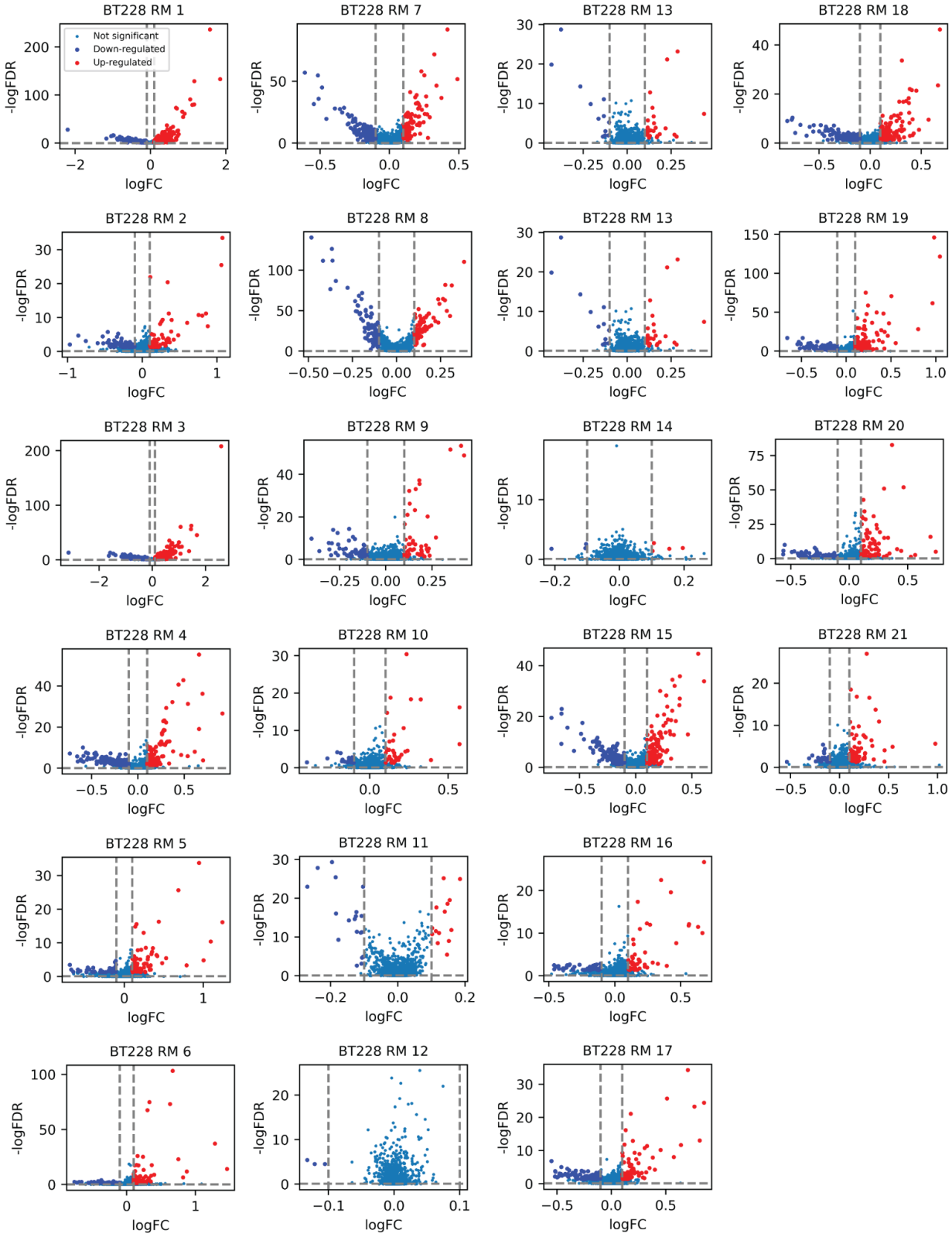


Supplemental Figure 12. Violin plots displaying the distribution of counterfactual sample-sample distances across drug-dose conditions for BT333. The sample-sample distances from a given cell's drug-dose condition to all other drug-dose conditions, including its own, are included as individual points in the violin plots. Hence, the number of cells x the number of drug-dose conditions data points are utilized. These plots exhibit the single-cell heterogeneity of drug-dose-specific induced transcriptional responses uncovered by MrVI.



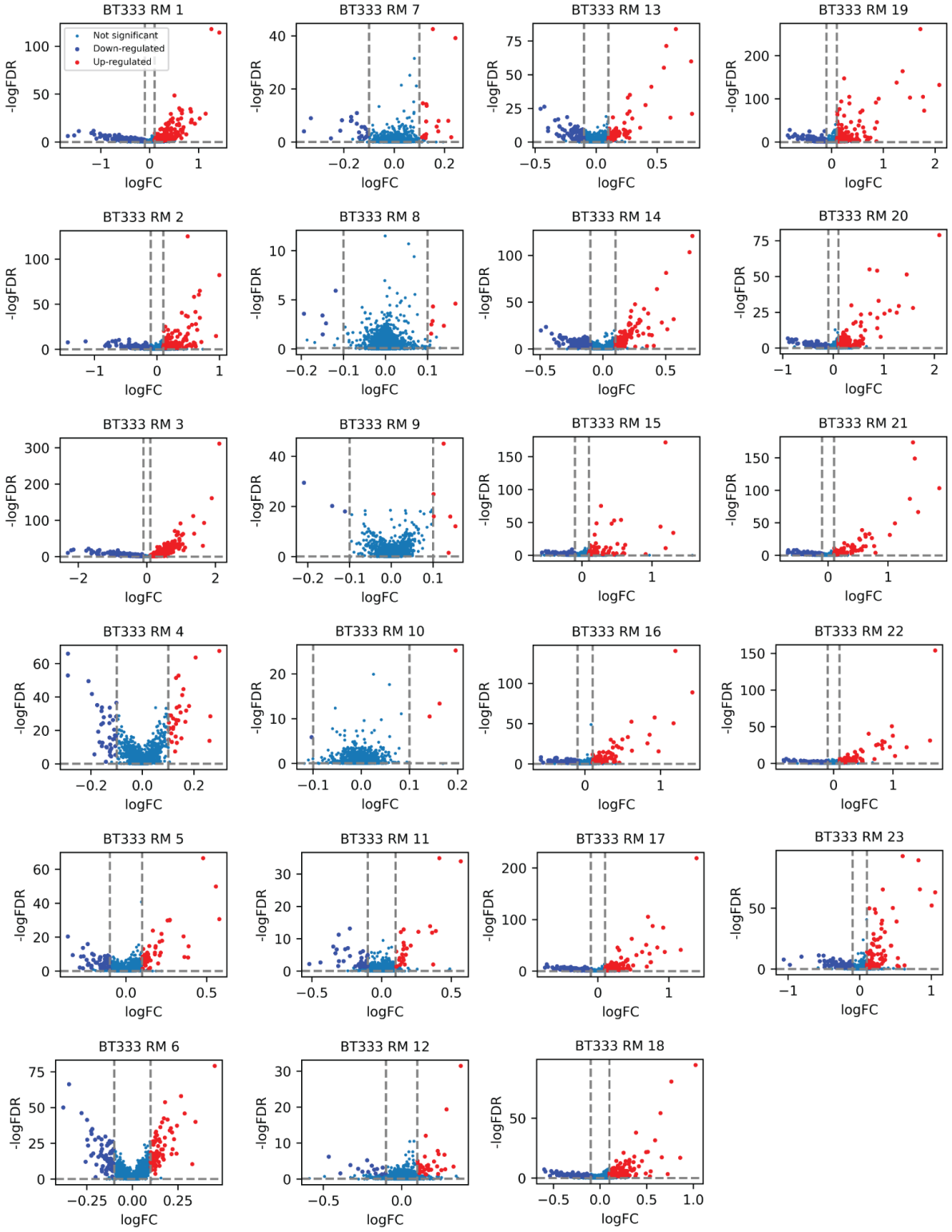
Supplemental Figure 13. Volcano plots to determine differentially expressed genes in response modules (RMs) for the BT112 patient-derived cell line (PDCL). The logFCs

(threshold = 0.1) were obtained with MrVI multivariate analysis and the FDR-adjusted p-values (threshold = 0.05) were obtained with pseudo-bulk general linear models.



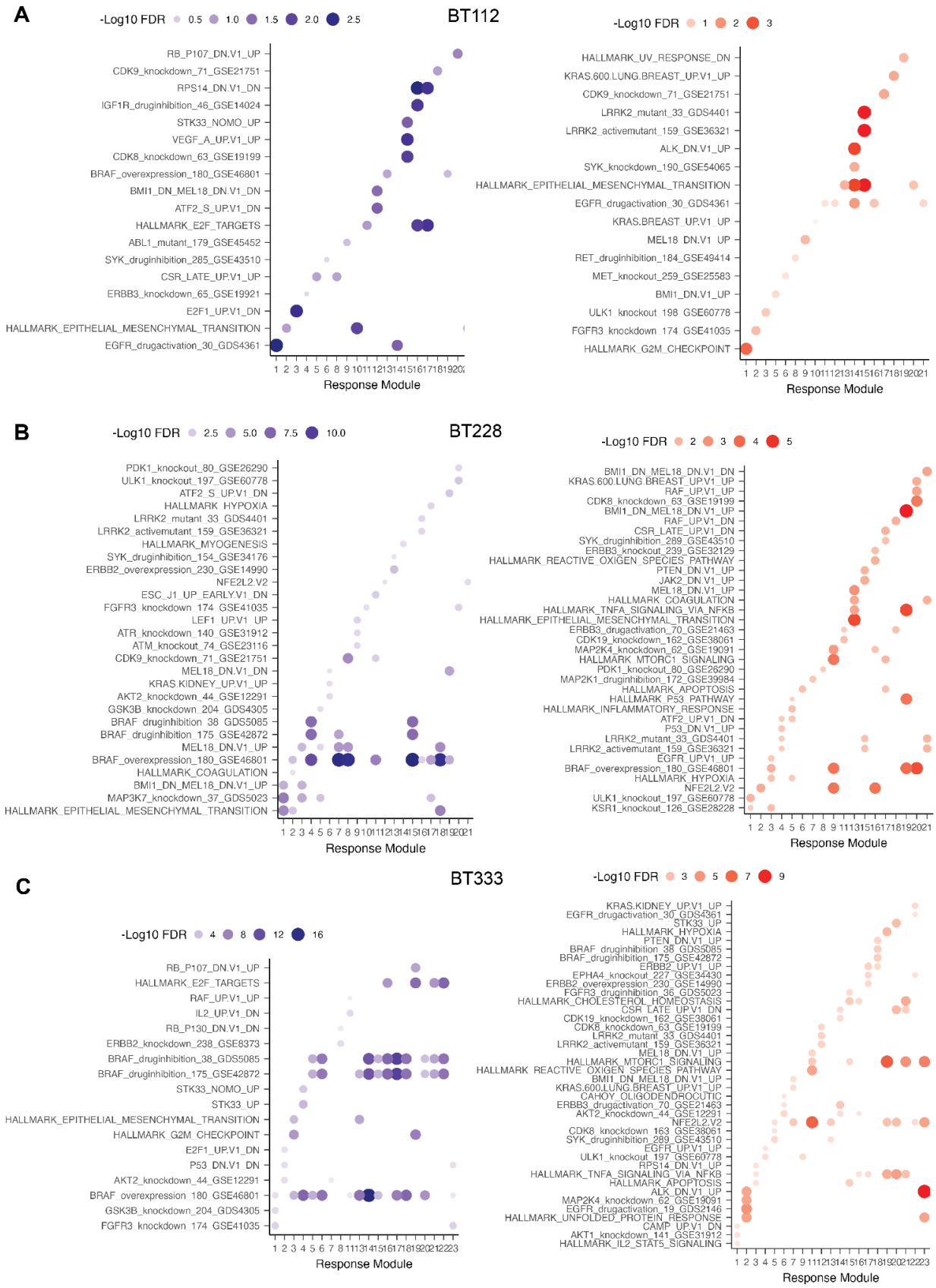
Supplemental Figure 14. Volcano plots to determine differentially expressed genes in RMs for the BT228 PDCL. The logFCs (threshold = 0.1) were obtained with MrVI multivariate analysis

and the FDR-adjusted p-values (threshold = 0.05) were obtained with pseudo-bulk general linear models.



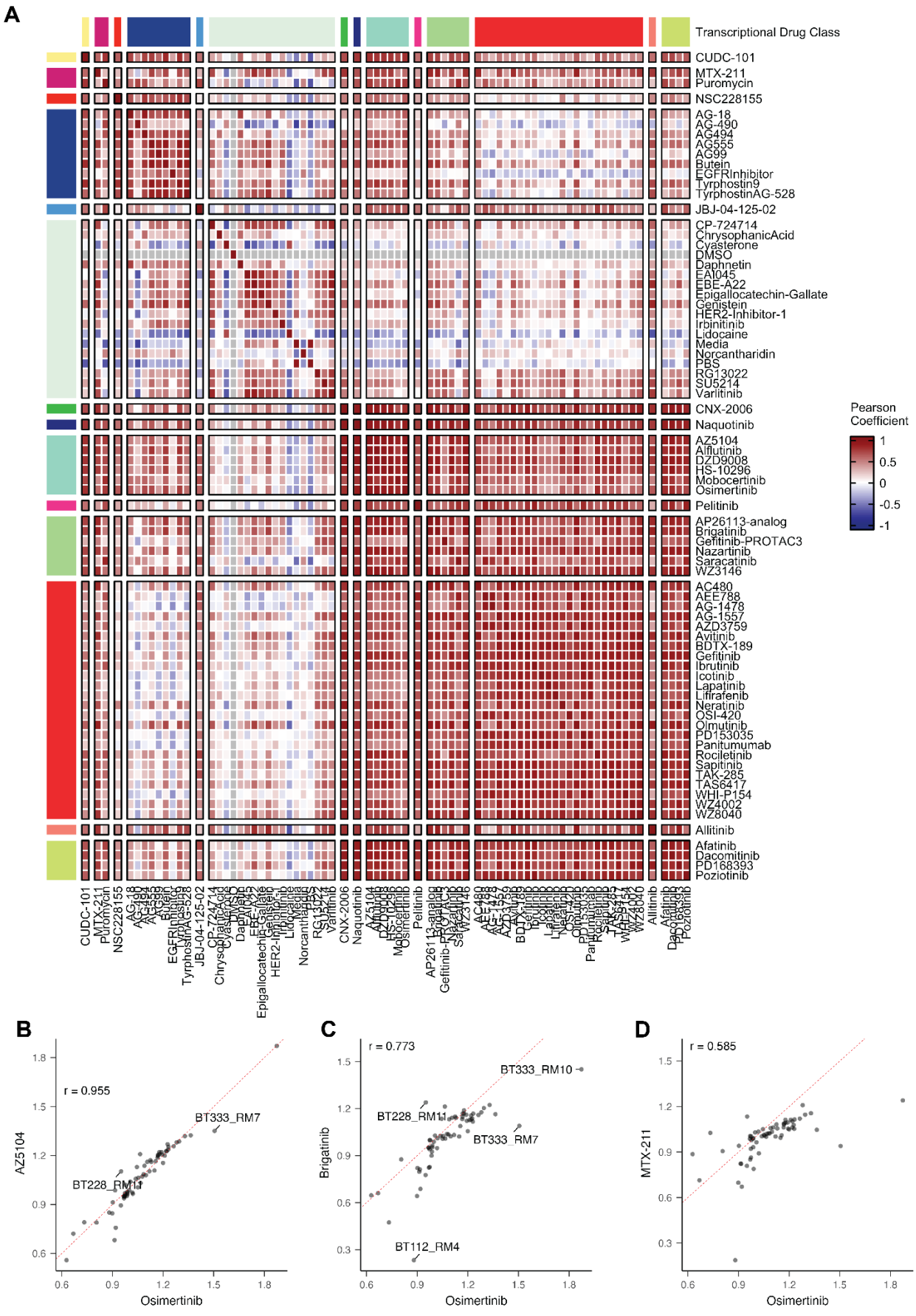
Supplemental Figure 15. Volcano plots to determine differentially expressed genes in RMs for the BT333 PDCL. The logFCs (threshold = 0.1) were obtained with MrVI multivariate analysis

and the FDR-adjusted p-values (threshold = 0.05) were obtained with pseudo-bulk general linear models.

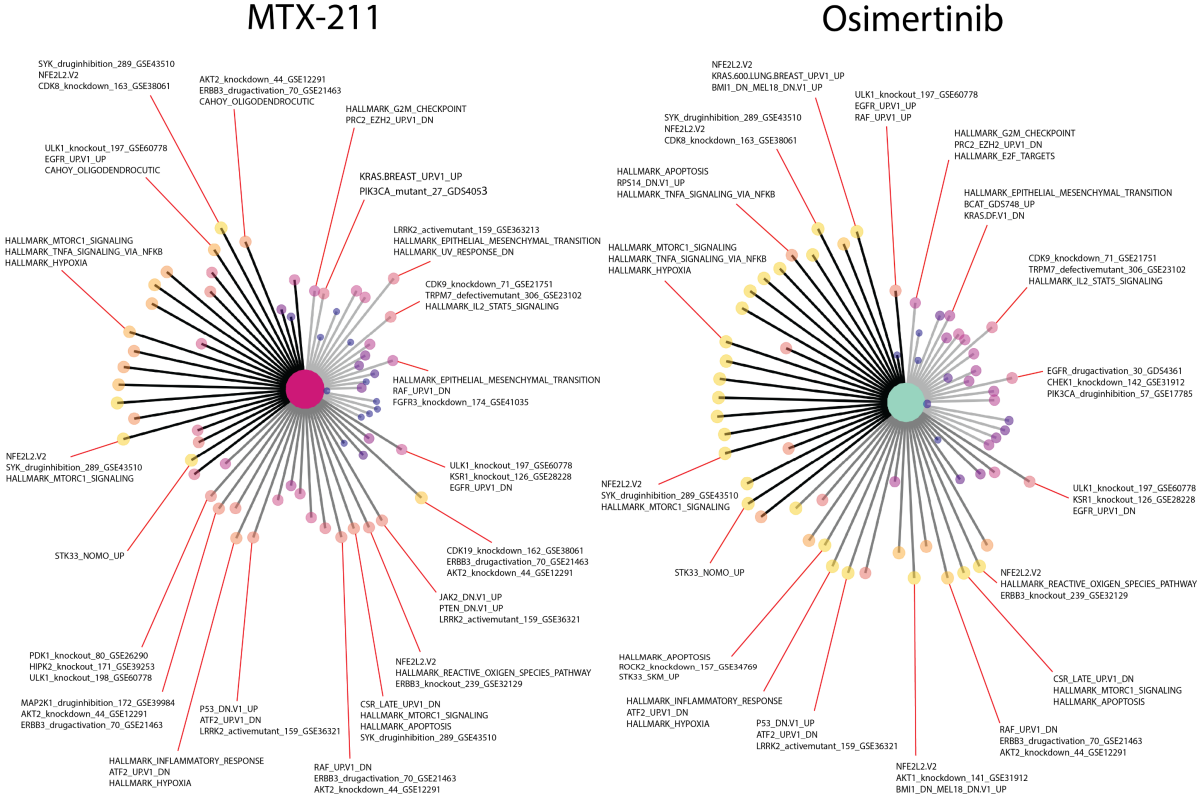


Supplemental Figure 16. GSEA results for upregulated and downregulated RM DEGs for **(A)** BT112, **(B)** BT228, **(C)** BT333. Of note, several BT112 RMs were not significantly enriched for the tested gene sets after multiple hypothesis correction, but top gene sets are displayed.

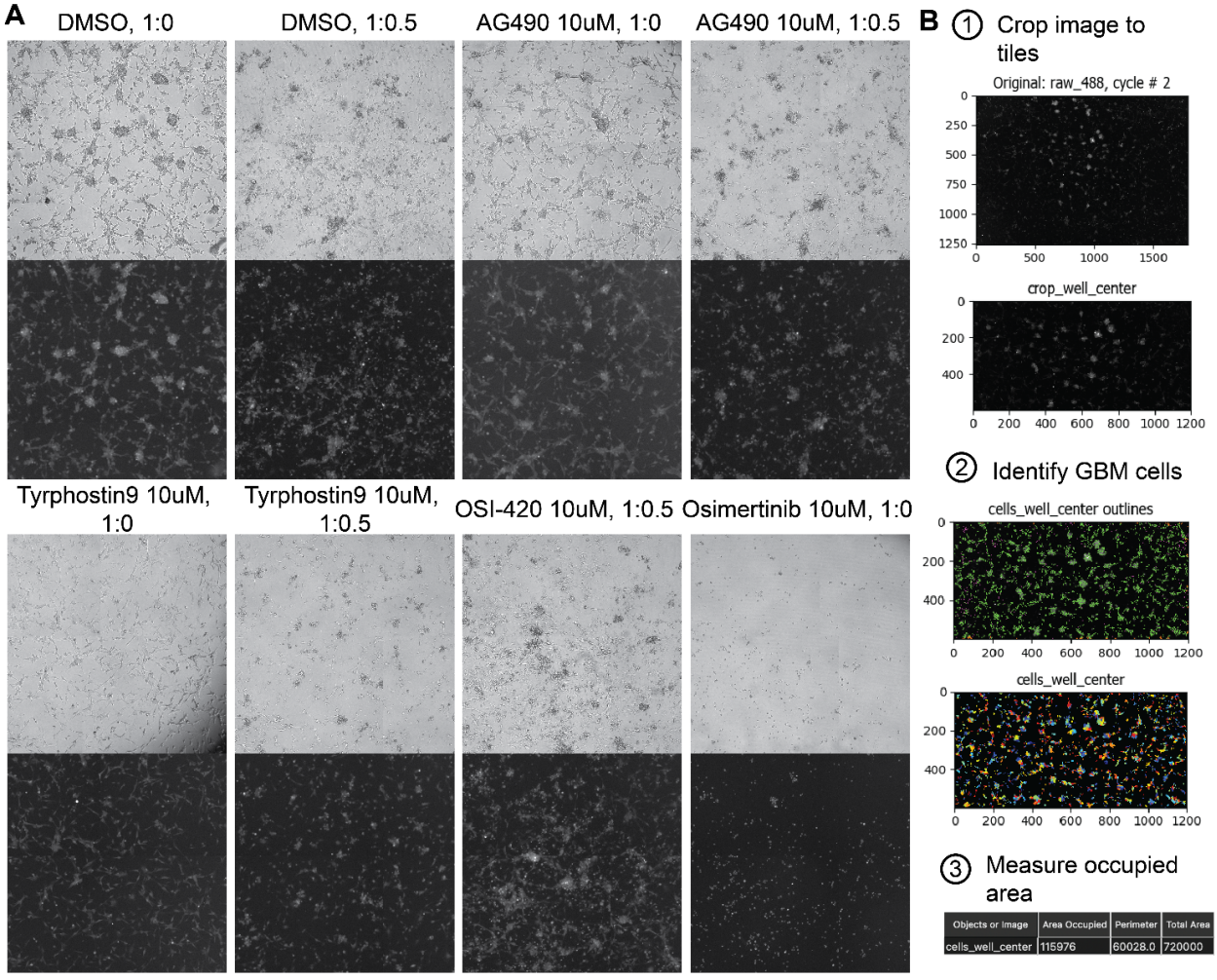
Supplemental Figure 17. Transcriptional drug classification of EGFRi visualized by lollipop plots. Briefly, for a given drug, each response module's respective up-regulated DE genes (identified by covariate-specific DE analysis) were aggregated, averaged, and log2 normalized to the mean of its corresponding DMSO control. This quantity is represented by the size and color of each lollipop. Additionally, because aggregate score is calculated at the single-cell level, heterogeneity in response module expression could be characterized as percent cells within a drug's high dose population with greater expression than DMSO control. This quantity is represented by the length of each lollipop. The center color of each plot represents its TDC membership.



Supplemental Figure 18. Drugs within TDCs are more similar than drugs in other TDCs by response module expression. A. Pairwise Pearson correlation of a 10uM drug's response module aggregate expression. **B-D.** Osimertinib RM expression is highly correlated with **(B)** AZ5104, a fellow member of TDC9. Osimertinib RM expression is less correlated with **(C)** Brigatinib and even less so with **(D)** MTX-211, which are part of TDC11 and TDC2, respectively.



Supplemental Figure 19. Lollipop plot examples labeled with representative gene sets. The quantities that determine the size and color of each lollipop are described in **Supp. Fig. 17**. GSEA enrichment analysis was performed with the up-regulated DE genes for each response module across PDCLs. Each lollipop is annotated with one to three of the most significantly enriched gene sets in the appropriate response module.



Supplemental Figure 20. Assessing GBM cytotoxic T-cell killing after pre-exposure to select EGFRi. **A.** Representative U87MG NY-ESO-1 brightfield (top) and fluorescent 488nm (bottom) 48hr images for select EGFRi pre-treatment and GBM:T-cell ratios, denoted as titles. Osimertinib images were included to demonstrate its drastic effect on U87MG cells even absent of $CD8^+$ T-cells. **B.** CellProfiler pipeline example to illustrate the image processing pipeline. Briefly, raw 488nm images were imported, cropped to well center, scanned to identify cells, and measured for area occupied by cells. **C.** Raw measurements of occupied CFSE area for 10uM EGFRi pre-treated U87MG NY-ESO-1 at 48hrs post-T-cell addition. It is important to note that osimertinib high dose treatment maximally decreased U87MG CFSE area and that addition of T-cells did not lead to any further decrease. **D.** Quantifications as in Fig. 5 for cells exposed to the specified inhibitor and a 1:0.25 ratio of GBM:T-cells. The color of the bars denote TDC membership, and the inhibitors in J are ordered by increasing T-cell effect, as decreasing mean relative area. Error bars represent standard deviation from the mean across replicates ($n_{DMSO,T\text{-}cell} = 34$, $n_{drug,T\text{-}cell} = 11$, $n_{Tyrphostin9,T\text{-}cell} = 9$).

Supplementary Tables

Supplementary Table S1: Pilot screen: Differentially expressed genes as a function of dose for each EGFR inhibitor.

Supplementary Table S2: Basal PDCL: Differentially expressed genes as a function of PDCL sub-clusters.

Supplementary Table S3: Chemicals used in large EGFR inhibitors screen.

Supplementary Table S4: Large-scale screen: Differentially expressed genes as a function of dose for each EGFR inhibitor.

Supplementary Table S5: Membership of each drug within each TDC

Supplementary Table S6: Large-scale screen: Differentially expressed genes as a function of MrVI-determined response module