**Drug Mechanism Enrichment Analysis: A tool to link molecular signatures with sensitivity to drug mechanisms of action**

Belinda B. Garana[1], Matthew P. Jeon[1], James H. Joly[1,†], and Nicholas A. Graham[1,2,3]

[1]Mork Family Department of Chemical Engineering and Materials Science,

[2]Norris Comprehensive Cancer Center,

[3]Leonard Davis School of Gerontology,

University of Southern California, Los Angeles, CA, USA

[†]Current address: Nautilus Biotechnology, San Carlos, CA, USA

Running title:

Drug Mechanism Enrichment Analysis

To whom correspondence should be addressed:

Nicholas A. Graham, University of Southern California, 3710 McClintock Ave., RTH 509, Los Angeles, CA 90089. Phone: 213-240-0449; E-mail: nagraham@usc.edu

## ABSTRACT

### Motivation

Targeted therapeutics have the potential for efficacy against tumors with minimal effects on normal tissues. However, predicting effective drugs from molecular signatures remains a challenge. Here, we present Drug Mechanism Enrichment Analysis (DMEA), a method that uses a transcriptomic signature to predict drug mechanism(s) of action to which a tumor cell may be sensitive or resistant. The method derives its power by aggregating data from many drugs with a shared mechanism of action.

### Results

We first tested the sensitivity of DMEA using synthetic data. We next validated that DMEA recapitulated known sensitivities to HMGCR, EGFR, and RAF inhibitors while also identifying drug mechanisms for resistant cancers. Finally, we predicted tissue-dependent drug sensitivity for tumors with high and low expression of the cystine/glutamate antiporter xCT. Collectively, DMEA is a novel bioinformatic tool that uses molecular signatures to predict targeted therapeutics sharing a common mechanism of action.

### Availability and implementation

DMEA is freely available to download as an R package at: https://github.com/BelindaBGarana/DMEA.

**Introduction**

There is a pressing need to identify effective targeted therapeutics for cancer. While targeted therapies have the potential for efficacy on tumor tissues with minimal effect on normal tissues, only fifteen percent of cancer patients are currently eligible for targeted therapies, and only half of eligible patients respond to treatment (Haslam *et al.*, 2021). Thus, improved methods to match tumors with effective therapeutics are desperately needed to realize the promise of precision oncology.

Molecularly targeted therapeutics are often chosen based on the overexpression or mutation of a single gene, usually the molecular target of the drug. For example, BRAF inhibitors like vemurafenib are given to patients with oncogenic, activating BRAF mutations. In BRAF-mutant melanoma, this approach has a response rate of over fifty percent (Sosman *et al.*, 2012). However, the presence of a mutation does not always indicate sensitivity to a given drug. In BRAF-mutant colon cancer, for example, only five percent of patients respond to BRAF inhibitor monotherapy (Corcoran *et al.*, 2018). This challenge has motivated the search for improved methods to identify biomarkers that predict drug sensitivity.

Advances in large-scale pharmacological screening and “omic” characterization of cancer cell lines have accelerated research into the molecular correlates of drug sensitivity (Garnett *et al.*, 2012; Iorio *et al.*, 2016; Basu *et al.*, 2013; Seashore-Ludlow *et al.*, 2015; Corsello *et al.*, 2020; Subramanian *et al.*, 2017; Barretina *et al.*, 2012; Ghandi *et al.*, 2019; Shoemaker, 2006). These databases of molecular profiles and drug responses from cancer cell lines have enabled the development of many algorithms for drug sensitivity prediction including elastic net regression, support vector regression, neural networks, random forests, multitask learning, and others (Parca *et al.*, 2019; Wei *et al.*, 2019; Suphavilai *et al.*, 2018; Yuan *et al.*, 2016; Choi *et al.*, 2020; Gonçalves *et al.*, 2020; Nguyen *et al.*, 2017). Although these models do not explicitly

include the molecular target of drugs, some algorithms accurately group drugs with the same mechanism of action without *a priori* knowledge (Suphavilai *et al.*, 2018; Yuan *et al.*, 2016). In fact, the similarity between sensitivity profiles for drugs with common molecular targets has enabled the discovery of mechanism of action for unannotated compounds (Rees *et al.*, 2016) and the estimation of on-target/off-target drug activity (Gonçalves *et al.*, 2020).

Because large-scale pharmacological screens often include multiple inhibitors with the same molecular target (e.g., EGFR, BRAF, etc.), we reasoned that a drug sensitivity prediction method that explicitly incorporates known mechanisms of action would result in robust predictions of sensitivity and resistance. In addition, we sought to develop an accessible method that can translate user-provided molecular signatures (i.e., not derived from the pharmacological screening databases) into drug sensitivity predictions. To address these needs, we present Drug Mechanism Enrichment Analysis (DMEA), a bioinformatic tool to predict drug mechanism(s) of action that will exhibit sensitivity or resistance for cells based on an input molecular signature. By aggregating information across drugs that share a common mechanism of action, DMEA increases confidence in the prediction of sensitive/resistant drug mechanisms, thereby providing a user-friendly, robust method to link molecular signatures with targeted therapeutics.

## ALGORITHM AND METHODS

### *Drug Mechanism Enrichment Analysis (DMEA)*

DMEA tests whether drugs that share a common mechanism of action are statistically enriched in a ranked list. Here, we have demonstrated the method using 1,351 drugs from the Cancer Dependency Map's PRISM dataset (Corsello *et al.*, 2020), but the method can be used on any ranked list of drugs where the drugs can be grouped by mechanism of action. For this study (Fig. 1), we have created ranked lists of drug sensitivity scores by 1) comparing two groups of interest (e.g., drug sensitive and drug resistant cells) to create a transcriptional signature of gene weights (i.e., $\log_2$(fold change) in gene expression, see section on Transcriptomic Data below); 2) calculating a weighted gene voting (WGV) molecular classifier score (Golub *et al.*, 1999) as the dot product between gene weights and gene expression values for 327 adherent cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) (Ghandi *et al.*, 2019); 3) calculating the Pearson correlation between the WGV score and PRISM drug sensitivity score (i.e., area under the curve (AUC) of cell viability versus drug concentration) for each of the 1,351 drugs; and 4) ranking 1,351 drugs by the Pearson correlation coefficient. Next, the ranked list is analyzed for statistically over-enriched drug mechanisms of action using an algorithm similar to Gene Set Enrichment Analysis (GSEA) (Subramanian *et al.*, 2005). That is, drug sets are assigned based on the known mechanism of action for each drug (Supp. Table 1), and then the enrichment score (ES) of each drug set is calculated as the maximum deviation from zero of a running-sum, weighted Kolmogorov-Smirnov-like statistic. The p-value is estimated using an empirical permutation test wherein the drugs are randomly assigned mechanism-of-action labels in 1,000 independent permutations to calculate a distribution of null enrichment scores (ES_null); the p-value is then calculated as the percentage of same-signed ES_null equal to or greater than the ES divided by the percentage of same-signed ES_null. The normalized enrichment score (NES) is then calculated by dividing the ES by the mean of the same-signed portion of the ES_null distribution. Finally, the q-value or false discovery rate is

calculated as the percentage of same-signed NES in the null distribution (i.e., NES_null) with NES equal or greater to the observed NES divided by the percentage of same-signed NES equal or greater. The R package for DMEA is freely available at: https://github.com/BelindaBGarana/DMEA.

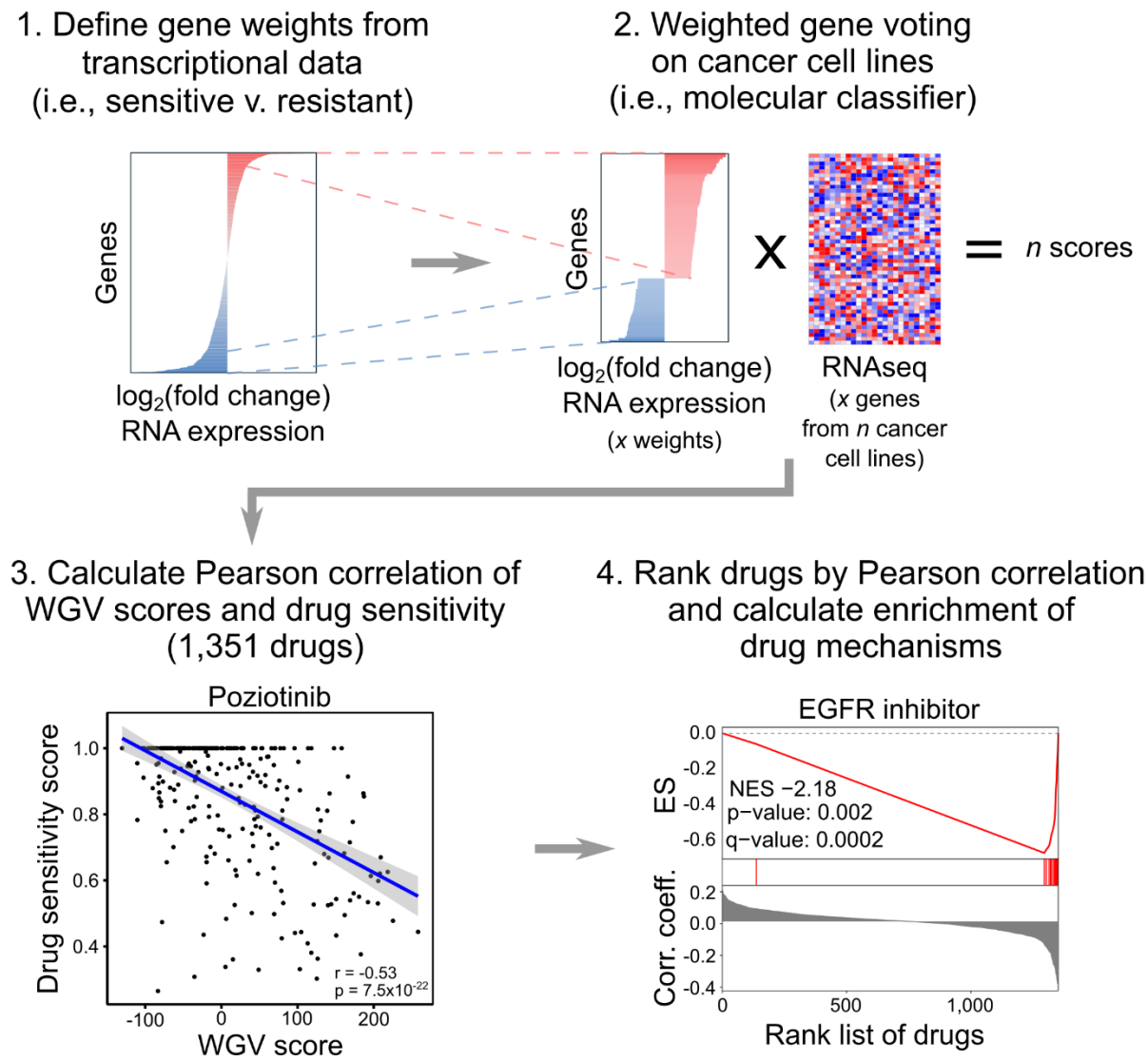



**Figure 1. Overview of Drug Mechanism Enrichment Analysis.** DMEA consists of four steps: First, gene weights are calculated as $\log_2$(fold change) values calculated between two groups (e.g., drug sensitive and drug resistant cells) from transcriptional data. Only the *x* genes with $\log_2$(fold change) with q-value < 0.05 are selected for subsequent steps (maximum 500 genes). Second, a weighted gene voting (WGV) molecular classifier score is calculated as the dot product between *x* gene weights and gene expression values from cancer cell lines (*x* genes by *n* cell lines). Here, we used 327 adherent cancer cell lines from the CCLE. Third, the Pearson

correlation is calculated between the WGV score and drug sensitivity score (i.e., area under the curve (AUC) of cell viability versus drug concentration from the PRISM database) for each of the 1,351 drugs. An example is shown for the EGFR inhibitor poziotinib. Fourth, the 1,351 drugs are ranked by Pearson correlation coefficient, and the ranked list is analyzed using a GSEA-like algorithm to identify drug mechanisms that are enriched at either end of the ranked list. In total, 85 mechanisms of action are tested (Supp. Table 1). An example for EGFR inhibitors is shown. Each red tick mark represents a drug annotated as an EGFR inhibitor in the waterfall plot of correlation coefficients.

*Drug Sensitivity Data*

Here, we used the PRISM dataset of 1,351 drug sensitivity scores (i.e., area under the curve (AUC) values for cell viability as a function of drug concentration) for 327 adherent cancer cell lines (Corsello *et al.*, 2020). For each drug, data from the most recent PRISM screen was used including screens HTS002, MTS005, MTS006, and MTS010. The drugs were grouped into 85 mechanisms of action according to their annotation in PRISM, with a minimum of six drugs per mechanism of action (Supp. Table 1).

*Synthetic Data*

To evaluate the sensitivity of DMEA, we first simulated gene expression data (19,144 genes) for 200 cell lines. The level of each transcript was sampled from a normal distribution with mean ($\mu$) of 0 and standard deviation ($\sigma$) of 0.5. We then simulated drug sensitivity data for 1,351 drugs from a bimodal mixture of two normal distributions ($\mu_1$=0.83, $\sigma_1$=0.08 and $\mu_2$=1.31, $\sigma_2$=0.08) with the lower distribution containing 72% of all drugs. These distributions were chosen to reflect the distributions of the CCLE RNAseq and PRISM drug sensitivity data (Supp. Fig. 1).

Next, we perturbed a set of 25 randomly sampled genes using a normal distribution gradient. The magnitude of the perturbation ranged from 0 (no perturbation) to 0.1 by 0.01. For example, with a perturbation magnitude of 0.1, the mean gene expression for the 25 perturbed genes in cell line 1 was $\mu$ = -0.1, and the mean sequentially increased from -0.1 to 0.099 by

0.001 (i.e., 0.2/200) for cell lines 2-200. This sampling method created a gradient of perturbations in the 200 cell lines ranging from largest negative perturbation (i.e., cell line 1) to largest positive perturbation (i.e., cell line 200). Then, we calculated WGV scores for each cell line by taking the dot product of the difference in average gene expression between the top and bottom 10 percent of cell lines (i.e, gene weights) and the synthetic gene expression values.

Similarly, we perturbed the drug sensitivity scores for a set of 10 randomly sampled drugs. Again, the magnitude of the perturbation ranged from 0 (no perturbation) to 0.1 by 0.01. For a perturbation of magnitude 0.1, the mean of the distribution of cell line 1 was perturbed by -0.1, with perturbation values for cell lines 2-199 sequentially increasing from -0.1 to 0.099 by 0.001. Then, we calculated the Pearson correlation between the WGV and drug sensitivity scores for each of the 1,351 drugs in the synthetic dataset. Finally, drugs were ranked by the Pearson correlation coefficient, and DMEA was performed as described above to measure the enrichment of the 10-drug synthetic drug set.

*Transcriptional Signatures of Known Drug Sensitivity*

We validated DMEA using molecular signatures of known drug sensitivity. First, we tested a signature of atorvastatin sensitivity using data from the PRISM and CCLE databases. Specifically, we classified the top 10% of cell lines based on AUC for the drug atorvastatin as resistant and the bottom 10% of cell lines as sensitive. Then, we calculated gene weights between these two classes to perform WGV and then DMEA. Next, we tested independent gene signatures of sensitivity to EGFR and RAF inhibitors. For prediction of EGFR inhibitor sensitivity, we used data sets GSE12790 (Hoeflich *et al.*, 2009), GSE31625 (Balko *et al.*, 2006), and (Coldren *et al.*, 2006). For GSE12790, cell lines were classified as sensitive to erlotinib (EC50 < 1 μM: HDQ-P1, CAL85-1, and HCC1806) or resistant to erlotinib (EC50 > 10 μM: CAL-51, CAL-120, MDA-MB-231, BT-20, HCC1569, EFM-192A, HCC1954, MDA-MB-453, BT474, HCC1428,

T47D, ZR-75-1, KPL-1, BT-483, MDA-MB-415, HCC1500, CAMA-1, and MCF7). For GSE31625, we used 17 transcriptomic profiles from 3 cell lines sensitive to erlotinib (H1650, H3255, and PC-9) and 12 profiles of 2 cell lines resistant to erlotinib (A549 and UKY-29). Finally, from Coldren et al., we used transcriptomic profiles of 5 cell lines sensitive to gefitinib (NCIH1650, HCC95, NCIH1975, NCIH1648, and NCIH2126) and 7 cell lines resistant to gefitinib (NCIH520, NCIH460, NCIH1299, HCC44, A549, NCIH1703, and HCC15). For prediction of RAF inhibitor sensitivity, we used dataset GSE66539 with paired biopsy samples from 3 patients before vemurafenib treatment and after emergence of vemurafenib resistance (Kemper *et al.*, 2016).

*Tissue-Dependent Drug Sensitivity for Tumors Based on xCT Expression*

For analysis of tumors based on expression of the glutamate/cystine antiporter xCT (*SLC7A11*), we used RNAseq data from twelve tumor types from the TCGA Research Network (https://www.cancer.gov/tcga) including: breast cancer (BRCA, n=1,212), kidney renal clear cell carcinoma (KIRC, n=606), lung adenocarcinoma (LUAD, n=576), head and neck squamous cell carcinoma (HNSC, n=566), lung squamous cell carcinoma (LUSC, n=552), prostate adenocarcinoma (PRAD, n=550), skin cutaneous melanoma (SKCM, n=473), uterine corpus endometrial carcinoma (UCEC, n=381), ovarian cancer (OV, n=307), low grade glioma (LGG, n=283), colorectal adenocarcinoma (COADREAD, n=263), and glioblastoma (GBM, n=171). We identified the top 10% (i.e., $SLC7A11^{hi}$) and bottom 10% (i.e., $SLC7A11^{low}$) of tumors based on *SLC7A11* gene expression, and then assigned gene weights as the $\log_2(SLC7A11^{hi}/SLC7A11^{low})$ fold change in gene expression for each gene using the DESeq2 analysis tool through GenePattern (https://www.genepattern.org) (Reich *et al.*, 2006).

For each transcriptomic data set, we used the R package limma (Ritchie *et al.*, 2015) to perform an eBayes statistical analysis. For genes with q-value < 0.05, the gene weights were

assigned as the $\log_2$(fold-change) in gene expression or difference of means between sensitive and resistant samples. In all analyses, a maximum of 500 gene weights were used for WGV to reduce computational time.

*Transcriptomic Data Used for Weighted Gene Voting*

For all DMEA examples, we used RNAseq data for 327 adherent cancer cell lines from CCLE (version 19Q4) (Ghandi *et al.*, 2019). To avoid overfitting, we removed any cell lines overlapping between the transcriptomic data set that was used to define WGV gene weights and the CCLE RNAseq dataset.

## RESULTS

### *Sensitivity Analysis*

To test the DMEA algorithm, we first performed a simulation study (see Methods). We generated gene expression and drug sensitivity scores for 200 cell lines by randomly selecting values from distributions that reflected the CCLE RNAseq and PRISM drug sensitivity data (Supp. Fig. 1). We then perturbated both the gene expression data and the drug sensitivity scores with a linearly increasing normally distributed random deviation to create a synthetic association between gene expression and drug sensitivity. To assess DMEA's sensitivity and specificity, we varied the magnitude of the random perturbation for both the gene expression data and drug sensitivity scores, ran 50 replicates, and then plotted the average normalized enrichment score (NES) (Fig. 1A) and the percent of replicates with significant enrichment (Fig. 1B) of the synthetic drug set across all perturbation combinations. As expected, the average NES of the synthetic drug set increased from zero when neither drug AUC nor RNA expression were perturbed to approximately 2.0 when both the drug AUC and RNA expression were perturbed by 0.10. Importantly, when the drug perturbation was small (<0.02), the tested drug set was never significantly enriched (0% of replicates) for all values of the RNA expression perturbation. Thus, DMEA is not prone to false positive results in this synthetic data set. To reach significance in 100% of replicates, the minimum perturbations for drug AUC and RNA expression were 0.08 and 0.07, respectively. These results illustrate that DMEA can successfully identify coordinated perturbations in gene expression and drug sensitivity data with high reproducibility.

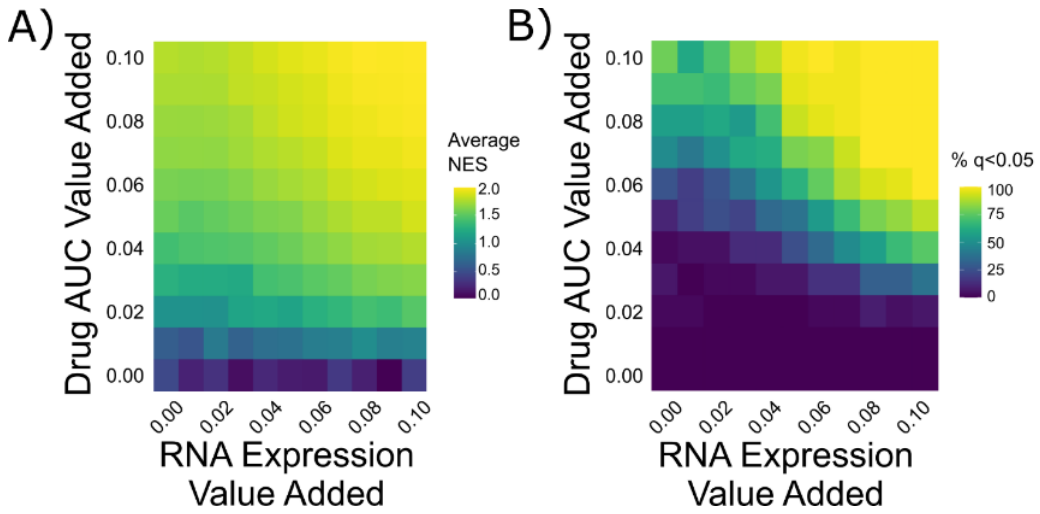



**Figure 2. Sensitivity analysis of DMEA pipeline using synthetic data.** Synthetic gene expression and drug sensitivity data was simulated with varying perturbations to RNA expression (x-axis) and drug sensitivity score (i.e., AUC, y-axis) (see Methods). A) Heatmap showing the average DMEA NES from 50 simulations for varying perturbations of the drug AUC and gene expression data. B) Heatmap showing the percent of DMEA replicates with FDR q-value < 0.05 from 50 simulations for varying perturbations of the drug AUC and gene expression data.

*DMEA correctly predicts sensitivity to HMGCR inhibitors*

After validating DMEA with synthetic data, we next tested DMEA using cancer cell lines common to both the CCLE (Ghandi *et al.*, 2019) and the PRISM (Corsello *et al.*, 2020) data sets. As our test case, we selected inhibitors of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase (HMGCR inhibitors), a class of six drugs (the minimum included in our study) that are widely used to lower cholesterol (Stancu and Sima, 2001). Using the drug atorvastatin, one of the most prescribed statins (Lippi *et al.*, 2019), we classified cancer cell lines as either sensitive (bottom 10% by AUC) or resistant (top 10% by AUC) to atorvastatin and then used RNAseq data from these cell lines to generate gene weights. Next, we performed WGV on the cancer cell lines not used to calculate gene weights (i.e., middle 80% by AUC) to define an HMGCRi sensitivity score. As expected, the HMGCRi inhibitor sensitivity score was significantly correlated with atorvastatin sensitivity (Fig. 3A). Because smaller AUC values indicate increased sensitivity, the observed correlation between HMGCRi and atorvastatin AUC was negative.

Next, we correlated WGV scores with drug sensitivity scores for the other 1,350 drugs in the PRISM data set, ranked drugs by the Pearson correlation, and used DMEA to identify enriched drug mechanisms of action (Supp. Table 2). Validating our approach, DMEA found that HMGCR inhibitors were significantly enriched in the rank list of drugs, with four of the top five ranked drugs being HMGCR inhibitors, including atorvastatin as the most negatively correlated drug (Fig. 3B). We next asked what other drug mechanisms of action were enriched in the HMGCRi drug rank list. This analysis revealed that ribonucleotide reductase inhibitors and PLK inhibitors were negatively enriched in the rank list, supporting that cell lines sensitive to HMGCR inhibitors are also sensitive to these drugs (Fig. 3C). Conversely, EGFR inhibitors and vitamin D receptor agonists were positively enriched in the rank list, supporting that cell lines resistant to HMGCR inhibitors will be sensitive to these drug mechanisms of action. Among the top 20 drugs in the rank list of 1,351 drugs, 15 were EGFR inhibitors including tucatinib, AZD8931, poziotinib, and neratinib (Fig. 3D). Altogether, these results support that DMEA can correctly identify sensitivity to HMGCR inhibitors and predict other drug mechanisms for sensitive and resistant cancer cell lines.

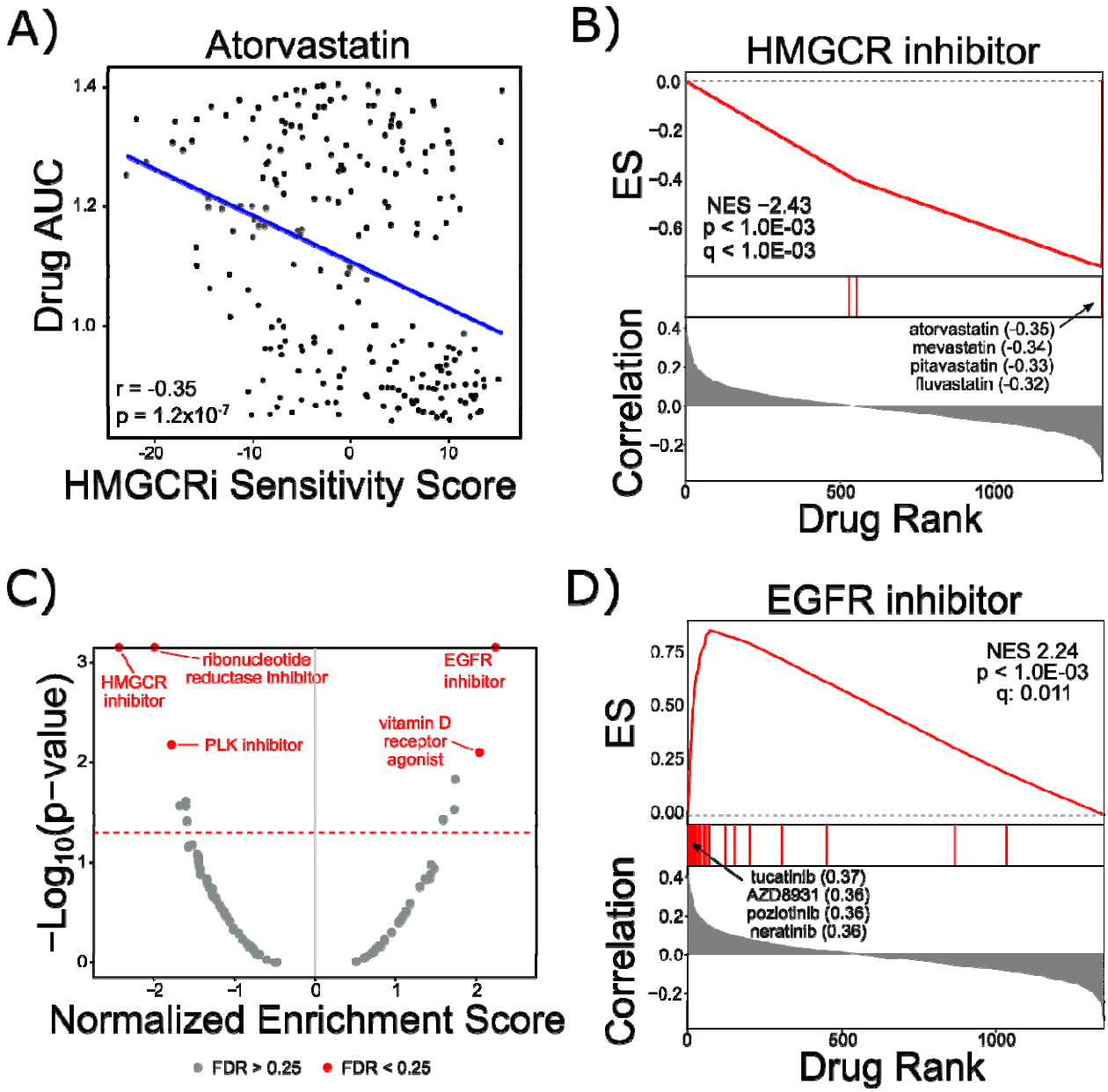



**Figure 3. DMEA correctly predicts sensitivity to HMGCR inhibitors.** Using the HMGCR inhibitor atorvastatin as a test case, an HMGCRi sensitivity score was calculated and then correlated with drug sensitivity data (i.e., AUC) for 1,351 drugs. Drugs were then ranked by correlation coefficient, and DMEA was performed. A) The HMGCRi sensitivity score was significantly correlated to atorvastatin sensitivity. Smaller AUC values indicate increased sensitivity to atorvastatin. B) Mountain plot showing that DMEA identified HMGCR inhibitors as enriched in the ranked list of drugs. The most negatively correlated HMGCR inhibitors are highlighted along with their correlation coefficients. C) Volcano plot of DMEA NES versus -$\log_{10}$(p-value) for all 85 tested drug mechanisms of action revealed negative enrichment of ribonucleotide reductase inhibitors and PLK inhibitors and positive enrichment of EGFR inhibitors and vitamin D receptor agonists in the ranked list of drugs. D) Mountain plot showing

enrichment of EGFR inhibitors in the ranked list of drugs. The most positively correlated EGFR inhibitors are highlighted along with their correlation coefficients.

*DMEA correctly predicts sensitivity to EGFR inhibitors*

We next sought to test DMEA using an independent (non-CCLE) transcriptional signature of drug sensitivity. To this end, we identified three transcriptomic signatures of sensitivity to EGFR inhibitors: 1) non-small cell lung cancer (NSCLC) cell lines treated with erlotinib (GSE31625) (Balko *et al.*, 2006); 2) breast cancer cell lines treated with erlotinib (GSE12790) (Hoeflich *et al.*, 2009); 3) and NSCLC cell lines treated with gefitinib (Coldren *et al.*, 2006). We used these signatures to define WGV scores for sensitivity to EGFR inhibitors. Notably, there was little overlap in the genes used for WGV (GSE12790 and Coldren *et al.* share zero genes, GSE12790 and GSE31625 share 11 genes, and GSE31625 and Coldren *et al.* share 14 genes). As expected, for all three EGFRi signatures, the WGV scores were significantly negatively correlated with sensitivity to the drug used to define the WGV signature (i.e., erlotinib or gefitinib) (Fig. 4A, Supp. Fig. 2A, Supp. Fig. 3A). Next, we correlated WGV scores with drug sensitivity scores for the other 1,350 drugs in the PRISM data set, ranked drugs by the Pearson correlation, and used DMEA to identify enriched drug mechanisms of action (Supp. Table 2). Validating our approach, DMEA found that EGFR inhibitors were significantly enriched in the rank list of drugs for each EGFRi signature (Fig. 4B, Supp. Fig. 2B, Supp. Fig. 3B). We next asked what other drug mechanisms of action were enriched in the EGFRi drug rank lists and whether these drug mechanisms of action were consistent across the three EGFRi signatures (Fig. 4C, Supp. Fig. 2C, Supp. Fig. 3C). Indeed, this analysis revealed that the three independent EGFRi signatures gave similar results (Fig. 4D) including the positive enrichment of HMGCR and MDM inhibitors (Fig. 4E,F, Supp. Fig. 2D, Supp. Fig. 3D). Using GSE12790, for example, among the top 20 drugs in the rank list of 1,351 drugs, 2 were HMGCR inhibitors (mevastatin, atorvastatin), and 4 were MDM inhibitors (idasanutlin, CGM097, nutlin-3, and AMG-232). This suggests that cell lines resistant to EGFR inhibitors will be sensitive to

inhibition of HMGCR and MDM. Taken together, these results support that DMEA can correctly identify sensitivity to EGFR inhibitors and predict other drug mechanisms for cancer cell lines resistant to EGFR inhibitors.

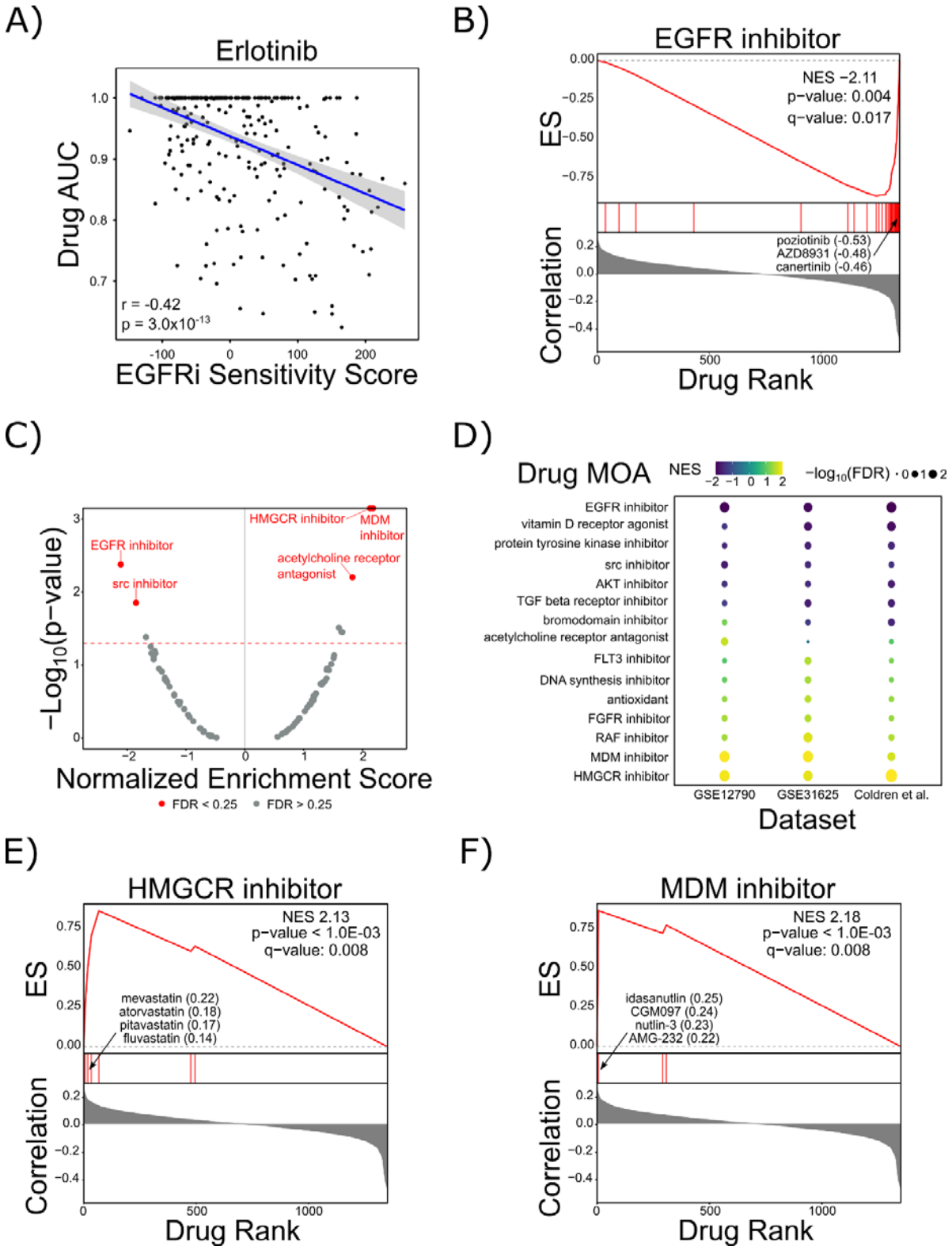



**Figure 4. DMEA correctly predicts sensitivity to EGFR inhibitors.** Using three gene expression signatures of EGFR inhibitor sensitivity (GSE12790 (Hoeflich *et al.*, 2009), GSE31625 (Balko *et al.*, 2006), and (Coldren *et al.*, 2006)), we calculated EGFRi sensitivity

scores. For each signature, the EGFRi sensitivity scores were correlated with drug sensitivity data (i.e., AUC) for 1,351 drugs. Drugs were then ranked by correlation coefficient, and DMEA was performed. A) The GSE12790 EGFRi sensitivity score was significantly correlated to erlotinib sensitivity. Smaller AUC values indicate increased sensitivity to the drug. B) Mountain plot showing that DMEA identified EGFR inhibitors as enriched in the ranked list of drugs using the GSE12790 EGFRi sensitivity score. The most negatively correlated EGFR inhibitors are highlighted along with their correlation coefficients. C) Volcano plot of DMEA NES versus -$\log_{10}$(p-value) for all 85 tested drug mechanisms of action revealed negative enrichment of Src inhibitors and positive enrichment of HMGCR inhibitors, MDM inhibitors, and acetylcholine receptor antagonists in the ranked list of drugs using the GSE12790 EGFRi sensitivity score. D) Dot plot showing consistency in DMEA results using the three different EGFRi sensitivity scores. The color and size of each circle represents the NES and -$\log_{10}$(FDR), respectively, of the indicated drug set. E,F) Mountain plots showing enrichment of HMGCR and MDM inhibitors in the ranked list of drugs using the GSE12790 EGFRi sensitivity score. The most positively correlated inhibitors are highlighted along with their correlation coefficients.

*DMEA correctly identifies acquired resistance to RAF inhibitors*

Many patients treated with molecularly targeted therapeutics suffer from acquired resistance. We thus sought to test DMEA using data from an independent transcriptomic signature of acquired drug resistance. Using RNAseq data from BRAF-mutant melanoma patients before treatment with the BRAF inhibitor vemurafenib and after acquired resistance (Kemper *et al.*, 2016), we defined a WGV signature of RAF inhibitor sensitivity. As expected, the RAFi WGV scores were significantly negatively correlated with vemurafenib AUC values (Fig. 5A). Next, we correlated RAFi WGV scores with drug sensitivity scores for the other 1,350 drugs in the PRISM data set, ranked drugs by the Pearson correlation, and used DMEA to identify enriched drug mechanisms of action (Supp. Table 2). Validating our approach, DMEA found that RAF inhibitors were significantly enriched in the rank list of drugs (Fig. 5B). Here, vemurafenib was the 35$^{th}$-most negatively correlated of 1,351 drugs, but the RAF inhibitor dabrafenib was the most negatively correlated drug. We next asked what other drug mechanisms of action were enriched in the RAFi drug rank lists. This analysis identified that inhibitors targeting MEK, MDM, nerve growth factor receptor, and HMGCR were negatively enriched in the rank list, supporting that tumors sensitive to RAF inhibitors would also be sensitive to these drugs (Fig. 3C). Conversely, inhibitors of HDAC, ERK, CHK, and SYK were positively enriched in the rank list,

suggesting that tumors resistant to RAF inhibitors will be sensitive to these drug mechanisms of action. For CHK, six of seven inhibitors in the PRISM data set were present in the top 40 most positively correlated drugs in the rank list (Fig. 5D). Altogether, these results support that DMEA can correctly identify acquired resistance to RAF inhibitors as well as predict other drug mechanisms for RAFi-resistant tumors.

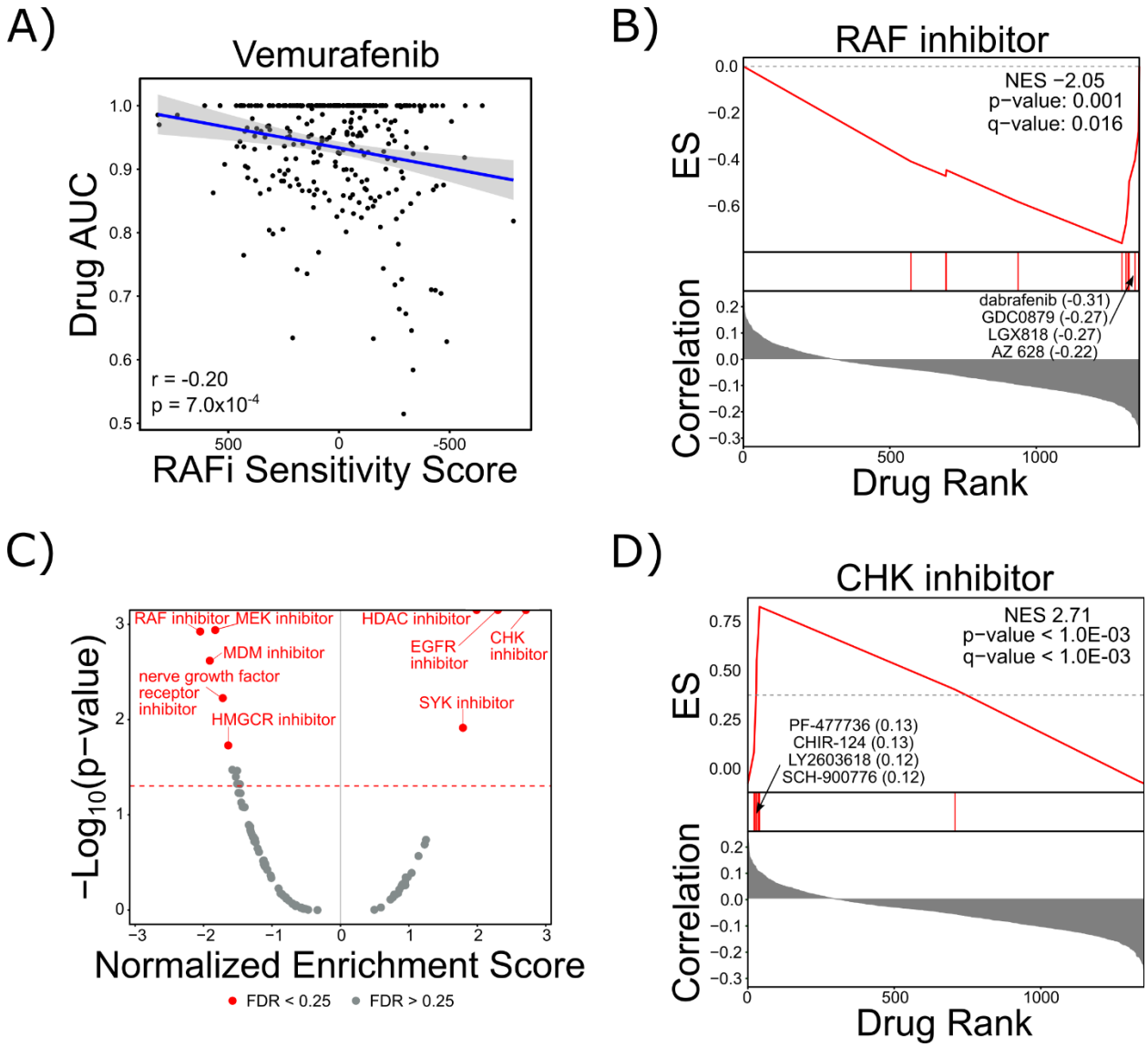



**Figure 5. DMEA correctly identifies acquired resistance to RAF inhibitors.** Using RNAseq data from paired biopsy samples of 3 BRAF-mutant melanoma patients before vemurafenib treatment and after emergence of vemurafenib resistance (Kemper *et al.*, 2016), we calculated a RAFi sensitivity score and then correlated the RAFi sensitivity score with drug sensitivity data

(i.e., AUC) for 1,351 drugs. Drugs were then ranked by correlation coefficient, and DMEA was performed. A) The RAFi sensitivity score was significantly correlated to vemurafenib sensitivity. Smaller AUC values indicate increased sensitivity to the drug. B) Mountain plot showing that DMEA identified RAF inhibitors as enriched in the ranked list of drugs. The most negatively correlated RAF inhibitors are highlighted along with their correlation coefficients. C) Volcano plot of DMEA NES versus -$\log_{10}$(p-value) for all 85 tested drug mechanisms of action revealed other drug mechanisms of action that were negatively or positively enriched in the ranked list of drugs. D) Mountain plot showing enrichment of CHK inhibitors in the rank list of drugs. The most positively correlated CHK inhibitors are highlighted along with their correlation coefficients.

*Predicting tissue-dependent drug sensitivity based on expression of the xCT glutamate/cystine antiporter*

Having validated that DMEA can identify sensitivity to HMGCR, EGFR, and RAF inhibitors, we next used the method to predict sensitive drug mechanisms for tumors with high and low expression of the xCT glutamate/cystine antiporter (*SLC7A11*). We chose xCT as our test case because it regulates cellular redox homeostasis, is associated with poor prognosis in several tumor types, and is an active drug discovery target (Lin *et al.*, 2020). We analyzed gene expression data from The Cancer Genome Atlas (TCGA) to generate xCT WGV scores for twelve tissue types by comparing tumors with high and low expression of xCT ($xCT^{hi}$ and $xCT^{lo}$, see Methods). The WGV scores for each tumor type were then correlated with PRISM drug sensitivity scores (AUC), drugs were ranked by their Pearson correlation estimates, and DMEA was used to identify enriched drug mechanism-of-action sets. Of 85 mechanisms of action tested, 44 were significantly enriched in at least one tissue type (FDR q-value < 0.25, Supp. Table 3). We then clustered NES values to compare the twelve tissue types (Fig. 6A). No drug mechanism was universally significant across all tumor types. For $xCT^{hi}$ tumors, EGFR inhibitors were significantly enriched in four tissue types (GBM, LGG, LUAD, and UCEC) with the strongest enrichment observed in GBM (Fig. 4B), suggesting that these $xCT^{hi}$ tumors would be susceptible to EGFR inhibitors. Conversely, for $xCT^{lo}$ tumors, bromodomain inhibitors were significantly enriched in COAD, KIRC, and SKCM, suggesting that $xCT^{lo}$ tumors of these types would be susceptible to bromodomain inhibitors. Taken together, this data demonstrates how

DMEA can be used to identify tissue-dependent drug sensitivity for tumors with high and low biomarker expression.

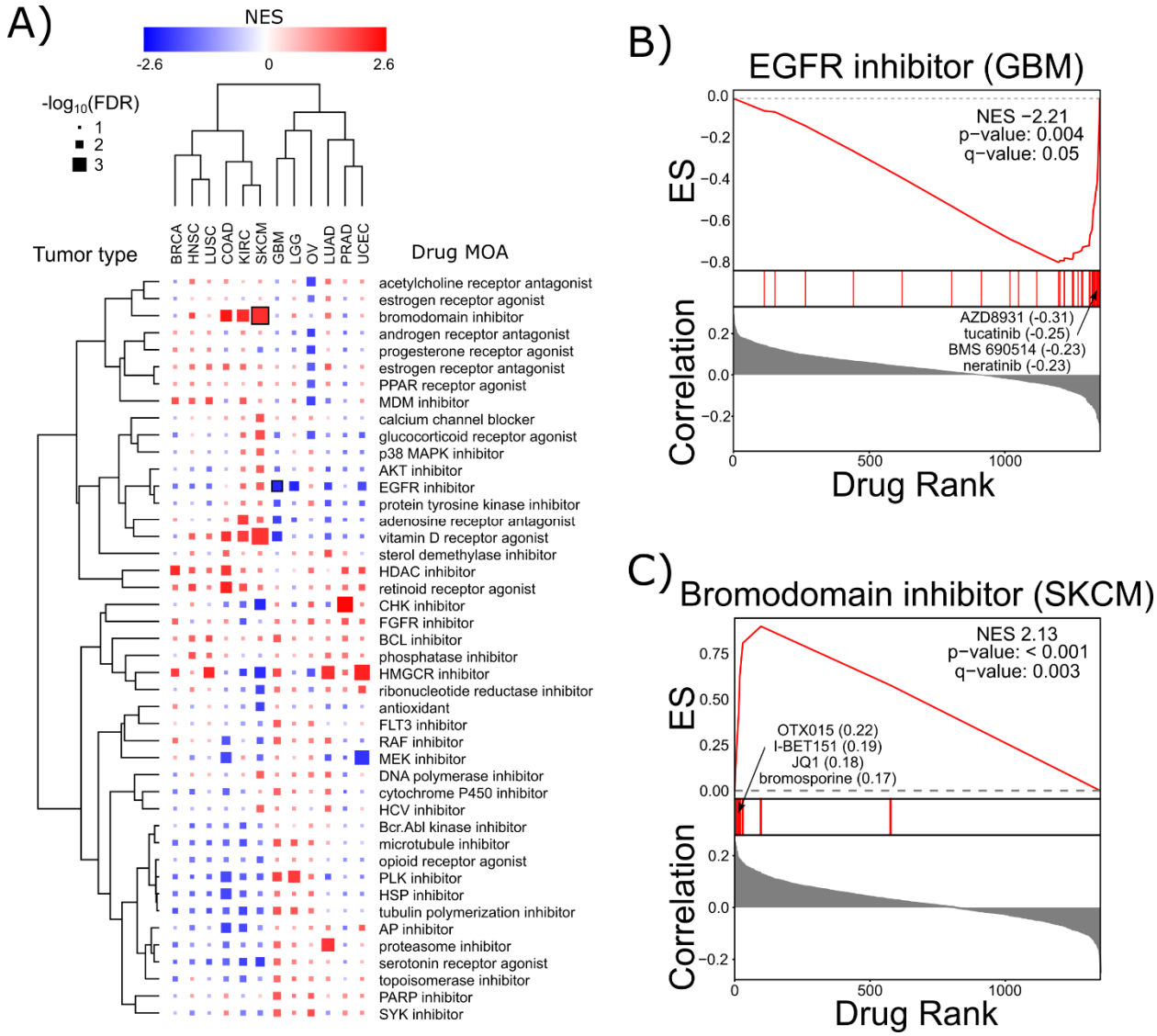



**Figure 6. DMEA predicts drug sensitivity for xCT$^{hi}$ and xCT$^{lo}$ tumors from twelve tissue types.** A) Hierarchical clustering of DMEA NES values across twelve tumor types based on tissue-specific xCT WGV signatures. COAD stands for COADREAD. Color indicates the direction and magnitude of enrichment (NES), and the size of the square indicates the -log$_{10}$(q-value). The examples highlighted in panels B and C are indicated by a black outline. B) Mountain plot showing enrichment of EGFR inhibitors in the rank list of drugs for xCT$^{hi}$ GBM. The most negatively correlated EGFR inhibitors are highlighted along with their correlation coefficients. C) Mountain plot showing enrichment of bromodomain inhibitors in the rank list of drugs for xCT$^{lo}$ SKCM. The most negatively correlated EGFR inhibitors are highlighted along with their correlation coefficients.

**DISCUSSION**

Molecularly targeted therapeutics have revolutionized cancer therapy. However, because as many as 85% of cancer patients are currently ineligible for targeted therapies (Haslam *et al.*, 2021), there remains an urgent need to match patients with effective drugs. Here, we introduce Drug Mechanism Enrichment Analysis (DMEA), a bioinformatic approach which uses an input transcriptional signature to predict drug mechanism(s) of action to which cancer cells may be sensitive or resistant. The main hypothesis underlying DMEA is that effective drugs with a common mechanism of action will be statistically enriched in a ranked list of predicted sensitivity scores. The power of DMEA thus lies in aggregating information from drugs that share a common mechanism of action, thereby increasing confidence in the algorithm's predictions. In addition, DMEA is a user-friendly method that translates molecular signatures (i.e., independent of the pharmacological screening databases) into drug sensitivity predictions. As such, DMEA provides a robust method to improve the selection of targeted therapeutics for cancer patients.

To validate DMEA, we have used both simulated data and published signatures of sensitivity to atorvastatin (HMGCR inhibitor), erlotinib and gefitinib (EGFR inhibitors), and vemurafenib (BRAF inhibitor). Notably, our HMCGRi signature was derived from the same databases used for drug sensitivity predictions (i.e., PRISM and CCLE, note that training and testing were conducted on separate cell lines), but our EGFRi and RAFi signatures were developed from independent publications (i.e., not CCLE or PRISM). In all tested cases, DMEA identified the tested drug mechanism as significantly enriched. In addition, DMEA identified drug mechanisms which may exhibit synergy with the tested drug mechanism against sensitive cells and/or efficacy against tumor cells resistant to the tested mechanism. One striking example is that cancer cells resistant to HMGCR inhibitors were predicted to be sensitive to EGFR inhibitors and vice versa (Figs. 3D and 4E). This finding is supported by work demonstrating that

HMGCR inhibitors can overcome resistance to EGFR inhibitors in NSCLC cells by inhibiting AKT (Hwang *et al.*, 2014; Chen *et al.*, 2013). Similarly, in $EGFR^+$ breast cancer cell lines resistant to EGFR inhibitors, inhibition of HMGCR sensitized cells to EGFR inhibitors by reducing the sequestration of EGFR in lipid rafts (Irwin *et al.*, 2011). In addition to the inverse relationship between EGFR and HMGCR inhibitors, DMEA also predicted that EGFR inhibitor-resistant cells would be sensitive to MDM inhibitors (Fig. 4F). Indeed, this bioinformatic result is supported by the observation that MDM2 mediates resistance to EGFR inhibitors in mouse models of NSCLC (Shen *et al.*, 2020). Furthermore, our analysis of BRAF inhibitor sensitivity suggested that tumors sensitive to BRAF inhibitors would also be sensitive to MEK inhibitors (Fig. 5C), an observation that is supported by clinical trials showing that combination treatment with BRAF and MEK inhibitors is more effective than inhibition of BRAF alone in BRAF-mutant melanoma patients (Long *et al.*, 2014). Finally, for cancers resistant to RAF inhibitors, DMEA identified CHK inhibitors as potentially beneficial, and in fact, CHK1 inhibition has shown efficacy in melanoma xenografts and cell lines that are resistant to RAF inhibitors (Hwang *et al.*, 2018). Taken together, these data support that DMEA identifies drug mechanisms beneficial for both combination treatment and drug resistant cells.

In addition to our validation experiments, we demonstrated that DMEA can identify tissue-dependent drug sensitivities in a pan-cancer analysis of $xCT^{hi}$ and $xCT^{lo}$ tumors (Fig. 6). In fact, two of the strongest findings from this analysis are supported by the xCT literature. First, DMEA suggested that $xCT^{hi}$ GBM, LGG, LUAD, and UCEC tumors would be sensitive to EGFR inhibitors, with the strongest enrichment observed in GBM. Indeed, in gliomas, EGFR interacts with and promotes the cell surface expression of xCT, thereby enhancing the antioxidant capacity of tumor cells (Tsuchihashi *et al.*, 2016). Second, DMEA suggested that $xCT^{lo}$ COAD, KIRC, and SKCM tumors will be sensitive to bromodomain inhibitors, with the strongest effect observed in melanoma. Interestingly, bromodomain inhibitors kill cancer cells through

ferroptosis (Sui *et al.*, 2019), a form of regulated cell death that is suppressed by xCT (Koppula *et al.*, 2021). Thus, $xCT^{lo}$ tumors may be more sensitive to bromodomain inhibitors because xCT is not present to suppress ferroptosis. Taken together, these results suggest that DMEA can serve as a discovery algorithm for inhibitors that will exhibit tissue-type and biomarker-specific toxicity.

In the studies presented here, we have used drug sensitivity data from PRISM (Corsello *et al.*, 2020). However, the DMEA approach should be applicable to other drug screening databases including the Cancer Therapeutics Response Portal (Seashore-Ludlow *et al.*, 2015; Basu *et al.*, 2013), the Genomics of Drug Sensitivity in Cancer database (Iorio *et al.*, 2016; Garnett *et al.*, 2012), and the NCI-60 cell line screening database (Shoemaker, 2006). In addition, DMEA is flexible in that it can accept any input ranked list of drugs provided that the drugs can be grouped by mechanism of action. Thus, DMEA may serve as an additional tool for evaluating the enrichment of drug mechanisms predicted to target certain cancer sub-types including dedifferentiated melanomas (Tsoi *et al.*, 2018) and therapy-resistant high-mesenchymal cell state (Viswanathan *et al.*, 2017).

One limitation of this study is that our findings are based on known drug mechanisms of action. However, drugs may have multiple targets, and in some cases, the putative mechanism of action for targeted therapeutics may be incorrect (Lin *et al.*, 2019). By aggregating information across multiple drugs, rather than considering each individual drug independently, DMEA may be less sensitive to misannotation of drug mechanisms. However, improved annotation of drug mechanisms, potentially through newer approaches including metabolomics (Holbrook-Smith *et al.*, 2022) and proteomics (Savitski *et al.*, 2014), will improve the power of DMEA.

In summary, Drug Mechanism Enrichment Analysis (DMEA) is a bioinformatic tool which can identify drug mechanisms with differential sensitivity for cancer cells characterized by a molecular signature. Our approach leverages the dramatic growth in public databases of transcriptomic profiles (e.g., CCLE) and drug sensitivity scores (e.g., PRISM) for cancer cell lines. The DMEA framework can be applied to many biological contexts including drugs targeting specific mutations, cancer sub-types, biomarker expression, acquired resistance, and others. In as little as 5 minutes when using up to 500 gene weights, DMEA can output a comprehensive analysis of drug mechanisms (including plots) for a given molecular signature of interest. DMEA is freely available to download as an R package at: https://github.com/BelindaBGarana/DMEA.

**FUNDING**

This work was supported by the Viterbi School of Engineering. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1842487. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

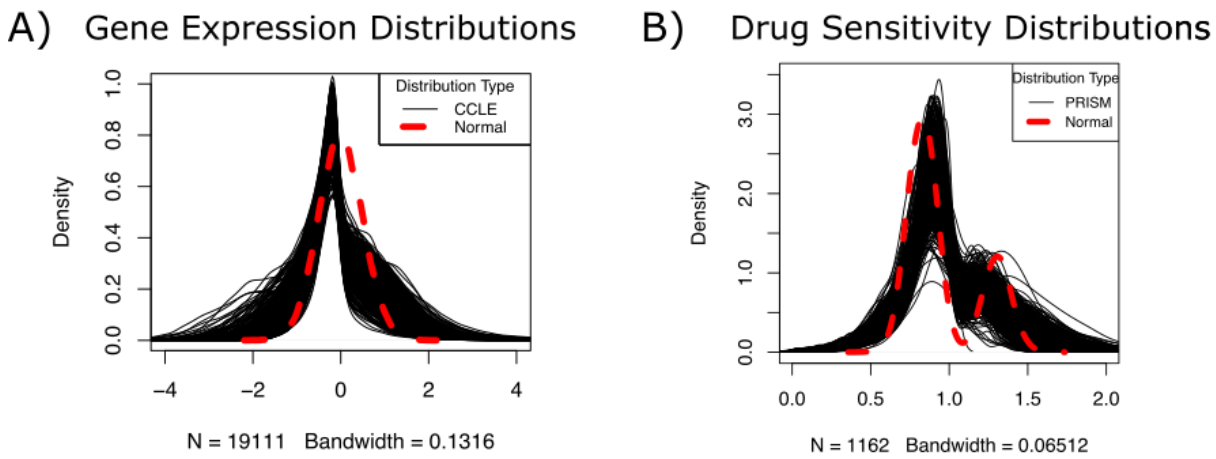
## SUPPLEMENTARY TABLES

**Supp. Table 1. Drug Mechanisms of Action.**

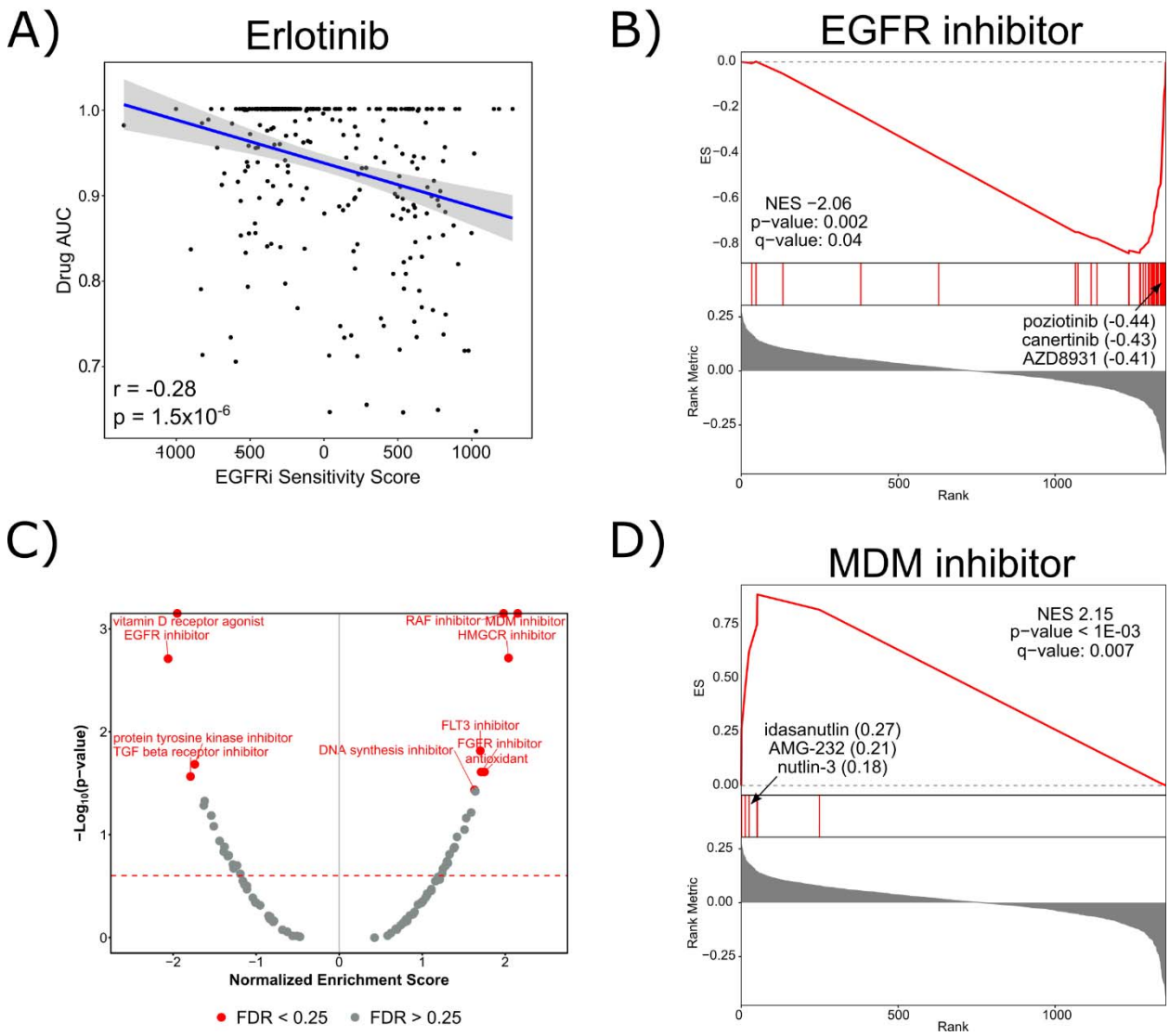
**Supp. Table 2. Validation of HMGCR, EGFR, and RAF sensitivities using DMEA.**

**Supp. Table 3. Drug sensitivity analysis of $xCT^{hi}$ and $xCT^{lo}$ tumors using DMEA.**

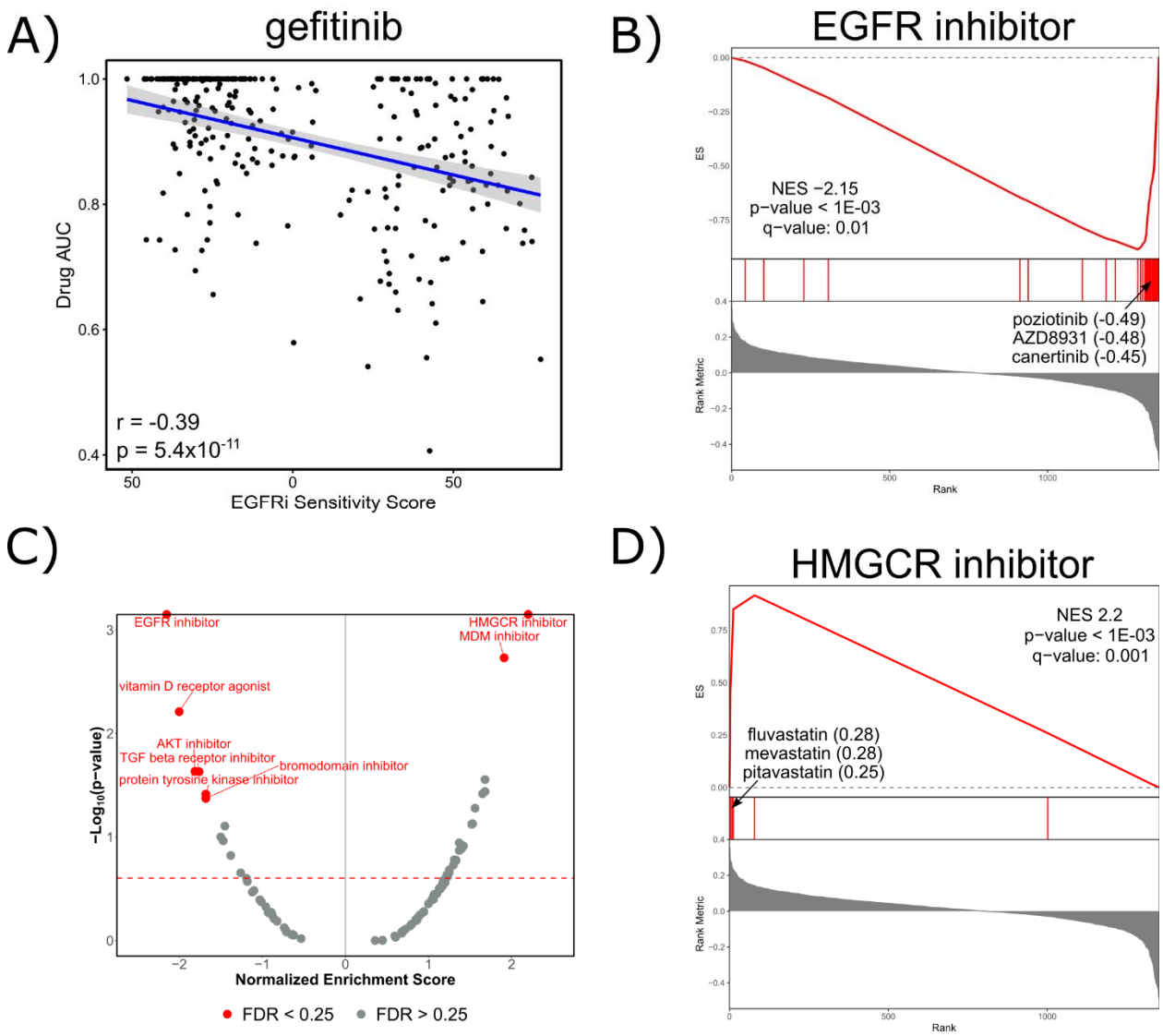
**SUPPLEMENTARY FIGURES**




**Supp. Figure 1. Gene expression and drug sensitivity distributions used for sensitivity analysis.** To test the DMEA algorithm, we first performed a simulation study (see Methods, Fig. 1). We generated gene expression and drug sensitivity scores for 200 cell lines by randomly selecting values from distributions that reflected the CCLE RNAseq and PRISM drug sensitivity data as shown in A) CCLE RNAseq data from 614 adherent cancer cell lines and B) PRISM drug sensitivity data (AUC values) from 482 cancer cell lines.




**Supp. Figure 2. DMEA correctly predicts sensitivity to EGFR inhibitors using GSE31625.** Using a gene expression signature of erlotinib sensitivity (GSE31625 (Balko *et al.*, 2006)), we calculated EGFRi sensitivity scores. The EGFRi sensitivity scores were correlated with drug sensitivity data (i.e., AUC) for 1,351 drugs. Drugs were then ranked by correlation coefficient, and DMEA was performed. A) The GSE31625 EGFRi sensitivity score was significantly correlated to erlotinib sensitivity. Smaller AUC values indicate increased sensitivity to the drug. B) Mountain plot showing that DMEA identified EGFR inhibitors as enriched in the ranked list of drugs using the GSE31625 EGFRi sensitivity score. The most negatively correlated EGFR inhibitors are highlighted along with their correlation coefficients. C) Volcano plot of DMEA NES versus -$\log_{10}$(p-value) for all 85 tested drug mechanisms of action revealed negative enrichment of vitamin D receptor agonists and positive enrichment of RAF, MDM, and HMGCR inhibitors in the ranked list of drugs using the GSE31625 EGFRi sensitivity score. D) Mountain plot showing enrichment of MDM inhibitors in the ranked list of drugs using the GSE31625 EGFRi sensitivity score. The most positively correlated inhibitors are highlighted along with their correlation coefficients.




**Supp. Figure 3. DMEA correctly predicts sensitivity to EGFR inhibitors using GSE12790.** Using a gene expression signature of gefitinib sensitivity (GSE12790 (Hoeflich *et al.*, 2009), we calculated EGFRi sensitivity scores. The EGFRi sensitivity scores were correlated with drug sensitivity data (i.e., AUC) for 1,351 drugs. Drugs were then ranked by correlation coefficient, and DMEA was performed. A) The GSE12790 EGFRi sensitivity score was significantly correlated to gefitinib sensitivity. Smaller AUC values indicate increased sensitivity to the drug. B) Mountain plot showing that DMEA identified EGFR inhibitors as enriched in the ranked list of drugs using the GSE12790 EGFRi sensitivity score. The most negatively correlated EGFR inhibitors are highlighted along with their correlation coefficients. C) Volcano plot of DMEA NES versus -$\log_{10}$(p-value) for all 85 tested drug mechanisms of action revealed negative enrichment of vitamin D receptor agonists and positive enrichment of HMGCR inhibitors and MDM inhibitors in the ranked list of drugs using the GSE12790 EGFRi sensitivity score. D) Mountain plot showing enrichment of HMGCR inhibitors in the ranked list of drugs using the GSE31625 EGFRi sensitivity score. The most positively correlated inhibitors are highlighted along with their correlation coefficients.

## REFERENCES


Balko,J.M. *et al.* (2006) Gene expression patterns that predict sensitivity to epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer cell lines and human lung tumors. *BMC Genomics*, **7**, 289.

Barretina,J. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.

Basu,A. *et al.* (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, **154**, 1151–1161.

Chen,J. *et al.* (2013) Atorvastatin overcomes gefitinib resistance in KRAS mutant human non-small cell lung carcinoma cells. *Cell Death Dis*, **4**, e814–e814.

Choi,J. *et al.* (2020) RefDNN: a reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci Rep*, **10**, 1861.

Coldren,C.D. *et al.* (2006) Baseline Gene Expression Predicts Sensitivity to Gefitinib in Non–Small Cell Lung Cancer Cell Lines. *Mol Cancer Res*, **4**, 521–528.

Corcoran,R.B. *et al.* (2018) Combined BRAF, EGFR, and MEK Inhibition in Patients with BRAFV600E-Mutant Colorectal Cancer. *Cancer Discov*, **8**, 428–443.

Corsello,S.M. *et al.* (2020) Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer*, **1**, 235–248.

Garnett,M.J. *et al.* (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, **483**, 570–575.

Ghandi,M. *et al.* (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, **569**, 503–508.

Golub,T.R. *et al.* (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*, **286**, 531–537.

Gonçalves,E. *et al.* (2020) Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens. *Molecular Systems Biology*, **16**, e9405.

Haslam,A. *et al.* (2021) Updated estimates of eligibility for and response to genome-targeted oncology drugs among US cancer patients, 2006-2020. *Ann Oncol*.

Hoeflich,K.P. *et al.* (2009) In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. *Clin Cancer Res*, **15**, 4649–4664.

Holbrook-Smith,D. *et al.* (2022) High-throughput metabolomics predicts drug-target relationships for eukaryotic proteins. *Mol Syst Biol*, **18**, e10767.

Hwang,B.-J. *et al.* (2018) Chk1 inhibition as a novel therapeutic strategy in melanoma. *Oncotarget*, **9**, 30450–30464.

Hwang,K.-E. *et al.* (2014) Effect of simvastatin on the resistance to EGFR tyrosine kinase inhibitors in a non-small cell lung cancer with the T790M mutation of EGFR. *Experimental Cell Research*, **323**, 288–296.

Iorio,F. *et al.* (2016) A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, **166**, 740–754.

Irwin,M.E. *et al.* (2011) Lipid raft localization of EGFR alters the response of cancer cells to the EGFR tyrosine kinase inhibitor gefitinib. *J. Cell. Physiol.*, **226**, 2316–2328.

Kemper,K. *et al.* (2016) BRAF(V600E) Kinase Domain Duplication Identified in Therapy-Refractory Melanoma Patient-Derived Xenografts. *Cell Rep*, **16**, 263–277.

Koppula,P. *et al.* (2021) Cystine transporter SLC7A11/xCT in cancer: ferroptosis, nutrient dependency, and cancer therapy. *Protein Cell*, **12**, 599–620.

Lin,A. *et al.* (2019) Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Science Translational Medicine*, **11**.

Lin,W. *et al.* (2020) SLC7A11/xCT in cancer: biological functions and therapeutic implications. *Am J Cancer Res*, **10**, 3106–3126.
Lippi,G. *et al.* (2019) Statins popularity: A global picture. *Br J Clin Pharmacol*, **85**, 1614–1615.
Long,G.V. *et al.* (2014) Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med*, **371**, 1877–1888.
Nguyen,L. *et al.* (2017) Systematic assessment of multi-gene predictors of pan-cancer cell line sensitivity to drugs exploiting gene expression data. *F1000Res*, **5**, ISCB Comm J-2927.
Parca,L. *et al.* (2019) Modeling cancer drug response through drug-specific informative genes. *Sci Rep*, **9**, 15222.
Rees,M.G. *et al.* (2016) Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*, **12**, 109–116.
Reich,M. *et al.* (2006) GenePattern 2.0. *Nat Genet*, **38**, 500–501.
Ritchie,M.E. *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**, e47–e47.
Savitski,M.M. *et al.* (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science*.
Seashore-Ludlow,B. *et al.* (2015) Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov*, **5**, 1210–1223.
Shen,H. *et al.* (2020) S6K1 blockade overcomes acquired resistance to EGFR-TKIs in non-small cell lung cancer. *Oncogene*, **39**, 7181–7195.
Shoemaker,R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer*, **6**, 813–823.
Sosman,J.A. *et al.* (2012) Survival in BRAF V600–Mutant Advanced Melanoma Treated with Vemurafenib. *New England Journal of Medicine*, **366**, 707–714.
Stancu,C. and Sima,A. (2001) Statins: mechanism of action and effects. *J Cell Mol Med*, **5**, 378–387.
Subramanian,A. *et al.* (2017) A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, **171**, 1437-1452.e17.
Subramanian,A. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, **102**, 15545–15550.
Sui,S. *et al.* (2019) Ferritinophagy is required for the induction of ferroptosis by the bromodomain protein BRD4 inhibitor (+)-JQ1 in cancer cells. *Cell Death Dis*, **10**, 1–17.
Suphavilai,C. *et al.* (2018) Predicting Cancer Drug Response using a Recommender System. *Bioinformatics*, **34**, 3907–3914.
Tsoi,J. *et al.* (2018) Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. *Cancer Cell*, **33**, 890-904.e5.
Tsuchihashi,K. *et al.* (2016) The EGF Receptor Promotes the Malignant Potential of Glioma by Regulating Amino Acid Transport System xc(-). *Cancer Res*, **76**, 2954–2963.
Viswanathan,V.S. *et al.* (2017) Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature*, **547**, 453–457.
Wei,D. *et al.* (2019) Comprehensive anticancer drug response prediction based on a simple cell line-drug complex network model. *BMC Bioinformatics*, **20**, 44.
Yuan,H. *et al.* (2016) Multitask learning improves prediction of cancer drug sensitivity. *Sci Rep*, **6**, 31619.