Gene expression has more power for predicting *in vitro* cancer cell vulnerabilities than genomics

Authors

Joshua M. Dempster[1], John Krill-Burger[1], Allison Warren[1], James M. McFarland[1], Todd R. Golub[1,2], and Aviad Tsherniak[1]

1. Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA
2. Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215, USA

Abstract

Achieving precision oncology requires accurate identification of targetable cancer vulnerabilities in patients. Generally, genomic features are regarded as the state-of-the-art method for stratifying patients for targeted therapies. In this work, we conduct the first rigorous comparison of DNA- and expression-based predictive models for viability across five datasets encompassing chemical and genetic perturbations. We find that expression consistently outperforms DNA for predicting vulnerabilities, including many currently stratified by canonical DNA markers. Contrary to their perception in the literature, the most successful expression-based models depend on few features and are amenable to biological interpretation. This work points to the importance of exploring more comprehensive expression profiling in clinical settings.

Introduction

Decades of work have confirmed cancer as a disease of the genome[1]. With this recognition has come the hope that the specific oncogenic mutations of cancer will confer equally specific vulnerabilities to therapeutic intervention. In the last decade, these vulnerabilities have been exhibited dramatically with compounds targeting known oncogenes such as B-RAF, EGFR, and ALK[2]. Driven by these early successes, the precision oncology paradigm has focused on sequencing patient tumors with the goal of determining the appropriate therapy[3]. However,

genomically-indicated therapies have yet to prove transformative in cancer care. Small proportions of patients have genomic indicators for a targeted therapy, and most that do show limited or no response[4].

Disappointing results with genome-based biomarkers have driven calls to look beyond the cancer genome to other possible indicators of cancer-specific vulnerabilities[5]. A major alternative for tumor characterization is the transcriptome. Although RNA is more difficult to maintain in the clinic than DNA, studies have found that gene expression supplies the most significant predictive features for patient prognosis[6]. Previous work on predicting cell viability after compound treatment suggests that expression may be more powerful than DNA features for predicting drug response[7], but expression-based models are widely treated as undesirable either because they are considered trivial proxies for tissue-type[8] or because they are seen as uninterpretable[9].

These suggestive studies point to the need for a comprehensive comparison of expression and genomic molecular features as predictors of cancer vulnerability and a deeper interrogation of the interpretability of expression models. Here we present the first such study across five large datasets of cancer cell viability including both genetic and chemical perturbations. We find that RNA-Seq expression outperforms DNA-derived features in predicting cell viability response in nearly all cases, including many perturbations with known genomic biomarkers. The best results are typically driven by a small number of interpretable expression features. Our findings suggest that both existing and new cancer targets are frequently better identified using RNA-seq gene expression than any combination of other cancer cell properties.

Results

Data and Methodology

We attempted to predict cell line viability in response to perturbations measured in five large *in vitro* cell viability studies, including CRISPR gene knockout (Cancer Dependency Map and Project Score[10]), arrayed oncology drug treatment experiments (GDSC[8]), pooled drug treatment experiments (PRISM Repurposing[11]), and RNAi knockdown studies (DEMETER2[12]). **(Supplementary Fig. 1a-b**). Many perturbations do not produce differential viability effects

across cell lines, indicating that no predictive model is likely to succeed. In order to assess performance specifically for cases where strong differential viability signal exists, we identified a subgroup of 100 perturbations in each dataset that appear to have the greatest biological signal based on their viability distributions across samples. We labeled these Strongly Selective Dependencies (SSDs, **Supplementary Data**).

Cell characterization was drawn from the Cancer Dependency Map and encompassed 181,951 total features including mutation calls from whole-exome sequencing, RNA-Seq expression (protein-coding genes) at the single-gene and gene-set level, reduced representation bisulfite sequencing (RBBS) methylation, gene fusions, and tissue annotations (**Supplementary Fig. 1c**). We refer to the union of all these measured properties as cell features. After filtering for cell lines with characterization for mutation and RNA-seq expression data, the perturbation datasets spanned a set of 1,432 unique cell lines with 241-649 cell lines per dataset.

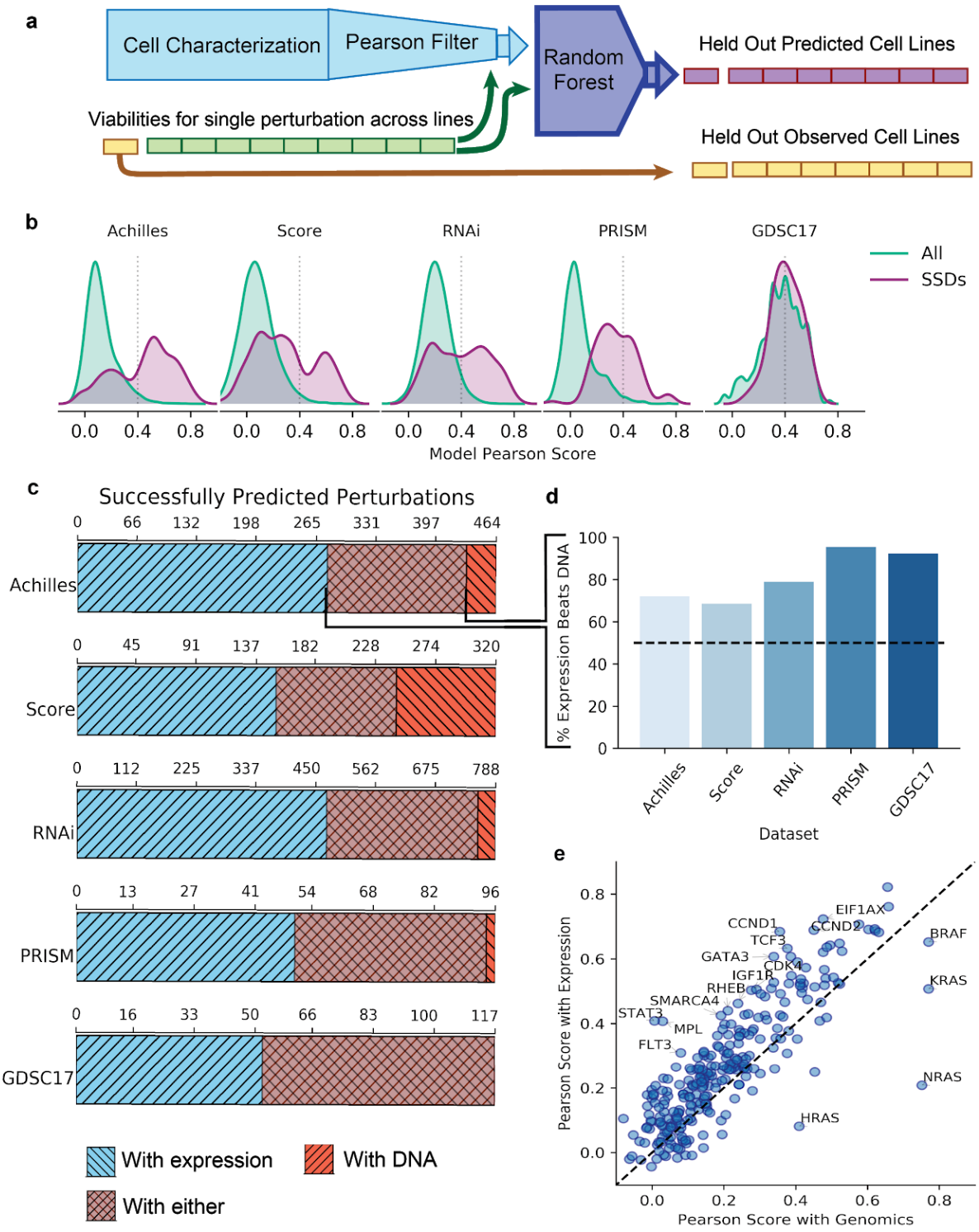

Fig. 1: Prediction performance with DNA and expression features. **a.** Outline of the predictive algorithm used. Models are scored by the Pearson correlation of the concatenated out-of-sample predictions with observed values. **b.** Distribution of model predictive performance using all features for all perturbations (green) or SSDs (maroon). Dotted lines indicate the threshold to call a result

successful (0.4). **c.** The number of perturbations successfully predicted only with expression features (blue), only with DNA features (red), or with either feature set (maroon). **d**. For perturbations with successful predictors with both the expression and DNA feature sets (overlap in **c**), how often the most successful prediction is made with expression. **e**. Predictive performance for CRISPR knockout of 278 OncoKB-labeled oncogenes in Achilles with either the DNA or expression feature set.

To evaluate the power of different cell properties for identifying the vulnerabilities of cancer cells, we trained a regression model individually for each perturbation. Our pipeline consisted of two components: a Pearson correlation filter to reduce the feature list to the top 1,000 features showing the strongest linear relationship to the viabilities, and a random forest model which produced the predictions using the selected 1,000 features (**Fig. 1a**). We found that not including a feature selection filter led to a slight drop in predictive performance due to the large number of irrelevant features included, along with a large increase in computation time. Model predictions were scored according to the Pearson correlation of the concatenated out-of-sample predictions calculated during cross-validation and the original viability scores, which we call the predictor's Pearson score. A comparison of this random forest predictor with another commonly used machine learning model (elastic net) found that they produced similar performance on predicting SSD perturbations, but with a consistent edge for the random forest model (88.75% (355) of 500 SSDs better predicted by random forest; **Supplementary Fig. 2**).

With a large number of samples, it is possible for a predictor to perform significantly better than chance while still performing too poorly to be useful in practice. Accordingly, we chose a Pearson score value (0.4) rather than statistical significance as the threshold of interest for examining successful predictors. The maximum Benjamini-Hochberg false discovery rate (FDR) for predictions reaching this threshold ranged from 0.0186 in the RNAi dataset to 2.98×10^{-25} in the Achilles CRISPR dataset.

RNA-Seq Expression Features Outperform DNA-based Features

To establish baseline performance, we first examined predictive performance across all perturbations, using both DNA-based and RNA-based features in addition to annotations of primary tissue. Predictive performance varies strongly with the chosen dataset (**Fig. 1b**). Models predicting GDSC17 drug sensitivity data, which is focused on anti-cancer compounds, had the best overall performance (median score 0.387; 46.3% of perturbations predicted with a score

greater than 0.4). The PRISM Repurposing dataset is comprised mostly of compounds with no known activity against cancer (3,583 out of 4,685 annotated compounds) and had the lowest overall scores (median 0.048 and 2.2% above 0.4). As expected, prediction performance is improved and has less extreme variation across datasets among SSDs: in the datasets, 30% (Project Score) to 65% (Achilles) of SSDs were predicted with Pearson scores above 0.4, with median Pearson scores ranging from 0.265 (Project Score) to 0.490 (Achilles).

We then directly compared the performance of models supplied only expression features, which included single gene RNA-Seq expression and single-sample gene set enrichment analysis (GSEA) of MSigDB[13] gene sets against models supplied only DNA-based features: mutation, methylation, copy number, and gene fusions (**Supplementary Fig. 1c**). Both feature sets included tissue annotations. There were more perturbations where gene expression was critical for predicting viability (1001 of 1785) compared with DNA-based features (144 of 1785) (**Fig 1c**). The remaining 745 perturbations could be predicted accurately with either feature set. Among these 745 perturbations, the expression predictor was most often the best, especially for drug data (**Fig 1d**).

The selective killing profile of some perturbations is likely related to tissue-of-origin-specific rather than cancer-specific dependency. Accordingly, we compared performance in the Achilles CRISPR dataset specifically for the 278 genes listed by OncoKB as “likely oncogenic” or “oncogenic”, and which had at least one non-expression-based indicator[14]. Even for the majority of these genes, expression features outperformed genomics for predicting viability after CRISPR knockout (**Fig. 1e**). The notable exceptions were BRAF, KRAS, NRAS (best predicted with their own mutation statuses), and SHOC2 (best predicted by NRAS mutation status). A similar pattern holds in the Project Score CRISPR and RNAi gene dependency datasets (**Supplementary Fig. 3**).

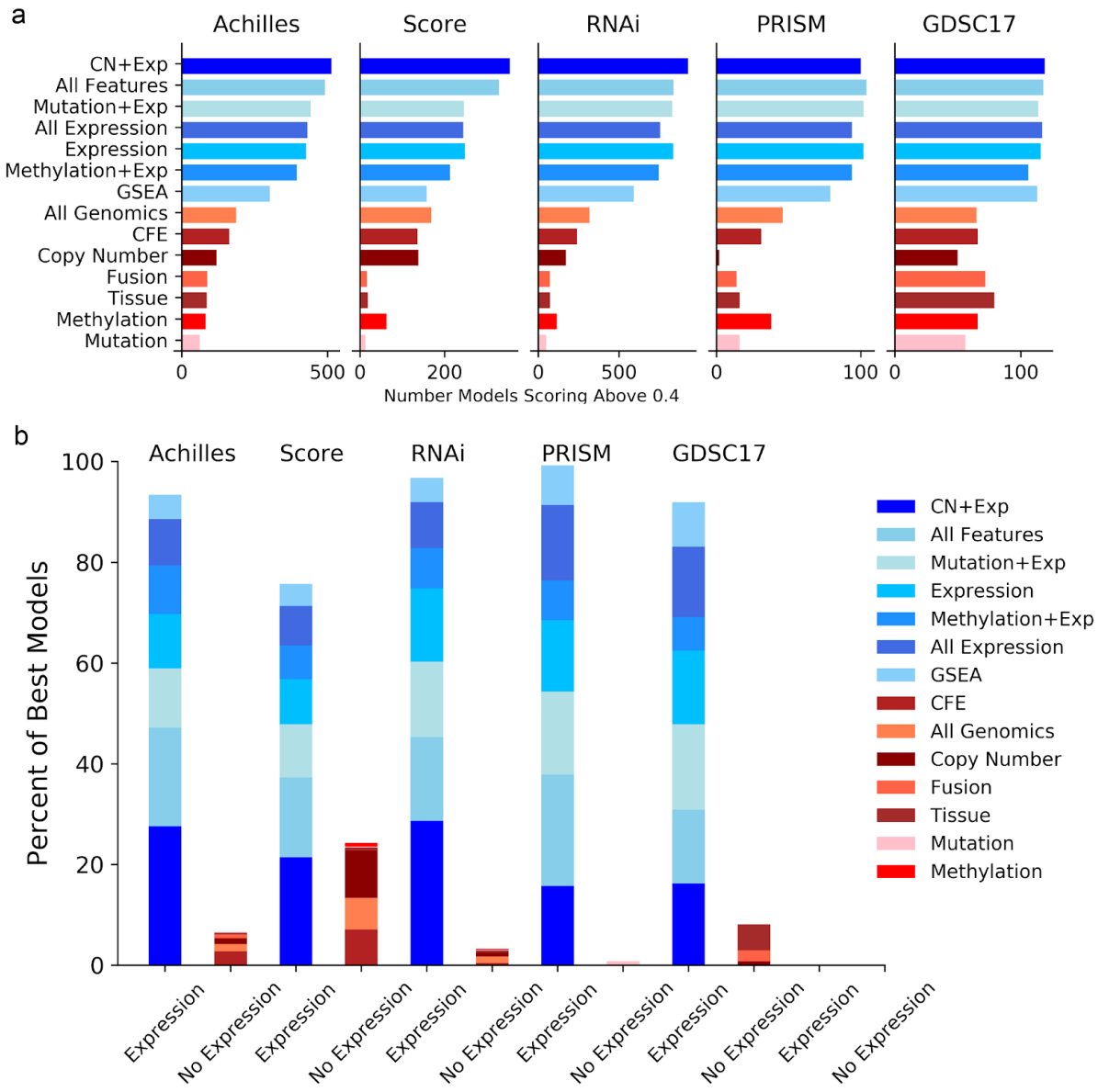

Fig. 2: Performance of individual feature combinations. **a.** Number of successful predictive models with different feature combinations. CFE stands for Cancer Functional Events, a curated set of DNA features. **b.** For perturbations with at least one successful predictor, how often each feature combination produces the best predictor for that perturbation.

Given that there is likely substantial redundancy between these DNA and RNA-based feature sets, we next asked what the value of adding one feature set to the other for improving predictions. To address this question, we first took the twenty CRISPR perturbations best predicted with DNA features (**Supplementary Fig. 4ab**) or expression (**Supplementary Fig 4g-l**) and evaluated the change in performance when both feature sets were used for prediction.

Adding expression improved the strongest DNA-based model scores by a median of 8.1% across datasets. In contrast, adding DNA features to the top expression models usually degraded performance by small amounts (due to the models overfitting slightly to noise in the datasets). Only RNAi showed positive median change in score when adding DNA features to expression (by 0.43%).

In order to further isolate the contributions of each feature type to the predictions, we next built models using each type of DNA feature individually or in combination with gene expression. All feature combinations included tissue annotations to ensure every model could identify tissue of origin. Across all data sets, any feature combination that included single-gene expression produced equal or more successful models than any feature set that did not (**Fig. 2a**). For a given perturbation with at least one successful model using any feature set, the best model included expression in 76.3% (Project Score) to 99.2% (PRISM) of cases (**Fig. 2b**). These results make it clear that expression features are the key elements for successfully predicting viability responses to most perturbations, and are beneficial even when DNA features suffice for successful prediction.

Specificity and Interpretability of Features

Models which utilize ‘diffuse’ sets of hundreds or thousands of features may be difficult to understand and act on. In contrast, models that use only a handful of features (sparse predictors) are generally easier to interpret, less vulnerable to overfitting, and require less intensive tumor characterization. We found that the most successful models trained with all features generally utilized a small number of features as indicated by the high importance they assigned to the top two features vs the rest (see Methods; **Fig. 3a**). To confirm that these apparently sparse models do not require many features to perform well, we examined how model performance varied with the number of features provided. For each training fold, we trained a model with the 1,000 features selected by Pearson correlation, ranked them by importance, and then trained using the top one, two, five, ten, or hundred features (**Supplementary Fig. 5**). For successful models where a single key feature held at least 10% of total feature importance, 62.0% of models using only the top one feature achieved Pearson scores at least 80% of the score using all features (**Fig. 3b**). Among successful models without a single key feature, only 14.0% of perturbations could be similarly well-predicted with their top

feature (**Fig. 3c**). We confirmed that these results were not a product of either the feature filtering strategy or the choice of random forest model by repeating this analysis with elastic net (EN) models for SSDs. We found that for perturbations where the random forest had identified a key feature, EN models using only the top feature achieved a median of 89.9% of their performance with all features (**Supplementary Fig. 6**). This shows that the most predictable cancer vulnerabilities often require only one feature for successful prediction.

Although predictive models using gene expression features are perceived as more diffuse and therefore less tractable than genomic predictors[9], we found did not find this to be the case for two important groups of perturbations: SSDs and successfully predicted perturbations. Sizeable fractions of both were dominated by a key single gene expression feature even when trained with all features (**Fig. 3de**). For the genetic dependency datasets only, we investigated how often key expression features were either the target gene's own expression, that of a paralog, or of a pathway or complex co-member (**Supplementary Table 1**). By these measures, key single gene expression features were found to have a known relationship to the target in as much as half of gene dependencies (50.1% for Achilles, 29.3% for Score, and 40.7% for RNAi).

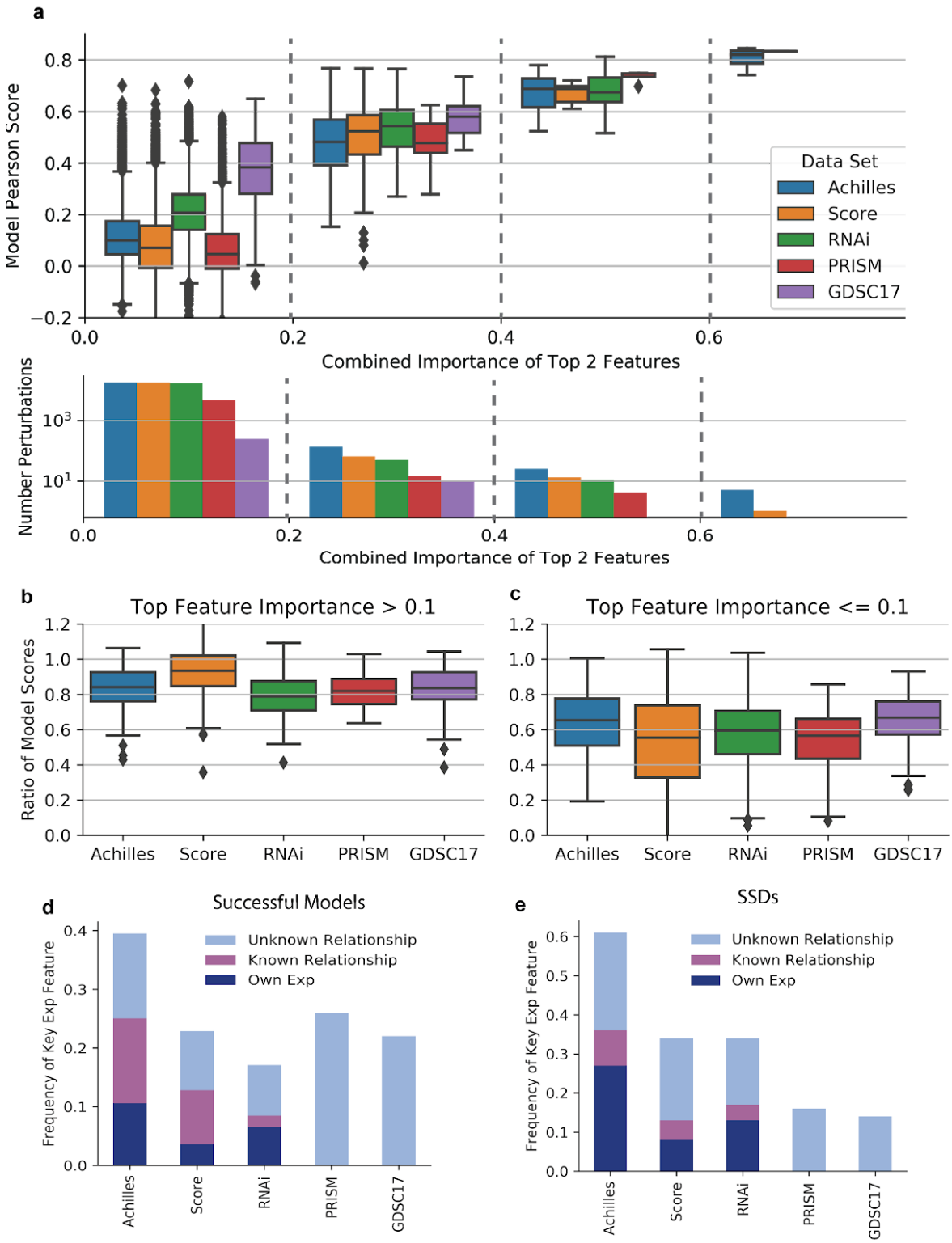

Fig. 3: Sparsity of models. **a.** For models using all available features, the relationship of the combined importance of the top 2 features as a percentage of total feature importance with model performance. Perturbations are binned by feature importance to the intervals [0, 0.02), [0.2, 0.4), [0.4, 0.6), and [0.6, 1]. **b.** The ratio of predictive model scores using the 1,000 features most correlated with the

perturbation from among all cell features, or using only one feature selected from the top 1,000 using random forest feature importance. Only perturbations with successful predictors with a single "key" feature with more than 10% of feature importance are included. **c.** Similar, but for predictors with no such key feature. **d**. For perturbations successfully predicted with all available features, how often the top feature is a single gene's expression with at least 10% of total feature importance. For genetic perturbations (but not compound perturbations), subcategories indicate the fraction of those known to be related to the target or the target itself. **e.** Similar results for SSD perturbations regardless of predictive performance.

Interpreting models

To demonstrate how sparse models can be made interpretable, we trained individual decision trees using only features with importance greater 0.08. In cases where only the top feature was used by the tree, we illustrated the relationship of the top feature with the target. Some examples are discussed below.

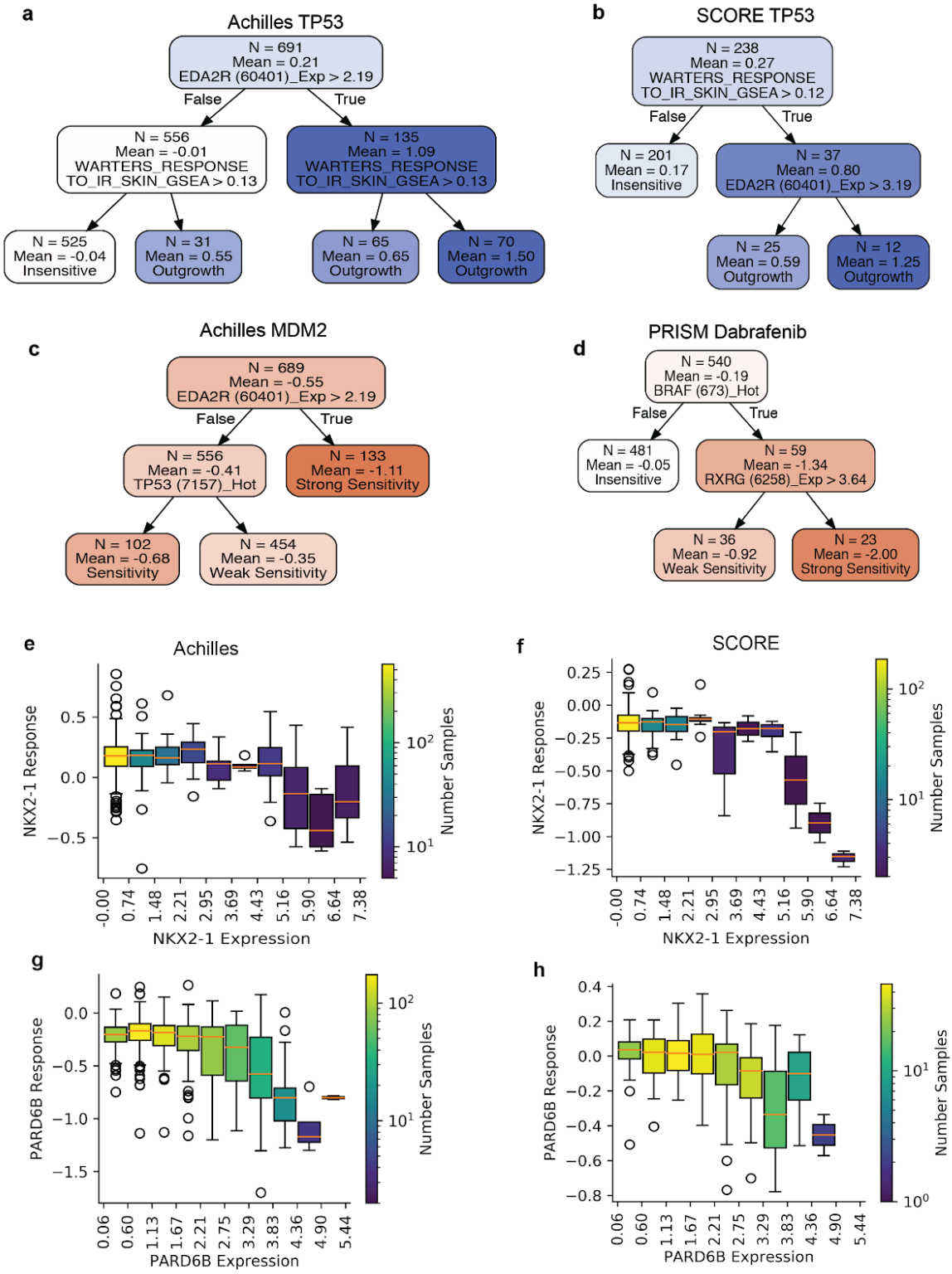

Fig. 4: Interpretable models. **a-c**. Decision trees for Achilles and Score viability post TP53 knockout and Achilles MDM2 knockout. **d**. Decision tree for PRISM dabrafenib response. **e-f**. Relationship between NXK2-1 effect on viability and its own expression in Achilles and Score. **g-h**. Similar for PARD6B.

The TP53 Pathway

TP53 mutation status is one of the key molecular characteristics of cancer across lineage types. Surprisingly, for both Achilles and Score data, the top two markers of viability response to TP53 KO were not TP53 mutation status but the expression of EDA2R (a p53 transcriptional target), along with a gene set defined by the transcriptional response of skin cells to ionizing radiation[15] **(Fig 3ab**). For RNAi-based suppression of TP53 the top predictor was a similar gene set defined by the response of primary blood mononuclear cells[16]. Both ionizing radiation gene sets include EDA2R. These results indicate that TP53 functional status is more cleanly inferred by these models from the transcription of its targets than directly from existing TP53 mutation calls. This may be due to failures in mutation calling or the complexity of possible causes of TP53 loss of function[17].

In cases where TP53 is functionally intact, cancers often depend on MDM2 to continue proliferation[18]. EDA2R expression was the top predictor of MDM2 dependency in Achilles (**Fig. 3c**) and was a top association in Project Score and RNAi data(**Supplementary Table 1**). There was a clear gradation of response to either TP53 or MDM2 knockout according to the magnitude of EDA2R expression observed in all of these datasets (**Supplementary Fig. 5**). Response to the MDM2 inhibitor nutlin-3 also exhibited negative correlation with EDA2R expression in GDSC17. The more potent MDM2 inhibitor idasanutlin was present only in the PRISM dataset, where it also exhibited a strong relationship with EDA2R expression (**Supplementary Table 1**). Given the significance of TP53 status in cancers, incorporating the expression of EDA2R and other TP53 targets for stratifying patients has considerable potential for clinical benefit.

BRAF Inhibitor Resistance

Many canonical oncogenic gain-of-function mutations occur in the RAS-RAF pathway. We showed earlier that mutations in members of this pathway (BRAF, NRAS, HRAS, and KRAS) stood out in providing a signal of dependency that cannot be recovered from gene expression. However, it does not follow that expression has no value for predicting response to compounds targeting members of this pathway. For example, an interpretable model of the BRAF inhibitor dabrafenib in PRISM showed it is first stratified by BRAF mutation status, but the strength of

response in BRAF-mutant lines is further stratified by RXRG expression (**Fig 3f**). The association of dabrafenib response with RXRG expression in BRAF-mutant lines was strong in both PRISM and GDSC17 (**Supplementary Table 1**). A similar result was obtained in PRISM with the BRAF inhibitor vemurafenib (**Supplementary Table 1**).

Untargeted Expression-Driven Dependencies

In addition to stratifying patients for existing therapies, gene expression features could be critical for identifying sensitive populations for many new selective dependencies. Among these are thyroid nuclear factor 1 (NKX2-1), identified in Project Score data as a strong dependency in the minority of non-small-cell lung cancer lines that overexpress it **(Supplementary Table 1**). Both dependent and non-dependent lines include a mix of adenocarcinoma and large undifferentiated cells, indicating that NKX2-1 expression is not merely a proxy for cell subtype but a critical biomarker of its own dependency status. Results were less strong but still highly significant in Achilles and RNAi data (**Supplementary Table 1**). NKX2-1 has previously been proposed as a prognostic biomarker in lung cancer[19]; these results suggest it may also be an effective target.

Par-6 family cell polarity regulator beta (PARD6B) is similarly stratified by its own expression. Cell lines with high expression show increased dependency in Achilles, Score, and RNAi (**Supplementary Table 1, Fig. 3gh**). The effect is particularly pronounced within the ovarian and pancreatic cell lines: all three datasets increase correlation when restricting cell lines to these lineages (**Supplementary Table 1, Supplementary Fig. 10**).

Discussion

Measurement of gene expression profiles is a potent and still under-utilized method for identifying cancer vulnerabilities, with superior performance over characterizing genomic features in both genetic and compound response prediction. The advantage of expression-based features over DNA-based features held consistently across five high-throughput experimental platforms using three different perturbation technologies; multiple different subsets of perturbations, including cancer genes with known genomic indications; and multiple combinations of feature sets. Critically, expression-based models need not be black boxes. Most high-performing expression models we identified could be converted to easily interpretable models suitable for stratifying patients.

Iorio *et al.* also found that single-gene expression outperformed DNA-based features, but attributed this to the fact that RNA encodes tissue of origin[8]. When they attempted prediction within lineages using DNA-based Cancer Functional Events (CFEs) they found greater performance than with expression features. However, Rydenfel *et al*. found that expression and proteomics were more predictive than DNA even within tumor types[20]. We believe the discrepancy between these two studies can be explained by the strong curation applied by Iorio *et al.* to their DNA features. They identify a total of 1,063 DNA-based CFEs seen in at least one cell line[8]. This curation is quite successful in enriching for established drug targets: for example, 20.4% (54) of the compounds assayed in GDSC17 target a gene represented among the 472 cancer genes identified by Iorio *et al*. It is therefore reasonable that curated, binarized DNA features outperform uncurated, continuous, genome-wide expression in the context of very small sample sizes and compounds specifically developed against known genomic drivers. In such cases, it is more likely for a model to identify a true relationship if the number of irrelevant features is reduced through curation. Developing a similar curation strategy for expression features is likely to be fruitful for finding models in the context of limited sample sizes.

A study by Rambow *et al.* found that RXRG expression increases in melanoma $BRAF^{V600E/K}$ patient-derived xenografts (PDXs) in response to anti-BRAF therapy and drives them into a neural crest stem cell state.[21] However, the authors identified this response as a key resistance mechanism to BRAF inhibitors, whereas we found in both PRISM and GDSC that elevated RXRG is a highly significant predictor of sensitivity to BRAF inhibition even within the setting of BRAF-mutated cancers. It may be that RXRG upregulation is a broadly protective response for stressed melanoma cells. In that case, constitutively high RXRG expression could indicate existing stress that leaves the cells with less reserve for additional challenges such as anti-BRAF therapy. Alternatively, the difference between studies may relate to the difference between PDXs and established cell lines or the difference between the *in vivo* and *in vitro* settings.

Although RNA-Seq is not yet a standard clinical diagnostic tool, it is rapidly advancing towards maturity. Widespread adoption of this technology has positive implications for assigning existing therapies to patients. However, this work also suggests that at present, a wealth of new

vulnerabilities can only be identified and exploited with expression data. Maximizing the arsenal of antineoplastic compounds will require integrating expression data in target development pipelines.

Methods

Preprocessing of Perturbation Data

CRISPR data were taken unaltered from the gene_effect file in avana_public_19Q2. RNAi data were taken from gene_effect in the combined RNAi dataset[22]. PRISM data were taken from Corsello *et al*.[11] GDSC17 data were downloaded using the GDSC data repository and processed using the gdscIC50 repository[23] following the “gdsc_17” vignette until data was normalized by dose such that negative controls were 0 and positive controls 100. For each compound, a single dose was chosen that had the greatest variance over cell lines for that compound, subject to the requirement that at least 100 cell lines have scores for that dosage in that compound. Cell lines were mapped to CCLE names using their COSMIC IDs. Score data were taken from CERES-processed gene effect scores[24].

Preprocessing of Cell Features

Cell features were taken from the Cancer Dependency Map 19Q2 dataset.[25] Cell lines that did not have all four of RNA-Seq expression, mutation, copy number, and disease annotations were dropped. Categorical features were expanded in one-hot encodings. Features with zero variance in the remaining cell lines were dropped. Continuous features were Z-scored individually. Finally, all remaining missing values were filled with 0. Thus, the number of cell lines used for evaluating perturbations is the same regardless of the combination of features chosen for training.

Mutation

Mutation data were divided into three cell-by-gene binary matrices with the following logic:

- Damaging: True if the gene has a deleterious mutation in the cell line.
- Hotspot: True if the gene has a non-deleterious mutation in a TCGA or COSMIC hotspot.
- Other: True if the gene has a mutation not falling in the other categories

GSEA

For each MSigDB gene set, the R method GSVA::gsva was used on TPM RNA-Seq expression for CCLE lines released as of 19Q1.

Cancer Functional Event Equivalents

To develop a curated set of genomic Cancer Functional Events similar to those identified by Iorio *et al.*[8], we did the following:

- We took the 461 genes identified by Iorio et al. as high confidence cancer driver genes (level A, B, or C) and filtered both damaging and other matrices to only these genes.
- We took the recurrent copy number alterations (CNA) identified by Iorio et. al. and estimated the mean copy number of the corresponding segment in each cell line, using the WES-prioritized segmented copy number file from CCLE. We then produced a line-by-segment continuous matrix with the mean copy number values.
- We took the informative CpG islands identified by Iorio *et al.*[8] and produced a line-by-island matrix with the RBBS methylation values

CFE features with no results for any cell lines in CCLE were dropped. Cell lines with any missing values in any of the remaining features were dropped when training models using CFEs only.

Identifying Interesting Perturbations

For CRISPR and RNAi data, we generated probabilities of dependency for each gene in each cell line using the methodology described in <Achilles bioarxiv>. Only genes with at least five lines greater than 0.5 probability of dependency and five lines less than 0.5 probability of dependency were retained for training.

We identified strongly selective dependencies (SSDs) in CRISPR, RNAi, and PRISM using the NormLRT score method developed by McDonald *et al.* [8,26] This test compares the quality of the fit of the perturbation's distribution with a normal and skewed-t distribution, with higher scores indicating less normally-distributed data. The 100 perturbations with the highest scores in each dataset were labeled SSDs for the purposes of the following analyses. As noted above, GDSC data is scaled and clipped to the interval [0, 100], and therefore tests for normality on these data

are inappropriate. Instead, we selected the 100 compounds with the greatest variance as SSDs for this dataset.

Standard Predictor Pipeline

For each perturbation evaluated, cell lines with missing perturbation scores or no feature values were removed. We then selected features belonging to the feature types being evaluated and split cell lines into ten equal folds, choosing one fold to leave out for evaluation and taking the rest for training.

Next, if the number of features was greater than 1,000, we filtered for the 1,000 features with highest absolute Pearson correlation to the perturbation scores in the training cell lines. We trained sklearn random forests with max depth eight, 100 trees, and a minimum of five cell lines per leaf using the training data and predicted perturbation viabilities for the held-out fold.

We repeated this process again, choosing a different fold to hold out evaluation, until we had out-of-sample predictions for all cell lines. We scored the predictor using the correlation of these out-of-sample predictions to the original perturbation scores.

For genes found to be predicted with Pearson correlation above 0.4, we repeated the training process, but this time additionally selected the top 1, 2, 5, 10, and 100 most importance features per fold and trained models on just those features, saving their predictions on the held-out fold.

Elastic Net Predictor Pipeline

For each dataset, we adjusted all post-perturbation viabilities and all features to have mean 0 and variance 1. To reduce the cost of training the models, we chose fixed hyperparameters for each dataset in the following way: we took a random set of 20 SSDs and used the python implementation of glmnet to find the optimal value of lambda (the overall strength of regularization) with the ratio between L1 and L2 priors held fixed at 0.5, and nine specified lambda values spaced logarithmically in the interval [0.05, 50]. The optimal value of lambda was highly bimodal, with SSDs successfully predicted tending to small lambda values while those that could not be predicted using the maximum available value of lambda. We chose the mode

of the smaller lambda values for training the dataset. This was 0.5 for Score data and 0.2 for all other datasets.

Training and scoring predictors proceeded as with the standard predictor pipeline, except that only SSDs were used and the magnitude of the coefficients was used to rank features instead of importance.

Feature Importance

To assess feature importance for a perturbation, we performed Pearson feature filtering and trained a random forest using all available cell lines with feature and perturbation data. Feature importance was taken from the resulting model using sklearn's default method, which measures gini importance. The resulting importance is normalized so that the sum of the importance of all features (of the thousand retained after Pearson filtering) is 1. Features with at least 0.1 importance were considered key features.

Interpretable Models

We formed interpretable models of selected perturbations using single decision trees. We took the top three features identified by random forest. For each of the three chosen features, if it was continuous, we trained a depth-1 sklearn RegressionTree to predict the perturbation and used the resulting optimal split to binarize the feature. We then trained an sklearn RegressionTree with max depth 4 and minimum samples per leaf 5, and the results visualized as in Figure 4. The trees were additionally regularized to require a minimum impurity decrease per split of 0.02 * *S*. *S* is a standard deviation found by taking all successfully predicted perturbations in the given dataset with a key feature, raveling all their viability measurements into single long vector, and calculating the standard deviation. If the trees only split on one feature, we present binned scatter plots as in e.g. **Fig. 4d** to illustrate the full relationship between the feature and the post-perturbation viability.

Competing Interests

T.R.G. is a paid consultant to GlaxoSmithKiline, is a co-founder of Sherlock Biosciences and FORMA Therapeutics, and receives sponsored research funding from BayerHealthCare, Calico

Life Sciences and Novo Ventures.. A.T. performs consulting for Tango Therapeutics. All other authors declare no competing interests. All authors were partially funded by the Cancer Dependency Map Consortium, but no consortium member was involved in or influenced this study.

Acknowledgments

The authors would like to thank Francesco Iorio and Emanuel Gonçalves for helpful commentary and suggestions.

References

1. Macconaill, L. E. & Garraway, L. A. Clinical implications of the cancer genome. *J. Clin. Oncol.* **28**, 5219–5228 (2010).
2. Haber, D. A., Gray, N. S. & Baselga, J. The evolving war on cancer. *Cell* **145**, 19–24 (2011).
3. Cheng, M. L., Berger, M. F., Hyman, D. M. & Solit, D. B. Clinical tumour sequencing for precision oncology: time for a universal strategy. *Nat. Rev. Cancer* **18**, 527–528 (2018).
4. Prasad, V. Perspective: The precision-oncology illusion. *Nature* **537**, S63 (2016).
5. Letai, A. Functional precision cancer medicine—moving beyond pure genomics. *Nat. Med.* **23**, 1028–1035 (2017).
6. Zhu, B. *et al.* Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Sci. Rep.* **7**, 16954 (2017).
7. Ali, M. & Aittokallio, T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys. Rev.* **11**, 31–39 (2019).
8. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740–754 (2016).

9. Aben, N., Vis, D. J., Michaut, M. & Wessels, L. F. A. TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types. *Bioinformatics* **32**, i413–i420 (2016).
10. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
11. Corsello, S. M. *et al.* Non-oncology drugs are a source of previously unappreciated anti-cancer activity. *Cancer Biology* 589 (2019).
12. McFarland, J. M. *et al.* Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat. Commun.* **9**, 4610 (2018).
13. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
14. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**, (2017).
15. Warters, R. L., Packard, A. T., Kramer, G. F., Gaffney, D. K. & Moos, P. J. Differential gene expression in primary human skin keratinocytes and fibroblasts in response to ionizing radiation. *Radiat. Res.* **172**, 82–95 (2009).
16. Macaeva, E. *et al.* Radiation-induced alternative transcription and splicing events and their applicability to practical biodosimetry. *Sci. Rep.* **6**, 19251 (2016).
17. Giacomelli, A. O. *et al.* Mutational processes shape the landscape of TP53 mutations in human cancer. *Nat. Genet.* **50**, 1381–1387 (2018).
18. Chène, P. Inhibiting the p53–MDM2 interaction: an important target for cancer therapy. *Nat. Rev. Cancer* **3**, 102–109 (2003).
19. Yang, L. *et al.* Nkx2-1: a novel tumor biomarker of lung cancer. *J. Zhejiang Univ. Sci. B* **13**,

855–866 (2012).

20. Rydenfelt, M., Wongchenko, M., Klinger, B., Yan, Y. & Blüthgen, N. The cancer cell proteome and transcriptome predicts sensitivity to targeted and cytotoxic drugs. *Life Sci Alliance* **2**, (2019).
21. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell* **174**, 843–855.e19 (2018).
22. DepMap, B. DEMETER 2 Combined RNAi. (2019) doi:10.6084/m9.figshare.9170975.v1.
23. CancerRxGene. CancerRxGene/gdscIC50. *GitHub* https://github.com/CancerRxGene/gdscIC50.
24. DepMap, B. Project SCORE processed with CERES. (2019) doi:10.6084/m9.figshare.9116732.v1.
25. DepMap, B. DepMap 19Q2 Public. (2019) doi:10.6084/m9.figshare.8061398.v1.
26. McDonald, E. R., 3rd *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577–592.e10 (2017).