nature biomedical engineering

Article

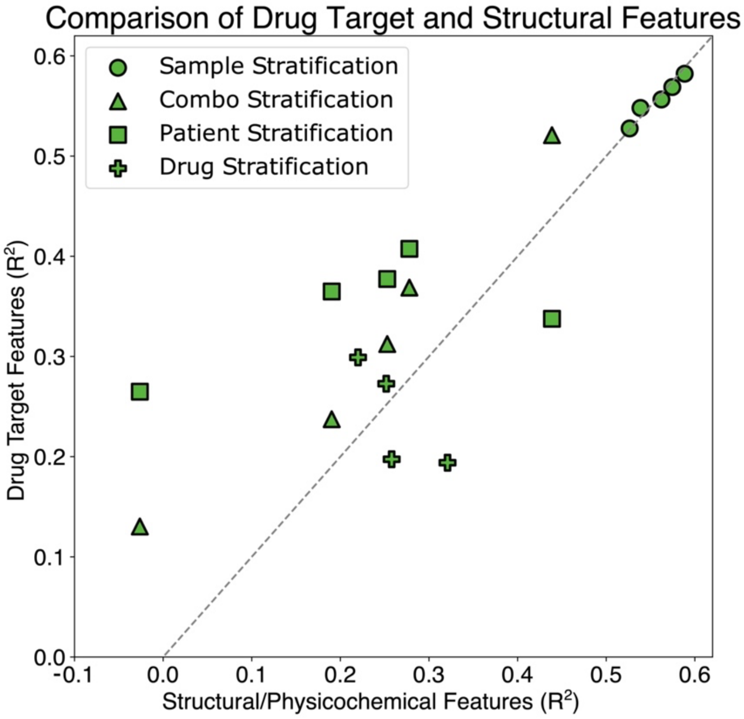
https://doi.org/10.1038/s41551-023-01034-0

Uncovering expression signatures of synergistic drug responses via ensembles of explainable machine-learning models

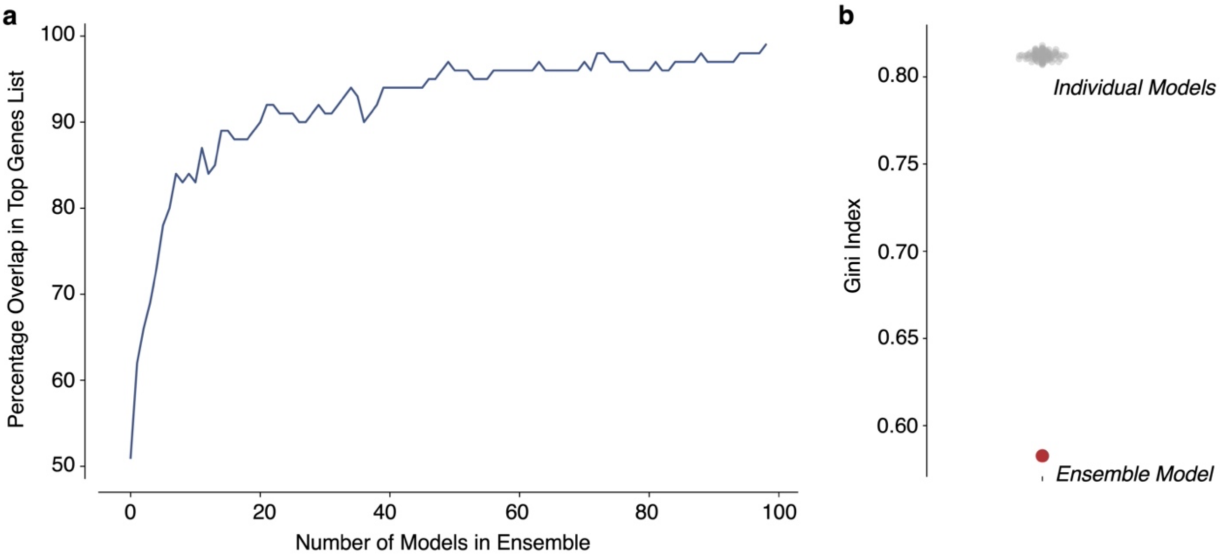
In the format provided by the authors and unedited

Supplementary Table 1 | Statistical comparisons associated with Main Fig. 3a-b. All p values are two-sided.

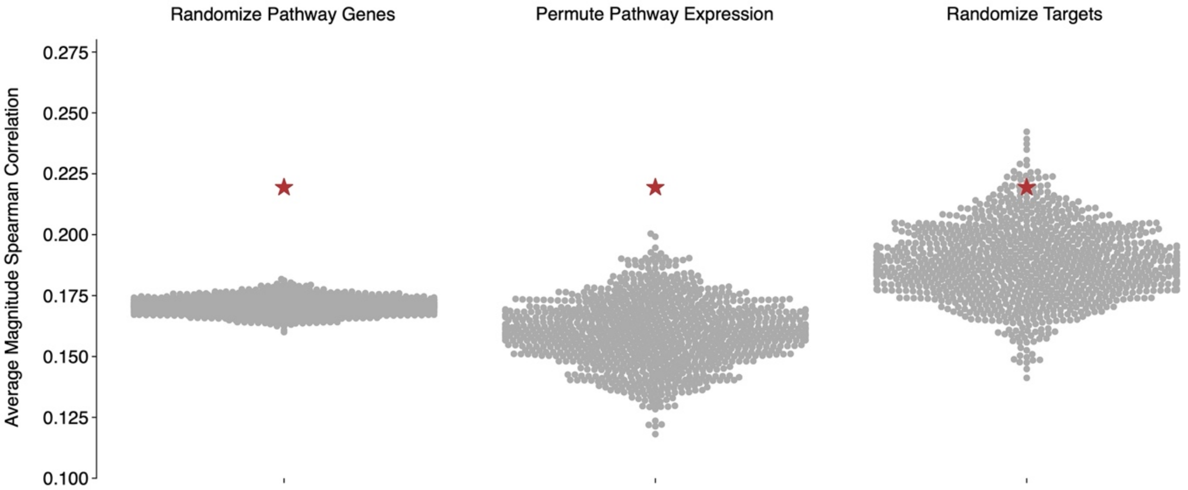
Model Type	**Dataset**	**Pearson's *r***	**p value**
All models	Step Function (AND)	–0.77	1.1×10^{-12}
Elastic Net	Step Function (AND)	0.19	0.43
Neural Network	Step Function (AND)	0.02	0.94
XGBoost	Step Function (AND)	–0.18	0.45
All models	Multiplicative	–0.82	1.2×10^{-15}
Elastic Net	Multiplicative	–0.11	0.65
Neural Network	Multiplicative	–0.22	0.35
XGBoost	Multiplicative	0.13	0.60



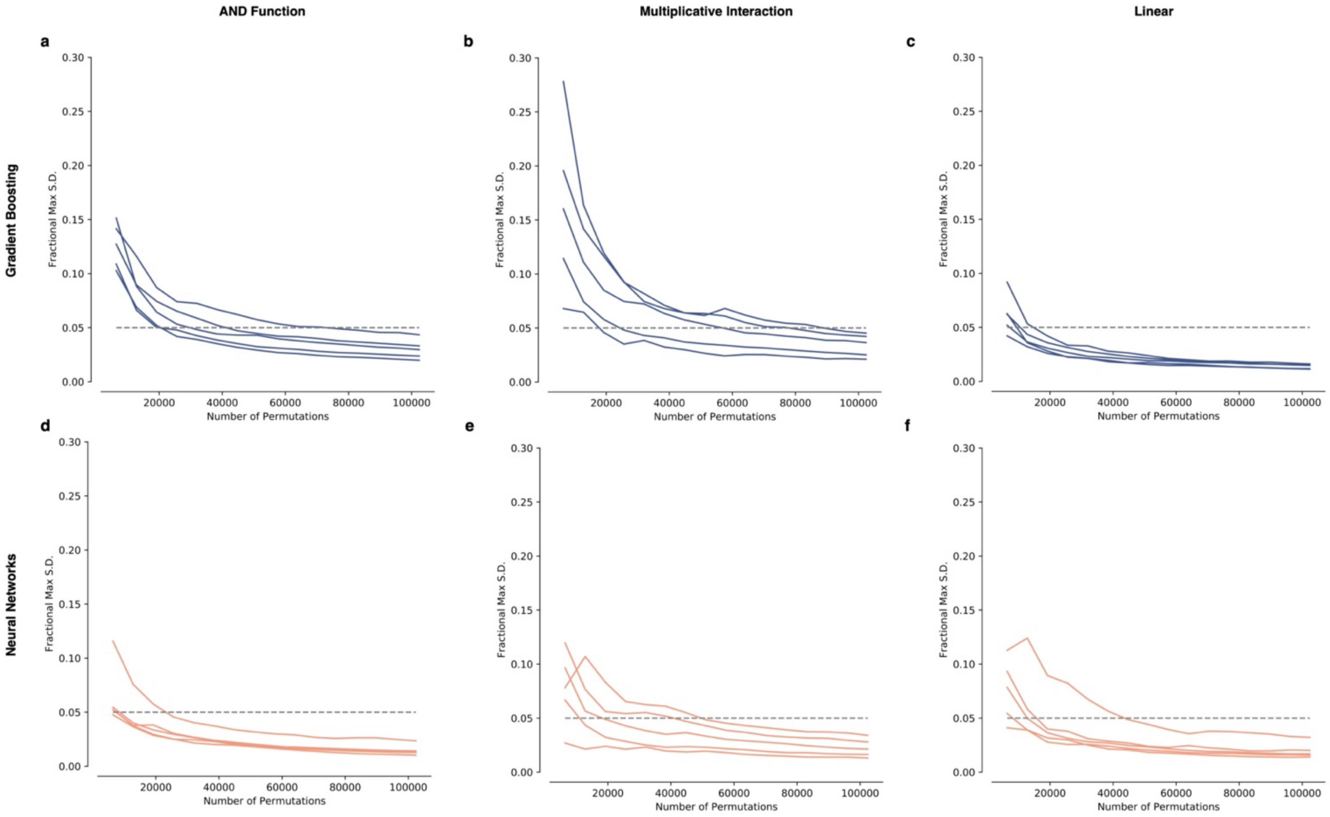
Supplementary Fig. 1 | Comparison of drug target and drug structure features. Each point in the plot represents an evaluation of an XGBoost model's performance on a different held-out split of the data, with the vertical axis representing the performance of a model with literature-derived drug target features, whereas the horizontal axis represents the performance of a model using structural/physicochemical features. Held-out test folds where drug-target features result in a model with higher predictive performance lead to points above the diagonal. Following the same procedure in Fig. 4, samples were stratified in four ways. In the first stratification, samples are randomly split into 5 different train test folds. In the second, samples are split on the basis of the drug combinations, so that held-out test folds contain novel drug combinations that were not present in the training data. In the third, samples are split on the basis of patients, so that held-out test folds contain patients that were not present in the training data. In the fourth, samples are split on the basis of individual drugs, so that held-out folds contain drugs that were not present in the training data. Across 17 of 20 held-out test folds, we find that drug target features lead to higher predictive performance than structural/physicochemical features.



Supplementary Fig. 2 | Beat AML data attribution characteristics. **a**,To ensure that a sufficient number of models were included in our final ensemble, we measured the percentage overlap in the final list of top 100 genes, and the cumulative top 100 genes list as additional XGBoost models were ensembled. **b**, Attribution vector sparseness for individual models trained on Beat AML dataset (grey) and attribution vector sparseness for our final ensemble model (red). A lower Gini Index indicates a sparser attribution vector.



Supplementary Fig. 3 | Hematopoietic differentiation expression signature replicates in external dataset. Using AML cancer cell line expression data and CRISPR genetic dependency scores from the DepMap database, we show that the average magnitude association of the S1 and D1 expression signatures with genetic dependency of the targets of the drugs in Fig. 5f (shown as red stars) are stronger than the average magnitude associations under three separate null distributions. The first null distribution (“Randomize Pathway Genes”) averages random genes rather than the actual genes in signatures S1 and D1. The second null distribution (“Permute Pathway Expression”) permutes the rows in the expression matrix. The third null distribution (“Randomize Targets”) measures the associations with random genetic dependency targets, rather than the actual targets of the drug combinations in Fig. 5f. Across all three null models, we find that the true expression signature is significantly associated with cancer cell line genetic dependency on the drug targets of the drugs in Fig. 5f (empirical p values 0.001, 0.001, and 0.026, respectively).



Supplementary Fig. 4 | Sampling convergence of SAGE values. To ensure convergence of the Shapley value estimates used in our benchmark, we measured the maximum standard deviation of the elements in our attribution vector (generated using SAGE) as a fraction of the total attribution (Fractional Max S.D.), plotted against the number of permutations sampled. We found that for all three outcome types (a step function generated using a Boolean AND function, a sum of pairwise multiplicative interactions, and a linear function), for both gradient boosting models (**a-c**) and neural networks (**d-f**) 102400 permutations were sufficient to drive the Fractional Max S.D. below 5%.