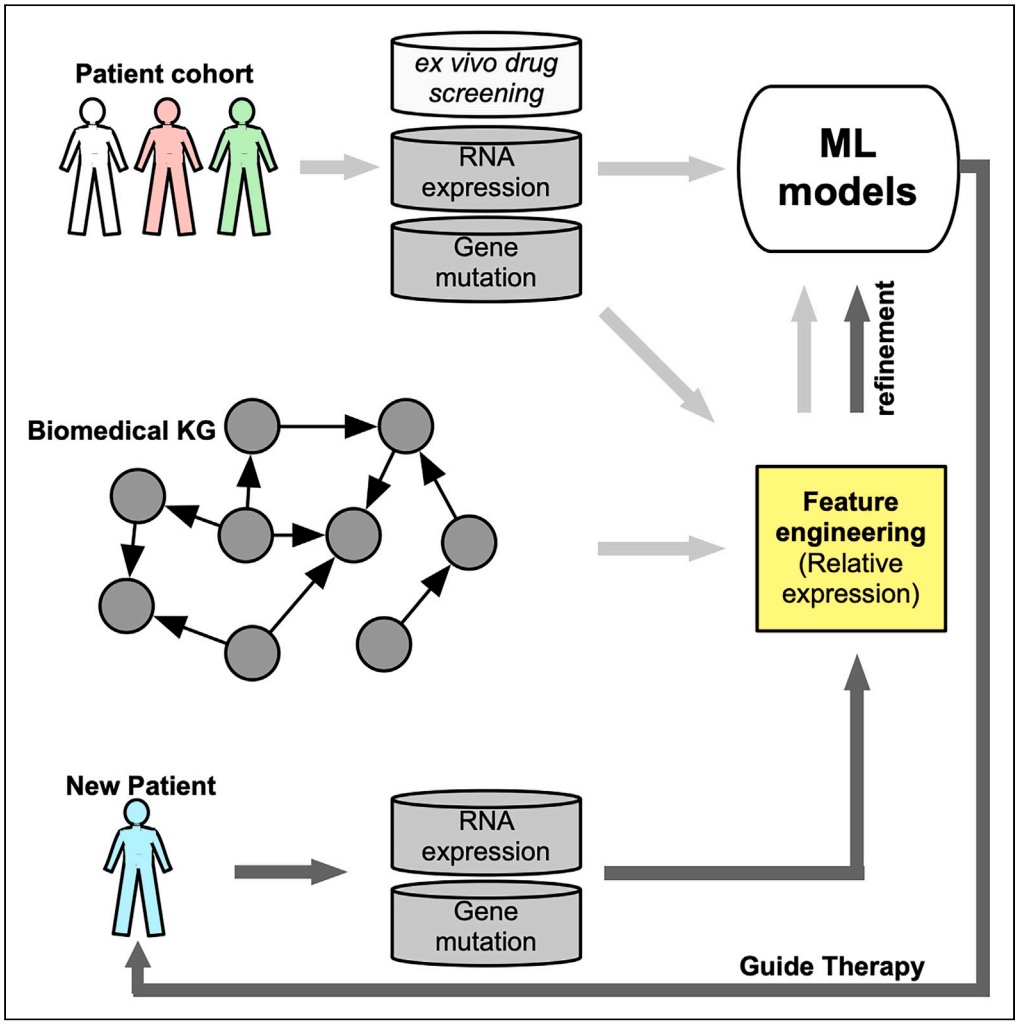
iScience

CellPress
OPEN ACCESS

Article

Knowledge graphs facilitate prediction of drug response for acute myeloid leukemia



Guangrong Qin, Yue Zhang, Jeffrey W. Tyner, Christopher J. Kemp, Ilya Shmulevich

guangrong.qin@isbscience.org

Highlights

Knowledge graphs facilitate feature engineering in machine learning

The engineered features can enhance accuracy for specific drugs

Our models provide predictive features for drug response in acute myeloid leukemia

Qin et al., iScience 27, 110755
September 20, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.110755

iScience

CellPress
OPEN ACCESS

Article

Knowledge graphs facilitate prediction of drug response for acute myeloid leukemia

Guangrong Qin,[1,5,*] Yue Zhang,[1] Jeffrey W. Tyner,[2] Christopher J. Kemp,[3] and Ilya Shmulevich[1,4]

SUMMARY

Acute myeloid leukemia (AML) is a highly aggressive and heterogeneous disease, underscoring the need for improved therapeutic options and methods to optimally predict responses. With the wealth of available data resources, including clinical features, multiomics analysis, and *ex vivo* drug screening from AML patients, development of drug response prediction models has become feasible. Knowledge graphs (KGs) embed the relationships between different entities or features, allowing for explanation of a wide breadth of drug sensitivity and resistance mechanisms. We designed AML drug response prediction models guided by KGs. Our models included engineered features, relative gene expression between marker genes for each drug and regulators (e.g., transcription factors). We identified relative gene expression of FGD4-MIR4519, NPC2-GATA2, and BCL2-NFKB2 as predictive features for venetoclax *ex vivo* drug response. The KG-guided models provided high accuracy in independent test sets, overcame potential platform batch effects, and provided candidate drug sensitivity biomarkers for further validation.

INTRODUCTION

Acute myeloid leukemia (AML) is a highly aggressive form of cancer with a low 5-year overall survival rate, primarily due to relapse and primary resistance to current standard therapeutic regimens.[1–4] Therapeutic outcomes remain particularly dismal for relapsed or refractory and older patients, who are often unfit for intensive therapies. There have been 10 U.S. Food and Drug Adminstration (FDA) approvals for new single-agents or drug combinations over the past 6 years. While these new regimens have resulted in improved response rates, nearly universal resistance has been observed. These clinical results underscore the need for better understanding of mechanisms of drug sensitivity and resistance coupled with methods to predict how a patient may respond to a given drug.

Numerous studies have yielded a wealth of data, encompassing clinical data from electronic medical records (e.g., clinical labs, treatments, and outcomes), multiomics data (e.g., RNA sequencing [RNA-seq] and whole exome sequencing), and *ex vivo* drug screening for patients with AML. These datasets include Beat AML,[5,6] Functional Precision Medicine Tumor Board (FPMTB),[7] and our previous study.[8] Additionally, other data resources containing AML cell line data, such as Genomics of Drug Sensitivity in Cancer[9] are also available. With the development of machine learning models and artificial intelligence, it has become possible to develop more robust predictive models for drug response prediction.

Various methods have been developed for predicting drug response, such as Logic Optimization for Binary Input to Continous Output (LOBICO),[9,10] random forest, LightGBM, and deep neural networks.[11] Traditional machine learning approaches typically use measurements from one sample as features to make predictions without considering existing knowledge or relationship among the features. A simulation model has been proposed to predict response to inhibitors to the Bromodomain and Extra-Terminal (BET) family of proteins using the Beat AML dataset.[12] Explainable drug sensitivity prediction through cancer pathways enrichment has been implemented in PathDSP.[13] AML specific drug response models have been reported, including using *ex vivo* drug sensitivity profiling in predicting patient survival,[14] and defining a drug sensitivity score (DSS) threshold that suggests sensitivity/resistance to venetoclax.[15,16] A cross-study analysis of drug response prediction in cancer cell lines suggested that a limitation still exists for prediction models: the ability of the model to generalize across studies,[11] which may be due to differences in viability assays or batch effects in the feature measurements. To overcome some of these challenges and improve prediction accuracy and generalizability, several aspects should be considered, including the use of prior knowledge or interactions among features, as well as defining features that can mitigate batch effects between training and test datasets. These challenges highlight the need to carefully select features and construct measurable features that can reduce batch effects for further model testing and clinical applications.

[1]Institute for Systems Biology, Seattle, WA 98109, USA
[2]Knight Cancer Institute, Oregon Health & Science University, Portland, OR 97239, USA
[3]Fred Hutchinson Cancer Center, Seattle, WA 98109, USA
[4]Deceased
[5]Lead contact
*Correspondence: guangrong.qin@isbscience.org
https://doi.org/10.1016/j.isci.2024.110755

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Knowledge graphs (KGs) encode extensive sets of associations among molecular, cellular, and clinical data and can correlate these features with activity of potential therapeutic agents. Nodes of KGs represent entities such as drugs, genes, clinical measurements, biological functions, and pathways, while edges represent known physical interactions, regulatory relationships between the entities, and statistical associations. Both nodes and edges stem from either accumulated knowledge or statistical inferences drawn from data. Consortium efforts have been made to integrate existing biomedical datasets and translate those data into insights[17] using a standardized way to represent KGs.[18]

Here we developed a novel approach to predict drug response for AML patient-derived samples using KG-guided feature engineering. Our approach leverages KGs to develop prediction models that use features derived from the graph, thereby accounting for prior knowledge and known interactions among the features. Additionally, we defined measurable features that can reduce batch effects for further model testing or clinical applications. To validate our approach, we applied it to the Beat AML wave 1/2 datasets (sample size of 672)[6] and tested it on the Beat AML wave 3/4 datasets (sample size of 297).[5] Our results demonstrate that the KG-guided feature engineering approach enhances the generalizability of the predictive models and provides robust features for drug sensitivity prediction.

RESULTS

Overview of the graph-guided prediction models

The primary objective is to predict *ex vivo* drug sensitivity for the purpose of guiding personalized therapy using molecular measurements from AML patients (Figure 1). Beyond using feature sets directly from molecular measurements, such as gene mutation, variant allele frequencies grouped by mutation sites for specific genes, and gene expression, we also utilize prior knowledge encoded in a graph and perform feature engineering based on the graph structure to generate relative gene expression features. This approach includes the following four steps: (1) AML KG extraction and feature selection, (2) feature engineering, (3) model construction, and (4) model validation. Work progress toward large biomedical KGs has been made in the Translator consortium.[17–19] However, to reproduce the complexity of large KGs and demonstrate this KG-derived feature engineering approach, we generated a minimal KG to make predictions.

We constructed AML-specific KGs from statistical approaches using public datasets or derived from public knowledge resources or literature. The statistical model derived KGs include (1) associations between gene expression and drug sensitivity derived from the Least Absolute Shrinkage and Selection Operator (LASSO) regression approach using Beat AML wave 1/2 dataset[6] (Table S1); (2) frequently mutated genes in AML[3,6,20]; (3) regulations between microRNAs and target genes from mirTarBase[21] and MSigDB[22]; and (4) regulations between transcription factors and target genes from the literature.[23] The regulatory effects between transcription factors/microRNAs and target genes in AML were derived from LASSO regression models with negative coefficients based on the Beat AML wave 1/2 dataset (Table S2). The underlying assumption of selecting the negatively correlated regulators is that the balance of gene expression between the signature genes and regulator genes will impact the state of cells, which further affect drug response.

For the feature engineering step, we calculated relative gene expression values between the genes that are associated with drug sensitivity and their regulators such as transcription factors or microRNAs. We first selected gene signatures that are predictive for the drugs in the KG. We next extended the graph by selecting regulators (transcription factors and miRNA) that may affect the expression of these drug-associated gene signatures. We then used the KG to facilitate feature engineering. We generated relative gene expression values as binarized values by comparing the expression levels of the predictive features and their negative regulatory partners (Figure 1). We also used housekeeping genes that show the lowest gene expression variance in different abundance levels to generate relative gene expression features by comparing the expression of genes that are associated with drug sensitivity and the housekeeping genes (Figures S1 and 1). The rationale of adding low variance housekeeping genes is to generate relative gene expression that may overcome batch effects. These steps define binarized features for further prediction models. We then used these engineered features in the machine learning models (XGBoost classifier) and used 90 percent of samples in Beat AML wave 1/2 as training sets, 10 percent of samples in Beat AML wave 1/2 as validation sets using a 10-fold cross validation approach, and used Beat AML wave 3/4 as independent test data.

Prediction of drug response using the features from the KGs

We used the gene mutations (Mut, see Table 1; Figure S2), variant allele frequency (VAF, see Table 1; Figure S2), and gene expression (RNA-seq) data (Expr), relative gene expression defined from the KGs (RelativeExpr) and the combined feature sets of gene mutations, VAF and relative gene expression from the gene pairs in the KG (Mut+VAF, Mut+RelativeExpr, Mut+VAF+RelativeExpr) to predict drug sensitivity. We used classification models to predict whether a sample/patient would be sensitive or resistant to a drug based on its inhibitory concentration (IC50) value from the *ex vivo* drug screening data in the Beat AML wave 1/2 dataset. We assumed that samples with IC50 values below a certain threshold indicated a sensitive response, while samples with IC50 values above that threshold indicated a resistant response. This assumption gives some advantages for the prediction models especially when researchers commonly give a cutoff of the tested drug concentration (commonly 10 μM as the Beat AML dataset). We evaluated prediction models using various IC50 thresholds (Q25: 25th percentile of IC50 in the Beat AML wave 1/2 dataset, Q50: median IC50 values, and Q75: 75th percentile if the threshold is smaller than 10 μM). We also considered an additional threshold 1 μM if Q75 value is below 1 μM to label the sensitive group or resistant group.

For instance, for venetoclax, we employed two candidate thresholds: 0.02 μM (Q25) and 1.45 μM (Q50). We then compared prediction results from XGboost classifier models using different feature sets and thresholds. Models with highest prediction accuracy in the independent dataset Beat AML wave 3/4 were those trained using the relative gene expression features, achieving a median balanced accuracy of 0.78 when classifying sensitive or resistant samples using the threshold of 1.45 μM (Figure 2A). The Wilcoxon test revealed no significant difference in balance accuracy among models trained with relative gene expression features with or without gene mutation or variant allele frequencies

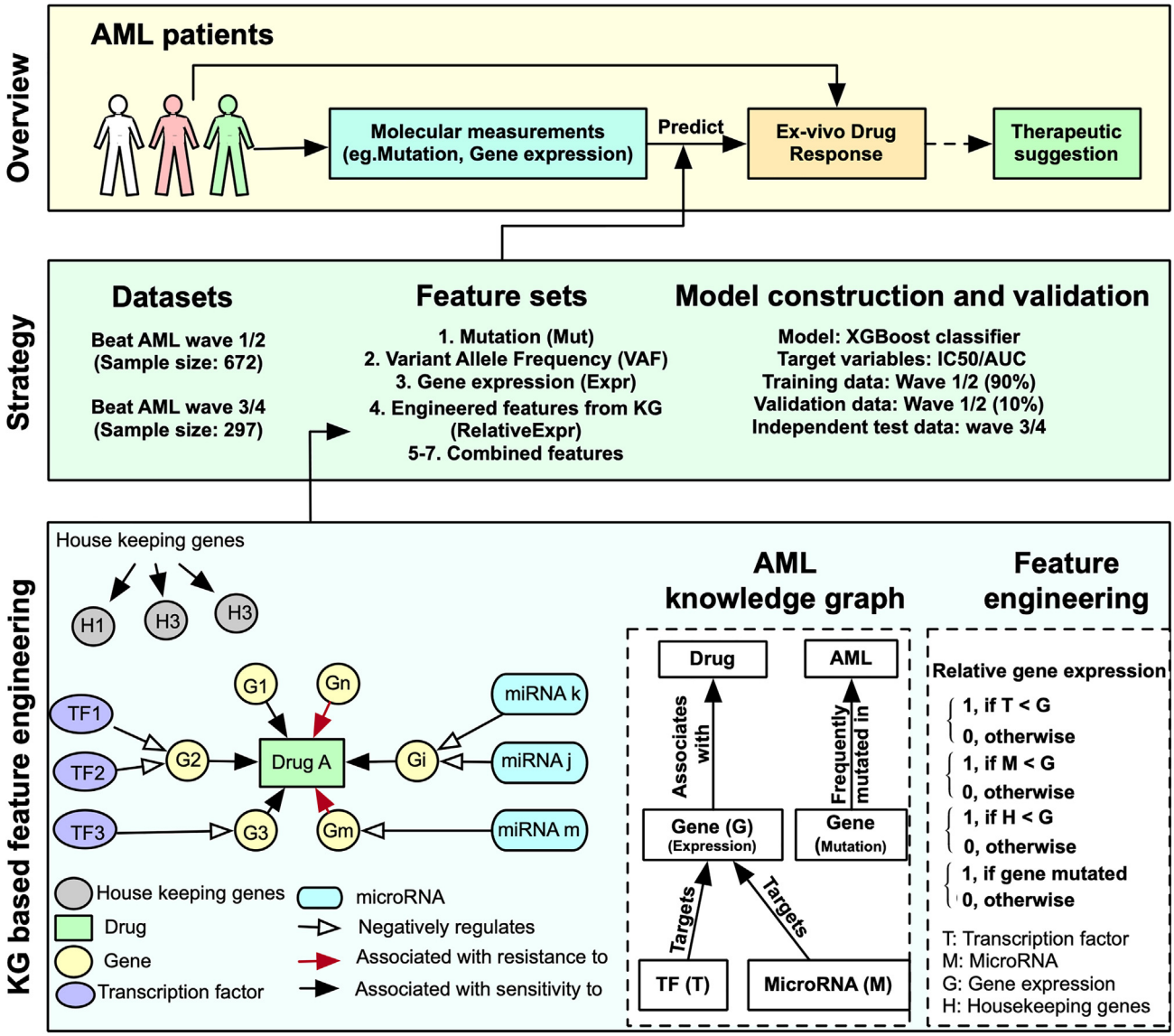


Figure 1. The workflow of knowledge graph-guided machine learning model for predicting drug response in AML

The goal of the work is to construct predictive models for the prediction of *ex vivo* drug response using the molecular measurements from AML patient-derived samples, which could provide therapeutic suggestions. We used the Beat AML wave 1/2 dataset as the training set, with 10-fold cross validation. Four types of feature sets were used to build the model, namely, (1) gene mutations; (2) variant allele frequency for selected mutation sites; (3) gene expression; and (4) engineered features from knowledge graphs. We also included combined features sets to make predictions, see Table 1. For the feature engineering step, we first constructed AML knowledge graphs that capture the relationships among gene expression, drugs, transcription factors, microRNAs, and frequently mutated genes in AML (white box on the left). We then performed feature engineering to generate binary features from knowledge graphs (white box on the right). Here, G represents the expression of signature genes, while T, M, and H represent the expression levels of transcription factors, microRNA, and housekeeping genes, respectively.

(Figure 2B). Models based on relative gene expression, defined through gene pairs from the KG, showed the best balanced accuracy and F1 score using a threshold of 1.45 μM in the independent testing set Beat AML wave 3/4 (Figures 2A and 2C). The gene pair based models exhibited significant differences in balanced accuracy or F1 score compared to those based on gene mutation or expression (Figures 2B and 2D). Our results underscore the importance of considering relative gene expression for drug response prediction of venetoclax. For the prediction of trametinib response, we found the models trained using the feature set combining mutation and relative gene expression yielded the highest balanced accuracy (Figure 2E). Similar to venetoclax, models incorporating relative expression features significantly improved the accuracy compared to those relying only on gene mutation or expression (Figure 2F). Using F1 score as another measurement of performance also supports the same conclusion (Figures 2G and 2H).

Table 1. Features used for the machine learning models

Category of features	Features	Number of features
Highly frequently mutated genes (>3% of the Beat AML samples) (Feature set 1: Mut)	TET2(1: mutated/0: wild type), STAG2, RUNX1, EZH2, JAK2, FLT3, SF3B1, GATA2, WT1, CEBPA, NPM1, KRAS, IDH1, BCOR, TP53, ASXL1, DNMT3A, U2AF1, PTPN11, SRSF2, NRAS, and IDH2	22
Variant allele frequency of highly frequently altered variant groups (altered in at least 1% of the Beat AML samples) (Feature set 2: VAF)	DNMT3A:p.R882 (DNMT3A:p.R882H/C), ASXL1:p.G645-646 (ASXL1:p.G645Vfs, ASXL1:p.G646Wfs, ASXL1:p.G645Wfs), FLT3:p.835-839 (FLT3:p.D835E/H/N/V/Y, FLT3.p.D839G, FLT3:p.I836del), IDH1:p.R132 (IDH1:p.R132C/G/H/L/S), IDH2:p.R140 (IDH2:p.R140L/Q/W), IDH2:p.R172 (IDH2:p.R172K), JAK2:p.V617 (JAK2:p.V617F), KIT:p.D816 (KIT:p.D816H/V/Y), KRAS:p.G12/13 (KRAS:p.G12A/C/D/R/V, KRAS:p.G13R), NPM1:p.W288-90(NPM1:p.W288Cfs*12, NPM1:p.W290Cfs*10, NPM1:p.W290Rfs*10), NRAS:p.G12/13(NRAS:p.G12A/C/D/R/S, NRAS:p.G13C/D/R/V) NRAS:p.Q61(NRAS:p.Q61H/K/L/P/R), SF3B1:p.K700E SF3B1:p.K666(SF3B1:p.K666E/M/N/Q/T), SRSF2:p.94-95(SRSF2:P95 H/L/R, SRSF2:P95_R102del, SRSF2:P94dup, SRSF2:P95delinsRA), U2AF1:p.S34(U2AF1:p.S34F/Y) U2AF1:p.156-157(U2AF1:p.Q157P/R, U2AF1:p.R156H) ZNF687:p.R939Pfs*36	18
RNA expression (genes with expression value [log(RPKM)] greater than 0 in at least 50%) (Feature set 3: Expr)	Expression of individual genes	17,691
Relative gene expression features from graph structure (Feature set 4: RelativeExpr)	Each drug has a different feature set	Vary among drugs
Combined feature set Mut+VAF	Feature Set 1 + Feature Set 2	40
Combined feature set Mut+RelativeExpr	Feature Set 1 + Feature Set 4	Vary among drugs
Combined feature set Mut+VAF+RelativeExpr	Feature Set 1 + Feature Set 2 + Feature Set 4	Vary among drugs

Validation of the prediction features for venetoclax

The prediction models employing relative gene expression for the prediction of venetoclax *ex vivo* drug sensitivity show two gene pairs presenting the highest importance scores, FGD4-MIR4519 and NPC2-GATA2 (Figure 3A). We also identified the relative gene expression between the drug target gene BCL2 and its regulators such as NFKB2 are among the top predictive features. The two groups of samples defined by the relative gene expression between FGD4 and MIR4519 show significant differences of log-transformed IC50 values in Beat AML wave 1/2 and Beat AML wave 3/4 (Figures 3B and 3C). Comparing the log-transformed drug IC50 values between the samples defined by the relative gene expression of NPC2 and GATA2, significant differences between the two groups in both the Beat AML wave 1/2 dataset and Beat AML wave 3/4 dataset have been observed ($p < 0.001$, Ranksum test) (Figures 3D and 3E). Using an independent dataset FPMTB,[7] we also validated that the two groups of samples defined by the relative expression of NPC2-GATA2 show differential sensitivity to venetoclax (Figure 3F). We also tested whether other newly defined features are statistically associated with drug sensitivity using the Wilcoxon sum rank test. We found features such as the relative gene expression of BCL2-NFKB2 are associated with sensitivity to venetoclax in multiple datasets, including Beat AML wave 1/2, Beat AML wave 3/4 and FPMTB (Figures 3G–3I). Some of these top predictive features can be explained by

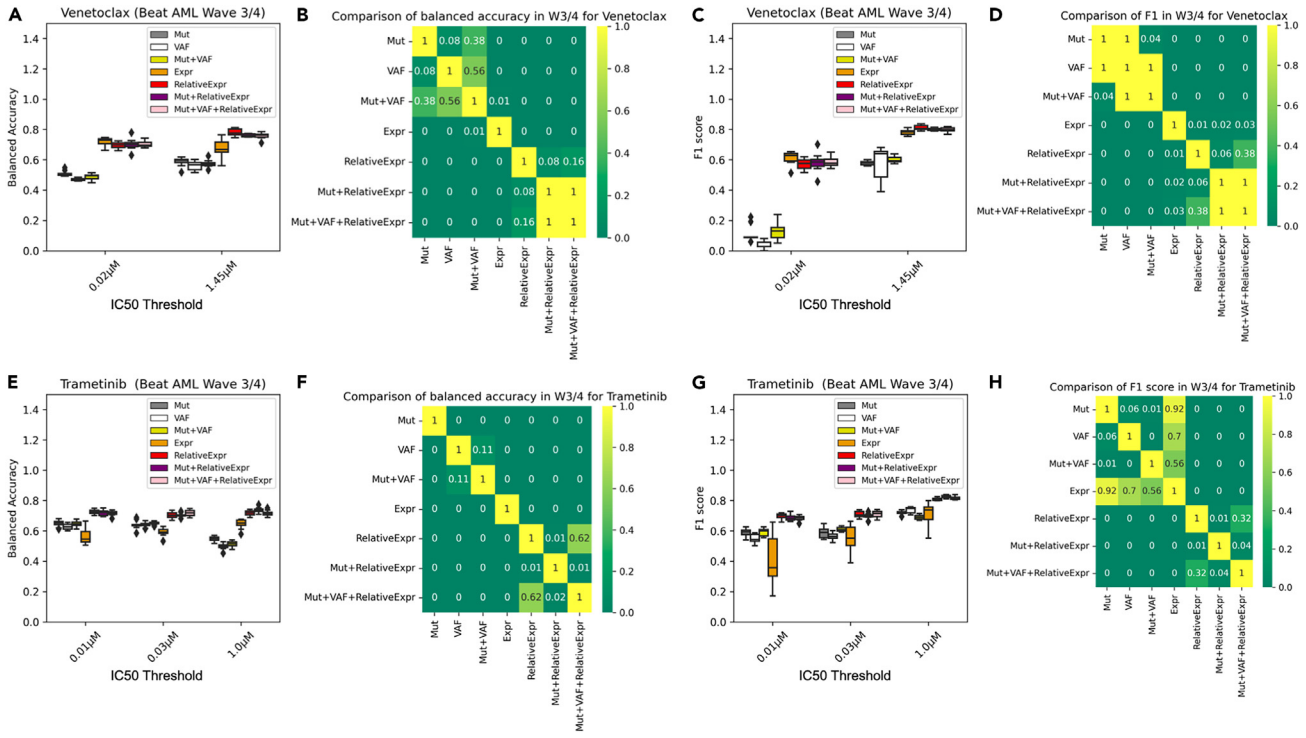


Figure 2. Prediction accuracy of the XGBoost classification models for the prediction of drug response of venetoclax and trametinib using different binarization thresholds and feature sets

(A, C, E, and G) Balanced accuracy and F1 scores of the XGBoost models using different feature sets to predict the *ex vivo* drug response of venetoclax and trametinib with different thresholds as cutoffs to define sensitive or resistant groups. The threshold used in the figures are the IC50 of the 25th percentile, 50th percentile or 1 μm. The feature sets used are defined in Table 1. The balanced accuracy and F1 scores shown here are based on the test dataset Beat AML wave 3/4.

(B, D, F, and H) Significance of difference (*p* value) of balanced accuracy and F1 scores between models based on different feature sets tested by Wilcoxon test.

existing literature. For example, our model predicts the relative gene expression between BCL2-NFKB2 are predictive of venetoclax sensitivity, as BCL2 is the known target protein for venetoclax, patients with higher expression of BCL2 would be expected to show higher sensitivity to venetoclax. For samples with NPC2 expression significantly higher than GATA2, IC50 of venetoclax is higher than samples with NPC2 expression lower than GATA2. This result suggests higher GATA2 expression indicates increased venetoclax sensitivity. GATA2 is suggested to act as a critical regulator of normal and leukemic stem cells. Ablation of GATA2 enforces a leukemic stem cell specific program of enhanced apoptosis, exemplified by attenuation of anti-apoptotic factor BCL2, and re-instigation of myeloid differentiation.[24] This suggests lower GATA2 may be associated with lower BCL2 expression and resistance to venetoclax. These results demonstrate the selected binarized features could be used as robust features to predict venetoclax *ex vivo* drug response.

Models based on relative gene expression provide higher accuracy in independent test datasets

We then extended our models to predict the *ex vivo* response for other drugs. Drugs or chemicals included in our models are those that show sensitivity in AML samples with a threshold of IC50 less than 50 nM in greater than ten samples in the Beat AML wave 1/2 dataset and which are also measured in Beat AML wave 3/4 dataset, resulting in 70 drugs. The top sensitive drugs include elesclomol (mitochondria), trametinib (Mitogen-activated protein kinase kinase [MEK] inhibitor), ponatinib (target BCR-ABL and multi-tyrosine kinase inhibitor), INK-128 (mTOR inhibitor), dasatinib (KIT inhibitor), panobinostat (HDAC inhibitor), JNJ-28312141 (receptor tyrosine kinase inhibitor), rapamycin (mTOR inhibitor), foretinib (multi-kinase inhibitor), quizartinib (FLT3 inhibitor), venetoclax (BCL2 inhibitor), sunitinib (tyrosine kinase inhibitor), dovitinib (multi-kinase; fibroblast growth factor receptor [FGFR] inhibitor), doramapimod (p38 mitogen-activated protein kinase [MAPK] inhibitor), and others. We then applied the XGBoost classifier approach to generate models to predict drug response using different feature sets, namely gene mutations, VAF, gene expression, and relative gene expression guided from the KG and three sets of combined feature sets as defined in Table 1. Our result shows most of the drugs show the best model accuracy based on the feature sets that include the relative gene expression guided by the KG in the independent test dataset Beat AML wave 3/4 as shown in Figure 4A. The threshold of IC50 that is used to define sensitive or resistant samples and the feature sets used for the best performing models are shown in Table S3. The drugs that show the best prediction accuracy based on gene mutation or gene expression are shown in Figure 4B, and drugs that show improved accuracy based on relative gene expression are shown in Figure 4C. Our results highlight those models based on either the relative gene

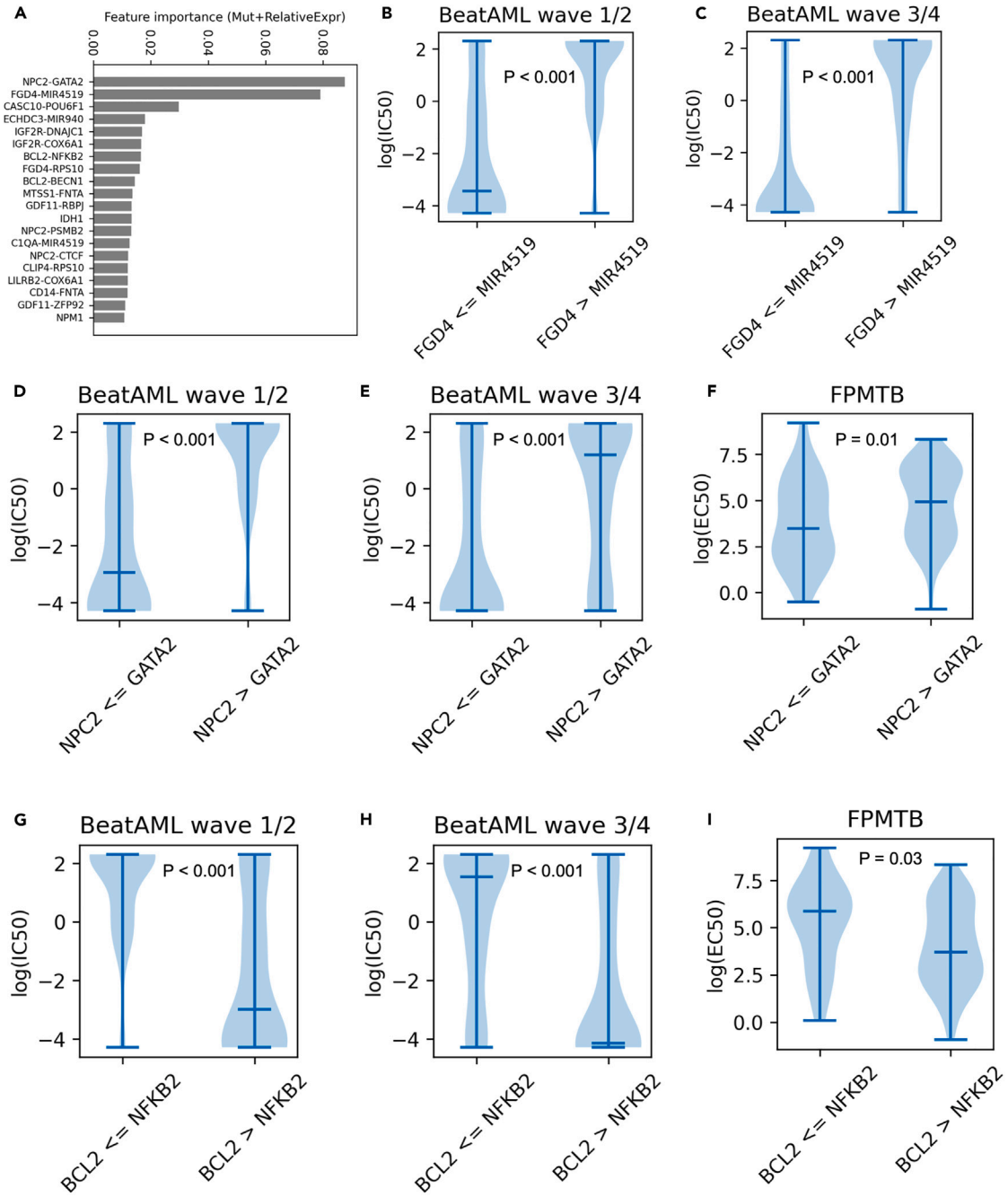


Figure 3. Top predictive features for venetoclax

(A) Features that are predictive to venetoclax with top importance scores in the XGBoost model.

(B and C) Violin plot of log-transformed IC50 values in the two groups of samples defined by the relative gene expression of FGD4 and MIR4519 in Beat AML wave 1/2 and wave 3/4. Rank-sum test was used to test the significance of difference between the two groups.

(D–F) Violin plot of log-transformed IC50 values in the two groups of samples defined by the relative gene expression of NPC2 and GATA2 in the Beat AML wave 1/2, Beat AML wave 3/4 and, FPMTB dataset.

(G–I) Violin plot of log-transformed IC50 or EC50 in the two groups defined by the relative gene expression of BCL2 and NFKB2 in the Beat AML wave 1/2, Beat AML wave 3/4, and FPMTB dataset.

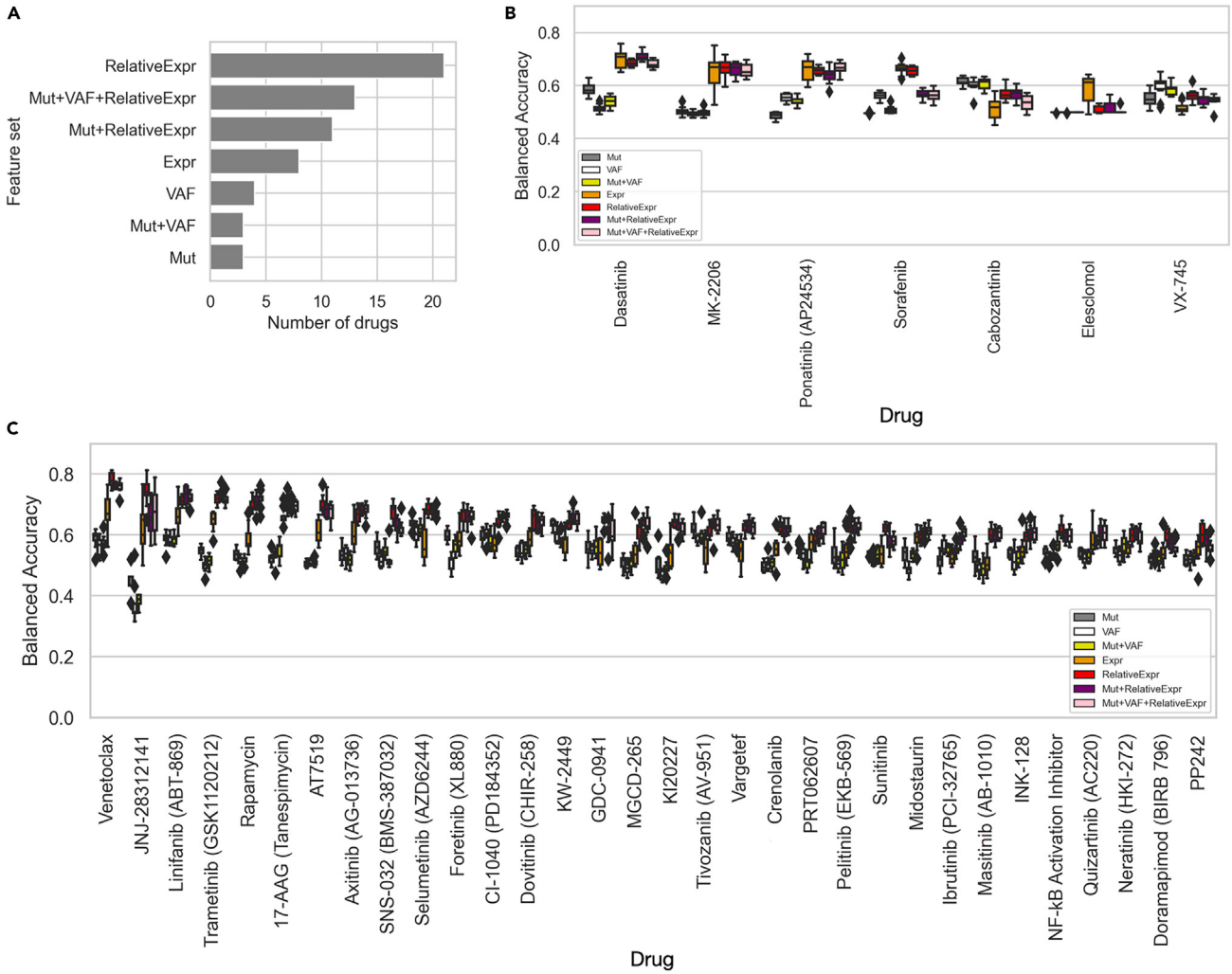


Figure 4. Prediction accuracy of models using different feature sets

(A) The number of drugs with XGBoost classifier models that show highest balanced accuracy in the independent test dataset (Beat AML wave 3/4) with each feature set.

(B) Balanced accuracy score in the independent test dataset Beat AML wave 3/4 for drugs with highest model accuracy using feature sets of Mut, VAF, Mut+VAF, or Expr.

(C) Balanced accuracy score in the independent test set Beat AML wave 3/4 for drugs showing highest model accuracy using feature sets with relative gene expression. The drugs shown in (B) and (C) are with median balanced model accuracy score grater than 0.6 using at least one feature set. The models are based on selected IC50 thresholds that show best accuracy in the test dataset.

expression-based feature sets or combined relative gene expression feature set and gene mutations provide higher accuracy in the independent datasets. The pairwise Wilcoxon test reveals a significant difference in model accuracy between those utilizing relative gene expression and those utilizing gene expression alone, as well as a significant difference between the models using relative gene expression and those incorporating gene mutations.

In addition to using classification models to predict drug sensitivity (using an IC50 threshold to binarize the output), we also used regression analysis to predict drug sensitivity (area under the curve, or AUC) using the same sets of features as the classification. We used a number of regression methods, including XGBoost regression, LASSO regression, and kernel ridge regression. We used kernel ridge regression because kernel ridge regression on gene expression data was one of the best-performing methods for predicting drug response in a previous study,[25] and the best performing on gene expression alone. Selecting only the drugs that had an average R-squared value on the independent test set of above 0.1, we found that the best-performing feature sets varied substantially by drug and regression method (Figure S3). For the XGBoost regression models, we observed most drugs perform better based on the feature sets with relative gene expression, which is consistent with the XGBoost classifiers as shown in Figure 4. For venetoclax, 17-AAG (tanespimycin), and axitinib, the relative expression features had the highest performance across all methods. In rapamycin and trametinib, gene expression features on LASSO and kernel ridge

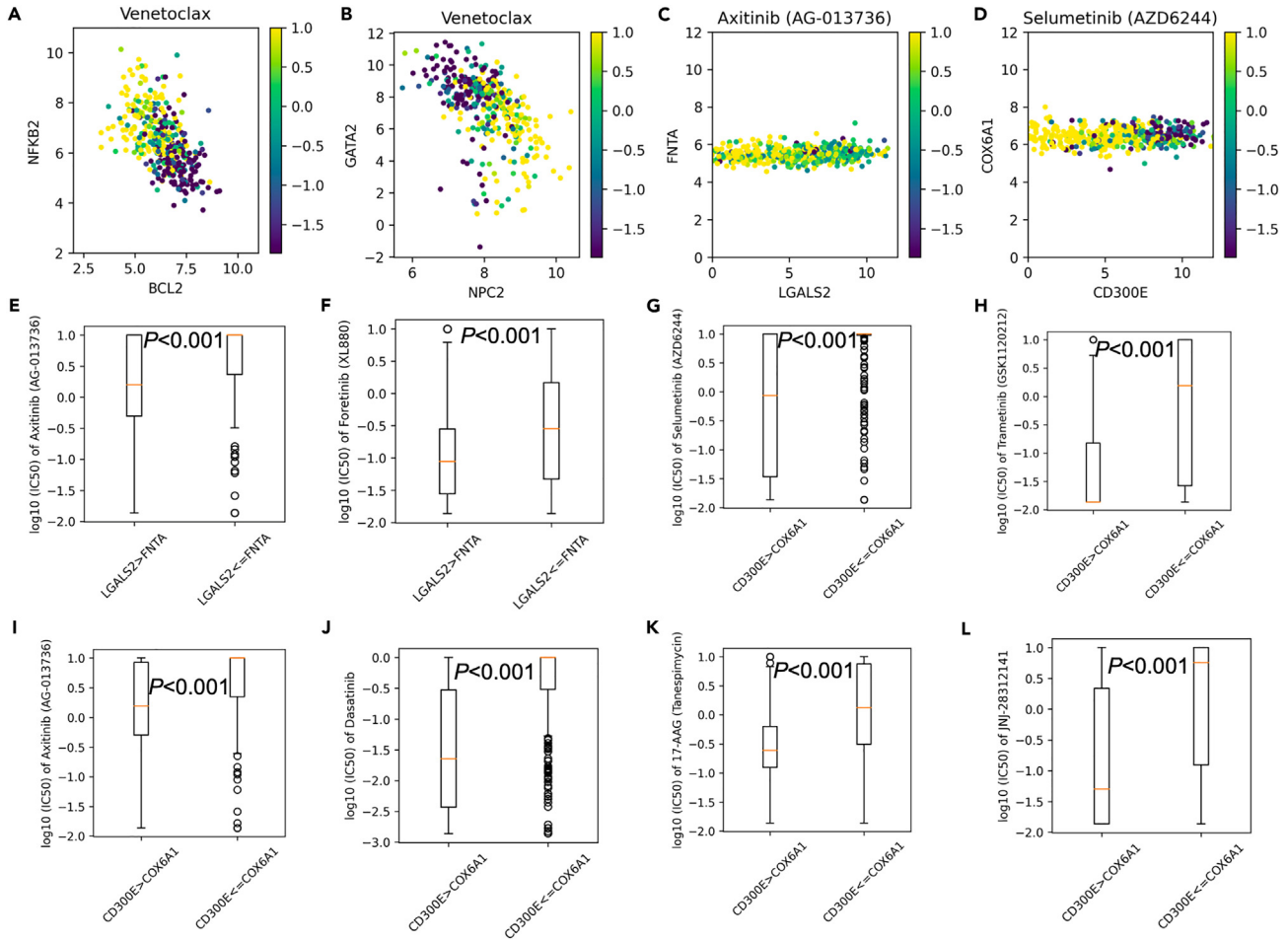


Figure 5. Highlighted predictive features

(A and B) Scatterplot of the gene expression of selected gene pairs that are predictive to drug response of venetoclax. The expression of the gene pairs shows negative correlations. x axis and y axis represent the log-transformed gene-level RNA-seq counts in the Beat AML wave 3/4 dataset. Each dot represents one sample (patient), colored by log(IC50) values. Yellow color indicates higher resistance, darker color indicates higher sensitivity.

(C and D) Scatterplot of the gene expression of selected gene pairs LGALS2-FNTA and CD300E-COX6A1 that are predictive to the response of axitinib and selumetinib respectively.

(E–L) Boxplots of the log10 transformed IC50 values for the samples defined by the relative gene expression features. Rank-Sum test was used to test the difference of each two groups.

regression had the highest performance. For the other drugs, the best-performing feature was inconclusive. In general, R-squared value was low (below 0.2) for most drugs other than venetoclax in the regression models.

We compared the predictive accuracy of our models with existing published models trained on cancer cell line data.[26] The previous prediction models are generally inaccurate when applied to the Beat AML dataset, using the same thresholds described earlier. The balanced accuracy was approximately 0.5 for all drugs that overlapped between the Beat AML dataset and the drugs in the pre-trained model.

Biomarkers associated with drug sensitivity or resistance for different drugs or chemicals in AML

We further extracted features that contribute to the predictive models for the response of different drugs. To explore which features would be associated with sensitivity or resistance to specific drugs, we selected the XGBoost classifiers models based on the feature set of relative gene expression with or without gene mutations or VAF that have predictive balanced accuracy of at least 0.7 in the independent dataset (Beat AML wave 3/4), then extracted the features with feature importance score greater than 0, and ranked the features by the average importance score.

Features with an average importance score greater than 0.01 are shown in Table S4. The relative expression of the gene pairs BCL2-NFKB2 and NPC2-GATA2 are highly predictive of venetoclax sensitivity in the Beat AML dataset (Figures 5A, 5B, and 3D–3I). We also found the relative gene expression between LGALS2 and FNTA are predictive to both axitinib and foretinib drug response. FNTA is a housekeeping gene with low expression variance in the Beat AML dataset (Figure 5C), while the expression of LGALS2 shows a much higher variance. An analysis

from a hierarchical differentiation tree of normal hematopoiesis with AMLs through BloodSpot[27] suggests LGALS2 is highly expressed in monocytes and FNTA does not show strong preference in cell types. This result indicates samples with higher proportions of monocytes may show higher sensitivity to axitinib and foretinib. We also found that the relative expression levels of CD300E and COX6A1 are predictive of sensitivity to multiple drugs, including selumetinib, trametinib, axitinib, dasatinib, tanespimycin, and JNJ-28312141 (Table S4; Figures 5D and 5G–5L). The expression of CD300E shows high variance and the expression of COX6A1 shows low variance (Figure 5D). Rank Sum tests on the log10-transformed IC50 values between the two groups, defined by the relative gene expression of CD300E and COX6A1 (CD300E > COX6A1 or CD300E $\leq$ COX6A1), show a significant difference in drug response (Figures 5G–5L). More specifically, higher CD300E expression is associated with higher sensitivity with these drugs. CD300E shows high correlation with CD14, which is also highly expressed with monotypes as suggested by BloodSpot.[27] These results suggest drugs such as selumetinib, trametinib, axitinib, dasatinib, tanespimycin, and JNJ-28312141 may perform better in monocytic AML; this is consistent with the observation that RAS mutation was correlated with monocyte-like state in the Beat AML dataset.[5] For the RAF/MEK/ERK inhibitor selumetinib (AZD6244), we found that NRAS mutation is among the top predictors for response prediction. More specifically, we have observed the VAF of gene mutation at NRAS.Q61 is among the top important features to predict drug response of selumetinib. This observation is consistent with our previous study, which suggests that RAS mutations are predictors for the MEK inhibitors such as selumetinib and trametinib,[8] using an independent cohort. With improved model accuracy by considering relative gene expression values guided by KGs, our results can help to identify robust biomarkers for drug response prediction beyond the features of gene mutations and variant allele frequencies.

DISCUSSION

Drug response prediction is a critical component of precision medicine, particularly for diseases that exhibit varied responses to therapeutic agents. Predictive signatures for drug response could provide guidance for clinical applications.

Our approach demonstrates the use of KGs for feature engineering in drug response prediction. By generating engineered features from the KG, we enhanced *ex vivo* drug sensitivity prediction. The engineered binarized gene expression features derived from individual patients help mitigate batch effects, reduce feature space, and produce highly accurate drug response prediction models. Using this approach, we built machine learning models for venetoclax that outperform the gene expression-based models especially for independent datasets. Our results highlight features such as the relative expression between NPC2 and GATA2 and the relative expression between BCL2 and NFKB2 are highly predictive of venetoclax drug response. This relationship has been confirmed using Beat AML wave 1/2, Beat AML wave 3/4 and FPMTB datasets. Applying this approach to other drugs, we also identified features that are predictive of *ex vivo* drug response. These features identified in our study could provide guidance for further development of biomarker assays to guide drug selection. Before applying to clinical practice, further validation using the clinical response at the patient level is still essential.

KGs encode existing knowledge or known relationships in a machine-readable format, allowing us to leverage them for downstream applications.[18] The KG can guide the selection of features and feature engineering to feed machine learning models. The selection of features will be depending on the context, which needs careful selection. For this method, two types of inputs are required: the selection of features based on existing knowledge and the selection of data for training the model. The selection of the features will be based on the biological hypothesis. For example, in our model, we hypothesize that relative gene expression between the transcription factors and down-stream target genes shapes the transcriptional states of the cells or sample, which may respond differently to certain drugs. The selection of gene expression data can reflect the transcriptional regulations that match the biological rationale. Our result demonstrates that KG-guided feature engineering provides a useful way to improve machine model accuracy, and provide predictive features for drug response prediction. With a consortium wide effort for KG development,[17] we expect the KG-facilitated machine learning to be applicable in many other use cases.

Our approach connects well-established biological KGs and empirical findings from multi-omics data and drug screening data from large AML datasets. We have designed two-layer models to select the drug sensitivity or resistance-associated features from the graph, and we have built prediction models using newly designed testable features of gene pairs. The model provides higher accuracy than the traditional machine learning models with only gene expression values as input, and provides a smaller list of features for testing. Furthermore, narrowing down the list of features can facilitate the design and development of assays for drug sensitivity testing.

Our models defined predictors of *ex vivo* drug sensitivity. For further clinical application, the ultimate aim will be to predict how patients responded to a given treatment. Beyond using drug sensitivity testing as a way of clinical diagnostics, our proposal would be that no diagnostic is sufficient as a standalone platform and that fine-tuning of omic analyses, such as with these KG-based gene expression ratios, would be useful to supplement all other clinical diagnostic tools, inclusive of CBCs, cytogenetics, morphological/cytochemical analysis, flow cytometry, DNA sequencing, and *ex vivo* drug sensitivity testing.

Several factors may affect the accuracy of drug response prediction models. Determining the sensitivity-resistance group by cutoff of IC50 values could be arbitrary due to variations in drug efficacy based on their pharmacological features. This complexity led us to establish different thresholds for the same drug. Our objective is to develop models that enhance accuracy, so we selected the best performing models. Even when we aggregate prediction accuracy across all the different thresholds, the models based on relative gene expression features still show a higher overall accuracy than other feature sets. Beyond the definition of the objective, the selection of models, features, and data are also key factors that contribute to model accuracy. However, the selection of features and data could be still be biased, and it may work well for some drugs but not be so effective for others drugs. Based on the mechanism of action for each drug, different features may be required for improving the prediction accuracy of drug response prediction. For example, for FLT3 inhibitors, the mutation status of FLT3 is important to take into consideration. For venetoclax, the relative gene expression from related genes provides informative features for

predictions. Other datasets and KGs such as protein expression and cell signaling profiles could be considered. In this paper we emphasized on the selection of molecular measurement features for better drug response prediction based on their molecular measurement, such as mutations, variant allele frequencies, and gene expression. Other features could be considered such as the proportion of cell types, and other clinical features could be useful for improving the performance of the drug response prediction models.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Guangrong Qin (guangrong.qin@isbscience.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This manuscript analyzes existing, publicly available data. We used public data resources from Beat AML project and FMTMB.[5–7] Knowledge sources including transcriptional factor—target pairs and microRNA—target pairs are derived from literatures and public knowledge resources.[21–23] All results generated in this study are shared in the supplemental data.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number U01CA217883 and U01CA282109 (C.J.K.) and R01CA270210 (I.S.). J.W.T. received funding from the Drug Sensitivity and Resistance Network, National Institutes of Health (NIH), National Cancer Institute (NCI) grant U54CA224019, the Cancer Target Discovery and Development Network grant U01CA217862, NCI award R01CA262758, the Mark Foundation For Cancer Research, and the Silver Family Foundation.

AUTHOR CONTRIBUTIONS

G.Q. led the computational analysis, playing a pivotal role in conceptual development, implementation, and the drafting and revision of the manuscript. Y.Z. contributed by implementing regression models and participating in manuscript drafting. J.W.T. was responsible for Beat AML data acquisition and manuscript revision. C.J.K. contributed to funding acquisition for the project and manuscript revision. I.S. played a key role in supervising the project, funding acquisition, and manuscript revision.

DECLARATION OF INTERESTS

J.W.T. has received research support from Acerta, Agios, Aptose, Array, AstraZeneca, Constellation, Genentech, Gilead, Incyte, Janssen, Kronos, Meryx, Petra, Schrodinger, Seattle Genetics, Syros, Takeda, and Tolero and serves on the advisory board for Recludix Pharma.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author(s) used Llama 3 in order to check grammar. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Data cumulation and processing
  - Graph construction
  - Feature engineering
  - XGboost classifiers to predict drug sensitivity or resistance
  - Selection of biomarkers for the prediction models
  - Regression models to predict the area under the curve
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Comparison of drug response defined by groups of samples based on predictive features
  - Statistical comparisons of predictive performance for different feature sets

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110755.

Received: November 20, 2023
Revised: May 4, 2024
Accepted: August 14, 2024
Published: August 20, 2024

REFERENCES

1. Dohner, H., Wei, A.H., Appelbaum, F.R., Craddock, C., DiNardo, C.D., Dombret, H., Ebert, B.L., Fenaux, P., Godley, L.A., Hasserjian, R.P., et al. (2022). Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood *140*, 1345–1377. https://doi.org/10.1182/blood.2022016867.
2. Pulte, D., Jansen, L., Castro, F.A., Krilaviciute, A., Katalinic, A., Barnes, B., Ressing, M., Holleczek, B., Luttmann, S., and Brenner, H.; GEKID Cancer Survival Working Group (2016). Survival in patients with acute myeloblastic leukemia in Germany and the United States: Major differences in survival in young adults. Int. J. Cancer *139*, 1289–1296. https://doi.org/10.1002/ijc.30186.
3. Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med. *374*, 2209–2221. https://doi.org/10.1056/NEJMoa1516192.
4. Patel, J.P., Gönen, M., Figueroa, M.E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O., et al. (2012). Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N. Engl. J. Med. *366*, 1079–1089. https://doi.org/10.1056/NEJMoa1112304.
5. Bottomly, D., Long, N., Schultz, A.R., Kurtz, S.E., Tognon, C.E., Johnson, K., Abel, M., Agarwal, A., Avaylon, S., Benton, E., et al. (2022). Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. Cancer Cell *40*, 850–864.e9. https://doi.org/10.1016/j.ccell.2022.07.002.
6. Tyner, J.W., Tognon, C.E., Bottomly, D., Wilmot, B., Kurtz, S.E., Savage, S.L., Long, N., Schultz, A.R., Traer, E., Abel, M., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. Nature *562*, 526–531. https://doi.org/10.1038/s41586-018-0623-z.
7. Malani, D., Kumar, A., Brück, O., Kontro, M., Yadav, B., Hellesøy, M., Kuusanmäki, H., Dufva, O., Kankainen, M., Eldfors, S., et al. (2022). Implementing a Functional Precision Medicine Tumor Board for Acute Myeloid Leukemia. Cancer Discov. *12*, 388–401. https://doi.org/10.1158/2159-8290.CD-21-0410.
8. Qin, G., Dai, J., Chien, S., Martins, T.J., Loera, B., Nguyen, Q.H., Oakes, M.L., Tercan, B., Aguilar, B., Hagen, L., et al. (2024). Mutation Patterns Predict Drug Sensitivity in Acute Myeloid Leukemia. Clin. Cancer Res. *30*, 2659–2671. https://doi.org/10.1158/1078-0432.CCR-23-1674.
9. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740–754. https://doi.org/10.1016/j.cell.2016.06.017.
10. Knijnenburg, T.A., Klau, G.W., Iorio, F., Garnett, M.J., McDermott, U., Shmulevich, I., and Wessels, L.F.A. (2016). Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. Sci. Rep. *6*, 36812. https://doi.org/10.1038/srep36812.
11. Xia, F., Allen, J., Balaprakash, P., Brettin, T., Garcia-Cardona, C., Clyde, A., Cohn, J., Doroshow, J., Duan, X., Dubinkina, V., et al. (2022). A cross-study analysis of drug response prediction in cancer cell lines. Briefings Bioinf. *23*, bbab356. https://doi.org/10.1093/bib/bbab356.
12. Drusbosky, L.M., Vidva, R., Gera, S., Lakshminarayana, A.V., Shyamasundar, V.P., Agrawal, A.K., Talawdekar, A., Abbasi, T., Vali, S., Tognon, C.E., et al. (2019). Predicting response to BET inhibitors using computational modeling: A BEAT AML project study. Leuk. Res. *77*, 42–50. https://doi.org/10.1016/j.leukres.2018.11.010.
13. Tang, Y.C., and Gottlieb, A. (2021). Explainable drug sensitivity prediction through cancer pathway enrichment. Sci. Rep. *11*, 3128. https://doi.org/10.1038/s41598-021-82612-7.
14. Andersen, A.N., Brodersen, A.M., Ayuda-Durán, P., Piechaczyk, L., Tadele, D.S., Baken, L., Fredriksen, J., Stoksflod, M., Lenartova, A., Fløisand, Y., et al. (2023). Clinical forecasting of acute myeloid leukemia using ex vivo drug-sensitivity profiling. Cell Rep. Methods *3*, 100654. https://doi.org/10.1016/j.crmeth.2023.100654.
15. Yadav, B., Pemovska, T., Szwajda, A., Kulesskiy, E., Kontro, M., Karjalainen, R., Majumder, M.M., Malani, D., Murumägi, A., Knowles, J., et al. (2014). Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. Sci. Rep. *4*, 5193. https://doi.org/10.1038/srep05193.
16. Kuusanmaki, H., Kytola, S., Vanttinen, I., Ruokoranta, T., Ranta, A., Huuhtanen, J., Suvela, M., Parsons, A., Holopainen, A., Partanen, A., et al. (2023). Ex vivo venetoclax sensitivity testing predicts treatment response in acute myeloid leukemia. Haematologica *108*, 1768–1781. https://doi.org/10.3324/haematol.2022.281692.
17. Fecho, K., Thessen, A.E., Baranzini, S.E., Bizon, C., Hadlock, J.J., Huang, S., Roper, R.T., Southall, N., Ta, C., Watkins, P.B., et al. (2022). Progress toward a universal biomedical data translator. Clin. Transl. Sci. *15*, 1838–1847. https://doi.org/10.1111/cts.13301.
18. Unni, D.R., Moxon, S.A.T., Bada, M., Brush, M., Bruskiewich, R., Caufield, J.H., Clemons, P.A., Dancik, V., Dumontier, M., Fecho, K., et al. (2022). Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. Clin. Transl. Sci. *15*, 1848–1855. https://doi.org/10.1111/cts.13302.
19. Morris, J.H., Soman, K., Akbas, R.E., Zhou, X., Smith, B., Meng, E.C., Huang, C.C., Cerono, G., Schenk, G., Rizk-Jackson, A., et al. (2023). The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information. Bioinformatics *39*, btad080. https://doi.org/10.1093/bioinformatics/btad080.
20. Cancer Genome Atlas Research Network, Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A.G., Hoadley, K., Triche, T.J., Jr., Laird, P.W., et al. (2013). Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074. https://doi.org/10.1056/NEJMoa1301689.
21. Huang, H.Y., Lin, Y.C.D., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J., Zuo, H., et al. (2022). miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. *50*, D222–D230. https://doi.org/10.1093/nar/gkab1079.
22. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260.
23. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375. https://doi.org/10.1101/gr.240663.118.
24. Menendez-Gonzalez, J.B., Vukovic, M., Abdelfattah, A., Saleh, L., Almotiri, A., Thomas, L.A., Agirre-Lizaso, A., Azevedo, A., Menezes, A.C., Tornillo, G., et al. (2019). Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia. Stem Cell Rep. *13*, 291–306. https://doi.org/10.1016/j.stemcr.2019.07.005.
25. Chen, J., and Zhang, L. (2021). A survey and systematic assessment of computational methods for drug response prediction. Briefings Bioinf. *22*, 232–246. https://doi.org/10.1093/bib/bbz164.
26. Chiu, Y.C., Chen, H.I.H., Zhang, T., Zhang, S., Gorthi, A., Wang, L.J., Huang, Y., and Chen, Y. (2019). Predicting drug response of tumors from integrated genomic profiles by deep neural networks. BMC Med. Genom. *12*, 18. https://doi.org/10.1186/s12920-018-0460-9.
27. Gislason, M.H., Demircan, G.S., Prachar, M., Furtwangler, B., Schwaller, J., Schoof, E.M., Porse, B.T., Rapin, N., and Bagger, F.O. (2024). BloodSpot 3.0: a database of gene and protein expression data in normal and malignant haematopoiesis. Nucleic Acids Res. *52*, D1138–D1142. https://doi.org/10.1093/nar/gkad993.
28. Hsiao, L.L., Dangond, F., Yoshida, T., Hong, R., Jensen, R.V., Misra, J., Dillon, W., Lee, K.F., Clark, K.E., Haverty, P., et al. (2001). A compendium of gene expression in normal human tissues. Physiol. Genom. *7*, 97–104. https://doi.org/10.1152/physiolgenomics.00040.2001.
29. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
30. Qin, G., Mallik, S., Mitra, R., Li, A., Jia, P., Eischen, C.M., and Zhao, Z. (2020). MicroRNA and transcription factor co-regulatory networks and subtype classification of seminoma and non-seminoma in testicular germ cell tumors. Sci. Rep. *10*, 852. https://doi.org/10.1038/s41598-020-57834-w.
31. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.
32. Tyner, J.W., Yang, W.F., Bankhead, A., 3rd, Fan, G., Fletcher, L.B., Bryant, J., Glover, J.M., Chang, B.H., Spurgeon, S.E., Fleming, W.H., et al. (2013). Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening. Cancer Res. *73*, 285–296. https://doi.org/10.1158/0008-5472.CAN-12-1906.

33. Kurtz, S.E., Eide, C.A., Kaempf, A., Khanna, V., Savage, S.L., Rofelty, A., English, I., Ho, H., Pandya, R., Bolosky, W.J., et al. (2017). Molecularly targeted drug combinations demonstrate selective effectiveness for myeloid- and lymphoid-derived hematologic malignancies. Proc. Natl. Acad. Sci. USA *114*, E7554–E7563. https://doi.org/10.1073/pnas.1703094114.

34. Kurtz, S.E., Eide, C.A., Kaempf, A., Mori, M., Tognon, C.E., Borate, U., Druker, B.J., and Tyner, J.W. (2018). Dual inhibition of JAK1/2 kinases and BCL2: a promising therapeutic strategy for acute myeloid leukemia. Leukemia *32*, 2025–2028. https://doi.org/10.1038/s41375-018-0225-7.

35. Kurtz, S.E., Eide, C.A., Kaempf, A., Long, N., Bottomly, D., Nikolova, O., Druker, B.J., McWeeney, S.K., Chang, B.H., Tyner, J.W., and Agarwal, A. (2022). Associating drug sensitivity with differentiation status identifies effective combinations for acute myeloid leukemia. Blood Adv. *6*, 3062–3067. https://doi.org/10.1182/bloodadvances.2021006307.

36. Edwards, D.K., 5th, Watanabe-Smith, K., Rofelty, A., Damnernsawad, A., Laderas, T., Lamble, A., Lind, E.F., Kaempf, A., Mori, M., Rosenberg, M., et al. (2019). CSF1R inhibitors exhibit antitumor activity in acute myeloid leukemia by blocking paracrine signals from support cells. Blood *133*, 588–599. https://doi.org/10.1182/blood-2018-03-838946.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Code	This manuscript	https://doi.org/10.5281/zenodo.12773035

METHOD DETAILS

Data cumulation and processing

Beat AML wave 1/2: Mutation data, RNASeq (RPKM) data, drug IC50 values, and drug response data for area under the curve (AUC) were downloaded from the Supplementary table from Beat AML study.[6] The consequences of the selected variants include frameshift variants, inframe deletions/insertions, internal tandem duplications, missense variants, protein altering variants, splice acceptor/donor variants, start loss, and stop gain/loss. Features extracted from the mutation data include the following: 1) whether one gene has variants with these selected functional alteration consequences; and 2) Variant allele frequency (VAF) for each variant. Two variant calling methods have been used in the Beat AML wave 1/2: Mutect and varscan. The VAF from the two approaches are similar. To assign one VAF for individual variants, we used the following procedure. If one variant was called from both approaches, we used the VAF from the Mutect call. If one variant was called from only one approach, we use the VAF from that approach. 3333 genes with selected variants were included in Beat AML wave 1/2. Among them, we selected genes with mutation in at least 3% samples as features, which includes 22 highly frequently mutated genes (Feature set 1 in Table 1). We used the VAF for variants that mutated in at least 1% of the samples for the model, and grouped them into 18 variant groups Feature set 2 in Table 1. We also considered the expression of genes with expression level (log(RPKM)) greater than 0 in at least 50% as features for the prediction models. We consider it as feature set 3. We used the Beat AML wave 3/4 dataset as an independent testing set,[6] and the processing of Beat AML wave 3/4 is the same as wave 1/2.

Graph construction

Selection of housekeeping genes to define relative gene expression features

Housekeeping genes were downloaded from a previous publication.[28] Uniport identifiers were converted to gene symbols. We initially selected genes that are expressed in the blood samples from GTEx dataset[29] and AML samples from Beat AML wave 1/2, resulting in a set of 205 housekeeping genes. To further narrow down the number of housekeeping genes, we then selected genes with expression variance smaller than 1 in both the GTEx dataset and Beat AML dataset [Figures S1A and S1B]. Among the low variance housekeeping genes, we selected representative genes for different expression levels using an interval of 1 for the mean log(RPKM) values [Figures S1C and S1D]. The final selected housekeeping genes are RPS10, FNTA, COX6A1, BECN1, SF3B2, PSMB2, AUP1, SRP14, HNRNPK, CCNI, RHOA, PABPC1, RPS11, TPT1, and FTL [Figure S1D]. These selected genes were then used for relative gene expression feature construction.

Gene expression to drug sensitivity KG construction

LASSO regression models were used to find the predictive features for gene expression to drug associations, offering a robust way to select gene expression features that are predictive for drug response. The models were trained using the Sklearn python package. The area under the curve (AUC) for each drug was predicted using gene expression data from the Beat AML wave 1/2. Genes with an expression value (log(RPKM)) greater than 0 in at least 50% of the samples were used as features for regression models. Each feature was normalized using z-score transformation ((Expression - mean)/STD), and the drug response data AUC was also normalized using z-score transformation before modeling. Features with non-zero regression coefficients were selected for knowledge graph construction. This step provides a graph that connects drugs and gene signatures which are potentially predictable to drug response.

Gene expression vs transcriptional regulatory KG construction

Transcription factors (TFs) and miRNAs are two important types of gene regulators.[30] We used microRNA and TFs annotated for each gene as potential features, and selected regulatory factors for each target gene in the context of AML using the Beat AML dataset wave 1/2. Common miRNA and target genes were extracted from miRTarBase[21] and MSigDB(v7.4).[31] Total of 770,183 microRNA and target pairs are included. When filtering the microRNA using the measured RNASeq data, only 55 microRNAs are overlapped and used in this study [Figure S4]. We selected TF-target gene pairs from a public resource, which presented an integrated TFs and target gene interactions with different types of evidence and confidences.[23] LASSO regression models were used to select the regulatory factors for each target gene using the Beat AML wave 1/2 dataset. Alpha is set to 0.1 to avoid overfitting of the model, and provide most relevant features to the expression of target genes. Before fitting to the model, both gene expression values and AUC values are z-score normalized. miRNA and TFs with non-zero coefficient were selected for knowledge graph construction. This step provides the regulatory KGs that connect the gene signatures and their upstream regulators (Table S2).

Feature engineering

The graphs constructed above provide the biological rationale of selecting and constructing features for drug sensitivity prediction, as it provides paths that connect miRNA/TFs to the target genes, then to the association of drug sensitivity. Two types of relative expression features were derived from the gene expression profile using the graph structure. The first one is the relative gene expression between the gene signatures (Gs) that are associated with sensitivity or resistance of drug response derived from the LASSO regression models (the first neighbors of a drug) and their regulators (Gr, the second neighbors of a drug). If Gs > Gr, we set the new feature Gs∼Gr = 1, otherwise 0. Only regulators which suggest negative regulatory effects are selected for analysis. The second relative gene expression feature is the relative gene expression between drug response associated gene signatures (Gs) and the house keep genes (Gh). If Gs > Gh, we set Gs∼Gh = 1, otherwise 0. The engineered features provide an opportunity to overcome batch effects since we expect the relative gene expression will be more robust than the absolute value of gene expression which may be altered by the measuring platforms or batches.

XGboost classifiers to predict drug sensitivity or resistance

XGboost classifier was used to classify whether a patient will show sensitivity or resistance to a drug defined by the IC50 threshold. XGBoost is an ensemble learning approach that combines the predictions of multiple base estimators to improve the model's performance. It also provides feature importance for further exploration. We used the Beat AML wave 1/2 dataset as the training set (90%), and validation set (10%),[6] and the Beat AML wave 3/4 dataset was used as the test dataset.[5] We first labeled the samples into the sensitive group if the IC50 is smaller than a threshold, and resistant group otherwise. The selection of IC50 might be arbitrary since the tolerance of drug dosage might be different in patients. We select several cutoff points including the 25th percentile, 50th percentile and 75th percentile if they are smaller than 10 μM. If the selected cutoff points are smaller than 1 μM, we also add 1 μM as a threshold. We then classify samples to the sensitive group or resistant group according to the selected threshold. The features were then generated as described above and are listed in Table 1. We used a stratified approach to split the samples to make sure both the training set and validation set have equitant proportions of sensitive samples and resistant samples. Ten-fold cross validation was used to validate the prediction model. We then tested the accuracy in the validation set and the independent testing set using balanced accuracy scores and F1 scores. The balanced accuracy is measured using the function of (sensitivity + specificity)/2. F1 score is calculated using the f1_score in the sklearn.metrics library. The F1 score and balanced accuracy metrics were used because they are standard metrics for evaluating classification quality when there is a class imbalance, as in our case.

Selection of biomarkers for the prediction models

Each XGBoost model provides a list of features with importance scores. From the models for each drug, we then selected the models which show balanced accuracy (ACC) in the validation set greater than 0.7, and aggregated the features by average the feature importance from the high-accuracy models. Features with average feature importance greater than 0.01 were selected for further visualization and analysis.

Regression models to predict the area under the curve

Regression models for predicting AUC (area under the dosage response curve) were trained using Beat AML wave 1/2 data and tested using Beat AML wave 3/4 data. We used a variety of models, including XGBoost regression, the LassoCV model (LASSO regression with internal cross-validation to determine the regularization parameter) from the scikit-learn python library, and kernel ridge regression (KernelRidge from scikit-learn), the last of which has been reported to be one of the best-performing methods for predicting drug response in a previous publication.[25] We used AUC as the independent value for the prediction task. AUC is a standard approach for evaluating dose response curve metrics. There are a variety of ways of handling AUC data such as drug sensitivity score (DSS).[15] We have used our AUC which has been used in numerous prior studies.[6,32–36] With 10-fold cross-validation, we trained ten different models on 90-10 splits of the Wave 1/2 patient data, and tested each model on the Wave 3/4 data. The data were processed and normalized in the same way as the classification predictions, with the same feature sets. To compare results across different feature sets and drugs, we used the R^2 score (coefficient of determination) for the observed versus predicted AUC, calculated with the scikit-learn metrics.r2_score function.

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison of drug response defined by groups of samples based on predictive features

Rank Sum test from the scipy.stats python library was used to test the significance of difference between the gene expression of one gene between the drug resistant group and sensitive group. Spearman correlation from scipy.stats library was used to calculate the correlation coefficient and P-value between the gene expression and drug IC50 values. The resulting P-value was adjusted using multitest.multipletest from the statsmodels.stats python library. Benjamini/Hochberg method used for adjustment of P-values. We considered the adjusted P-value smaller than 0.05 from multiple tests as a threshold value of significance. Pearson correlations were calculated between the expression of different genes using the Pandas python package.

Statistical comparisons of predictive performance for different feature sets

In order to compare the relative performance of different feature sets, we used the Wilcoxon signed-rank test to compare the balanced accuracies across different feature sets. For every drug, we calculated a p-value using the Wilcoxon signed-rank test for every pair of feature sets, where each sample is the balanced accuracy on the independent test set of a cross-validation run. This gives us 6 distinct p-values per drug. We then performed an FDR correction using the Benjamini-Hochberg procedure.