

## ARTICLE OPEN



# Prediction model for drug response of acute myeloid leukemia patients

Quang Thinh Trac<sup>1</sup>, Yudi Pawitan<sup>1</sup>, Tian Mou<sup>2</sup>, Tom Erkers<sup>3</sup>, Päivi Östling<sup>3,4</sup>, Anna Bohlin<sup>5</sup>, Albin Österroos<sup>6</sup>, Mattias Vesterlund<sup>3</sup>, Rozbeh Jafari<sup>3</sup>, Ioannis Siavelis<sup>3</sup>, Helena Bäckvall<sup>3</sup>, Santeri Kiviluoto<sup>3</sup>, Lukas M. Orre<sup>3</sup>, Mattias Rantalainen<sup>1</sup>, Janne Lehtiö<sup>3</sup>, Sören Lehmann<sup>5,6</sup>, Olli Kallioniemi<sup>3,4</sup> and Trung Nghia Vu<sup>1✉</sup>

Despite some encouraging successes, predicting the therapy response of acute myeloid leukemia (AML) patients remains highly challenging due to tumor heterogeneity. Here we aim to develop and validate MDREAM, a robust ensemble-based prediction model for drug response in AML based on an integration of omics data, including mutations and gene expression, and large-scale drug testing. Briefly, MDREAM is first trained in the BeatAML cohort ( $n = 278$ ), and then validated in the BeatAML ( $n = 183$ ) and two external cohorts, including a Swedish AML cohort ( $n = 45$ ) and a relapsed/refractory acute leukemia cohort ( $n = 12$ ). The final prediction is based on 122 ensemble models, each corresponding to a drug. A confidence score metric is used to convey the uncertainty of predictions; among predictions with a confidence score  $>0.75$ , the validated proportion of good responders is 77%. The Spearman correlations between the predicted and the observed drug response are 0.68 (95% CI: [0.64, 0.68]) in the BeatAML validation set,  $-0.49$  (95% CI:  $[-0.53, -0.44]$ ) in the Swedish cohort and 0.59 (95% CI: [0.51, 0.67]) in the relapsed/refractory cohort. A web-based implementation of MDREAM is publicly available at <https://www.meb.ki.se/shiny/truvu/MDREAM/>.

npj Precision Oncology (2023)7:32; <https://doi.org/10.1038/s41698-023-00374-z>

## INTRODUCTION

Acute myeloid leukemia (AML) is an aggressive blood cancer with high genetic heterogeneity and varying responses to therapy. Traditionally, targeted cancer therapies, including AML, are designed based on genetic biomarkers. In the classification of the World Health Organization (WHO) in 2008<sup>1</sup>, genomic features such as fusion genes, including PML-RARA, RUNX1-RUNX1T1, DEK-NUP214, CBFB-MYH11, and mutations of CEBPA and NPM1, determined molecular subtypes in adult AML. A more recent study<sup>2</sup> proposed 11 molecular subtypes, each associated with specific diagnostic and clinical features. The first targeted therapy of AML used all-trans retinoic acid (ATRA) targeting to PML-RARA fusion of Acute Promyelocytic Leukemia (APL), a subtype of AML characterized by the translocation of the retinoic acid receptor  $\alpha$  (RARA) on chromosome 17 and promyelocytic leukemia (PML) gene on chromosome 15<sup>3</sup>. Recently, Midostaurin, a fms-like tyrosine kinase 3 (FLT3) inhibitor, was approved for use in AML with FLT3 mutations in 2017<sup>4</sup>.

Despite some encouraging successes, predicting therapy responses of AML patients remains highly challenging<sup>5</sup>. The patterns of the mutations and genetic abnormalities within and across patients are highly complex, and only a few targeted therapies have been developed for AML. The genetic complexity can also reduce the efficacy of a given therapy or make the patient stratification inaccurate. For example,  $>40\%$  of AML patients with FLT3+ fail to respond to Midostaurin, but  $>30\%$  of FLT3- cases potentially respond to the drug<sup>6</sup>. In clinical practice, because of the complexity of AML and its potential treatments, the selection of therapy is often subjective, decided based on the experience or intuition of the clinician<sup>7</sup>. These suggest there is still a need for an effective computational prediction model for the drug response of individual AML patients.

To date, there are a few studies on the methodologies for drug-response prediction using omics data for AML. In 2016, Ammad-uddin and colleagues developed a kernelized Bayesian matrix factorization method<sup>8</sup>, but it was validated in a small set of AML cell lines ( $n = 6$ ). Similarly, Gerdes et al.<sup>9</sup> introduced a machine-learning approach to predict and rank anti-cancer drugs on 26 AML cell lines using proteomics data. Lee et al. presented the MERGE score<sup>10</sup> to identify gene expression markers and predict drug sensitivity of AML cases. However, it was also trained and validated in a limited number of AML cases ( $n = 34$  AML patients and  $n = 14$  AML cell lines), which cannot capture the wide AML heterogeneities.

In this study, we develop MDREAM (Monotherapy Drug Response prediction for AML) using integrated omics data. MDREAM is trained and first validated using the BeatAML cohort ( $n = 278$  for training and 183 for validation)<sup>11</sup>. The prediction is based on 122 ensemble models, each corresponding to a drug. The results show MDREAM is well validated in the BeatAML validation set and in the external validation set using an in-house cohort (Clinseq) including 45 Swedish AML patients, and also a public acute leukemia dataset (LeeAML) ( $n = 12$ )<sup>10</sup>. For the prediction of each patient, we introduce a confidence score to measure the prediction uncertainty of individual drugs, guiding the practical use of the prediction for the patient. MDREAM is implemented in a publicly available web application at <https://www.meb.ki.se/shiny/truvu/MDREAM/>.

## RESULTS

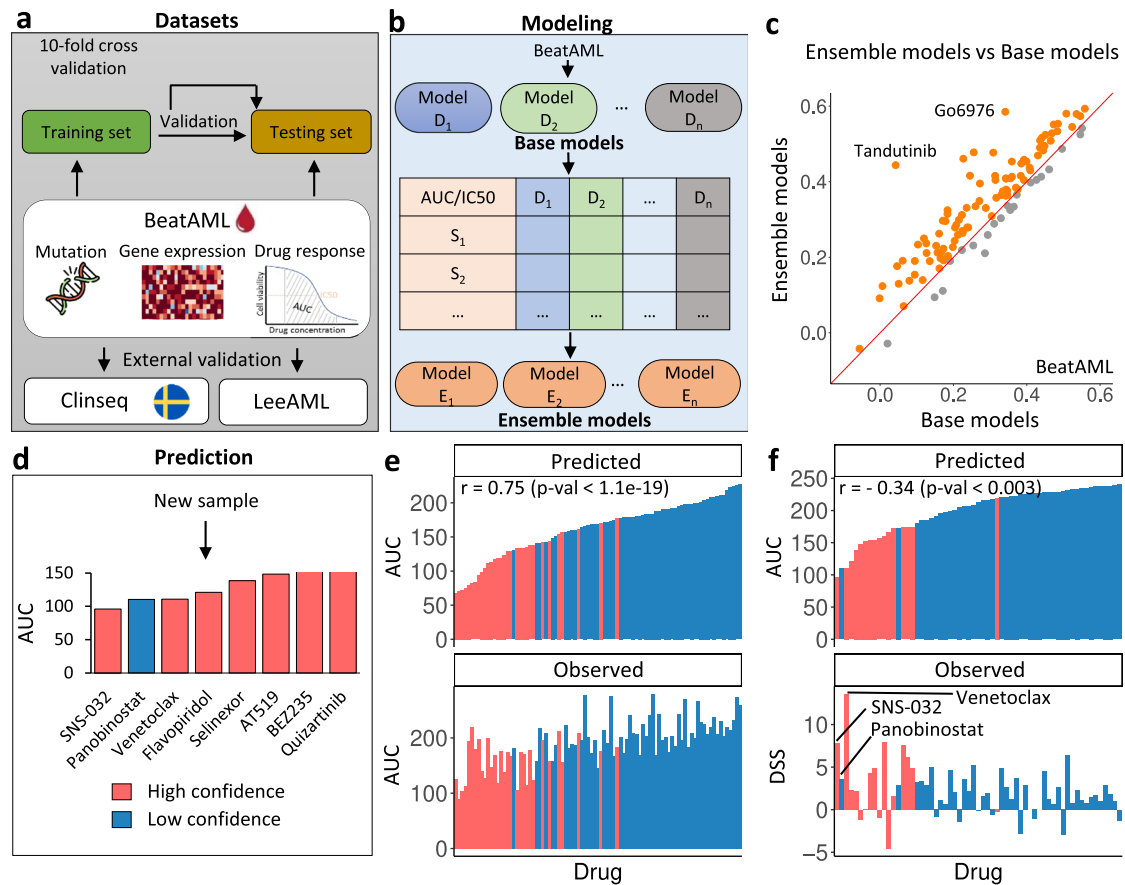
### AML-specific drug response prediction model

Figure 1 describes the overview of MDREAM. Briefly, MDREAM takes input omics data, including gene expression and mutation

<sup>1</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>2</sup>School of Biomedical Engineering, Shenzhen University, Shenzhen, China.

<sup>3</sup>Department of Oncology Pathology, Karolinska Institutet, Science for Life Laboratory, Stockholm, Sweden. <sup>4</sup>Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. <sup>5</sup>Department of Medicine Huddinge, Karolinska Institutet, Unit for Hematology, Karolinska University Hospital Huddinge, Stockholm, Sweden.

<sup>6</sup>Department of Medical Sciences, Hematology, Uppsala University Hospital, Uppsala, Sweden. ✉email: [trungnghia.vu@ki.se](mailto:trungnghia.vu@ki.se)



**Fig. 1 Overview of MDREAM.** **a** Omics data of the BeatAML cohort are used to train and validate the performance of the model. We further use the Clinseq cohort and the LeeAML cohort as two external validation datasets. **b** The ensemble-based prediction model uses the stacking approach including two layers: (1) the first layer contains 122 base models for prediction ( $D_1, D_2, \dots, D_n$ ), each is built from the data of a single drug; (2) the second layer consists of 122 ensemble models ( $E_1, E_2, \dots, E_n$ ) corresponding to 122 drugs. Each ensemble model is built based on the prediction output of the first layer. An advantage of ensemble learning is that it can exploit the correlation of drug sensitivity between drugs. **c** Comparison of performance between the base models (x-axis) and the ensemble models (y-axis) in the BeatAML validation set using the correlation between the predicted and observed response of individual drugs. Orange points above the diagonal line indicate higher correlation—better performance—achieved by the ensemble models. **d** Prediction results with information on confidence score for a new patient sample from the external validation dataset. This plot shows the results of the patient from the ClinSeq cohort with a median Pearson correlation between predicted AUC and observed drug sensitivity. The drugs with low AUC and high confidence (confidence score  $>0.75$ ) might be of interest for further investigation in the treatment of that patient. **e, f** Waterfall plots of the predicted AUC and observed drug sensitivities of the patient with median Pearson correlation from the BeatAML validation set (**e**) and the Clinseq dataset (**f**). Note that the negative correlation in **f** panel is as expected since the AUC and drug sensitivity score (DSS) have opposite directions: lower AUC and higher DSS correspond to better drug response.

profiles, and drug sensitivity data from the BeatAML cohort for training the drug-response prediction model, then the model is validated using both internal and external datasets, Fig. 1a. A total of 461 AML patients and their ex-vivo drug sensitivities from 122 small molecule inhibitor from the BeatAML cohort are included in the analysis. Two prediction models are built separately for two available metrics of drug sensitivity in the cohort:  $IC_{50}$  and area under the curve (AUC), where a lower  $IC_{50}$ /AUC indicates a better drug response.

The model of MDREAM is built using the stacking approach, Fig. 1b, using relevant biological features extracted from the omics data; see further details in “Prediction model and genomic features for drug response of AML”. The ensemble models of the stacking approach are generally more advanced than the base models because they can utilize the aggregation of base models to improve the robustness of the prediction. The ensemble approach also fits the problem of drug response prediction in this study. Indeed, it is common that a drug is not screened across all patients, which can limit the performance of the base model. Furthermore, many drugs sharing similar gene targets can have

highly correlated responses, which can also be efficiently utilized by the ensemble models. Figure 1c shows the comparison of prediction performance between the ensemble models and the base models in the BeatAML validation set. Each point presents the correlation between the predicted and observed response for each single drug. Orange points above the diagonal line indicate drugs with higher correlation (97/122 drugs, 80%), indicating the ensemble models to be better than the base models. The similar comparison in Clinseq also shows the higher performance of the ensemble models for the majority of drugs (50/76 drugs, 66%) compared to the base models (see Supplementary Fig. 1).

For a new patient sample, MDREAM reports the predicted  $IC_{50}$ /AUC of 122 drugs and also provides the confidence score (CS) for each prediction, Fig. 1d. The value of CS reflects the consistency of the prediction across bootstrap replicates of the training data, see more details in the section “Confidence score for drug-response prediction”. Figure 1e shows the prediction performance of the model for the patient from the BeatAML testing set with a median correlation between prediction and observation ( $r = 0.75$ ,  $p$  value  $< 1.1\text{e-}19$ ). Each drug is presented by a bar and the color of the

point represents its confidence score. The y-axis of the two panels are ranked by the drug response prediction. Drugs predicted to have a good response, and with a high confidence score appear consistently on the left side of the panels, e.g., elesclomol, quizatinib, and foretinib.

Similarly, Fig. 1f shows the individual with a median correlation from the Clinseq cohort ( $r = -0.34$ ,  $p$  value  $< 0.003$ ). The drug sensitivity in the Clinseq cohort is measured using drug sensitivity score (DSS)<sup>12</sup>, where a higher DSS indicates a better drug response, so we expect a negative correlation between the predicted AUC and the observed DSS. We note that the data used in Fig. 1d and f are from the same patient. Among the top three predicted drugs, sns-032 and venetoclax achieved the lowest predicted response at AUC = 96.0 (95% CI: [94.2, 125.2]) and 110.9 (95% CI: [90.9, 129.7]) with high confidence scores of 1.0 and 0.99 respectively, so might be worth further investigation for the patient. Ponobiostat shows a non-confident prediction with low confidence score, indicating it is unlikely that the patient would be a good responder to the drug as predicted; indeed, the observed DSS for the patient was low.

The AML-context-driven features used for the drug-response prediction are described in the section “Prediction model and genomic features for drug response of AML”. To explore the contribution of these features to the prediction, we apply the variable importance method of Fisher et al.<sup>13</sup> for the SVM base model of every drug. The method produces a score for each feature, presenting the contribution of that feature to the prediction performance of the drug. The top ten most important features of each drug are provided in Supplementary Data 1. Taking Venetoclax as an example, the method reports TSPAN10, NFIA, AARD, DACH1, AGR2, AATK-AS1, BCL2, DHRS2, LRP1, and MYO7A as the top important features for the prediction. Among those, BCL2 is the primary target of that drug. However, further investigation of the other important features of Venetoclax and other drugs is beyond the scope of this study.

### Validation of MDREAM in the BeatAML cohort

We first evaluate the performance of MDREAM using an independent testing set and 10-fold cross-validation of the training set; see Fig. 1a. The BeatAML patients are split based on their center IDs into the training set (60%,  $n = 278$ ) and an independent testing set (40%,  $n = 183$ ). The boxplots in Fig. 2a summarize the cross-validation results in the training set. The x-axis indicates the index of each of the ten folds of the cross-validation. The y-axis presents the correlation between predictions and observations of individual drugs. The median correlation of each fold ranges from 0.26 to 0.52, and the median correlation across tenfolds is 0.36.

To predict the BeatAML testing set, we use the whole training set to build the prediction model. For this testing set, the median correlation between the predicted and observed values for individual drugs is 0.35, which is close to the above cross-validation result, indicating no over-fitting issue in the training model. The correlation data of 76 drugs overlapping with the Clinseq cohort are shown in Fig. 2b and also reported in Supplementary Data 2. In this figure, each point presents a single drug with the corresponding correlation value in the x-axis. The histogram on top of the figure shows the correlation distribution of these drugs. As the results, among 76 overlapped drugs, trametinib achieves the highest correlation ( $r = 0.63$ ), while venetoclax also reports good prediction ( $r = 0.52$ ).

Figure 2c shows a high overall correlation between the predicted AUC and the observed AUC of the testing set ( $r = 0.68$ , 95% CI: [0.64, 0.68]). This plot includes all data points of the testing set, each point represents a (patient, drug)-pair. The patient with the median correlation ( $r = 0.75$ ) between the predicted and observed values across drugs is depicted in Fig. 1e,

while the distribution of the correlation for all individual patients is provided in Supplementary Fig. 2. We further compare MDREAM with a simple model (drug-average model) which uses the average of responses of each drug from the training data. Thus, for a single drug, the drug-average model uses the average value for all patients. The correlation between the predicted values from the drug-average model and the observed values is calculated for each patient in the validated set. As a result, see Supplementary Fig. 2, MDREAM achieves better performance than the drug-average model for most patients (62%, orange points below the diagonal line). The prediction results of trametinib and venetoclax are highlighted in Fig. 2d, e.

For the prediction of  $IC_{50}$ , we obtained slightly lower median correlations of individual drugs,  $r = 0.35$  and  $r = 0.33$  for the cross-validation and the testing set, respectively. This can be due to a drawback in the  $IC_{50}$  data that a majority (>40%) have a constant upper boundary ( $IC_{50} = 10$ ). Further details of the results for  $IC_{50}$  are provided in Supplementary Data 2.

### Validation of MDREAM in the Clinseq cohort

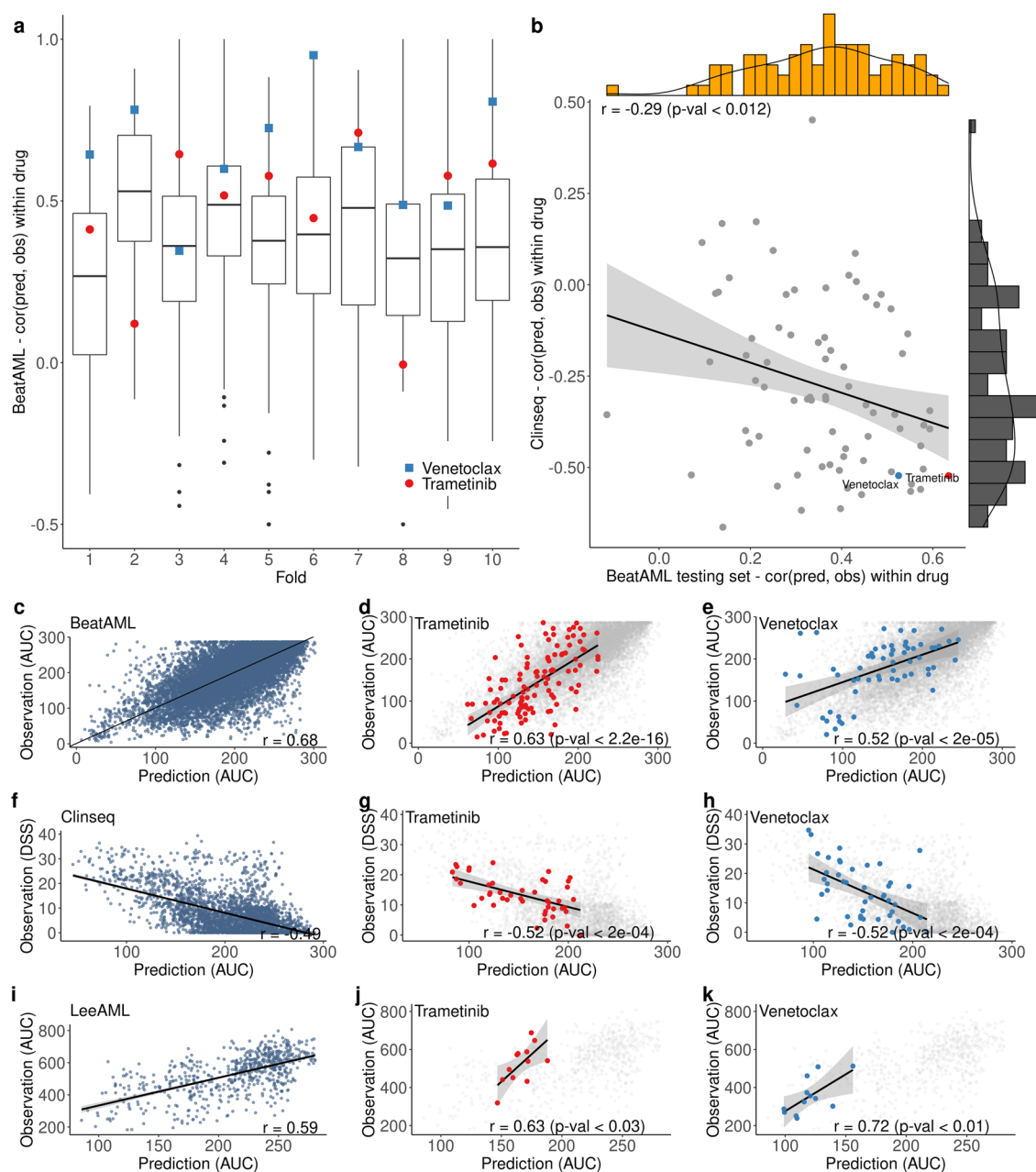
We include all samples of the BeatAML cohort ( $n = 461$ ) to build MDREAM and use the Clinseq cohort for validation. The Clinseq cohort is based on 45 Swedish AML patients and 76 drugs overlapping with 122 drugs in the BeatAML cohort. There are a total of  $n = 3420$  drug response data, which are measured using Drug Sensitivity Score (DSS)<sup>12</sup>. The summary of the correlation for individual drugs is presented by the histogram at the right-most side of Fig. 2b. The median correlation across all individual drugs is  $-0.32$ , slightly less than the results of the BeatAML cohort ( $r = 0.35$ ). We bootstrap the BeatAML data ( $n = 100$  times), then apply the same method for model building and prediction to get the bootstrap correlation values, and finally obtain a 95% confidence interval of  $-0.33$  to  $-0.25$  for the correlation (Supplementary Fig. 3).

Figure 2f presents a well validation of MDREAM for the Clinseq cohort with the overall Spearman correlation of  $-0.49$  (95% CI:  $[-0.53, -0.44]$ ) between the predicted AUC and the observed DSS across all data points. The patient with the median correlation ( $r = -0.34$ ) between the predicted and observed values across drugs is depicted in Fig. 1f, while the distribution of the correlation for all individual patients is provided in Supplementary Fig. 2. MDREAM also obtains better performance in 71% patients than the drug-average model, see Supplementary Fig. 2.

Figure 2b also shows a good concordance of the prediction results of 76 drugs between the two cohorts ( $r = -0.29$ ). In addition, the two drugs Trametinib and Venetoclax that have good results in the BeatAML cohort also show good performance in the Clinseq cohort, Fig. 2g, h. It is worth noting that the drug response metrics are different and independent between the two cohorts, suggesting the robustness of MDREAM. For the prediction based on  $IC_{50}$ , the correlation of all data points ( $r = -0.51$ ) and the median correlation of individual drugs ( $r = -0.29$ ) are comparable to the AUC-based results (See Supplementary Data 2).

### Validation of MDREAM in the LeeAML cohort

We apply a similar procedure to validate MDREAM in the LeeAML dataset<sup>10</sup>. The dataset contains genomic data from 12 relapsed/refractory acute leukemia patients and 624 drug response profiles from 52 drugs that overlap with the 122 drugs in the BeatAML dataset. Similar to the validation in the BeatAML cohort (Fig. 2c) and the Clinseq cohort (Fig. 2f), Fig. 2i illustrates a high overall correlation between the predicted AUC and observed AUC of 624 (drug, patient) points from the LeeAML dataset ( $r = 0.59$ , 95% CI: [0.51, 0.67]). The distribution of the correlation over all drugs for all individual patients (the median of 0.63) is provided in Supplementary Fig. 2. In comparison with the drug-average model, MDREAM gains better correlation for all patients, see



**Fig. 2 Performance of MDREAM in BeatAML, Clinseq, and LeeAML.** **a** Boxplots show the correlation of individual drugs from the tenfold cross-validation in the training set. Each boxplot displays the interquartile range (IQR) between the 25th percentile (the lower boundary) and the 75th percentile (the upper boundary). The center line of the box presents the median, and the whiskers are within the 1.5 IQR value. The points (blue or red) in each boxplot represent the correlation results of venetoclax and trametinib. **b** Prediction performances of the BeatAML testing set (x-axis) versus the Clinseq dataset (y-axis) across 76 overlapping drugs using the correlation between predicted values and observed values. The drugs at the bottom right indicate the drugs with good predictions in both cohorts. **c–e** Predicted AUCs vs observed AUCs of all data points (**c**) and individual drugs, including Trametinib (**d**) and Venetoclax (**e**) from the BeatAML testing set. **f–k** Similar scatter plots for predicted AUCs vs observed DSSs of the Clinseq dataset and LeeAML dataset, respectively.

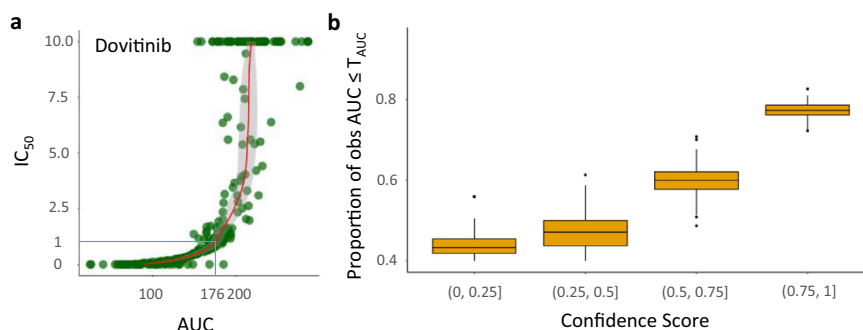
Supplementary Fig. 2. The correlations for all individual drugs are also shown in Supplementary Fig. 4. Trametinib and venetoclax, which have good performances in the BeatAML cohort, also archive top results in the LeeAML cohort, Fig. 2j, k. A negative correlation of some drugs in the LeeAML cohort might be due to the small number of samples ( $n = 12$ ).

#### Uncertainty assessment using a confidence score

Given data from a new patient, theoretically, the prediction model can report the predicted responses of 122 drugs; those predicted

to have good responses may be considered further. But, how confident are we in the prediction? In the BeatAML cohort,  $IC_{50}$  ranges from 0 to 10, where lower  $IC_{50}$  indicates better drug response. So we first define “good response” in terms of  $IC_{50} \leq 1$  and translate the threshold for the AUC (to get  $T_{AUC}$ ) based on the drug-specific relationship between  $IC_{50}$  and AUC; see Fig. 3a. The confidence score (CS) of each prediction is then computed using the bootstrap samples; see section “Confidence score for drug-response prediction”. The confidence score ranges from 0 to 1, and a higher confidence score indicates higher consistency in





**Fig. 3** Confidence score for drug response prediction. **a** Relationship between AUC and  $IC_{50}$  for the drug dovitinib in the BeatAML dataset; this is used to determine the threshold of good response based on the AUC ( $T_{AUC}^D$ ). **b**. The proportion of validated good responders (observed  $AUC \leq T_{AUC}^D$  in the BeatAML testing set) increases with higher confidence scores. Each boxplot displays the interquartile range (IQR) between the 25th percentile (the lower boundary) and the 75th percentile (the upper boundary). The center line of the box presents the median, and the whiskers are within the 1.5 IQR value.  $T_{AUC}^D$  is the threshold of AUC for good drug responses in drug D. The x-axis represents four categories of confidence scores.

obtaining a good-response prediction across the bootstrap replications of the data.

We expect predictions with higher confidence scores to have a higher level of validation. Indeed, Fig. 3b shows the proportion of good-response predictions that are validated in the BeatAML testing set increases with the confidence score. In this plot, the x-axis presents four increasing categories of confidence scores, and the y-axis shows the proportion of validated good-response. From the plot, the median of validated good-response proportions among predictions with a confidence score of less than 0.50 is less than 50%. If we keep only predictions with a confidence score  $>0.75$ , the validated proportion reaches around 77%. We found a similar trend for the results based on  $IC_{50}$ , Supplementary Fig. 5.

The example in Fig. 1c shows an application of the confidence score to identify drugs with low and high confidence of an individual in the Clinseq cohort. Among top three predicted good-response drugs, only sns-032 and venetoclax show high confidence (score  $>0.75$ ), which might be considered further individualized therapy. Figure 1e shows the prediction results of all drugs and their confidence scores of the same patient in Fig. 1c. A similar plot for a patient of the BeatAML testing set is found in Fig. 1d.

To investigate the advantage of the confidence score (CS) over the approach using the most extreme predictions, we compare the results of K drugs with a high confidence score (CS  $>0.75$ ) with top K extreme-prediction drugs ranked by the predicted AUCs for each patient in the BeatAML testing set. We summarize the correlations between the predicted and observed AUCs for each group of all patients for comparison in Supplementary Fig. 6. The plot shows that the group of drugs selected by the confidence score (the orange boxplot) obtains a higher correlation than the group of drugs from the most extreme prediction (the green boxplot). Thus, ranking drugs by the confidence score is better than ranking by the most extreme predictions.

### Drug response in relation to AML subtypes

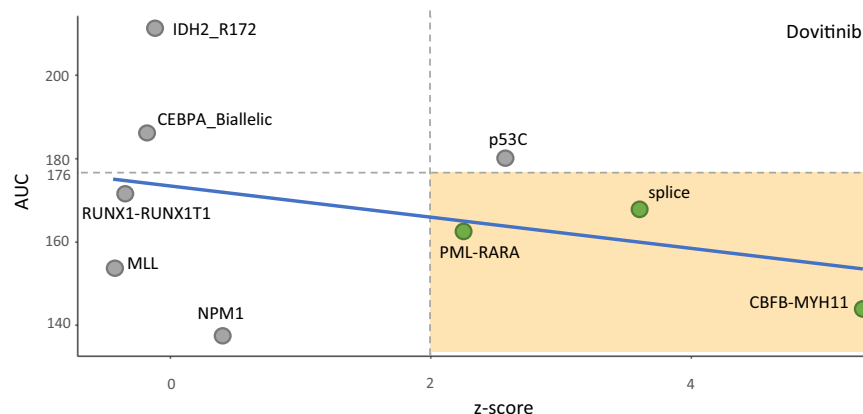
AML shows highly heterogeneous clinical phenotypes and molecular characteristics, which greatly complicate its clinical management. In a recent work<sup>2</sup>, Papaemmanuil and colleagues propose a classification of 11 distinct AML molecular subtypes, suggesting molecular driving mechanisms of the disease. Many molecular subtypes are concordant with known clinical subtypes, such as those associated with PML-RARA, RUNX1-RUNX1T1, or CBFB-MYH11 fusions, etc. Characterizing the relationship between these molecular subtypes and the drugs might inform us of the molecular mechanism of a drug in a responsive patient.

We investigate the association between the molecular subtypes and drugs by assessing the functional interaction between drug-target genes and the genes that are specific to the subtypes. The molecular subtype-specific genes are collected from the recent work of Mou et al.<sup>14</sup>, and the target genes of each drug are collected from DrugBank database<sup>15</sup>. For each pair of (subtype, drug), we apply the network enrichment analysis (NEA)<sup>16</sup> to obtain the enrichment score (z-score) between the two genesets; see further details in the section “AML molecular subtypes in relation to drug response”. A higher z-score ( $>1.96$ ) indicates a stronger functional interaction between the target genes of the drug and the subtype-specific genes of the subtype.

For a drug, the enrichment scores can vary across subtypes and correlate with the sensitivity of the drug. Figure 4 shows such an example from the drug dovitinib of the BeatAML cohort. In this plot, each point represents an AML subtype, the x-axis presents the z-score, and the y-axis expresses the median AUC of patients in each subtype. The Spearman correlation between AUC and z-score is  $-0.23$ . The subtypes with a z-score greater than 1.96 (the vertical dash line), including PML-RARA, splice, CBFB-MYH11, and p53C are significant ( $p$  value  $<0.05$ ). Among those, dovitinib has a good response to the first three subtypes under the horizontal dash line. The plot indicates that dovitinib has a good response for the patients in PML-RARA, splice and CBFB-MYH11 subtypes, and the good response could potentially be explained by the functional interaction between its target genes and subtype-specific genes of the subtypes. Note that dovitinib also shows a good response for RUNX1-RUNX1T1, MLL, and NPM1 subtypes; however, the plot tells us that we cannot explain why the drug works for those subtypes based on the interaction between the target genes and the subtype-specific genes. This subtype-drug interaction analysis is integrated into MDREAM; a classifier to predict the subtype of a new patient is also provided; see section “Prediction model and genomic features for drug response of AML”.

### MDREAM in comparison with existing methods

An early version of MDREAM was developed during our participation in the CTD-squared BeatAML DREAM challenge<sup>17</sup>. The DREAM challenge is a rigorous challenge for a competition of 28 international research groups working intensively for a long time (6 months) with extensive validation to evaluate competing methods. The challenge had two sub-challenges: (i) monotherapy drug response prediction and (ii) clinical response prediction. MDREAM ranked third in the drug response prediction sub-challenge, though statistically, there was little difference between the top four teams; see the final leaderboard of the DREAM



**Fig. 4 AUC versus z-score for different AML molecular subtypes.** Each point represents samples of an individual subtype group. The y-axis is the median value of drug response (AUC) of each subtype group. The z-score (x-axis) is a statistic of network enrichment analysis (NEA), which is used to assess the interaction between the subtype-specific genes and the drug-target genes in the functional gene network. A higher z-score indicates a stronger interaction than expected in a random permutation of the network. The blue line is the linear regression line. The yellow rectangle suggests some interesting subtypes, which respond to the drug potentially via the functional interaction between the subtype-specific and the drug-target genes.

challenge (our team was called CBA)<sup>17</sup>. It is noted that both MDREAM in this study and the competing methods in the challenge use the same data provided from the BeatAML article<sup>11</sup>.

Ammad-ud-din et al. applied Kernelized Bayesian Matrix Factorization (KBMF) model for prediction and computed the overall correlation between the predicted and the observed response measured from wet-lab validation of 48 drug-response profiles from eight drugs and six AML cell lines<sup>8</sup>. Their study reported a validated overall correlation of 0.44, compared to our results of 0.68, -0.49, and 0.59 for the validation of MDREAM in the BeatAML, Clinseq, and LeeAML cohorts, respectively. When applying the KBMF model to train and test on the same datasets used for MDREAM in this study, the median correlations across individual drugs show that MDREAM also outperforms the KBMF method in both BeatAML testing set ( $r = 0.35$  vs  $0.24$ ) and Clinseq dataset ( $r = -0.32$  vs  $-0.05$ ), see Supplementary Fig. 7. We exclude the LeeAML dataset from the analysis due to its small sample size ( $n = 12$ ). In the BeatAML testing set, 106/122 drugs (87%, orange points) have a better performance in MDREAM. Similarly, MDREAM also has the majority of the drugs (80%) with better correlation than KBMF does.

We further compare the features used by MDREAM with the clinically relevant gene biomarkers of 7 AML cell types from Zeng et al.<sup>18</sup>. Particularly, we collect the top five genes specific to each cell type which are provided in the article. Next, we replace the current data features of MDREAM with the gene expression of these gene biomarkers to build a cell type-markers-based model for comparison with MDREAM. Thus, the cell type-markers-based model and MDREAM use exactly the same prediction framework, so their results can be used for comparison of the importance of features used in the models. Supplementary Figure 8a compares the performance of MDREAM vs the cell type-markers-based model using the BeatAML testing set. Each point represents the Spearman correlation between the predicted AUC and the observed AUC values of an individual drug (a higher positive correlation is better). A large proportion of points (102 of 122, 84%) are below the diagonal line, indicating that MDREAM has better performance than the cell type-markers-based model in most drugs. A similar result is also found for the Clinseq dataset (see Supplementary Fig. 8b). In this plot, each point indicates the Spearman correlation between predicted (AUC) and observed (DSS) values of an individual drug where a lower negative correlation indicates better response. In 63 of 76 drugs (83%), MDREAM shows a better performance. We further investigate

whether these cell-type gene biomarkers bring added value to the performance of MDREAM. To do this, we combine the features of both the cell type-markers-based model and MDREAM to build a combined model (MDREAM + cell type-markers model). The high correlations for both datasets in Supplementary Fig. 9 indicate that adding the cell-type gene biomarkers does not significantly impact the MDREAM performance.

## DISCUSSION

We have developed and validated MDREAM, a computational method for drug response prediction in AML based on integrating omics data, including mutations and gene expression, and large-scale drug response assay data. We also proposed a confidence score to convey the uncertainty of the predictions. MDREAM is trained using the BeatAML cohort, and validated in the BeatAML cohort and two external cohorts. We further investigate AML subtype genes in association with drug sensitivity.

The establishment of pharmacogenomic datasets (e.g., NCI-60 database<sup>19</sup>; Genomics of Drug Sensitivity in Cancer (GDSC) project<sup>20</sup>; Cancer Cell Line Encyclopedia (CCLE) project<sup>21</sup>) enables the development of drug response prediction methods in pan-cancer. Most of these pan-cancer methods train predictors by using the gene expression profiles of cancer cell lines, which has been shown as the most informative source for drug response prediction<sup>22,23</sup>. However, these pan-cancer drug response predictors are usually not optimal for specific cancer, since each cancer has a limited number of cell lines which cannot comprehensively capture the heterogeneity and diversity of cancer. Thus, new efforts have been focused on developing cancer-specific drug response predictions for several cancers, including breast cancer<sup>24,25</sup> and colorectal cancer<sup>26,27</sup>. AML is also targeted under the effort of developing drug response prediction methods and applications of few studies<sup>8–10</sup>.

MDREAM would provide useful information supplement to simple clinical guidelines. For example, we attempt to explore the advantage of MDREAM over Midostaurin for patients with FLT3 mutation. We select 29 patients carrying FLT3 mutation from the BeatAML testing set for the evaluation. The observed drug responses of these patients using Midostaurin are compared to those using Top-1 drug and Top-5 drugs with the highest confidence scores discovered by MDREAM. The results in Supplementary Fig. 10 show that the observed AUCs of both Top-1 drug and Top-5 drugs are significantly lower than those of

Midostaurin, indicating that MDREAM can suggest drugs with a better response than Midostaurin. Since the genetic complexity of AML leads to >40% of AML patients with FLT3 mutation fail to respond to Midostaurin<sup>6</sup>, the suggested drugs from MDREAM might provide useful information for clinical application.

As its main strength, MDREAM had been previously compared and seen to perform well against a large number of competing models. It is built using a large cohort of AML patients, and validated in three distinct cohorts. The measurements of drug sensitivities in the BeatAML and the Clinseq cohorts are different: BeatAML cohort reports AUC and  $IC_{50}$  for drug response, while Clinseq uses the drug sensitivity score (DSS). The comparable validation results in the three datasets suggest the robustness of our prediction model.

Our study also has a number of limitations. First, the proposed method focuses only on the monotherapy setting, while the practical treatments of AML are usually compound combinations, e.g., the addition of FLT3 inhibitors (gilteritinib, sorafenib, quizartinib, and others) to the intensive chemotherapy or the combination of IDH inhibitors (ivosidenib and enasidenib) and venetoclax<sup>7</sup>. Second, the current approach of the prediction model does not allow easily further investigation into the molecular mechanism of drugs. However, in MDREAM we also introduce the characterization of the relationship between drug and molecular subtypes, which might suggest the molecular mechanism of a drug in a responsive patient. Third, the two external cohorts have a limited number of samples, but we hope we can increase the sample size in the future. Finally, BeatAML, Clinseq, and LeeAML cohorts only share a small proportion of drugs, which leads to the lack of validation of many important drugs. Over time we expect to overcome this weakness as we validate MDREAM in future patients.

The computational model of this study is based on the support vector machine (SVM) algorithm<sup>28</sup> and the stacking technique<sup>29</sup>. Our future direction is applying deep learning algorithms. In recent years, deep learning has shown many potential advantages in drug discovery and drug response prediction<sup>5,30,31</sup>. A notable advantage is that deep learning can automatically extract meaningful features or can integrate prior knowledge to improve overall performance. For example, transfer learning utilizes prior knowledge from pre-trained models to improve the performance of prediction models in new datasets<sup>32</sup>. Mourragui et al. propose an approach to transfer drug response predictors on cell lines and patient-derived xenografts to human tumors<sup>33</sup>. Another potential direction is extending MDREAM for drug synergy. Some recent studies show the effectiveness of drug combinations compared to monotherapy in hypertension<sup>34,35</sup>. In AML, Jafari and colleagues recently developed bipartite network models to design combination therapies<sup>36</sup>.

## METHODS

### Prediction model and genomic features for drug response of AML

The ensemble-based prediction model of MDREAM is built based on the stacking approach<sup>29</sup>. It consists of two layers: one for the base models  $M_{base}^{D_i}$  and another for the ensemble models  $M_{ensemble}^{D_i}$ , where  $D_i$  is one of 122 drugs from BeatAML; see Fig. 1b. This approach can help one drug getting extra information from other drugs to improve the robustness of the prediction by utilizing the correlation of drug responses between drugs, for example, those share similar set of gene targets in the second layer. A base model  $M_{base}^{D_i}$  of drug  $D_i$  inputs genomic features from gene expression and mutation profiles into a machine-learning method to build a prediction model for the drug. The ensemble model  $M_{ensemble}^{D_i}$  for drug  $D_i$  collects the prediction results from all of the base models across drugs to build another prediction

model for drug  $D_i$ . The outputs of the ensemble models are the final prediction results of MDREAM.

The genomic features to train the base models are selected using gene expression and mutation profiles. First, the gene expression of each sample is normalized by median centering and unit variance (median = 0, sd = 1). We then focus on AML-context-driven genes, including

- AML subtype-specific genes: we utilize the findings from Mou et al.<sup>14</sup>, which provides a list of genes specific to 11 AML molecular subtypes. For each molecular subtype, we keep 15 top genes.
- Pathway activation score (PAS): The previous study<sup>37</sup> introduces PAS to present a tumor-specific pathway activity level that is relevant to a specific drug and shows an association between PAS and drug sensitivity in AML. We calculate the PAS using the gene expression of the somatic mutations for the prediction model.
- Drug-target genes: Primary target genes of each drug are collected from DrugBank database<sup>15</sup>. Furthermore, we include genes from upstream and downstream of the target genes from regulatory network databases<sup>38–41</sup>. Theoretically, the response of a drug may be associated with gene expression of these genes.
- Mutated genes: driver mutations are targeted by multiple inhibitors; therefore, we also collect the mutations of patients to the prediction model. Gene expression of the mutated genes is used as the input for the model.
- Other AML-relevant genes: we further collect genes relevant to AML from previous studies<sup>2,42</sup>, FLT3 signatures<sup>43</sup>, and TP53 signatures<sup>44</sup>.

We also consider data-driven genes, which are selected based on the information from the training data, including

- Highly varying genes: 100 genes with the highest variance across the training patients.
- Correlated genes: 100 genes most correlated with the drug response.

We filter out genes which are highly correlated with other genes ( $r > 0.90$  Spearman correlation) to get the final features set.

As the predictor, we use the support vector machine (SVM)<sup>28</sup> with the radial kernel using the e1071 package, R version 3.6.3. The SVM is fast and flexible, but other machine-learning methods, such as the random forest<sup>45</sup> or elastic net<sup>46</sup> can be used. No parameter tuning for the SVM is used. The default parameter setting of SVM in the e1071 R-package is selected to work for most datasets in general. During the challenge, we tried out multiple prediction models such as random forests, elastic net, etc., as well as parameter tuning for those models. We discovered that for MDREAM, tuning the SVM parameters did not make a significant difference in overall performance but required a huge computational resource. We found that the performance mainly depends on choosing meaningful biological features rather than on tuning the parameters and selecting the models. We checked and found that the overall prediction results did not depend very much on the machine-learning method and parameter tuning (which requires a huge computational resource), but more on the set of meaningful biological features (data not shown). In the web application, we also use a similar geneset from the MDREAM model and the SVM to build the subtype prediction model for new samples.

### Confidence score for drug-response prediction

To compute the confidence score (CS) of a drug response prediction, we first define “good response” based on a threshold of drug sensitivity. In the BeatAML cohort,  $IC_{50} \leq T_{IC_{50}}$  or  $AUC \leq T_{AUC}^D$  are defined as good response, where  $T_{IC_{50}}$  and  $T_{AUC}^D$  are the

thresholds for drug  $D$ . We select  $T_{IC50} = 1$  as the threshold for  $IC_{50}$  prediction, but this can be adjusted in the MDREAM application. For the prediction of AUC, the threshold  $T_{AUC}^D$  of drug  $D$  is estimated using a non-parametric regression model fitting the  $IC_{50}$  and AUC values of the drug. Figure 3a illustrates the relationship between  $IC_{50}$  (x-axis) and AUC (y-axis) of  $D = \text{dovitinib}$  in the BeatAML cohort, which is used to estimate the  $T_{AUC}^{\text{Dovitinib}}$ . The fitted model (red curve) shows that  $T_{AUC}^{\text{Dovitinib}} = 176$  corresponds to  $T_{IC50} = 1$ .

Given  $Pr$  a predicted AUC of drug  $D$  for patient  $P$ ,  $CS$  is the proportion of good-response prediction for (drug  $D$ , patient  $P$ ) across bootstrap replicates of the training data. Specifically, we bootstrap the training set  $N$  times to build  $N$  bootstrap prediction models  $M_i$ , where  $i = (1, \dots, N)$ . Then,  $N$  predictions  $\{Pr_1, Pr_2, \dots, Pr_N\}$  of drug  $D$  for patient  $P$  from the bootstrap models are collected.  $CS$  is computed by

$$CS(Pr) = \frac{\#\{Pr_i \leq T_{AUC}^D\}}{N}. \quad (1)$$

The calculation of  $CS$  for  $IC_{50}$  is similar but use  $T_{IC50}^D$  for the drug sensitivity threshold. In the current study, we use  $N = 100$  bootstrap replications.

### AML molecular subtypes in relation to drug response

The AML molecular subtypes are somewhat well-known among AML oncologists. So its connection to a predicted drug response is useful information to suggest possible biological reasons for the effectiveness of the drug. In this study, we focus on 11 AML distinct molecular subtypes from the recent study of Papaemmanuil et al.<sup>2</sup>. In the web application of MDREAM, the molecular subtype of a patient is either provided in advance by the user or predicted by a prediction model as described in the section "Prediction model and genomic features for drug response of AML".

In the analysis of the relationship between AML molecular subtypes and drug response, we use the network enrichment analysis (NEA)<sup>16</sup>. NEA assesses the functional network connectivity between two genesets. It extends the traditional geneset enrichment analyses (GSEA) with informative topological information in terms of gene interaction networks. This analysis utilizes a comprehensive network containing ~1.4 million functional interactions between 16,299 distinct HUP0 genes. In this context, we measure the functional connectivity between subtype-specific genes and drug-target genes.

For the first geneset, we utilize the top 25 genes of the 11 molecular subtypes reported by Mou and colleagues<sup>14</sup>. For the second geneset, we obtain a list of drug-target genes derived from the Genomics of Drug Sensitivity in Cancer (GDSC) cohort<sup>20</sup> and DrugBank database<sup>15</sup>. And then, NEA simplifies the assessment of the association by defining an enrichment score as:

$$z = \frac{d_{AF} - \bar{d}_{AF}}{\sigma_{AF}} \quad (2)$$

where  $d_{AF}$  is the number of links between the two genesets;  $\bar{d}_{AF}$  and  $\sigma_{AF}$  are the mean and standard deviation of  $d_{AF}$  respectively, which are estimated on a randomized network under the null hypothesis of no drug-subtype interaction. Then, we investigate the correlation between z-score of drug-subtype interaction and drug sensitivity AUC.

### Datasets

**BeatAML cohort.** The BeatAML project<sup>11</sup> provides both omics data of AML, including gene expression and mutation profile, as well as drug-response assay data, from 461 patients. Gene expressions (count-per-million, CPM) of 26,086 hg19 genes and mutations of 302 genes are available. The gene expression data

are normalized centrally by the median, and we keep only mutations which are recurrent in at least two samples. The drug sensitivity data consist of 47,650 records from 528 AML patients across 122 compounds. A total of 32,263 records from 337 patients with RNA-Seq data are used for this study. Both measurement metrics of drug response, including  $IC_{50}$  and AUC are included.

For evaluating the prediction model, we split data into a training set ( $n = 278$ ) and a testing set ( $n = 183$ ) according to the center ID of the samples, such that no center IDs overlap between the two sets. The samples of each set are listed in Supplementary Data 3. Further evaluation is implemented using 10-fold cross-validation on the training set.

**Clinseq cohort.** This is a Swedish AML cohort<sup>47</sup>, which comprises 315 AML patients from 1997 to 2014. The RNA-seq samples were sequenced by Illumina HiSeq 2500 platform and gene expressions of 20,000 genes were estimated by XAEM<sup>48</sup> using hg19 annotation. The drug data contains the drug sensitivity of 528 drugs on 45 patients measured by drug sensitivity score (DSS)<sup>12</sup>. A total of 76 drugs shared between Clinseq and BeatAML is included in the analysis.

**LeeAML cohort.** This cohort is based on a pilot clinical trial on relapsed or refractory acute leukemia patients in the United States from 2015 to 2021, which included drug sensitivity profiles and gene expressions of 54 patients<sup>49</sup>. Our study uses only the published gene expressions and drug data from 12 patients<sup>10</sup>. Two RNA-seq replicates for each patient were prepared using Illumina TruSeq stranded mRNA kit genes. The final gene expression (in fragments per kilobase of transcript per million mapped fragments, FPKM) of 20,060 genes for each patient was averaged from Cufflink's output of two replicates. The drug data includes 1872 drug sensitivity (AUC) profiles of 156 drugs.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

The public data were available from their original studies: Gene Expression Omnibus (GEO) under accession numbers GSE108003 for the LeeAML dataset<sup>10</sup>, dbGaP accession ID phs001657.v1.p1 for the BeatAML dataset<sup>11</sup>. The demographic information and somatic mutations of the Clinseq dataset<sup>47</sup> are publicly available in the Clinseq AML repository at Zenodo <https://doi.org/10.5281/zenodo.292986>. The drug data from the in-house Clinseq dataset<sup>47</sup>, used here for model validation, is not publicly available at present, as it is part of an ongoing main study that generates the data. For further details about the Clinseq drug data, please contact the corresponding author.

### CODE AVAILABILITY

The web-based implementation of MDREAM is publicly available at <https://www.meb.ki.se/shiny/truvu/MDREAM/>. The pipeline for training MDREAM is available at <https://www.github.com/tracquangthinh/MDREAM>.

Received: 6 October 2022; Accepted: 13 March 2023;

Published online: 24 March 2023

### REFERENCES

1. Vardiman, J. W. et al. The 2008 revision of the world health organization (who) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937–951 (2009).
2. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
3. Meng-er, H. et al. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* **72**, 567–572 (1988).



4. Stone, R. M. et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a *flt3* mutation. *N. Engl. J. Med.* **377**, 454–464 (2017).
5. Adam, G. et al. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis. Oncol.* **4**, 1–10 (2020).
6. Fischer, T. et al. Phase IIB trial of oral midostaurin (*pkc412*), the *fms*-like tyrosine kinase 3 receptor (*flt3*) and multi-targeted kinase inhibitor, in patients with acute myeloid leukemia and high-risk myelodysplastic syndrome with either wild-type or mutated *flt3*. *J. Clin. Oncol.* **28**, 4339 (2010).
7. Kantarjian, H. et al. Acute myeloid leukemia: current progress and future directions. *Blood Cancer J.* **11**, 1–25 (2021).
8. Ammad-Ud-Din, M. et al. Drug response prediction by inferring pathway-response associations with kernelized bayesian matrix factorization. *Bioinformatics* **32**, i455–i463 (2016).
9. Gerdes, H. et al. Drug ranking using machine learning systematically predicts the efficacy of anti-cancer drugs. *Nat. Commun.* **12**, 1–15 (2021).
10. Lee, S.-I. et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat. Commun.* **9**, 1–13 (2018).
11. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
12. Yadav, B. et al. Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci. Rep.* **4**, 1–10 (2014).
13. Fisher, A., Rudin, C. & Dominici, F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **20**, 177 (2019).
14. Mou, T. et al. The transcriptome-wide landscape of molecular subtype-specific mRNA expression profiles in acute myeloid leukemia. *Am. J. Hematol.* **96**, 580–588 (2021).
15. Wishart, D. S. et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
16. Alexeyenko, A. et al. Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC Bioinformatics* **13**, 1–11 (2012).
17. DREAM-Challenge. CTD-squared BeatAML DREAM challenge. <https://www.synapse.org/#Synapse:syn20940518/wiki/> (2020).
18. Zeng, A. G. et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat. Med.* **28**, 1212–1223 (2022).
19. Shoemaker, R. H. The nci60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6**, 813–823 (2006).
20. Yang, W. et al. Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2012).
21. Ghandi, M. et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508 (2019).
22. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, 1–12 (2014).
23. Ding, Z., Zu, S. & Gu, J. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics* **32**, 2891–2895 (2016).
24. Daemen, A. et al. Modeling precision treatment of breast cancer. *Genome Biol.* **14**, 1–14 (2013).
25. Costello, J. C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).
26. Kong, J. et al. Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. *Nat. Commun.* **11**, 1–13 (2020).
27. Ooft, S. N. et al. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci. Transl. Med.* **11**, eaay2574 (2019).
28. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* 144–152 (Association for Computing Machinery, 1992).
29. Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
30. Chiu, Y.-C. et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genomics* **12**, 143–155 (2019).
31. Kuenzi, B. M. et al. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* **38**, 672–684 (2020).
32. Weiss, K., Khoshgoftaar, T. M. & Wang, D. A survey of transfer learning. *J. Big Data* **3**, 1–40 (2016).
33. Mourragui, S. M. et al. Predicting patient response with models trained on cell lines and patient-derived xenografts by nonlinear transfer learning. *Proc. Natl Acad. Sci. USA* **118**, e2106682118 (2021).
34. Marinier, K. et al. Effectiveness of two-drug therapy versus monotherapy as initial regimen in hypertension: a propensity score-matched cohort study in the uk clinical practice research datalink. *Pharmacoepidemiol. Drug Saf.* **28**, 1572–1582 (2019).
35. Thomopoulos, C., Bazoukis, G., Grassi, G., Tsioufis, C. & Mancía, G. Monotherapy vs combination treatments of different complexity: a meta-analysis of blood pressure lowering randomized outcome trials. *J. Hypertens.* **39**, 846–855 (2021).
36. Jafari, M. et al. Bipartite network models to design combination therapies in acute myeloid leukaemia. *Nat. Commun.* **13**, 1–12 (2022).
37. Trac, Q. T., Zhou, T., Pawitan, Y. & Vu, T. N. Discovery of druggable cancer-specific pathways with application in acute myeloid leukemia. *GigaScience* **11**, giac091 (2022).
38. Bovolenta, L. A., Acencio, M. L. & Lemke, N. Htridb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* **13**, 1–10 (2012).
39. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
40. Karolchik, D. et al. The ucsc table browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
41. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. Phosphosite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 (2004).
42. Metzeler, K. H. et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* **112**, 4193–4201 (2008).
43. Bullinger, L. et al. An FLT3 gene-expression signature predicts clinical outcome in normal karyotype AML. *Blood* **111**, 4490–4495 (2008).
44. Coutant, C. et al. Distinct p53 gene signatures are needed to predict prognosis and response to chemotherapy in er-positive and er-negative breast cancers: distinct p53 gene signatures are needed in breast cancer. *Clin. Cancer Res.* **17**, 2591–2601 (2011).
45. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
46. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 301–320 (2005).
47. Wang, M. et al. Validation of risk stratification models in acute myeloid leukemia using sequencing-based molecular profiling. *Leukemia* **31**, 2029–2036 (2017).
48. Deng, W. et al. Alternating em algorithm for a bilinear model in isoform quantification from rna-seq data. *Bioinformatics* **36**, 805–812 (2020).
49. Becker, P. S. et al. A multi-omic precision medicine clinical trial in acute leukemia. *Blood* **134**, 1269 (2019).

## ACKNOWLEDGEMENTS

This work was partially supported by funding from the Swedish Research Council (VR), CancerFonden, and the Swedish Foundation for Strategic Research (SSF). The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) in Uppsala, which is partially funded by the Swedish Research Council through grant agreement no. 2018-05973. We acknowledge the investigators of the BeatAML project, Oregon Health & Science University, USA for the data access. We also thank the patients who contributed their data used in this research.

## AUTHOR CONTRIBUTIONS

T.N.V. and Y.P. initiated and oversaw the study; Q.T.T. performed data analysis; Q.T.T., Y.P., T.M., M.R., T.N.V. contributed to method development and manuscript writing. M.V., R.J., A.B., A.O., I.S., H.B., T.E., S.K., P.O., L.M.O., O.K., S.L., and J.L. contributed to manuscript writing and the Clinseq cohort including the acquisition and processing of patient samples, the generation of drug data, and the collection of clinical data. All authors read and approved the final manuscript.

## FUNDING

Open access funding provided by Karolinska Institute.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-023-00374-z>.

**Correspondence** and requests for materials should be addressed to Trung Nghia Vu.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023