Deconvolution of ex-vivo drug screening data and bulk tissue expression predicts the abundance and viability of cancer cell subpopulations

May 10, 2023

Alexandre Coudray[1], Romain Forey[1], Benjamin Bejar Haro[1,2], Filipe Martins[1,3], Joana Carlevaro-Fita[1], Shaoline Sheppard[1], Sandra Eloise Offner[1], Gioele La Manno[1,*], Guillaume Obozinski[1,2,*], Didier Trono[1,*]
1 EPFL - Ecole Polytechnique Fédérale de Lausanne. 1015 Ecublens, Switzerland
2 SDSC - Swiss Data Science Center. 1015 Ecublens, Switzerland
3 Centre Hospitalier Universitaire Vaudois (CHUV), Hematology Service and Central laboratory, Oncology department, 1011 Lausanne, Switzerland.
* Corresponding authors

Abstract

Ex-vivo drug sensitivity screening (DSS) allows the prediction of cancer treatment effectiveness in a personalized fashion. However, it only provides a readout on mixtures of cells, potentially occulting important information on clinically relevant cell subtypes. To address this shortcoming, we developed a machine-learning framework to deconvolute bulk RNA expression matched with bulk drug sensitivity into cell subtype composition and cell subtype drug sensitivity. We first determined that our method could decipher the cellular composition of bulk samples with top-ranking accuracy compared to state-of-the-art deconvolution methods. We then optimized an algorithm capable of estimating cell subtype- and single-cell-specific drug sensitivity, which we evaluated by performing *in-vitro* drug studies and in-depth simulations. We then applied our deconvolution strategy to Acute Myeloid Leukemia (AML) context using the beatAML cohort dataset, currently the most extensive database of *ex-vivo* DSS. We generated a landscape of cell subtype-specific drug sensitivity and focused on four therapeutic compounds predicted to target leukemic stem cells: crenalotinib, AZD1480, bosutinib, and venetoclax. We defined their efficacy at the single-cell level and characterized a population of venetoclax-resistant cancer stem-like cells. Our work provides an attractive new computational tool for drug development and precision medicine.

Introduction

Ex-vivo drug sensitivity screening (DSS) allows for the parallel testing of hundreds of compounds on cells extracted from a human tumor, potentially paving the way to personalized therapeutic decisions (Kohtamäki et al., 2022; de Mel et al., 2020; Spinner et al., 2020). However, as currently performed, DSS assesses the susceptibility of cell mixtures and not of individual cell subtypes, hence

fails to provide information that might be of utmost clinical importance, for instance on the drug sensitivity of rare cancer cells with stem-like properties. Coupling *ex-vivo* DSS with cell purification techniques such as fluorescence-activated cell sorting (FACS) can partly address this shortcoming, but sorting requires large quantities of cells, is limited by the availability of specific cell surface markers, and the procedure itself can impair the viability of some cell subtypes (Hu et al., 2016). Therefore, computational methods capable of deconvoluting *ex-vivo* DSS would constitute attractive alternatives, as they would alleviate the need for cell subtypes purification.

Acute Myeloid Leukemia (AML) is a cancer type typically displaying extensive cellular heterogeneity, with the bone marrow of AML patients harboring a mixture of more than twenty cell subtypes, several of which are directly relevant to the oncogenic process (van Galen et al., 2019). In this setting, FACS typically fails to distinguish all individual cell subtypes, which can instead be identified by single-cell RNA sequencing (scRNA-seq), an increasingly appreciated source of biomarkers to complement somatic mutations (Kim et al., 2021). In AML samples, scRNA-seq can notably detect the rare leukemic stem cells (LSCs) (Stetson et al., 2021), likely responsible for relapse and decreased survival (Vetrie et al., 2020; Tan et al., 2020; Velten et al., 2021). Hence, new strategies to detect and target LSCs would be of great interest to improve AML patients' outcomes.

Here, we describe new methods to simultaneously infer cell subtype composition and deconvolute *ex-vivo* DSS into cell subtype- or single-cell-specific drug sensitivity maps. We first developed CLIMB, which outperformed a panel of eight deconvolution methods to decipher cell type composition of AML bulk samples. We then added CLIFF, which performs the deconvolution of drug sensitivity data. We benchmarked these methods through the *in vitro* analysis of mixtures of cancer cell lines, as well as *in silico* using pseudo-bulk samples derived from AML patients. We further documented CLIMB's accuracy on six cohorts of patients across two cancer types. We then deployed CLIMB and CLIFF to predict drug sensitivity at the cell-subtype, and single-cell resolution based on the currently largest publicly available dataset of *ex-vivo* DSS: the beatAML cohort (Tyner et al., 2018). This led us to predict gilteritinib, selinexor, erlotinib, and venetoclax drugs to have the highest efficiency against HSC-like cancer cells. Using the hereby identified single-cell drug sensitivity data for these four compounds, we identified a subpopulation of venetoclax-resistant HSC-like population. We further characterized this HSC-like population by showing significant upregulation of PRDM2, THOC2, and KIAA0368, suggesting that these genes could be predictive of venetoclax resistance.

Results

Experimental design

We developed two complementary computational methods to integrate bulk RNA-seq and *ex-vivo* DSS data (Methods, Supplemental Methods). First is CLIMB, which dissects the cellular composition of bulk samples (Fig.1, upper part). Second, we developed CLIFF, which integrates this information with bulk drug sensitivity to infer the cell subtypes' drug susceptibility (Fig. 1, bottom). CLIFF relies on a new mathematical model using cell-subtype-associated latent variables, which we resolve with an Expectation-Maximization algorithm. We tested CLIMB/CLIFF dual approach by generating in vitro mixtures of HL60, SUDHL4, K562, and THP1 human leukemic cell lines (Fig. 2, 4). We then benchmarked bulk deconvolution followed by drug sensitivity prediction using pseudo-bulk samples made from single-cell RNA in AML context (van Galen et al., 2019) (Fig. 3, 5). Ultimately, we applied our calibrated models to six cohort of patients, the beatAML cohort (Tyner et al., 2018), the Leucegene cohort (Leucegene), and the TCGA-AML cohort (Ley et al., 2013) in the context of AML, and the Syscol cohort (Bramsen et al., 2017), and two TCGA datasets

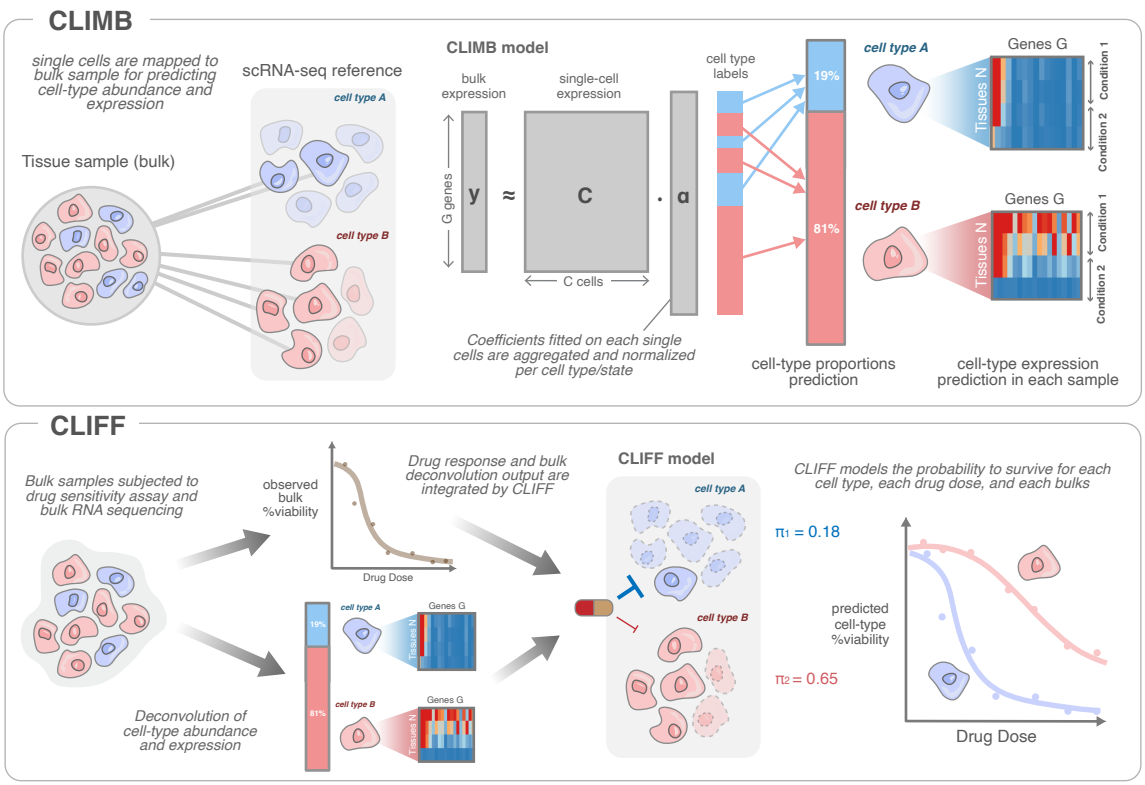


Figure 1: Experimental design. (**Top**) CLIMB performs bulk deconvolution with the full raw count matrix from scRNA-seq, fitting coefficients on each single-cells, aggregated at the cell-subtype level to deconvolute bulk expression into cell-subtype proportions and expression. (**Bottom**) The two-steps approach is presented here. First, a bulk deconvolution method is applied to estimate cell subtype proportion in each mixed sample. Then, deconvoluted cell type abundance and expression are integrated with drug sensitivity data to predict cell type drug sensitivity. CLIFF is a probabilistic model that resolves latent variables corresponding to cell type survival probability.

(Corces et al., 2018) (READ and COAD), in the context of colorectal cancer (CRC).

CLIMB finds the best combination of single cells to reconstruct bulk expression. The latter is achieved by resolving a linear regression directly on the raw single-cell expression data, where gene are used as observations and cells as features. It takes as input a matrix of bulk RNA-seq expression and a scRNA-seq reference dataset from the same biological context, and yields a prediction of cell-subtype proportions and cell-subtype expression (Fig. 1, upper part). On the other hand, CLIFF takes as input the cell-subtype proportions estimated by any bulk deconvolution methods, a scRNA-seq reference dataset, and bulk drug sensitivity data, to output a prediction of cell-type drug sensitivity, with a gene signature trained on the input expression data (Fig. 1, bottom). Thus, CLIFF can directly use CLIMB's output and integrate it with bulk drug sensitivity data. Of note, we use two assumptions to simplify CLIFF model: first, we assume that all bulk samples share the same cell-subtype expression, and second, we assume that cells of the same cell-subtypes share the same drug sensitivity.

Benchmark of bulk deconvolution methods using mixtures of cancer cell lines

CLIMB infers cell-subtype proportions through a linear combination of single-cell expression. By aggregating coefficients by cell subtype, it predicts the composition of a cell mix (Fig. 2A). We generated *in vitro* pools containing various ratios of HL60, SUDHL4, K562, and THP1 human leukemic cells, analyzed by standard bulk RNA sequencing. We could thus use these pre-determined ratios as ground truth to compare the predictive value of CLIMB (Fig. 2B) and a panel of bulk deconvolution methods: BayesPrism (Chu et al., 2022), CiberSortX (Newman et al., 2019), TAPE (Chen et al., 2022), Scaden (Menden et al., 2020), MUSIC (Wang et al., 2019), BisqueRNA (Jew et al., 2020), and NNLS (Wang et al., 2019), using random proportions as a negative control. We then assessed single-cell RNA expression of a mixture of cell lines, and used it as reference for all deconvolution methods with the same gene and cell selection. In this context, CLIMB reached a higher accuracy than other deconvolution methods based on all four tested metrics: the Pearson correlation coefficient (PCC, Fig. 2E), the Spearman correlation coefficient (SCC, SFig. 1B), the root mean square error (RMSE, Fig. 2F), and the R-square value (R2, SFig. 1C). We also considered scores according to different *angles*, for each cell line independently ('per cell line'), or for each bulk samples ('per sample', Fig. 2E, F, SFig. 2B, C). Notably, CLIMB was more accurate than other methods at the sample level in terms of RMSE and R-square (Fig. 2F and SFig. 2C, middle panel) but not for PCC and SCC metrics where Scaden could reach higher accuracy (Fig. 2E and SFig. 2B). Moreover, we observed a better PCC and R-square for BayesPrism and MUSIC at the cell line level (Fig. 2E and SFig. 2C, right panel), which was not confirmed by RMSE (Fig. 2E, right panel) nor SCC (SFig. 1B, right panel). The overall ranking among four metrics and three *angles* (average over 12 ranks), placed CLIMB first, and Scaden second best deconvolution methods (Fig. 2D). CLIMB produced significantly better correlation (Fig. 2E, left panel) and ranking than other deconvolution methods except Scaden. Thus, CLIMB and Scaden produced top-ranking accuracies to deconvolute bulk RNA-seq samples of *in-vitro* cell mixes.

Benchmarking bulk deconvolution methods using pseudo-bulk samples made from scRNA-seq

To expand the results obtained in vitro, we performed deconvolution of pseudo-bulk samples made from scRNA-seq data, a context where cell subtype proportions are known. We used two independent scRNA-seq datasets in the context of AML (Van Galen (van Galen et al., 2019), and Naldini (Naldini et al., 2023)) to perform pseudo-bulk deconvolution analysis. To reduce the overfitting of the reference dataset, we used half samples to build scRNA-seq reference and half to build pseudo-bulk samples (both in Van Galen>Van Galen, Fig. 3B, SFig. F-J, and Naldini>Naldini, SFig. K-O). Performing Van Galen > Van Galen analysis, we observed CLIMB to be the most accurate method in terms of PCC, SCC, and R-square metrics (Fig. 3B, SFig. F, H, I) but not RMSE (Scaden and BayesPrism ranked first and second respectively, Fig. 3B, SFig. G). Overall, CLIMB was the best performing method in Van Galen > Van Galen analysis (SFig. J), but not in Naldini > Naldini analysis where it ranked 3/9 after Scaden and CiberSortX (SFig. O). We then performed deconvolution of pseudo-bulk samples across datasets, using Van Galen as a scRNA-seq reference to deconvolute pseudo-bulk samples from Naldini's dataset (Van Galen > Naldini). This way, we mimic a realistic batch effect across datasets. In this realistic context, CLIMB was the best performing method according to all four metrics (PCC, RMSE, SCC, and R-square, Fig. 3A, SFig. A-D). CLIMB obtained the best average rank over four metrics and three analyses (red dots in Fig. 3C). Interestingly, the ranking of CLIMB improved upon performing deconvolution across datasets (blue

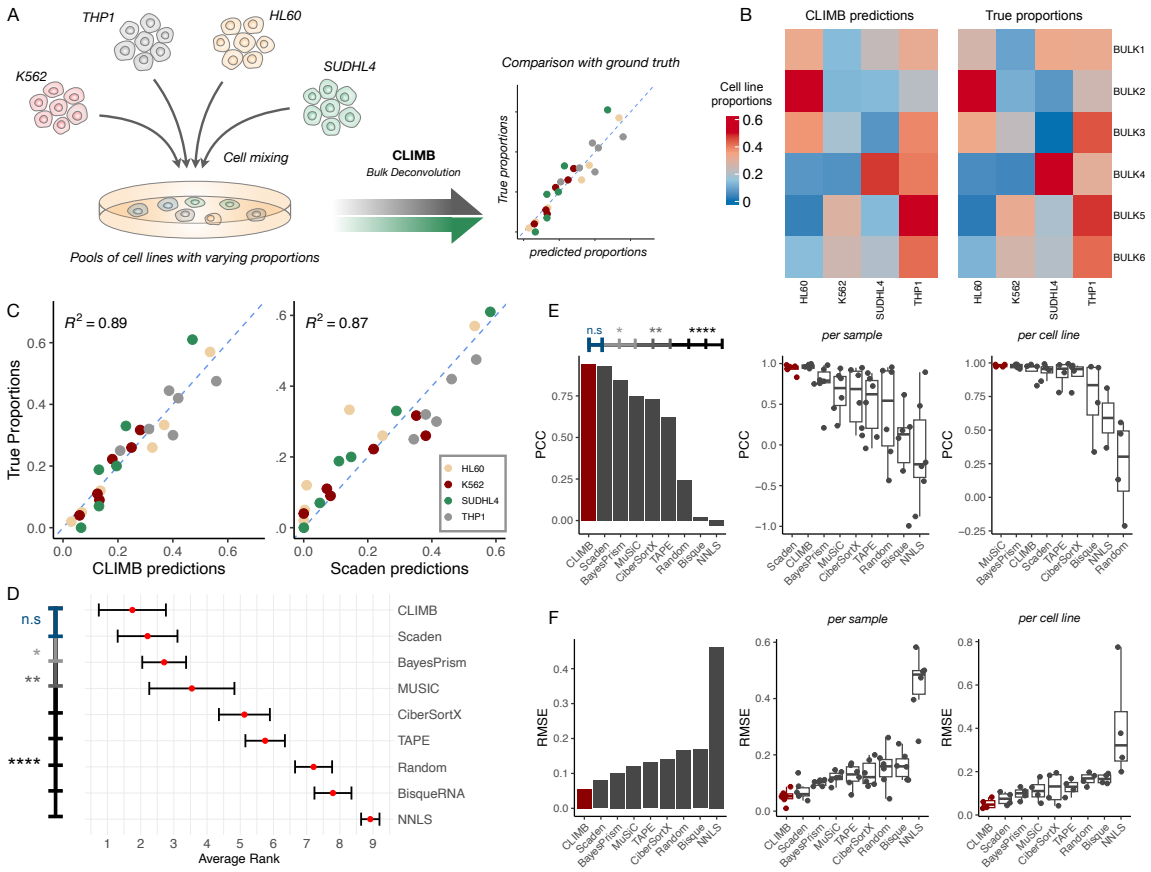


Figure 2: Deconvolution of cell line abundance based on bulk RNA sequencing data, in an *in-vitro* experiment with ground truth proportions available. (**A**) Sketch of the experimental design of the *in-vitro* experiment with mixtures of human leukemic cell lines. (**B**) Heatmap showing cell line proportions predicted by CLIMB (left) and measured cell line proportions (ground truth, right). (**C**) Scatter plot of true versus predicted cell line abundance, as predicted by CLIMB and Scaden. Colors indicates cell lines (6 bulks and 4 cell lines). (**D**) Overall results shown as ranks, taking four metrics into account (PCC, SCC, RMSE, and R-square), and three 'angles' (overall, per sample, per cell line). Statistical test between CLIMB and every other methods using Wilcoxon rank test. (**E**) Comparative deconvolution accuracy given as Pearson correlation coefficient (PCC) between predicted and true cell line proportions. Statistical comparison between CLIMB and other methods, using Pearson and Filon method implemented in *cocor* R package. Left barplot shows PCC *overall* (24 true versus 24 predicted data point), middle boxplot shows PCC *per sample* (4 true versus 4 predicted cell line abundance, for 6 bulk samples), right boxplot shows PCC *per cell line* (6 true versus 6 predicted data points, for 4 cell lines). (**F**) Comparative deconvolution accuracy given as root mean square error (RMSE) between predicted and true cell line proportions. left, middle, and right plot, as in (**E**).

dot versus red dot in Fig. 3C), which suggests a better resilience of CLIMB predictions to batch effects. We further validated the advantage of CLIMB using 200 pseudo-bulk samples (Methods) used to simulate drug sensitivity data in Fig. 5. In this context, CLIMB was consistently the best method based on four metrics (PCC, SCC, RMSE, and R-square, SFig. A-E). Another interesting observation was the seemingly better accuracy of CLIMB to predict the relative proportion of cells from the HSC-like subtype (Fig. 3D, SFig. F). Lastly, CLIMB displayed superiority over other methods in pseudo-bulks analysis from three additional scRNA-seq datasets: another scRNA-seq

AML dataset, from Wu et al. (Wu et al., 2020) and a colorectal cancer dataset, from Lee et al. (Lee et al., 2020), with CLIMB ranking in both settings as best method for four metrics (SFig. A-E); and a set of 101 simulated pancreatic islets bulk samples (Wang et al., 2019), where CLIMB ranked third but with a high accuracy close to the top methods (SFig. I, J). Thus, CLIMB reaches top-ranking accuracy in the context of pseudo-bulk analyses not only in AML but also in other contexts.

Deconvolution of three cohorts of AML patients

We then deconvoluted cell-type proportions based on bulk RNA-seq samples from three cohorts of AML: the beatAML cohort (Tyner et al., 2018), the Leucegene cohort (Leucegene), and the TCGA-AML cohort (Ley et al., 2013), using only bone marrow samples to fit our reference scRNA-seq (van Galen et al., 2019), and using only cells under pre-treatment conditions as a reference. Due to unknown cellular composition, we employed a panel of fourteen metrics (Methods) to compare with proportions predicted by our panel of deconvolution methods. Overall, CLIMB obtained the best average rank over fourteen metrics and three datasets (Fig. 3E). Applying CLIMB, we defined a landscape of cell subtype proportions in AML, with Mono-like and GMP-like cells seemingly to most frequent cancer cells (Fig. 3H). Overall, cancer cells seem to compose the majority of cells in AML bone marrows, except in a minority of patients with enriched normal monocytes (far right in Fig. 3H). Correlation between deconvoluted proportions and cell subtype signatures defined by Van Galen et al. (van Galen et al., 2019) placed TAPE, Scaden, and CLIMB, as best performing methods in BeatAML, Leucegene, and TCGA dataset, respectively (SFig. A), with CLIMB obtaining the best average ranking over three datasets (SFig. B). We then compared cell subtype proportions with the French-American-British (FAB) classification, which further placed CLIMB above other deconvolution methods (SFig. C,D). Notably, CLIMB showed the strongest enrichment of GMP-like cell subtype in PML-RARA samples (classified as M3 by FAB), an AML subtype known for its enrichment of GMP-like cancer cells (SFig. E). We then compared cancer cell proportions estimated by deconvolution with clinical blast count, which placed Scaden, BayesPrism and CLIMB as top performers (SFig. F). Interestingly, CLIMB showed a strong advantage for evaluating HSC-like proportions based on three metrics ('HSC-like vs LSC17', 'M0 vs HSC-like', and 'VG_sign_HSC-like', Fig. 3F), notably producing in all three datasets the strongest correlation with the 17-gene signature of leukemic stem cells defined by Stanley et al. (Ng et al., 2016) ('HSC-like vs LSC17' score, Fig. 3F-G). As expected, we found HSC-like proportions to display the highest correlation with the LSC17 signature (Fig. 3G). Moreover, when comparing the distribution of predicted cell subtypes proportions among all patients with their distribution among Van Galen dataset's patients, CLIMB-generated predictions had the highest concordance (SFig. G, I). Looking at similarities between proportions estimated by all deconvolution methods, we found CLIMB, Scaden, and CiberSortX predictions to be similar, whereas BayesPrism and MUSIC formed a separate cluster (SFig. H). Overall, CLIMB and Scaden were seemingly the best performing methods to deconvolute cell-subtype proportions in AML, with CLIMB seemingly better at predicting HSC-like cell proportions, a cell subtype with critical clinical implications.

Bulk deconvolution of colorectal cancer samples

In order to evaluate further the performance of CLIMB, we extended our analysis to the context of colorectal cancer (CRC). We conducted pseudo-bulk deconvolution into 31 different cell subtypes using Lee et al. scRNA-seq data (Lee et al., 2020), splitting cells as previously done in half to generate scRNA-seq reference and pseudo-bulk samples. CLIMB was found to be the most accurate deconvolution method for each four metric tested to compare predicted proportions with ground

Figure 3: Deconvolution of cell subtype composition in Acute Myeloid Leukemia (**A**) Pseudo-bulk analysis using Van Galen dataset as reference to deconvolute Naldini's pseudo-bulks. Barplots show PCC/RMSE for all data points (true against predicted). Boxplots shows one PCC/RMSE value for each pseudo-bulk sample (comparing 21 cell subtype's predictions). (**B**) Pseudo-bulk analysis using half of Van Galen samples as reference, to deconvolute the leftover samples assembled into pseudo-bulks. Barplot and boxplot as in (**A**). (**C**) Merged results from three pseudo-bulk deconvolution analysis. (Left) Circles colored with RMSE value, and sized according to PCC metric. (Right) Average rank over RMSE, PCC, SCC, and R-square metrics. Red dots/standard deviation is the average over three analysis. Blue dot is the average over Van Galen>Naldini analysis (maximum batch effect). (**D**) Predictive accuracy for HSC-like cell subtype, comparison 200 simulated pseudo-bulk to ground truth proportions (scatter plot). (left) Barplot showing R^2 values (method colored). (**E**) Deconvolution accuracy over three cohorts of AML patients, using ranks over fourteen metrics (Methods). Circles position depicts metrics' average rank over three datasets. Size only used for visualization. Statistical test of CLIMB versus 3 other best methods used Wilcoxon rank test. (**F**) Same as (**E**), but showing only three metrics related to HSC-like cell subtype. (**G**) PCC between cell-type proportions predicted by CLIMB and LSC17 score defined by Stanley et al. (Ng et al., 2016) (**H**) Cell-subtype proportions inferred with CLIMB, for bone marrow in BeatAML, Leucegene, and TCGA-AML cohorts (N=628). Below, scores associated with seven signatures (LSC17, and six Van Galen signatures), and FAB labels.

truth (PCC, SCC, RMSE, and R-square, SFig. A-E). We then applied our panel of deconvolution methods to three cohorts of CRC patients: the Syscol cohort (Bramsen et al., 2017) (287 patients), and two TCGA datasets (Corces et al., 2018) (READ and COAD, 436 patients). In order to estimate deconvolution accuracy, we used eight metrics as a proxy for cell subtype content and compared them with deconvoluted proportions (Methods). In the Syscol cohort, Bayesprism was found to be the most effective method, followed by CLIMB (SFig. F, G). Similarly, in the TCGA cohorts, Scaden was the most effective method followed by BayesPrism and CLIMB (SFig. H, I). Thus, CLIMB produced accurate results in the context of CRC, particularly for deconvoluting pseudo-bulk samples.

Deconvolution of cell type expression and evaluation of simulated over-expression

CLIMB is a deconvolution method that allows predicting gene expression at *high resolution*, meaning predicting cell-type expression in each bulk sample. Among the panel of deconvolution methods tested, we found only Bayesprism to be able to produce deconvoluted expression at high-resolution without filtering genes. We thus compared CLIMB's and Bayesprism's gene expression deconvolution accuracy based on pseudo-bulk samples made with Van Galen dataset (van Galen et al., 2019), a context where true cell-type expression is known. When we compare predicted cell-type expression with ground truth, expression deconvoluted by CLIMB showed higher correlation (PCC=0.86, SFig. A) than expression deconvoluted by BayesPrism (PCC=0.72, SFig. B). We then induced *in-silico* an over-expression of a hundred random genes in the HSC-like cell subtype and in half of the pseudo-bulk samples, to interrogate deconvolution methods' ability to recover differentially expressed (DE) genes in the correct cell subtype and in the correct pseudo-bulk samples. The deconvoluted expression of the 100 induced genes was better predicted by CLIMB than Bayesprism (red dots, SFig. A, B). We then computed the expression fold change (FC) between pseudo-bulk samples in wild type (WT) versus DE conditions (SFig. C, D). While CLIMB attributed high FC to HSC-like cells (SFig. E), Bayesprism attributed high FC to the wrong cell types (mainly cDC-like, SFig. F). We then performed DE analysis between WT and DE conditions using DESeq2 method (Michael I Love, 2014). DE analysis of true cell type expression recovered 99/100 true DE genes. We then used the same approach with CLIMB and Bayesprism deconvoluted expression. Using a ROC curve, we assessed the ability of DESeq2 p-values to correctly rank DE genes (SFig. G, H), which further showed an advantage of CLIMB over Bayesprism. Overall, CLIMB appears to be more accurate than BayesPrism to detect DE genes in HSC-like cell subtype.

Drug sensitivity deconvolution using mixtures of cell lines generated *in-vitro*

We next aimed at integrating drug sensitivity data with CLIMB bulk deconvolution output. For this purpose, we developed CLIFF, which deconvolutes the drug sensitivity of cell subtypes (Fig. 4A, Methods, Supplemental Methods). We thus performed deconvolution of six *in vitro* mixes containing four human leukemia cell lines as used in Fig. 2 (HL60, SUDHL4, K562, and THP1) exposed in parallel to seven drugs commonly used in AML consolidation therapies (5-Aza-2'- deoxycytidine, cytarabine, imatinib, nilotinib, doxorubicin, rapamycin, mitomycin C). To assess the accuracy of CLIFF deconvolution, we determined cell viability of the four cell lines exposed individually to the same drug panel, doses, and time points. CLIFF showed a PCC of 0.79 between the deconvoluted and the experimentally-assessed cell line drug sensitivity, considering all time points and drugs jointly (Fig. 4B). We then investigated the deconvolution accuracy for different cell lines and drugs (Fig. 4C-E). Overall, CLIFF was more accurate in predicting K562 sensitivity (PCC of

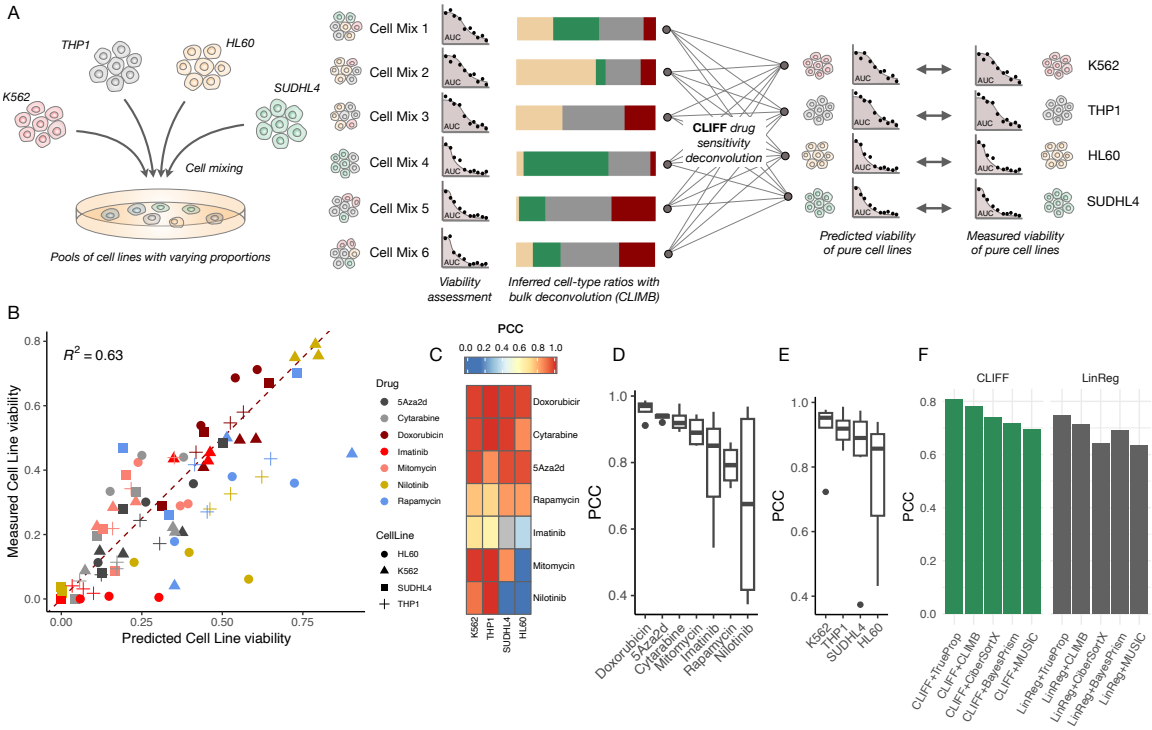


Figure 4: Cell line drug sensitivity deconvolution in an *in-vitro* experiment with available ground truth. (**A**) Experimental design of the *in-vitro* experiment. Cell lines are mixed with known ratios, and their abundance estimated through deconvolution. Deconvoluted abundances are then integrated with bulk drug sensitivity to deconvolute drug sensitivity of individual cell lines. Predictions are then compared to experimentally-assessed drug sensitivity of pure cell lines. (**B**) Comparison between predicted cell line drug sensitivity summarized as AUC values (x-axis) and true cell line sensitivity summarized as AUC values (y-axis). Three timepoints, four cell lines (shape) and seven drugs (colors) are shown (84 points) (**C**) Heatmap showing predictive accuracy as PCCs for each drug and cell lines. (**D**) Deconvolution accuracy as PCC for each drug. (**E**) Deconvolution accuracy as PCC for each cell line. (**F**) Comparaison of PCCs for CLIFF method versus a linear regression model (LinReg), using cell line proportions estimated by different deconvolution methods as input (CLIMB, CiberSortX, BayesPrism, or MUSIC). True cell line proportion was used as a positive control (CLIFF+TrueProp).

0.92, Fig. 4E) and doxorubicin and 5-Aza compounds (PCC of 0.96 and 0.94, Fig. 4D). We then compared CLIFF's accuracy when coupled with CLIMB or with cell subtype proportions estimated by BayesPrism, CiberSortX, MUSIC, using the known true cell line proportions (TrueProp) as a positive control (Fig. 4F). CLIFF+CLIMB provided the highest accuracy (PCC of 0.79) close to our positive control CLIFF+TrueProp (using true cell line proportion, PCC of 0.82), and was followed by CLIFF+CiberSortX, CLIFF+BayesPrism, and CLIFF+MUSIC (PCCs of 0.74, 0.71, and 0.69, respectively, Fig. 4F). We then implemented a linear regression (LinReg) to predict cell subtype drug sensitivity as done by Karakaslar et al. (E Onur Karakaslar, 2022) (LinReg+MUSIC), as an alternative algorithm to CLIFF. Overall, CLIFF outperformed LinReg, irrespective of which cell subtype proportion was used (Fig. 4F). In particular, CLIFF+CLIMB provided a better accuracy (PCC of 0.79) than LinReg+MUSIC (PCC of 0.627). Thus, in a set-up of six samples containing four cell lines, CLIFF can accurately deconvolute drug sensitivity, and works best in combination with cell subtype proportions estimated with CLIMB.

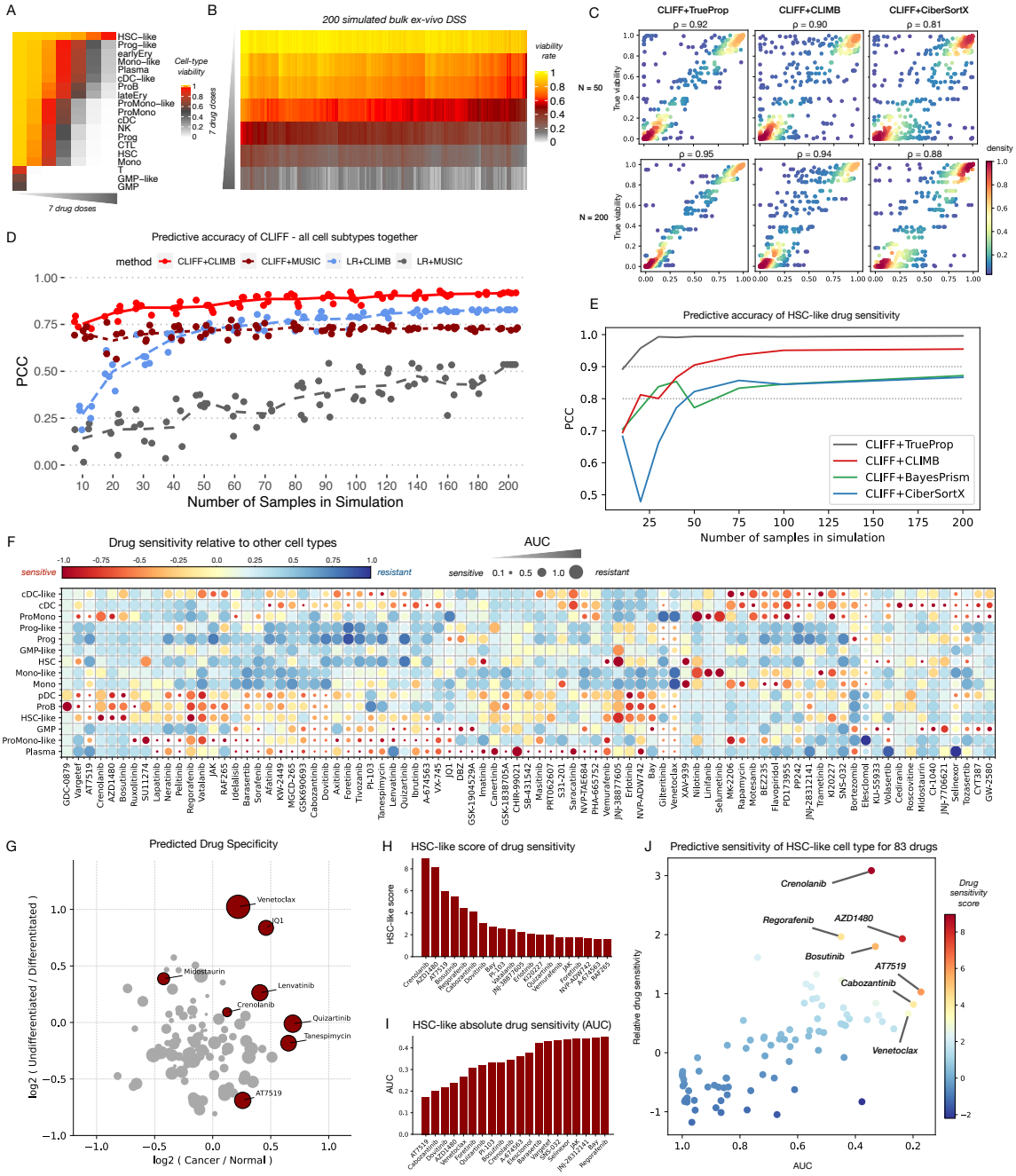


Figure 5: Deconvolution of *ex-vivo* drug sensitivity data into cell subtype drug sensitivity in Acute Myeloid Leukemia. (**A**) Simulated cell-subtype drug sensitivity. (**B**) Drug sensitivity simulation on 200 pseudo-bulk samples, built with cell-subtype drug sensitivities shown in (**A**). (**C**) Scatter plot comparing true and predicted cell-subtype drug sensitivity in simulation. Colors indicates dot density. (**D**) Comparison of PCC between true and predicted cell-subtype drug sensitivity in simulation, for different number of samples as input an different methods (colors). Five dots correspond to five independent simulations. (**E**) Line plot showing the PCC between true and predicted HSC-like sensitivity, using diverse deconvoluted cell subtype proportions. (**F**) Deconvoluted drug sensitivity of cell subtype in the beatAML cohort dataset. Size of the dots relates to the raw AUC value computed from 7 predicted data points (smaller = more sensitive). Colors relates to standardize AUC (column-wise) showing the relative drug sensitivity among cell subtypes. (**G**) (y-axis) Differentiated (Mono/ProMono/cDC) versus undifferentiated (HSC/Prog/GMP) myeloid cells' drug sensitivity (grouped). (x-axis) Normal versus cancer myeloid cells' drug sensitivity. (**H**) HSC-like drug sensitivity from AUC and standardized AUC (Methods). (**I**) AUC values for HSC-like cell subtype (small = sensitive). (**J**) Drug sensitvity of HSC-like. (x-axis) AUC, (y-axis) standardized AUC, with the color showing the score made from both measures.

Simulation of *ex-vivo* drug sensitivity assay data and prediction of drug sensitivity at cell-subtype level

We then benchmarked CLIFF's accuracy when many more cell subtypes are mixed together. Ground truth is unavailable when performing deconvolution of beatAML samples, we thus performed a simulation aimed at mimicking the beatAML bone marrow samples. We used Van Galen dataset to simulate 200 pseudo-bulk RNA-seq, paired with simulated drug sensitivity data. Accordingly, the bulk DSS data in Fig. 5B represents CLIFF's input used to predict cell subtype drug sensitivity depicted in Fig. 5A. We also tested the influence of input size on CLIFF's accuracy by using an increasing number of bulk samples (Fig. 5C, D). We found CLIFF predictions to be highly accurate when using true cell line proportions and >20 samples (PCC > 0.9 and RMSE < 0.2, Fig. 5C and SFig. B, C). With >40 input bulk samples, PCC reached 0.93. Since cell subtype proportions estimated through bulk deconvolution methods negatively impact CLIFF's accuracy, we compared CLIFF+CLIMB, CLIFF+BayesPrism, CLIFF+CiberSortX, and CLIFF using true proportions as a positive control (CLIFF+TrueProp). We found that CLIFF+CLIMB generally improved the accuracy over CLIFF+BayesPrism and CLIFF+CiberSortX (Fig. 5C, SFig. B, C), both in terms of PCC (SFig. B) and RMSE (SFig. C). CLIFF+CLIMB could reach PCC >0.9 with >50 input bulk samples. CLIFF+BayesPrism and CLIFF+CiberSortX reached a plateau of maximal PCC around 0.85 for >50 input bulk samples (SFig. B). We then ran our simulation using Linear Regression (LR) as an alternative to CLIFF method, and found it to be less accurate than CLIFF no matter the number of bulk samples used as input (CLIFF+ CLIMB vs LR+ CLIMB, or CLIFF+ MUSIC vs LR+MUSIC, Fig. 5D). In particular, we observed LR+MUSIC to produce poor accuracy even when using 200 bulk samples. We next examined the accuracy for predicting the drug sensitivity of HSC-like cells, attributing to HSC-like subtype a high viability mimicking drug resistance (Fig. 5A). CLIFF+TrueProp showed high accuracy with a maximal plateau of PCC around 0.98 for >30 input bulk samples. CLIFF+CLIMB accuracy for HSC-like sensitivity prediction plateaued at 0.95 with >50 input bulk samples, while CLIFF+BayesPrism and CLIFF+CiberSortX plateaued at 0.85 (Fig. 5E). Therefore, CLIFF+CLIMB requires only 50 cell mixes to deconvolute 19 cell types accurately, and can predict therein the drug sensitivity of a HSC-like resistant subpopulation.

Deconvolution of cell subtype drug sensitivity in the beatAML cohort

These results prompted us to select only drugs with at least 50 available ex-vivo DSS in the beatAML cohort. We then used cell subtype proportions estimated by CLIMB, using only cells under pre-treatment conditions as a reference for both CLIMB and CLIFF algorithms (van Galen et al., 2019). Applying CLIFF drug sensitivity deconvolution on the beatAML cohort revealed a heterogeneous landscape of drug sensitivity between cell subtypes (Fig. 5F). We delineated groups of drugs preferentially targeting undifferentiated cancer cell subtypes (Fig. 5G), including venetoclax, JQ1, lenvatinib, and crenolanib. In contrast, nilotinib, linifanib, and selumetinib were predicted to target differentiated Mono-like leukemic cells, and selinexor, rapamycin, and trametinib to aim preferentially at cDC-like cancer cells. Venetoclax appeared to be the most specific drug targeting undifferentiated cells, while quizartinib and Jak inhibitor I appeared to be less efficient against the latter, but to be potentially less toxic against healthy bone marrow cells (Fig. 5G). Focusing on HSC-like cells, we ranked drugs according to their putative efficiency on this particular cell subpopulation (Methods, Fig. 5H-J). We found crenolanib and AZD1480 to have the highest putative efficiency against this cell subset, followed by AT7519, bosutinib, regorafenib, cabozantinib, and venetoclax. Altogether, combining CLIMB and CLIFF succeeded in unveiling a landscape of drug sensitivity at the cell-subtype level in AML, together with a ranking of drugs based on their putative

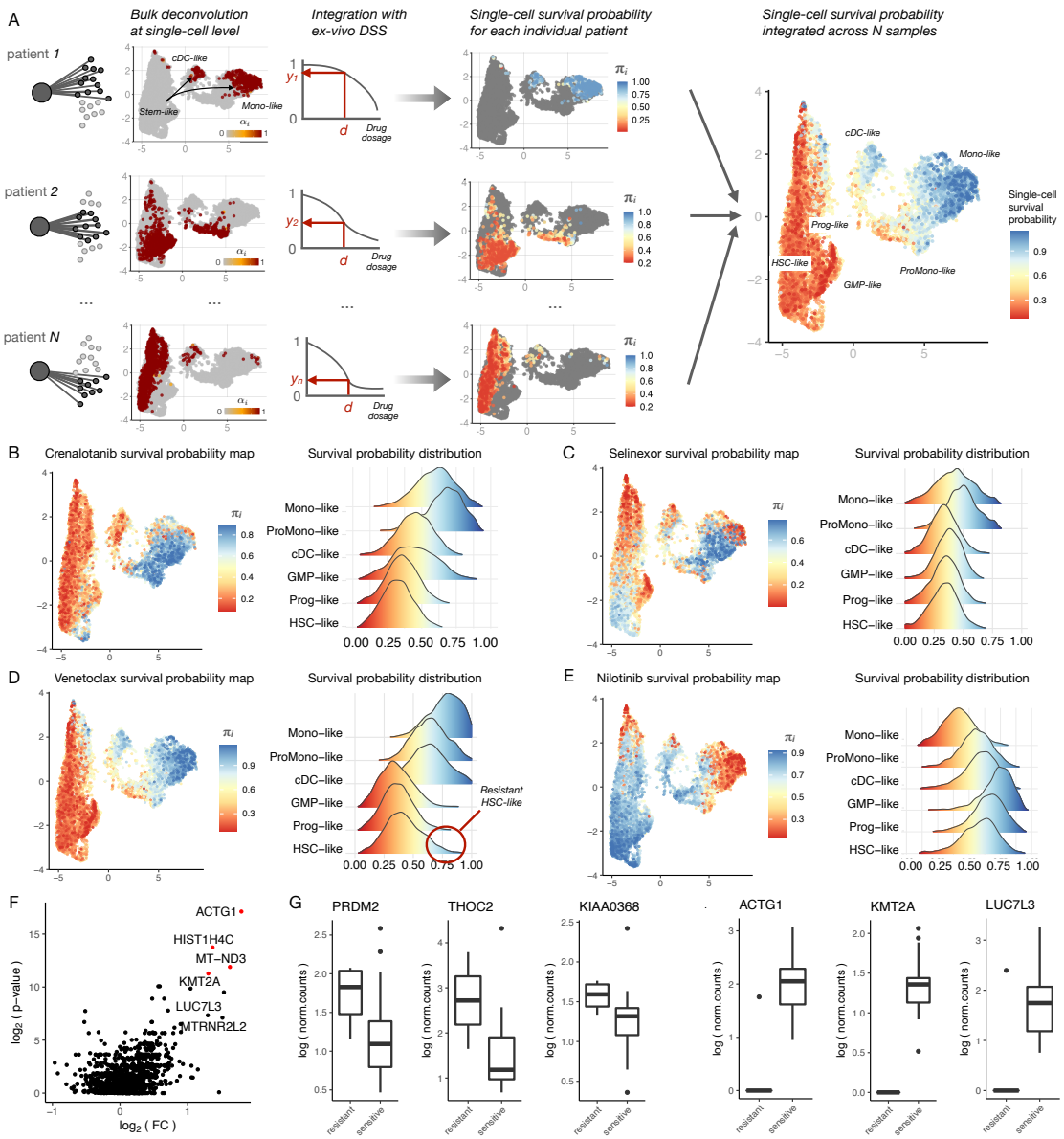


Figure 6: Drug sensitivity prediction at single-cell resolution. (**A**) Experimental design. The framework relies on CLIMB deconvolution which produces coefficients for each single cell (left), and is integrated with *ex-vivo* drug screening data (DSS) to produce single-cell survival probabilities (right). (**B**) (left) Single-cell survival probability for Crenalotanib, using UMAP projections of cancer cells from Van Galen et al. (van Galen et al., 2019), (right) Survival probability distributions among six cancer cell subtypes. (**C**) same as (**B**) showing selinexor single-cell drug sensitivity. (**D**) same as (**B**) showing venetoclax single-cell drug sensitivity. On the right, circled in red, the upper tail of HSC-like cells predicted to be resistant to venetoclax treatment. (**E**) same as (**B**), showing nilotinib single-cell drug sensitivity. (**F**) Volcano plot showing p-value versus log fold change for comparison of gene expression between venetoclax-resistant/ventoclax-sensitive cells. (**G**) Genes over-expressed in venetoclax-resistant population of HSC-like cells. (**H**) Genes over-expressed in venetoclax-sensitive population of HSC-like cells.

efficiency to target the HSC-like population.

Prediction of drug sensitivity at single-cell resolution identifies resistant cancer cell subpopulations

We further developed CLIFF-SC, which computes viability prediction at single-cell resolution (Fig. 6A, Methods, Supplemental Methods), using as input coefficients that CLIMB fits on every single cell. We focused on three drugs predicted to target the HSC-like compartment: crenalotanib, venetoclax, and selinexor, and one drug with an opposite pattern, nilotinib, which targets Monocyte-like cancer cells (Fig. 5F). We visualized single-cell survival probability on 2D projections (UMAP) using all cancer cells from Van Galen dataset (van Galen et al., 2019), which showed considerable heterogeneity in distributions across the four drugs and the six cancer cell subtypes (Fig. 6B-E). We observed that the ProMono-like cell subtype was apparently more resistant to crenalotanib and selinexor. Interestingly, we observed a long upper tail of seemingly resistant HSC-like cells for venetoclax compound (Fig. 6D, right panel, circled in red). We thus decided to study in more detail this population of resistant HSC-like. We performed differential expression analysis between resistant and sensitive HSC-like populations, which were clustered using single-cell venetoclax sensitivity predictions. The comparison revealed more upregulated genes in the sensitive population (Fig. 6F). Nevertheless, we could highlight three significantly upregulated genes in the resistant HSC-like populations: PRDM2, THOC2, and KIAA0368 (Fig. 6G). We also identified three genes most strongly upregulated in drug-sensitive HSC-like cells: ACTG1, KMT2A, and LUC7L3 (Fig. 6H). These data suggest that expression of these genes could be predictive of venetoclax sensitivity at single-cell level.

Discussion

The beatAML consortium has performed the largest *ex-vivo* drug sensitivity screening (DSS) so far conducted in any cancer type (Tyner et al., 2018), opening the door to deconvoluting cell type-specific drug sensitivity. In the present study, we propose a method to achieve this goal, as a potential substitute to costly studies commonly needed to assess cell population level drug sensitivity in large cohorts of patients. Interestingly, Bottomly et al. found that drug sensitivity is mainly governed by AML cell differentiation state (Bottomly et al., 2022), while Zeng et al. additionally observed that proportions of HSC-like cell-subtype estimated with CiberSortX correlate with drug response and clinical outcome (Zeng et al., 2022). Another interesting study by Karakaslar et al. (E Onur Karakaslar, 2022) recently showed that the proportion of Monocyte-like cancer cells was predictive of venetoclax resistance in the BeatAML cohort. However, these studies relied on cell-subtype proportions estimated with either a scoring system (Bottomly et al., 2022), CiberSortX (Zeng et al., 2022), or MUSIC (Wang et al., 2019) deconvolution, which we demonstrate to be suboptimal to infer cell type proportions in the context of AML. Furthermore, the approaches they used cannot predict the drug sensitivity of cell subtypes in new datasets. In contrast, the hereby described deconvolution algorithms are tailored to AML, with a machine-learning framework geared to predict the cell type and single-cell drug sensitivity from new, not previously analyzed scRNA-seq datasets.

Bulk deconvolution allows dissecting the cell type composition of samples from large cohorts of patients at a low cost. Here, we describe a new algorithm to achieve this goal and find it to perform better than current standard methods. The superior performance of CLIMB could be linked to its use of the full single-cell expression matrix as input instead of averaged expression per subtype, as the latter may lead to discarding cells with high dropout and low signal-to-noise

ratio. Many deconvolution methods assume that each bulk sample shares the same cell subtype expression, which can be inaccurate. In particular, we expect a high variability between patients for cancer cell subtypes. In this regard, CLIMB probably better grasps the underlying distribution of gene expression in individual cell subtypes. Also, we show that CLIMB probably better handle batch effects, as it produced more accurate results in pseudo-bulk analysis when used across two independent datasets. As CLIMB resolves an independent problem for each bulk sample, it is not affected by sample size. In line with this observation, we tend to observe an increased advantage in settings where few bulk samples were available (e.g. our in-vitro experiment with only six bulk samples). Interestingly, CLIMB carries a substantial edge for computing proportions of HSC-like cells, a cell subtype of primary interest as likely responsible for poor prognosis and relapse in AML patients. By improving the estimation of HSC-like subtype proportions in bulk samples, we also improved predictions of the drug sensitivity of this critical cell type. We believe CLIMB will likely be helpful in a wide range of biological contexts, particularly for deconvoluting the cellularity of cancer samples.

We also developed CLIFF, which extracts cell type- and single-cell-specific drug sensitivity from bulk DSS and RNA-seq data. By considering latent variables and resolving an Expectation-Maximization algorithm, CLIFF can predict the drug sensitivity of specific sub-populations of cancer cells (Method, Supplemental Methods). CLIFF is a new deconvolution algorithm able to increase the resolution of drug sensitivity data by deconvoluting its signal over cell subtypes defined by scRNA-seq. We demonstrated it first with engineered mixes of leukemic cell lines and then with the beatAML dataset, where it revealed gilteritinib, selinexor, erlotinib, and venetoclax to be efficient at targeting HSC-like cells. CLIFF also led us to delineate a subpopulation of HSC-like cells seemingly resistant to venetoclax (Fig. 6), which displayed significantly higher expression of PRDM2, THOC2, and KIAA0368 than their drug-sensitive counterparts, pointing to these genes as potential biomarkers of venetoclax resistance. More generally, HSC-like venetoclax-resistant cells exhibited more downregulated genes than HSC-like sensitive cells, consistent with quiescent HSC-like cells as defined by Zeng et al. (Zeng et al., 2022). Using the signature revealed by analysis of the beatAML dataset, CLIFF might be able to predict the presence of venetoclax-resistant HSC-like cells in any new patient sample. Plasma cells (Fig. 5C) were also predicted to be sensitive to the JAK inhibitor tofacitinib, which has indeed been reported to inhibit B cell development (Rizzi et al., 2017), and to NF-kB inhibitors such as BMS-345541, not a completely unexpected finding since B-cell receptor signaling activates the NF-kB pathway (Burke et al., 2003).

The limitation of our approach is that it does not incorporate data from somatic mutations. Although somatic mutations are a well-known source of biomarkers for drug sensitivity, they do not allow to differentiate sensitivity between cell subtypes using a deconvolution model. More generally, single-cell predictions made by CLIFF would need further experimental validation in disease contexts, for instance comparing a cohort of patients responsive or not to venetoclax therapy. By integrating multiple data types, the combined approach described here aims to improve cellular composition predictions while dissecting drug sensitivity at the cell sub-populational level. As single-cell RNA-seq will likely become increasingly utilized for diagnosis purposes, there is a need to develop appropriate computational methods to extract all the information this breakthrough technology can deliver.

METHODS

Prototypical bulk deconvolution model

The goal of a bulk deconvolution model is to fit coefficients $\mathrm{w_k}$ corresponding to each cell-subtype k through a linear model. Fitted coefficients directly relates to cell-subtype proportions. Thus, bulk deconvolution solves the following problem :

$$\mathbf{x}_\mathrm{n} \approx \mathbf{S} \cdot \mathbf{w}_\mathrm{n} \tag{1}$$

with $\mathbf{x}_\mathrm{n}$ representing a vector of gene expression of length G, for a given mixture n. The matrix $\mathbf{S}$ is a G by K matrix containing the average expression for each cell-subtype k and each gene g. To obtain proportions $\mathrm{w_{nk}}$ that sum up to 1, the model can either impose a constraint $\sum_\mathrm{k} \mathrm{w_{nk}} = 1$, or the coefficients $\mathrm{w_{nk}}$ can be normalized after fitting. In this model, the genes are our observations, and corresponds to genes shared between $\mathbf{x}_\mathrm{n}$ and $\mathbf{S}$ or to a subset of marker genes. Note that every mixture n are resolved with an independent linear model, and that the average expression of cell-subtypes is usually derived from a scRNA-seq that does not contain mixture n. Thus, we assume that $\mathbf{S} \approx \mathbf{S}_\mathrm{n}$. Of note, this simple model do not apply for advanced deconvolution algorithms like CiberSortX, BayesPrism, Scaden, BLADE, or TAPE.

CLIMB - Cell-level linear model of bulk deconvolution

We developed CLIMB method, which proposes an innovative approach to solve a bulk deconvolution problem. Instead of solving Equation 1, we fit coefficients α_i for each single-cell i, and thus we find the best linear combination of single-cell $\mathrm{c_i}$ to predict bulk expression $\mathbf{x}_\mathrm{n}$. Thus, we use the relation $\mathbf{x}_\mathrm{n} \sim \mathbf{C} \cdot \boldsymbol{\alpha}_\mathrm{n}$, shown below in more details :

$$\begin{bmatrix} \mathrm{x_{n1}} \\ \vdots \\ \mathrm{x_{nG}} \end{bmatrix} \propto \underbrace{\begin{bmatrix} \mathrm{c}_1^1 & \dots & \mathrm{c}_1^\mathrm{C} \\ \vdots & \ddots & \vdots \\ \mathrm{c}_\mathrm{G}^1 & \dots & \mathrm{c}_\mathrm{G}^\mathrm{C} \end{bmatrix}}_{\mathbf{C}^{(\mathrm{G}\times\mathrm{C})}} \cdot \begin{bmatrix} \alpha_\mathrm{n}^1 \\ \vdots \\ \alpha_\mathrm{n}^\mathrm{C} \end{bmatrix} \tag{2}$$

where the vector $\mathbf{x}_\mathrm{n}$ represent the bulk expression matrix for a mixture n, $\mathbf{C}$ is the raw counts matrix from scRNA-seq, and $\boldsymbol{\alpha}_\mathrm{n}$ is the vector of coefficients fitted for each single-cell c for a given mixture n. We introduce an indicator vector $\boldsymbol{\tau}_\mathrm{k}$, which is of length C and equal to 1 whenever cell i is of type k. We then group and normalize these coefficients α_ni at the cell-subtype level to obtain the cell-subtype proportions $\mathrm{w_{nk}}$ for each cell-subtype k and mixture n :

$$\mathrm{w_{nk}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{C}} \tau_\mathrm{ki} \cdot \alpha_\mathrm{ni}/\theta_\mathrm{i}}{\sum_{\mathrm{i}=1}^{\mathrm{C}} \alpha_\mathrm{ni}/\theta_\mathrm{i}} \tag{3}$$

With α_ni being the coefficients fitted with Glmnet library, which relies on a generalized linear model to fit a linear regression, using a coordinate descent algorithm (Friedman et al., 2010). θ_i being a normalization factor corresponding to the total expression of cell i :

$$\theta_\mathrm{i} = \sum_{\mathrm{g}=1}^{\mathrm{G}} \mathrm{C_{gi}} \tag{4}$$

Thus, CLIMB loss function is defined by :

$$\mathrm{L}(\boldsymbol{\alpha}_{\mathrm{n}}|\mathbf{C},\mathbf{x}_{\mathrm{n}}) = \min_{\boldsymbol{\alpha}} \frac{1}{2\mathrm{G}} |\mathbf{x}_{\mathrm{n}} - \mathbf{C}\cdot\boldsymbol{\alpha}_{\mathrm{n}}|^2 \quad \text{, with } \alpha_{\mathrm{ni}} \in [0,+\infty] \tag{5}$$

Thus, we impose a non-negativity constraint on the coefficients, and no regularization. Glmnet were thus used with options lower.limits = 0, lambda = 0, and scale = TRUE. We indeed observed no improvements of the deconvolution with neither l_1 nor l_2 regularization.

Deconvolution of cell-type expression

CLIMB allows the prediction of cancer cell expression in a two step manner. Let first define our target *high-resolution* expression $\hat{\mathbf{S}}_{(\mathrm{ngk})}$, which is the cell-type-specific and bulk sample-specific expression for each cell-type k. We assume that for a given group of samples, the *grouped* expression $\hat{\mathbf{S}}_{(\mathrm{gk})}$ relates to the *high-resolution* expression with the following formula:

$$\mathrm{s}_{\mathrm{gk}}^{(\mathrm{grouped})} = \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{s}_{\mathrm{ngk}}^{(\mathrm{high-res})} \tag{6}$$

First, we take advantage of the bulk to single-cell mapping to directly derive the cell-type expression from the bulk-to-single-cell deconvolution or *mapping*. We can simply sum up the expression of cells mapped by CLIMB to get a *high-resolution* cell-type-specific and bulk-specific expression $\mathrm{s}_{\mathrm{ngk}}^{(\mathrm{mapping})}$:

$$\mathrm{s}_{\mathrm{ngk}}^{(\mathrm{mapping})} = \sum_{\mathrm{i}=1}^{\mathrm{C}} \alpha_{\mathrm{in}} \cdot \mathrm{c}_{\mathrm{ig}} \cdot \tau_{\mathrm{ik}} \tag{7}$$

with α_{in} being the coefficient fitted by GLMNET on each single-cell, linking each cell i to each bulk n. c_{ig} denotes the single-cell expression of cell i and gene g, and τ_{ik} is an indicator vector equal to 1 if cell i is of type k.

Although this is potentially sufficient in cases where scRNA-seq reference dataset maps all cell states present in bulk samples, it can be insufficient to detect genes deferentially-expressed (DE) at cell-type level. As DE genes will likely induce error on the cell-type expression deconvolution, we use the error itself to fit our model. We can also make the reasonable assumption that DE genes only originates in cancer cells. Thus, for a given gene g and a given bulk n, we have:

$$\mathrm{y}_{\mathrm{ng}} \sim \hat{\mathrm{y}}_{\mathrm{ng}}^{(\mathrm{mapping})} + \epsilon_{\mathrm{ng}} \quad \rightarrow \quad \mathrm{y}_{\mathrm{ng}} - \hat{\mathrm{y}}_{\mathrm{ng}}^{(\mathrm{mapping})} = \epsilon_{\mathrm{ng}} \tag{8}$$

Thus, we use $\epsilon_{\mathrm{ng}} = \mathrm{y}_{\mathrm{ng}} - \hat{\mathrm{y}}_{\mathrm{ng}}^{(\mathrm{mapping})}$ as input for a model to predict cancer cell-type expression. We decompose our error ϵ_{ng} between the error originated from DE genes $\epsilon_{\mathrm{ng}}^{(\mathrm{DE})}$, and an additional error term $\epsilon_{\mathrm{ng}}^{*}$. Similar as other methods [ref, ref], we use the information from all bulk samples N to deconvolute the high-resolution cell-type specific expression $\hat{\mathrm{s}}_{\mathrm{ngk}}^{(\mathrm{DE})}$. We consider that $\epsilon_{\mathrm{ng}}^{(\mathrm{DE})}$ only originates from cancer cells $\mathrm{C}^{(\mathrm{C})}$ of type $\mathrm{K}^{(\mathrm{C})}$, thus

$$\epsilon_{ng} \sim \underbrace{\sum_{k=1}^{K^{(C)}} \left(w_{nk} \cdot \hat{s}_{ngk}^{(DE)} \right)}_{\epsilon_{ng}^{(DE)}} + \epsilon_{ng}^{*} = \sum_{i=1}^{C^{(C)}} \left(\alpha_{in} \cdot \hat{c}_{ig} \right) + \epsilon_{ng}^{*} = \mathbf{A} \cdot \hat{\mathbf{c}}_{g} + \epsilon_{ng}^{*} \tag{9}$$

Where $\hat{\mathbf{c}}_g$ is the vector of coefficients we wish to predict. We used Glmnet with options lower.limits = 0, lambda = 0, and scale = TRUE, fitting the following loss function:

Our final prediction of cell-type expression for a cancer cell-type k, a given bulk n, and a given gene g is:

$$\hat{s}_{ngk}^{(high-res)} = \underbrace{\hat{s}_{ngk}^{(mapping)}}_{\text{from scRNA-seq}} + \underbrace{\hat{s}_{ngk}^{(DE)}}_{\text{learned}} = \sum_{i=1}^{C} \alpha_{in} \cdot (c_{ig} + \hat{c}_{ig}) \cdot \tau_{ik} \tag{10}$$

Following the assumptions made above, we set $\hat{c}_{ig}$ to zero for normal cells, which imply that $\hat{s}_{ngk}^{(high-res)} = \hat{s}_{ngk}^{(mapping)}$ for normal cell types.

Differential Expression analysis of deconvoluted cell-type expression

High-resolution deconvolution of cell-type expression allows to perform differential expression (DE) analysis between two conditions, whenever two groups are found is our mixture samples. The deconvoluted expression will be raw counts, or CPM values (normalized at sample level) if CPM values are inputed for mixtures. Nevertheless, these count values will be strongely biased by cell-type proportions. One solution is to normalize the cell-type expression *per cell*, that is to compute an average expression per cell. Another way is to assess DE analysis taking into account cell-type proportion as a covariate, which is the approach we emplyed. We used predicted CPM counts provided by CLIMB to compare two groups using DESeq2 library [ref], using cell-type proportions as covariate.

CLIFF - Cell fate inference for drug sensitivity deconvolution

Here we propose to infer the drug sensitivity of individual cell subtypes using only drug sensitivity data observed on mixed cell populations, considering K different cell populations in N mixed samples, with proportions than can be defined as $\mathbf{w}_n = [w_{n1}, ..., w_{nK}]$, with $w_{nk} \in [0, 1]$ and $\sum_{k=1}^{K} w_{nk} = 1$. Drug sensitivity observed in the bulk samples we wish to demix are represented by a matrix $\mathbf{Y}^{N \times D}$, with $y_{nd} \in [0, 1]$ being the empirical surviving rate of cells, for sample n and drug concentration d. Therefore the target matrix we wish to predict is $\hat{\mathbf{Y}}^{N \times D}$, and the predicted values of drug sensitivity is assumed to depend on the underlying drug sensitivity of cell subtypes $\pi_{ndk} \in [0, 1]$ for patient n, drug concentration d and cell-subtype k, represented as a surviving rate of cells.

To model drug sensitivity, we assume that single-cell survival is represented as a binary output $\delta_{nid} \in \{0, 1\}$ that follows a Bernoulli distribution. Thus, every single-cell i in patient n exposed to a drug dose d, can either survive ($\delta_{nid} = 1$) or die ($\delta_{nid} = 0$). Thus a cell i has a probability $P(\delta_{nid} = 1)$ to survive and $1 - P(\delta_{nid} = 1)$ to die. Let's say we are interested to know the survival

probability $P(\delta_{nid} = 1)$, given a drug dose. We could resolve the log-likelihood of the Bernoulli distribution for all the cells of a single sample n :

$$\mathcal{L}_n = \sum_{i=1}^{C_n} [\delta_{nid} \cdot \log P(\delta_{nid} = 1) + (1 - \delta_{nid}) \cdot \log(1 - P(\delta_{nid} = 1))] \tag{11}$$

It turns out that in our problem, δ_{nid} cannot be directly observed. We are indeed only able to observe the overall ratio of cells that survives among all cells C_n for a given patient n, which we define as y_{nd}. It relates to the values δ_{nid} with :

$$y_{nd} = \frac{1}{C_n} \sum_{i=1}^{C_n} \delta_{nid} \tag{12}$$

Thus, by multiplying (11) by $1/C_n$, if we assume that every cell i share the same average probability $P(\delta_{nid} = 1)$, we can re-write (11) and obtain the following log-likelihood :

$$y_{nd} \cdot \log P(\delta_{nid} = 1) + (1 - y_{nd}) \cdot \log(1 - P(\delta_{nid} = 1)) \tag{13}$$

As stated above, the core of the problem is to assume different survival probability for our single-cells. To slightly simplify this, we can consider a cell-subtype-specific survival probability. Thus, we consider the survival probability observed in our mixed population of cells $P(\delta_{nid} = 1)$, as a mixture of K probabilities $P(\delta_{nid}^{(k)} = 1)$, one for each cell subtypes and each drug doses (we will come later on the drug dose component). Let introduce π_{kdn}, a variable describing the survival probability of cell-subtype k:

$$\pi_{kdn} = P(\delta_{nid}^{(k)} = 1) \tag{14}$$

We also define ν_{nd}, which corresponds to the expected number of cells surviving in patient n with drug dose d, and assume it follows a Binomial distribution, with y_{nd} being the ratio of surviving cells among cells in C_n. We further define $\hat{y}_{nd}$ as an average of cell-subtype survival probabilities π_{kdn} constituting the bulk, weighted with cell-subtype proportions w_{nk}:

$$\nu_{nd} = C_n \cdot \underbrace{\sum_{k=1}^{K} w_{nk} \cdot \pi_{kdn}}_{\hat{y}_{nd}} \tag{15}$$

with w_{nk} the cell-subtype proportions. $\hat{y}_{nd}$ is the predicted cell surviving rate in the sample n, and thus the target we wish to predict through a cross-validation procedure. We can say that if we let :

$$\hat{y}_{nd} = \sum_{k} w_{nk} \cdot \pi_{kdn} \text{ we have that } \hat{y}_{nd} \sim \mathrm{Ber}(\hat{y}_{nd}). \tag{16}$$

Additionally, we impose the cell-subtype-specific survival probabilities π_{kdn} to be non-linearly related with cell-subtype gene expression $\mathbf{s}_{kn}$ through a sigmoid function:

$$\pi_{\mathrm{kdn}} = \sigma(\beta^{\mathrm{T}}\mathbf{s}_{\mathrm{kn}} + \theta_{\mathrm{d}}) \tag{17}$$

Note that the sigmoid in (17) also depends on θ_{d}, which is an intersect associated with drug dose d, inducing a shift on the sigmoid output. To clarify the notation, we set that $\tilde{\boldsymbol{\beta}} = \{\boldsymbol{\beta}, \boldsymbol{\theta}\}$, with $\boldsymbol{\theta}$ being a vector containing an intersect associated with each drug dose d.

There are two main assumptions we use to simplify the problem. First, we assume that every sample n share the same cell-subtype expression s_{gk}, for a given gene g and cell-subtype k. Thus, $\mathrm{s}_{\mathrm{gk}} \approx \mathrm{E}_{\mathrm{N}}[\mathrm{s}_{\mathrm{gkn}}]$. Second, we assume that cells of the same cell-subtypes across the N samples share the same sensitivities π_{kd}. We thus assume that π_{kd} correspond to the expected drug sensitivity over N samples, and therefore that $\pi_{\mathrm{kd}} \approx \mathrm{E}_{\mathrm{N}}[\pi_{\mathrm{kdn}}]$.

Thus, the overall log-likelihood for our problem is :

$$\mathcal{L}(\tilde{\boldsymbol{\beta}} \mid \mathbf{Y}, \mathbf{W}, \mathbf{S}) = \max_{\tilde{\boldsymbol{\beta}}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} \left[\mathrm{y}_{\mathrm{nd}} \log \left(\sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{w}_{\mathrm{nk}} \cdot \pi_{\mathrm{kd}}(\tilde{\boldsymbol{\beta}}) \right) + (1 - \mathrm{y}_{\mathrm{nd}}) \log \left(1 - \sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{w}_{\mathrm{nk}} \cdot \pi_{\mathrm{kd}}(\tilde{\boldsymbol{\beta}}) \right) \right] \tag{18}$$

As this log-likelihood expression contains the log of a sum, it is difficult to optimize with respect to $\tilde{\boldsymbol{\beta}}$. Therefore, we use an Expectation-Maximization to iteratively define lower bound functions that we optimize with respect to $\tilde{\boldsymbol{\beta}}$, which allows solving (18).

Expectation-Maximization algorithm to solve CLIFF

By using the concavity of the log through Jensen inequality, we can define a lower bound of (18) (more details on this in the Supplemental Methods) :

$$\begin{aligned} \mathcal{L}_{\mathrm{B}} = & \sum_{\mathrm{d,n}}^{\mathrm{D,N}} \sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{q}_{\mathrm{ndk}}^{(1)} \cdot \mathrm{y}_{\mathrm{nd}} \cdot \log(\pi_{\mathrm{kd}}(\tilde{\boldsymbol{\beta}})) + \mathrm{q}_{\mathrm{ndk}}^{(0)} \cdot (1 - \mathrm{y}_{\mathrm{nd}}) \cdot \log(1 - \pi_{\mathrm{kd}}(\tilde{\boldsymbol{\beta}})) \\ & - \sum_{\mathrm{d,n}}^{\mathrm{D,N}} \left[\mathrm{y}_{\mathrm{nd}} \cdot \mathrm{KL}(\mathbf{q}_{\mathrm{nd}}^{(1)} \| \mathbf{w}_{\mathrm{n}}) + (1 - \mathrm{y}_{\mathrm{nd}}) \cdot \mathrm{KL}(\mathbf{q}_{\mathrm{nd}}^{(0)} \| \mathbf{w}_{\mathrm{n}}) \right] \end{aligned} \tag{19}$$

In order to solve (18), we will iterate over the Expectation step or E-step where we update $\mathrm{q}_{\mathrm{ndk}}^{(0)}$ and $\mathrm{q}_{\mathrm{ndk}}^{(1)}$ in (19), and the Maximization step or M-step, where we maximize (19) with respect to $\tilde{\boldsymbol{\beta}}_{\mathrm{d}}$. We can therefore define our Maximization algorithm as follow :

E-step

During E-step, we define our lower bound function at step t, $\mathcal{L}_{\mathrm{B}}^{(\mathrm{t})}$ by defining $\mathrm{q}_{\mathrm{ndk}}^{(0),(\mathrm{t})}$ and $\mathrm{q}_{\mathrm{ndk}}^{(1),(\mathrm{t})}$ using $\pi_{\mathrm{dk}}^{(\mathrm{t-1})}$ defined at previous step :

$$\begin{aligned} \mathrm{q}_{\mathrm{ndk}}^{(1),(\mathrm{t})} &= \frac{\mathrm{w}_{\mathrm{nk}} \cdot \pi_{\mathrm{dk}}^{(\mathrm{t-1})}}{\sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{w}_{\mathrm{nk}} \cdot \pi_{\mathrm{dk}}^{(\mathrm{t-1})}} \\ \mathrm{q}_{\mathrm{ndk}}^{(0),(\mathrm{t})} &= \frac{\mathrm{w}_{\mathrm{nk}} \cdot (1 - \pi_{\mathrm{dk}}^{(\mathrm{t-1})})}{\sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{w}_{\mathrm{nk}} \cdot (1 - \pi_{\mathrm{dk}}^{(\mathrm{t-1})})} \end{aligned} \tag{20}$$

This gives us the lower bound function of (19), tight to the original log-likelihood (18) when $\mathrm{q}_{\mathrm{ndk}}^{(0),(\mathrm{t})}$ and $\mathrm{q}_{\mathrm{ndk}}^{(1),(\mathrm{t})}$ satisfies (20). For a precise description of how $\mathrm{q}_{\mathrm{ndk}}^{(0),(\mathrm{t})}$ and $\mathrm{q}_{\mathrm{ndk}}^{(1),(\mathrm{t})}$ were defined, see Supplemental Methods.

M-step

During M-step, we optimize (19) with respect to $\tilde{\boldsymbol{\beta}}$, considering the variables $\mathbf{q}^{(0),(\mathrm{t})}$ and $\mathbf{q}^{(0),(\mathrm{t})}$ defined during E-step as constants. Also, the two last term of (19) do not depend on $\tilde{\boldsymbol{\beta}}$, thus we can solve the following problem which will produce the same optimal $\tilde{\boldsymbol{\beta}}$ coefficients than if solving (19) :

$$\mathcal{L}_{\mathrm{B}} = \max_{\tilde{\boldsymbol{\beta}}} \sum_{\mathrm{d,n}}^{\mathrm{D,N}} \sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{q}_{\mathrm{ndk}}^{(1)} \cdot \mathrm{y}_{\mathrm{nd}} \cdot \log(\pi_{\mathrm{kd}}(\tilde{\boldsymbol{\beta}}_{\mathrm{d}})) + \mathrm{q}_{\mathrm{ndk}}^{(0)} \cdot (1 - \mathrm{y}_{\mathrm{nd}}) \cdot \log(1 - \pi_{\mathrm{kd}}(\tilde{\boldsymbol{\beta}}_{\mathrm{d}})) \tag{21}$$

Interestingly, (21) can be solved with standard libraries using a weighted logistic regression model, where we define weights as being $\boldsymbol{\gamma}_{\mathrm{dk}}^{(1)} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{q}_{\mathrm{ndk}}^{(1)} \cdot \mathrm{y}_{\mathrm{nd}}$ and $\boldsymbol{\gamma}_{\mathrm{dk}}^{(0)} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{q}_{\mathrm{ndk}}^{(0)} \cdot (1 - \mathrm{y}_{\mathrm{nd}})$.

We optimize $\tilde{\boldsymbol{\beta}}$ coefficients through a ten-fold cross-validation (taking 90% of the samples N in the training set, and 10% in the testing set), and using early-stopping to terminate the EM algorithm. Optimized $\tilde{\boldsymbol{\beta}}$ coefficients allow to compute optimized values for π_{kd}, our estimate of cell-subtype viability rate.

CLIMB-SC - bulk deconvolution at single-cell level

The bulk deconvolution performed by CLIMB has the particularity of estimating coefficient directly on single-cells from a reference scRNA-seq dataset. Instead of transforming these coefficients into cell-type proportions as in Methods (3), we can compute normalized coefficients $\tilde{\alpha}_{\mathrm{ni}}$ with the following formula :

$$\tilde{\alpha}_{\mathrm{ni}} = \frac{\alpha_{\mathrm{ni}}}{\sum_{\mathrm{i=1}}^{\mathrm{C}} \alpha_{\mathrm{ni}}} \tag{22}$$

With α_{ni} being the coefficients fitted with GLMNET library. We therefore call this *single-cell bulk deconvolution*, which links every bulk RNA-seq n to every single-cells i from a scRNA-seq dataset, through what we will call *single-cell proportions* $\tilde{\alpha}_{\mathrm{ni}}$. These *single-cell proportions* are thus normalized coefficients fitted by CLIMB, as shown above in (22). Thus we obtain that $\sum_{\mathrm{i=1}}^{\mathrm{C}} \tilde{\alpha}_{\mathrm{ni}} = 1$. As $\tilde{\alpha}_{\mathrm{ni}}$ sum up to 1, they can be used by CLIFF as if they were cell-type proportions.

CLIFF-SC for drug sensitivity inference at single-cell level

By slightly modifying CLIFF method and by linking it with the ouput of CLIMB-SC, we propose CLIFF-SC, a method providing drug sensitivity estimates at single-cell resolution. If the input dataset is composed of paired bulk RNA-seq and drug sensitivity screening data, a so-called *bulk deconvolution at single-cell level* is performed using a reference scRNA-seq dataset (cf. CLIMB-SC above).

Thus, as in (16), we model the observed ratio of cells $\mathrm{y_{nd}}$ that survives in a mixed population of cells as an average of single-cell survival probabilities π_{id}, associated with the bulk sample n, weighted by the coefficients $\tilde{\alpha}_{\mathrm{ni}}$:

$$\hat{\mathrm{y}}_{\mathrm{nd}} = \sum_{\mathrm{i}=1}^{\mathrm{C}} \tilde{\alpha}_{\mathrm{ni}} \cdot \pi_{\mathrm{id}} \tag{23}$$

With C being the number of cells contained in a scRNA-seq reference datasets. Thus, we model the single-cell drug sensitivity π_{id}, which is also modeled with a sigmoid function, similar to (17) :

$$\pi_{\mathrm{id}} = \sigma(\beta^{\mathrm{T}} \mathbf{c}_{\mathrm{i}} + \theta_{\mathrm{d}}) \tag{24}$$

Where $\mathbf{c}_{\mathrm{i}}$ is the ith row of the single-cell expression matrix $\mathbf{C}$, thus $\mathbf{c}_{\mathrm{i}}$ is a vector of length G, representing the gene expression of cell i. Similarly as in the cell-type-centered model, we can derive a log-likelihood function :

$$\mathcal{L}(\tilde{\boldsymbol{\beta}} \mid \mathbf{Y}, \mathbf{W}, \mathbf{C}) = \max_{\tilde{\boldsymbol{\beta}}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \sum_{\mathrm{d}=1}^{\mathrm{D}} \left[\mathrm{y_{nd}} \log \left(\sum_{\mathrm{i}=1}^{\mathrm{C}} \tilde{\alpha}_{\mathrm{ni}} \cdot \pi_{\mathrm{id}}(\tilde{\boldsymbol{\beta}}) \right) + (1 - \mathrm{y_{nd}}) \log \left(1 - \sum_{\mathrm{i}=1}^{\mathrm{C}} \tilde{\alpha}_{\mathrm{ni}} \cdot \pi_{\mathrm{id}}(\tilde{\boldsymbol{\beta}}) \right) \right] \tag{25}$$

A lower bound can be derived from (25), which gives :

$$\begin{aligned} \mathcal{L}_{\mathrm{B}} = & \sum_{\mathrm{d,n}}^{\mathrm{D,N}} \sum_{\mathrm{i}=1}^{\mathrm{C}} \mathrm{q}_{\mathrm{ndi}}^{(1)} \cdot \mathrm{y_{nd}} \cdot \log(\pi_{\mathrm{id}}(\tilde{\boldsymbol{\beta}})) + \mathrm{q}_{\mathrm{ndi}}^{(0)} \cdot (1 - \mathrm{y_{nd}}) \cdot \log(1 - \pi_{\mathrm{id}}(\tilde{\boldsymbol{\beta}})) \\ & - \sum_{\mathrm{d,n}}^{\mathrm{D,N}} \left[\mathrm{y_{nd}} \cdot \mathrm{KL}(\mathbf{q}_{\mathrm{nd}}^{(1)} \| \mathbf{w}_{\mathrm{n}}) + (1 - \mathrm{y_{nd}}) \cdot \mathrm{KL}(\mathbf{q}_{\mathrm{nd}}^{(0)} \| \mathbf{w}_{\mathrm{n}}) \right] \end{aligned} \tag{26}$$

Then, the Expectation-Maximization algorithm can initialized with all coefficients $\tilde{\boldsymbol{\beta}}$ at 0. This implies that $\pi_{\mathrm{id}}^{(\mathrm{t}=0)} = 0.5$ for all i, d. During E-step, we compute values for $\mathrm{q}_{\mathrm{ndi}}^{(0)}$ and $\mathrm{q}_{\mathrm{ndi}}^{(1)}$ variables :

$$\begin{aligned} \mathrm{q}_{\mathrm{ndi}}^{(1),(\mathrm{t})} &= \frac{\tilde{\alpha}_{\mathrm{ni}} \cdot \pi_{\mathrm{id}}^{(\mathrm{t}-1)}}{\sum_{\mathrm{i}=1}^{\mathrm{C}} \tilde{\alpha}_{\mathrm{ni}} \cdot \pi_{\mathrm{di}}^{(\mathrm{t}-1)}} \\ \mathrm{q}_{\mathrm{ndi}}^{(0),(\mathrm{t})} &= \frac{\tilde{\alpha}_{\mathrm{ni}} \cdot (1 - \pi_{\mathrm{id}}^{(\mathrm{t}-1)})}{\sum_{\mathrm{i}=1}^{\mathrm{C}} \tilde{\alpha}_{\mathrm{ni}} \cdot (1 - \pi_{\mathrm{id}}^{(\mathrm{t}-1)})} \end{aligned} \tag{27}$$

Then, we optimize $\tilde{\boldsymbol{\beta}}$ coefficients at each step (t), during the so-called M-step :

$$\mathcal{L}_{\mathrm{B}} = \max_{\boldsymbol{\beta},\theta} \sum_{\mathrm{d,n}}^{\mathrm{D,N}} \sum_{\mathrm{i=1}}^{\mathrm{C}} \mathrm{q}_{\mathrm{ndi}}^{(1)} \cdot \mathrm{y}_{\mathrm{nd}} \cdot \log(\pi_{\mathrm{id}}(\tilde{\boldsymbol{\beta}})) + \mathrm{q}_{\mathrm{ndi}}^{(0)} \cdot (1 - \mathrm{y}_{\mathrm{nd}}) \cdot \log(1 - \pi_{\mathrm{id}}(\tilde{\boldsymbol{\beta}})) \tag{28}$$

And we run the EM algorithm in a similar way than shown above for CLIFF.

Evaluation of deconvolution methods in Acute Myeloid Leukemia through fourteen metrics

To evaluate the performance of the deconvolution algorithms in the context of real cancer samples, we added as many metrics as possible. In summary, we compared cell-type proportion estimated by deconvolution algorithms correlated with previously defined signatures of cell-types (using Spearman correlation coefficients). We used the correlation between HSC-like with the signature of leukemic stem cells, the LSC17 score (Ng et al., 2016) ('HSC-like vs LSC17' score), or the correlation between the cell-type signature score defined by (van Galen et al., 2019) and the corresponding cell type content ('VG_sign_HSC-like', 'VG_sign_Prog-like', 'VG_sign_GMP-like', 'VG_sign_ProMono-like', 'VG_sign_Mono-like', 'VG_sign_cDC-like' [ref]). Signature score were obtained by averaging logCPM values of the 50-gene signatures. Then, we derived a score using the Area under the ROC curve (AUROC) to compare FAB classification binary labels (e.g. 'PML-RARA' vs 'other types') with cell-type proportions estimated by deconvolution algorithms, expected to be over-represented in the corresponding FAB subgroup ('M0 vs HSC-like', 'PML-RARA vs GMP-like', 'M5 vs Mono-like', 'M6 vs Erythroblasts'). We thus evaluate the predictive power of cell-type proportions to "classify" the labels of their associated FAB class through an AUROC score. Then, we use cancer cell content evaluated by microscopy/histology and correlate it with the deconvoluted proportions summed over all cancer cell types ('%Blast prediciton'). We also used qualitative value of surface marker expression to derived a score (only available for BeatAML cohort, described below). As we can assume that patients in our cohorts should roughly display the same ranking of cell-type abundance, we used a spearman correlation coefficient to compare the average proportions with the patients in our scRNA-seq reference ('scRNA-seq concordance').

Evaluation of deconvolution methods in Colo-Rectal Cancer through nine metrics

To evaluate deconvolution methods in the context of colorectal cancer, we used different angles. Lee et al. (Lee et al., 2020) defined 31 cell subtypes, including 4 can subtypes corresponding to the four CMS classification types (CMS1-4). Thus, we correlated the cell subtypes proportions estimated by deconvolution methods with the probability given by CMSclassifier (Guinney et al., 2015) to belong to each CRC cancer subtype. We removed CMS4 as no deconvolution method could obtain a significant correlation (probably because only 14 cells were labeled 'CMS4' in Lee et al. dataset (Lee et al., 2020)). We then used estimate of stromal, immune and cancer cell content given by ESTIMATE algorithm (Yoshihara et al., 2013), correlated with deconvoluted proportions of stromal, immune, and cancer cell content, as an additional metric ('immune', 'stromal', 'cancer'). We then correlated cancer cell proportions with estimations from microscopy/histology, and computed concordance of cell-type proportions with scRNA-seq data, as described in previous section.

Simulation of pseudo-bulk samples from scRNA-seq

We generated pseudo-bulk samples by summing up single-cell expression vectors per gene to obtain an artificially generated "pseudo"-bulk RNA-seq. As the precise amount of cells from each cell subtype added to the mix is known, we thus precisely know the cell-subtype proportions of the mixture and can benchmark deconvolution algorithms efficiently. We generated two datasets of pseudo-bulk in this work, both based on single-cells found in the Van Galen et al. dataset.

First, we used all single cells from the same sample (thus, the same patient), which we summed up to obtain a patient-specific pseudo-bulk RNA-seq sample. Thus, we define a variable ϕ_{ni} as a binary output variable indicating if a cell i belongs to patient n. Thus, a pseudo-bulk samples is a vector of length G containing the gene expression summed over all cells from patient n :

$$\tilde{\mathbf{x}}_{\mathrm{n}} = \sum_{\mathrm{i}=1}^{\mathrm{C}} \phi_{\mathrm{ni}} \cdot \mathbf{c}_{\mathrm{i}} \tag{29}$$

where $\mathbf{c}_{\mathrm{i}}$ is a vector of length G representing the gene expression of cell i. We can therefore easily get the cell-subtype proportion w_{nk}, by simply counting the number of cells from each cell subtypes and dividing by the total number of cells in patient n. For this we introduce an indicator vector $\boldsymbol{\tau}_{\mathrm{k}}$, which is of length C and equal to 1 whenever cell i is of type k.

$$\mathrm{w}_{\mathrm{nk}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{C}} \phi_{\mathrm{ni}} \cdot \tau_{\mathrm{ik}}}{\sum_{\mathrm{i}=1}^{\mathrm{C}} \phi_{\mathrm{ni}}} \tag{30}$$

Van Galen et al. dataset offers only a limited number of patients (16 AML patients at day 0 pre-treatment). To simulate an *in-silico* a dataset of similar size to the beatAML cohort, we generated 200 pseudo-bulk samples from the Van Galen et al. dataset allowing a random mix of cells from multiple patients. First, we generated 200 random cell-subtype proportions using a Dirichlet distribution. Thus, we assume that cell-subtype proportions $\mathbf{w}_{\mathrm{n}} \sim \mathrm{Dir}(\boldsymbol{\alpha})$, with $\boldsymbol{\alpha}$ being a vector of length K, containing one value for each cell subtype k. We fitted $\boldsymbol{\alpha}$ using cell-subtype proportions in Van Galen et al. dataset with the R library *feralaes/dampack*. Then, we generated cell-subtype proportions for 200 pseudo-bulks, with the R library *DirichletReg* and the $\boldsymbol{\alpha}$ previously fitted. Cells from Van Galen et al. dataset were selected with probabilities corresponding to cell-subtype proportions. To use a realistic number of cells to generate our pseudo-bulk samples, we used a gamma distribution fitted on Van Galen et al. dataset using *fitdistr* function from R library *MASS*. In the end, we generated a matrix of pseudo-bulk expression $\bar{\mathbf{X}}$ of size $(\mathrm{N} \times \mathrm{G})$ with $\mathrm{N} = 200$, together with associated true cell-subtype proportions $\bar{\mathbf{W}}$ of size $(\mathrm{N} \times \mathrm{K})$.

To simulate batch effect, we separated Van Galen et al. dataset between a "reference" (used as a scRNA-seq), and a "target" dataset (used to generate pseudo-bulk), by using the 16 patients from Van Galen et al. at day 0. When batch level is 100%, target/reference datasets are composed of different patients. When batch level is 0%, target/reference datasets are composed of random cells coming from any patients. When 0%<Batch<100%, we mixed the two target/reference datasets abovementionned with different ratios, generating different levels of Batch effect between 0% and 100%.

Simulation of drug sensitivity screening data

To simulate drug screening data, we assumed that the drug sensitivity observed on a mix of cells corresponds to the sum of the cell-subtype drug sensitivity constituting its bulk weighted by their respective proportions. Additionally, we added Gaussian noise with a signal-to-noise ratio of 100. Thus, the simulated survival rate for pseudo-bulk n, in cell-subtype k exposed to drug dose d was generated with the following formula, given the proportions :

$$\bar{\pi}_{\mathrm{kd}} = \frac{1}{1 + \exp(\zeta_{\mathrm{f}} \cdot \mathrm{d} + \mathrm{c})} \tag{31}$$

with ζ_{f} being a randomly generated number, with $\zeta_{\mathrm{f}} \in [1, 2]$ if $\mathrm{f} = 1$ (low sensibility cell-subtypes - resisting cells) and $\zeta_{\mathrm{f}} \in [5, 6]$ if $\mathrm{f} = 2$ (high sensibility cell-subtypes). Here we did not use the drug dose d, only the integer number related to drug dose $\mathrm{d} = \{0, 1, ..., 6\}$, as we considered seven drug doses. We arbitrarily defined three cell subtypes with high sensitivity and the rest having low sensitivity. The constant c added was set to $\mathrm{c} = -8$ as it provided visually convincing results. Thus $\bar{\pi}_{\mathrm{kd}}$ follows a sigmoid function that depends on drug dose d but in a way that is slightly different than in our model. Then, we generated the survival rate at the bulk level using the same formula as in our model but adding gaussian noise to it :

$$\bar{\mathrm{y}}_{\mathrm{nd}} = \sum_{\mathrm{k}=1}^{\mathrm{K}} \bar{\mathrm{w}}_{\mathrm{nk}} \cdot \bar{\pi}_{\mathrm{kd}} + \epsilon \quad \text{with} \quad \epsilon \in \mathcal{N}(0, \sigma^2) \tag{32}$$

and σ was chosen to generate a signal-to-noise ratio arbitrarily chosen at 100.

Area under the curve for summarizing drug sensitivity of multiple data points

As CLIFF produces D predictions for each cell subtype, and CLIFF-SC produces D predictions for every single cell, it is desired to summarize these D data points into a single value. The goal is to summarize cell-subtype or single-cell drug sensitivity. To summarize drug sensitivity from D different drug doses, we normalize drug doses so that the minimum dose = 0 and the maximum dose = 1. As the cell survival rate is $\in [0, 1]$, we can compute an area under the curve (AUC) which will be $\in [0, 1]$. Conversely, we get an AUC of 1.0 if all cell survives and an AUC of 0.0 if all cells die, taking every dose into account. We computed the AUC using the function *computeAUC* from the R package *PharmacoGx*.

Marker score for testing bulk deconvolution

Assessing the accuracy of bulk deconvolution of the beatAML cohort is challenging as the true cell-subtype proportions are unknown. Among the clinical data available for the patients in the beatAML cohort, one can find qualitative information about surface marker expression. We indeed have access to a label indicating if patient n is positive for a given marker m. The presence/absence of surface markers in a sample is informative about its cell-subtype contents, as specific markers are only expressed on the surface of certain cell subtypes. Thus, we decided to build a score made from the comparison of two groups of patients: the patients *positive* for a given marker m (samples labeled positive in the clinic), and the patients *negative* for the marker m (no mention of the marker for this patient). Then, we compare the cell-subtype proportions associated with marker m between

positive/negative patients and record the difference as a t-statistic. Thus, for a given marker m, we define $N_m^{(+)}$ positive patients and $N_m^{(-)}$ negative patients, and we compare them with the following formula:

$$t_m = \frac{\bar{X}_m^{(+)} - \bar{X}_m^{(-)}}{s_p \sqrt{\frac{1}{N_m^{(+)}} + \frac{1}{N_m^{(-)}}}} \tag{33}$$

where s_p is the pooled standard deviation [ref]. $\bar{X}_m^{(+)}$ and $\bar{X}_m^{(-)}$ define the sum of all cell-subtypes K_m expected to be associated with marker m :

$$\begin{aligned} \bar{X}_m^{(+)} &= \frac{1}{N_m^{(+)}} \sum_{n=1}^{N_m^{(+)}} \sum_{k=1}^{K_m} w_{nk} \\ \bar{X}_m^{(-)} &= \frac{1}{N_m^{(-)}} \sum_{n=1}^{N_m^{(-)}} \sum_{k=1}^{K_m} w_{nk} \end{aligned} \tag{34}$$

In this situation, if the cell-subtype proportions are well estimated, and the marker is indeed expressed in the expected cell subtypes, we expect the score the be positive and high whenever the cell-subtype proportions are estimated correctly. Thus t_m gives a metric that can score bulk deconvolution in a context where ground truth is unavailable. For a given bulk deconvolution method, we, therefore, sum all t_m across M different markers to obtain our final score, that we call marker score :

$$\text{Marker Score } = \sum_{m=1}^{M} t_m \tag{35}$$

In this study, we used 13 different markers to compute the scores : CD33, CD4, CD13, CD11b, CD14, CD34, CD38, CD58, CD123, CD15,CD45, HLA-DR, and CD56. We chose this set of markers because they had at least ten positive and ten negative patients in the beatAML cohort.

Culture conditions of cancer cell lines

Human cancer cell lines (K562, HL60, THP1, SUDHL4) were purchased from the American Type Culture Collection (ATCC; Rockville, MD, USA) and cultured at 37°C in a humidified 5% CO2 environment in RPMI 1640 (HCC38) supplemented with 2mM L-glutamine, 2 g/L D-glucose, and 10% fetal bovine serum (FBS; Gibco ThermoFisher).

Pharmaceutical compounds

Stock solutions for every agent (Table2) were prepared using DMSO (Sigma-Aldrich; stored at 80°C), further diluted in H2O to the appropriate concentration, plated in Greiner 96-well flat bottom (REF) and the plates were stored at 20°C. The pharmaceutical compounds were screened at eight concentrations (Log concentration M = -9.205031395, -8.699881554, -8.130127907, -7.70415926, -7.209309101, -6.70415926, -6.125806332, -5.699837685, -5.204987526, -4.699837685) using a Log2-fold dilution series with matched DMSO concentration vehicle controls.

CatalogID	Name	Supplier
A3656	5-Aza-2'-deoxycytidine	Sigma
C3350000	Cytarabine	Sigma
CDS022173	Imatinib	Sigma
CDS023093	Nilotinib	Sigma
HY-15142	Doxorubicin (hydrochloride)	MedChemExpress
O9512	Oxaliplatin	Sigma
R0395	Rapamycin	Sigma
Y0000378	Mitomycin	Sigma

Table 1: Table of eight coumpounds used in the *in-vitro* drug sensitivity experiment.

Bulk proportion for bulk RNAseq and drug sensitivity

Six bulks were created by mixing the four cell lines with the following proportions. Pure and mixed cell populations were subjected to RNA extraction and drug sensitivity immediately after mixing.

Cell line	Mixture1	Mixture2	Mixture3	Mixture4	Mixture5	Mixture6
SUDHL4	0.33	0.07	0.00	0.61	0.188	0.20
K562	0.09	0.11	0.222	0.04	0.317	0.26
HL60	0.26	0.57	0.333	0.05	0.0198	0.12
THP1	0.32	0.25	0.444	0.30	0.475	0.42

Table 2: Table of cell line proportions added in tubes manually to generate the mixtures of cell lines.

Cell viability assay

Cells were plated in the 96-well plates containing chemical compounds previously equilibrated at 37C, at a density of 1×104 cells per well in 100µl culture medium. Cell viability was assessed after drug treatment for 24, 48, 72 or 96 in duplicates hours by incubating the cells with PrestoBlue™ Cell Viability Reagent, (Thermo Fisher Scientific - Life Technologies A13261). PrestoBlue was at room temperature before adding to the cells in each well with. The cells were incubated with 10µl PrestoBlue (10% of cell culture volume) for two hours at 37°C. The fluorescence was measured by the bottom with a 560nm excitation filter and a 590nm emission filter in a BioTek Neo HTS microplate. The acquisitions were performed with the following parameters: Gain: 54, Light Source Xenon Flash, Lamp Energy: High, Read Speed: Normal, Delay 0 msec, Mesurements/Data Point: 10, Read Height 5.25mm.

RNA extraction and sequencing

RNA-seq Total RNA from cell lines was isolated with NucleoSpin™ RNA Plus kit (Machery-Nagel). cDNA was prepared with Maxima Reverse Transcriptase (Thermo Scientific). Sequencing libraries were performed with Illumina Truseq Stranded mRNA LT kit. Reads were mapped to the human (hg19) genome using Hisat2. Counts on genes were generated using featureCounts. Normalization for sequencing depth has been done using the counts on genes as library size using the TMM method as implemented in the limma package of Bioconductor.

Single Cell RNAseq (MULTIome-Seq)

Single cell RNAseq was performed using 10x Genomics Chromium Single Cell Controller following the 10x Genomics Multiome (scRNA/scATAC-seq) demonstrated protocol: Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338). To note, experimental parameters specific to the experiment were 3,000 for the Targeted Nuclei Recovery and 5 minutes for the lysis time.

Single-cell RNA-seq analysis

Cellranger-arc was used to obtain counts on genes using default parameters. Counts were obtained using cellranger using hg19. Only uniquely mapped reads on genes that were expressed in at least 1% of the samples were kept. Seurat's SCTransform was used to normalize the data and correct for mitochondrial percentage and total number of reads biases.

Data Availability

The data produced for the in-vitro experiment shown in Fig. 2 and Fig. 4 can be found on a Gene Expression Omnibus repository (GEO): GSE217924 (superserie). The bulk RNA-seq of the mixtures is available on the GSE217919 repository and the single-cell RNA-seq data on the GSE217922 repository. The rest of the study was relying on re-analysis of existing data, openly available at locations cited in the Reference section. Installation guides and examples are provided.

Code Availability

The code is freely accessible on GitHub. CLIMB code is accessible on `https://github.com/alexdray86/CLIMB` repository with the documentation available on `https://alexdray86.github.io/CLIMB/build/index.html` website. Similarly, CLIFF is accessible on `https://github.com/alexdray86/CLIFF` repository with its documentation available on `https://alexdray86.github.io/CLIFF/build/index.html`.

References

D. Bottomly, N. Long, A. R. Schultz, S. E. Kurtz, C. E. Tognon, K. Johnson, M. Abel, A. Agarwal, S. Avaylon, E. Benton, A. Blucher, U. Borate, T. P. Braun, J. Brown, J. Bryant, R. Burke, A. Carlos, B. H. Chang, H. J. Cho, S. Christy, C. Coblentz, A. M. Cohen, A. d'Almeida, R. Cook, A. Danilov, K. T. Dao, M. Degnin, J. Dibb, C. A. Eide, I. English, S. Hagler, H. Harrelson, R. Henson, H. Ho, S. K. Joshi, B. Junio, A. Kaempf, Y. Kosaka, T. Laderas, M. Lawhead, H. Lee, J. T. Leonard, C. Lin, E. F. Lind, S. Q. Liu, P. Lo, M. M. Loriaux, S. Luty, J. E. Maxson, T. Macey, J. Martinez, J. Minnier, A. Monteblanco, M. Mori, Q. Morrow, D. Nelson, J. Ramsdill, A. Rofelty, A. Rogers, K. A. Romine, P. Ryabinin, J. N. Saultz, D. A. Sampson, S. L. Savage, R. Schuff, R. Searles, R. L. Smith, S. E. Spurgeon, T. Sweeney, R. T. Swords, A. Thapa, K. Thiel-Klare, E. Traer, J. Wagner, B. Wilmot, J. Wolf, G. Wu, A. Yates, H. Zhang, C. R. Cogle, R. H. Collins, M. W. Deininger, C. S. Hourigan, C. T. Jordan, T. L. Lin, M. E. Martinez, R. R. Pallapati, D. A. Pollyea, A. D. Pomicter, J. M. Watts, S. J. Weir, B. J. Druker, S. K. McWeeney, and J. W. Tyner. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell*, 40(8):850–864, Aug 2022.

J. B. Bramsen, M. H. Rasmussen, H. Ongen, T. B. Mattesen, M. W. rntoft, S. S. ttir, J. Sandoval, T. Laguna, S. Vang, B. ster, P. Lamy, M. R. Madsen, S. Laurberg, M. Esteller, E. T. Dermitzakis, T. F. rntoft, and C. L. Andersen. Molecular-Subtype-Specific Biomarkers Improve Prediction of Prognosis in Colorectal Cancer. *Cell Rep*, 19(6):1268–1280, May 2017.

J. R. Burke, M. A. Pattoli, K. R. Gregor, P. J. Brassil, J. F. MacMaster, K. W. McIntyre, X. Yang, V. S. Iotzova, W. Clarke, J. Strnad, Y. Qiu, and F. C. Zusi. BMS-345541 is a highly selective inhibitor of I kappa B kinase that binds at an allosteric site of the enzyme and blocks NF-kappa B-dependent transcription in mice. *J Biol Chem*, 278(3):1450–1456, Jan 2003.

Y. Chen, Y. Wang, Y. Chen, Y. Cheng, Y. Wei, Y. Li, J. Wang, Y. Wei, T. F. Chan, and Y. Li. Deep autoencoder for interpretable tissue-adaptive deconvolution and cell-type-specific gene analysis. *Nat Commun*, 13(1):6735, Nov 2022.

T. Chu, Z. Wang, D. Pe'er, and C. G. Danko. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer*, 3(4):505–517, Apr 2022.

M. R. Corces, J. M. Granja, S. Shams, B. H. Louie, J. A. Seoane, W. Zhou, T. C. Silva, C. Groeneveld, C. K. Wong, S. W. Cho, A. T. Satpathy, M. R. Mumbach, K. A. Hoadley, A. G. Robertson, N. C. Sheffield, I. Felau, M. A. A. Castro, B. P. Berman, L. M. Staudt, J. C. Zenklusen, P. W. Laird, C. Curtis, W. J. Greenleaf, H. Y. Chang, R. Akbani, C. C. Benz, E. A. Boyle, B. M. Broom, A. D. Cherniack, B. Craft, J. A. Demchok, A. S. Doane, O. Elemento, M. L. Ferguson, M. J. Goldman, D. N. Hayes, J. He, T. Hinoue, M. Imielinski, S. J. M. Jones, A. Kemal, T. A. Knijnenburg, A. Korkut, D. C. Lin, Y. Liu, M. K. A. Mensah, G. B. Mills, V. P. Reuter, A. Schultz, H. Shen, J. P. Smith, R. Tarnuzzer, S. Trefflich, Z. Wang, J. N. Weinstein, L. C. Westlake, J. Xu, L. Yang, C. Yau, Y. Zhao, and J. Zhu. The chromatin accessibility landscape of primary human cancers. *Science*, 362(6413), Oct 2018.

S. de Mel, M. B. M. Rashid, X. Y. Zhang, J. Goh, C. T. Lee, L. M. Poon, E. H. L. Chan, X. Liu, W. J. Chng, Y. L. Chee, J. Lee, Y. C. Yuen, J. Q. Lim, B. K. H. Chia, Y. Laurensia, D. Huang, W. L. Pang, D. M. Z. Cheah, E. K. Y. Wong, C. K. Ong, T. Tang, S. T. Lim, S. B. Ng, S. Y. Tan, H. Y. Loi, L. K. Tan, E. K. Chow, and A. D. Jeyasekharan. Application of an ex-vivo drug sensitivity platform towards achieving complete remission in a refractory T-cell lymphoma. *Blood Cancer J*, 10(1):9, 01 2020.

E. S.-L. P. A. v. V. M. Z. J. J. v. D. A. M. O. C. J. H. P. v. B. H. V. M. J. R. M. G. E. B. v. d. A. E Onur Karakaslar, Jeppe Severens. Single cell guided deconvolution of bulk transcriptomics recapitulates differentiation stages of acute myeloid leukemia and predicts drug response. *Bioarxiv*, 2022.

J. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*, 33(1):1–22, 2010.

J. Guinney, R. Dienstmann, X. Wang, A. s, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B. M. Bot, J. S. Morris, I. M. Simon, S. Gerster, E. Fessler, F. De Sousa E Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G. C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J. W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S. H. Friend, P. Laurent-Puig, J. P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, and S. Tejpar. The consensus molecular subtypes of colorectal cancer. *Nat Med*, 21(11):1350–1356, Nov 2015.

P. Hu, W. Zhang, H. Xin, and G. Deng. Single Cell Isolation and Analysis. *Front Cell Dev Biol*, 4: 116, 2016.

B. Jew, M. Alvarez, E. Rahmani, Z. Miao, A. Ko, K. M. Garske, J. H. Sul, K. H. Pietiläinen, P. Pajukanta, and E. Halperin. Nat CommunAccurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat Commun*, 11(1):1971, 04 2020.

N. Kim, H. H. Eum, and H. O. Lee. Clinical Perspectives of Single-Cell RNA Sequencing. *Biomolecules*, 11(8), 08 2021.

L. Kohtamäki, M. Arjama, S. Mäkelä, P. Ianevski, K. Välimäki, S. Juteau, S. Ilmonen, D. Ungureanu, O. Kallioniemi, A. Murumägi, and M. Hernberg. High-throughput ex vivo drug testing identifies potential drugs and drug combinations for NRAS-positive malignant melanoma. *Transl Oncol*, 15 (1):101290, Jan 2022.

H. O. Lee, Y. Hong, H. E. Etlioglu, Y. B. Cho, V. Pomella, B. Van den Bosch, J. Vanhecke, S. Verbandt, H. Hong, J. W. Min, N. Kim, H. H. Eum, J. Qian, B. Boeckx, D. Lambrechts, P. Tsantoulis, G. De Hertogh, W. Chung, T. Lee, M. An, H. T. Shin, J. G. Joung, M. H. Jung, G. Ko, P. Wirapati, S. H. Kim, H. C. Kim, S. H. Yun, I. B. H. Tan, B. Ranjan, W. Y. Lee, T. Y. Kim, J. K. Choi, Y. J. Kim, S. Prabhakar, S. Tejpar, and W. Y. Park. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet*, 52(6): 594–603, Jun 2020.

Leucegene. *https://data.leucegene.iric.ca/*.

T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing, G. Eley, S. Alonso, B. Ayala, J. Baboud, M. Backus, S. P. Barletta, D. L. Berton, A. L. Chu, S. Girshik, M. A. Jensen, A. Kahn, P. Kothiyal, M. C. Nicholls, T. D. Pihl, D. A. Pot, R. Raman, R. N. Sanbhadti, E. E. Snyder, D. Srinivasan, J. Walton, Y. Wan, Z. Wang, J. P. Issa, M. Le Beau, M. Carroll, H. Kantarjian, S. Kornblau, M. S. Bootwalla, P. H. Lai, H. Shen, D. J. Van Den Berg, D. J. Weisenberger, D. C. Link, M. J. Walter, B. A. Ozenberger, E. R. Mardis, P. Westervelt, T. A. Graubert, J. F. DiPersio, and R. K. Wilson. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*, 368(22):2059–2074, May 2013.

K. Menden, M. Marouf, S. Oller, A. Dalmia, D. S. Magruder, K. Kloiber, P. Heutink, and S. Bonn. Deep learning-based cell composition analysis from tissue expression profiles. *Sci Adv*, 6(30): eaba2619, Jul 2020.

W. H. . S. A. Michael I Love. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 2014.

M. M. Naldini, G. Casirati, M. Barcella, P. M. V. Rancoita, A. Cosentino, C. Caserta, F. Pavesi, E. Zonari, G. Desantis, D. Gilioli, M. G. Carrabba, L. Vago, M. Bernardi, R. Di Micco, C. Di Serio, I. Merelli, M. Volpin, E. Montini, F. Ciceri, and B. Gentner. Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia. *Nat Commun*, 14(1):1285, Mar 2023.

A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, and A. A. Alizadeh. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*, 37(7):773–782, 07 2019.

S. W. Ng, A. Mitchell, J. A. Kennedy, W. C. Chen, J. McLeod, N. Ibrahimova, A. Arruda, A. Popescu, V. Gupta, A. D. Schimmer, A. C. Schuh, K. W. Yee, L. Bullinger, T. Herold, D. rlich, T. chner, W. Hiddemann, W. E. Berdel, B. rmann, M. Cheok, C. Preudhomme, H. Dombret, K. Metzeler, C. Buske, B. wenberg, P. J. Valk, P. W. Zandstra, M. D. Minden, J. E. Dick, and J. C. Wang. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*, 540(7633):433–437, Dec 2016.

M. Rizzi, R. Lorenzetti, K. Fischer, J. Staniek, I. Janowska, A. Troilo, V. Strohmeier, M. Erlacher, M. Kunze, B. Bannert, D. Kyburz, R. E. Voll, N. Venhoff, and J. Thiel. vivo. *J Autoimmun*, 77: 55–66, Feb 2017.

M. A. Spinner, A. Aleshin, M. T. Santaguida, S. A. Schaffert, J. L. Zehnder, A. S. Patterson, C. Gekas, D. Heiser, and P. L. Greenberg. Ex vivo drug screening defines novel drug sensitivity patterns for informing personalized therapy in myeloid neoplasms. *Blood Adv*, 4(12):2768–2778, 06 2020.

L. C. Stetson, D. Balasubramanian, S. P. Ribeiro, T. Stefan, K. Gupta, X. Xu, S. Fourati, A. Roe, Z. Jackson, R. Schauner, A. Sharma, B. Tamilselvan, S. Li, M. de Lima, T. H. Hwang, R. Balderas, Y. Saunthararajah, J. Maciejewski, T. LaFramboise, J. S. Barnholtz-Sloan, R. P. Sekaly, and D. N. Wald. Single cell RNA sequencing of AML initiating cells reveals RNA-based evolution during disease progression. *Leukemia*, 35(10):2799–2812, 10 2021.

Y. Tan, Q. Wu, and F. Zhou. Crit Rev Oncol HematolTargeting acute myeloid leukemia stem cells: Current therapies in development and potential strategies with new dimensions. *Crit Rev Oncol Hematol*, 152:102993, Aug 2020.

J. W. Tyner, C. E. Tognon, D. Bottomly, B. Wilmot, S. E. Kurtz, S. L. Savage, N. Long, A. R. Schultz, E. Traer, M. Abel, A. Agarwal, A. Blucher, U. Borate, J. Bryant, R. Burke, A. Carlos, R. Carpenter, J. Carroll, B. H. Chang, C. Coblentz, A. d'Almeida, R. Cook, A. Danilov, K. T. Dao, M. Degnin, D. Devine, J. Dibb, D. K. Edwards, C. A. Eide, I. English, J. Glover, R. Henson, H. Ho, A. Jemal, K. Johnson, R. Johnson, B. Junio, A. Kaempf, J. Leonard, C. Lin, S. Q. Liu, P. Lo, M. M. Loriaux, S. Luty, T. Macey, J. MacManiman, J. Martinez, M. Mori, D. Nelson, C. Nichols, J. Peters, J. Ramsdill, A. Rofelty, R. Schuff, R. Searles, E. Segerdell, R. L. Smith, S. E. Spurgeon, T. Sweeney, A. Thapa, C. Visser, J. Wagner, K. Watanabe-Smith, K. Werth, J. Wolf, L. White, A. Yates, H. Zhang, C. R. Cogle, R. H. Collins, D. C. Connolly, M. W. Deininger, L. Drusbosky, C. S. Hourigan, C. T. Jordan, P. Kropf, T. L. Lin, M. E. Martinez, B. C. Medeiros, R. R. Pallapati, D. A. Pollyea, R. T. Swords, J. M. Watts, S. J. Weir, D. L. Wiest, R. M. Winters, S. K. McWeeney, and B. J. Druker. Functional genomic landscape of acute myeloid leukaemia. *Nature*, 562(7728): 526–531, 10 2018. URL `https://www.nature.com/articles/s41586-018-0623-z`.

P. van Galen, V. Hovestadt, M. H. Wadsworth Ii, T. K. Hughes, G. K. Griffin, S. Battaglia, J. A. Verga, J. Stephansky, T. J. Pastika, J. Lombardi Story, G. S. Pinkus, O. Pozdnyakova, I. Galinsky, R. M. Stone, T. A. Graubert, A. K. Shalek, J. C. Aster, A. A. Lane, and B. E. Bernstein. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*, 176(6):1265–1281, 03 2019. URL `https://www.sciencedirect.com/science/article/pii/S0092867419300947?via%3Dihub`.

L. Velten, B. A. Story, P. Hernández-Malmierca, S. Raffel, D. R. Leonce, J. Milbank, M. Paulsen, A. Demir, C. Szu-Tu, R. Frömel, C. Lutz, D. Nowak, J. C. Jann, C. Pabst, T. Boch, W. K. Hofmann, C. Müller-Tidow, A. Trumpp, S. Haas, and L. M. Steinmetz. Nat CommunIdentification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. *Nat Commun*, 12(1):1366, 03 2021.

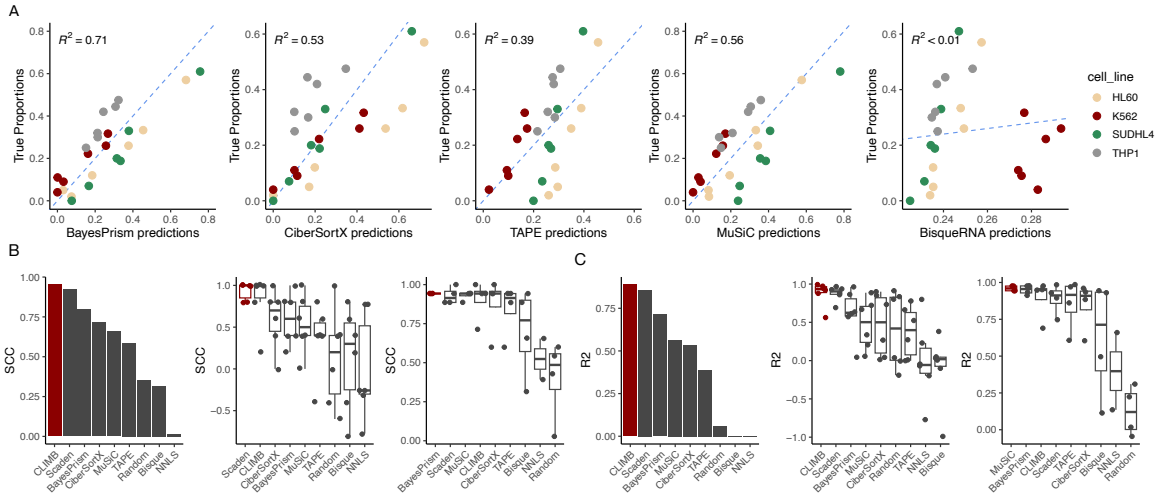
D. Vetrie, G. V. Helgason, and M. Copland. Nat Rev CancerThe leukaemia stem cell: similarities, differences and clinical prospects in CML and AML. *Nat Rev Cancer*, 20(3):158–173, 03 2020.

X. Wang, J. Park, K. Susztak, N. R. Zhang, and M. Li. Nat CommunBulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun*, 10(1):380, 01 2019.

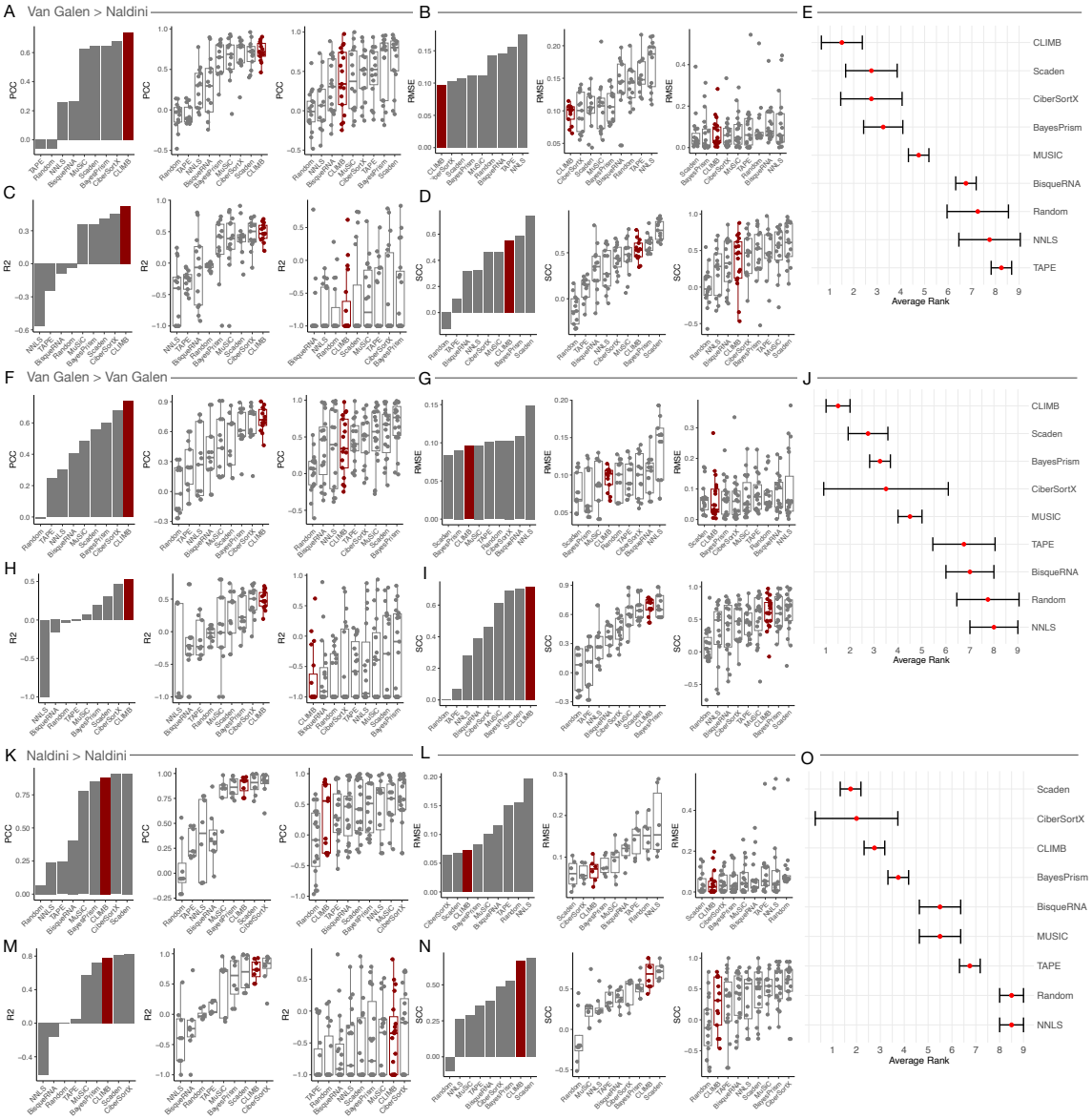
J. Wu, Y. Xiao, J. Sun, H. Sun, H. Chen, Y. Zhu, H. Fu, C. Yu, W. E, S. Lai, L. Ma, J. Li, L. Fei, M. Jiang, J. Wang, F. Ye, R. Wang, Z. Zhou, G. Zhang, T. Zhang, Q. Ding, Z. Wang, S. Hao, L. Liu, W. Zheng, J. He, W. Huang, Y. Wang, J. Xie, T. Li, T. Cheng, X. Han, H. Huang, and G. Guo. A single-cell survey of cellular hierarchy in acute myeloid leukemia. *J Hematol Oncol*, 13 (1):128, 09 2020.

K. Yoshihara, M. Shahmoradgoli, E. nez, R. Vegesna, H. Kim, W. Torres-Garcia, V. o, H. Shen, P. W. Laird, D. A. Levine, S. L. Carter, G. Getz, K. Stemke-Hale, G. B. Mills, and R. G. Verhaak. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*, 4:2612, 2013.

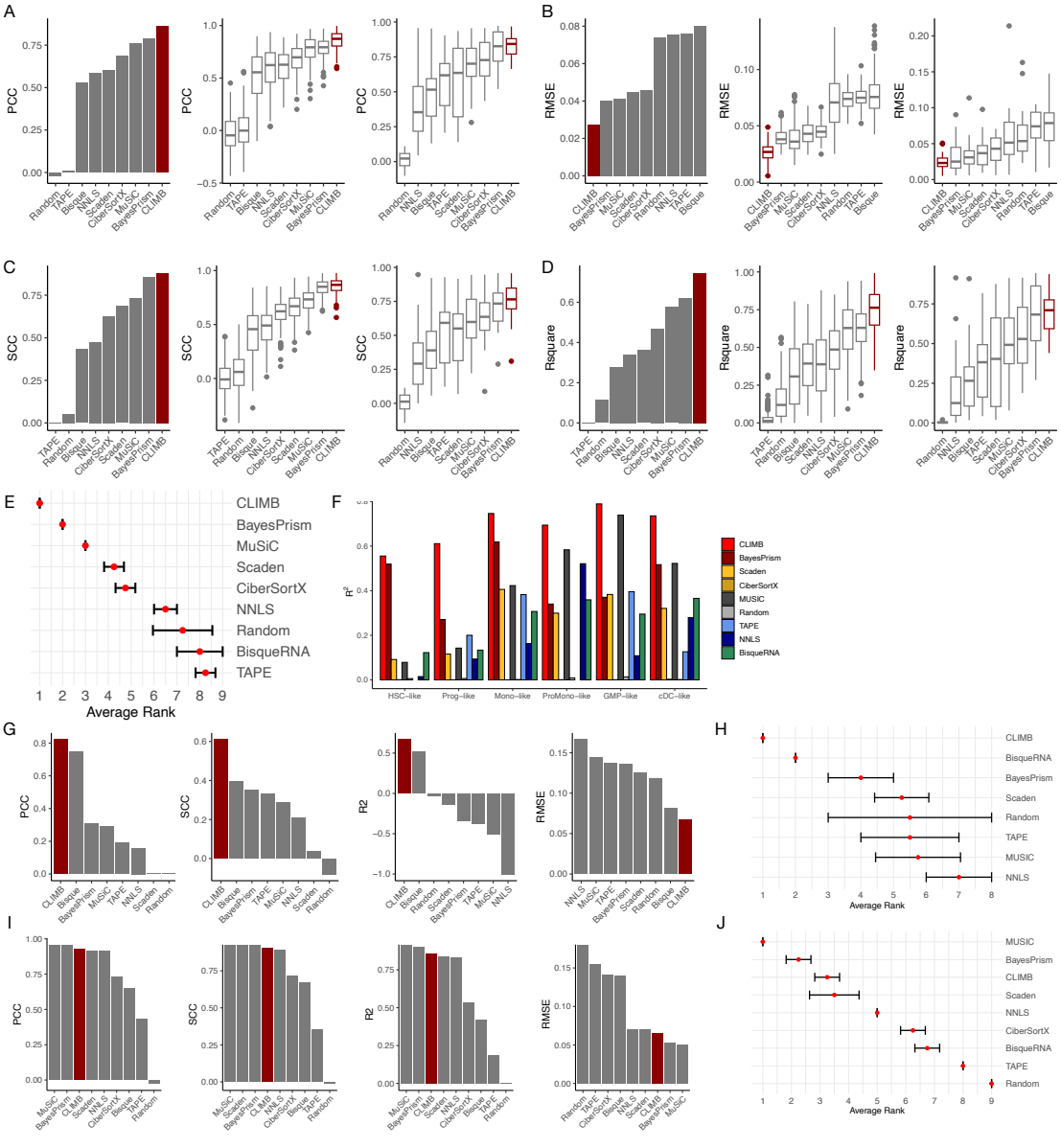
A. G. X. Zeng, S. Bansal, L. Jin, A. Mitchell, W. C. Chen, H. A. Abbas, M. Chan-Seng-Yue, V. Voisin, P. van Galen, A. Tierens, M. Cheok, C. Preudhomme, H. Dombret, N. Daver, P. A. Futreal, M. D. Minden, J. A. Kennedy, J. C. Y. Wang, and J. E. Dick. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med*, 28(6):1212–1223, 06 2022.


A
R^2 = 0.71
R^2 = 0.53
R^2 = 0.39
R^2 = 0.56
R^2 < 0.01
True Proportions
BayesPrism predictions
CiberSortX predictions
TAPE predictions
MuSiC predictions
BisqueRNA predictions
cell_line
HL60
K562
SUDHL4
THP1
B
C
SCC
R2
CLIMB
Scaden
BayesPrism
CiberSortX
MuSiC
TAPE
Random
Bisque
NNLS

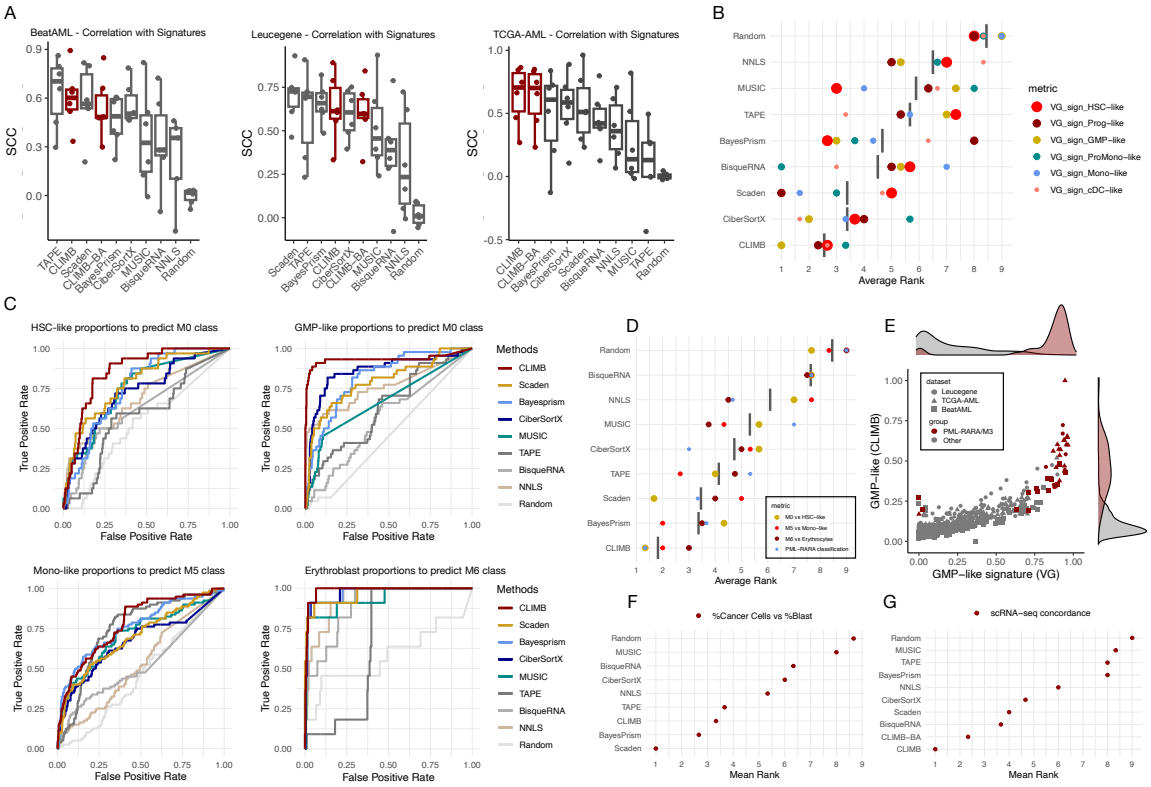
[p!]


A Van Galen > Naldini
B
E
C
D
F Van Galen > Van Galen
G
J
H
I
K Naldini > Naldini
L
O
PCC
RMSE
R2
SCC
CLIMB
Scaden
CiberSortX
BayesPrism
MUSIC
BisqueRNA
Random
NNLS
TAPE
Average Rank

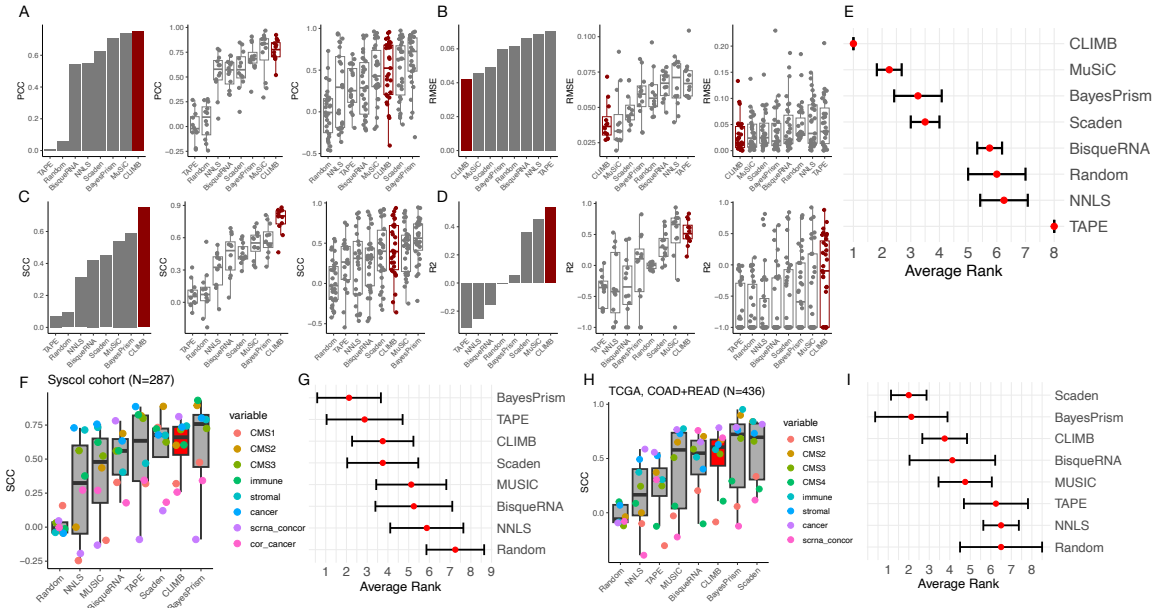
[p!]


A
B
C
D
E
F
G
H
I
J
PCC
RMSE
SCC
Rsquare
R2
Average Rank
CLIMB
BayesPrism
MuSiC
Scaden
CiberSortX
NNLS
Random
BisqueRNA
TAPE
MUSIC
HSC−like
Prog−like
Mono−like
ProMono−like
GMP−like
cDC−like

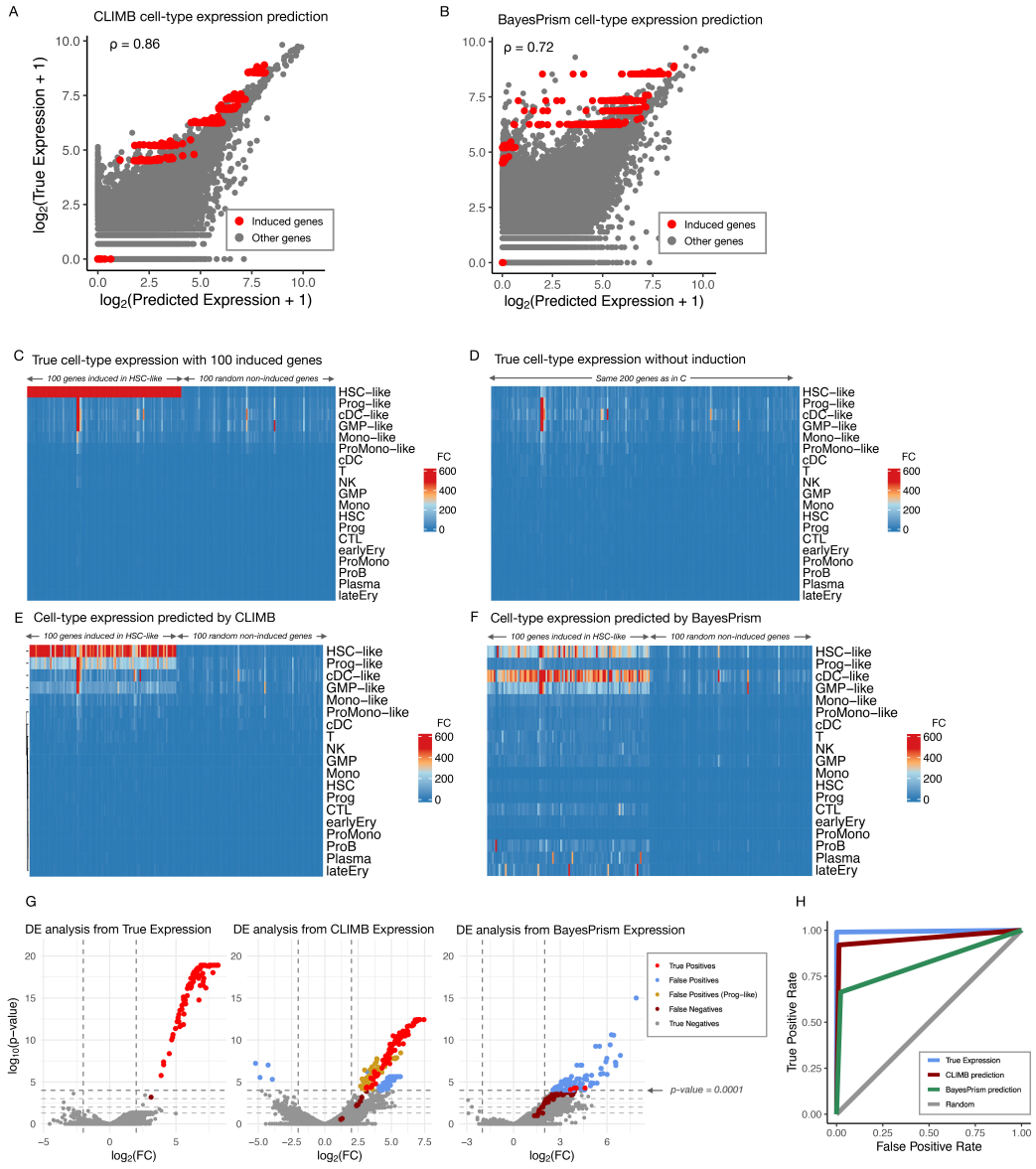
[p!]


A
BeatAML - Correlation with Signatures
Leucegene - Correlation with Signatures
TCGA-AML - Correlation with Signatures
SCC
B
Average Rank
metric
VG_sign_HSC−like
VG_sign_Prog−like
VG_sign_GMP−like
VG_sign_ProMono−like
VG_sign_Mono−like
VG_sign_cDC−like
C
HSC-like proportions to predict M0 class
GMP-like proportions to predict M0 class
Mono-like proportions to predict M5 class
Erythroblast proportions to predict M6 class
True Positive Rate
False Positive Rate
Methods
CLIMB
Scaden
Bayesprism
CiberSortX
MUSIC
TAPE
BisqueRNA
NNLS
Random
D
Average Rank
E
GMP−like (CLIMB)
GMP−like signature (VG)
dataset
Leucegene
TCGA-AML
BeatAML
group
PML−RARA/M3
Other
F
%Cancer Cells vs %Blast
Mean Rank
G
scRNA−seq concordance
Mean Rank
H
PCC
I
mean
Van Galen et al.
J
Marker Score (Significance)
n.s

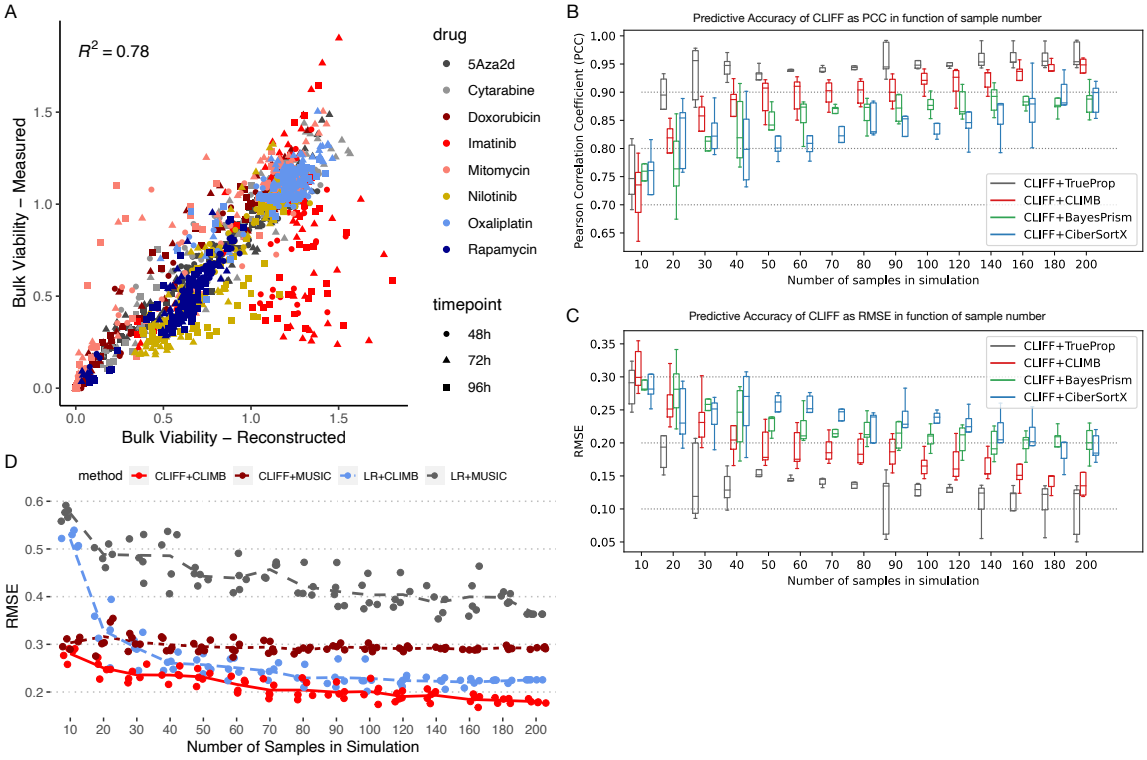
[p!]


A
B
C
D
E
F
G
H
I
PCC
RMSE
SCC
R2
CLIMB
MuSiC
BayesPrism
Scaden
BisqueRNA
Random
NNLS
TAPE
Average Rank
Syscol cohort (N=287)
TCGA, COAD+READ (N=436)
variable
CMS1
CMS2
CMS3
CMS4
immune
stromal
cancer
scrna_concor
cor_cancer
MUSIC

[p!]


A
CLIMB cell-type expression prediction
ρ = 0.86
log2(True Expression + 1)
log2(Predicted Expression + 1)
Induced genes
Other genes
B
BayesPrism cell-type expression prediction
ρ = 0.72
log2(Predicted Expression + 1)
Induced genes
Other genes
C
True cell-type expression with 100 induced genes
100 genes induced in HSC-like
100 random non-induced genes
D
True cell-type expression without induction
Same 200 genes as in C
E
Cell-type expression predicted by CLIMB
F
Cell-type expression predicted by BayesPrism
HSC–like
Prog–like
cDC–like
GMP–like
Mono–like
ProMono–like
cDC
T
NK
GMP
Mono
HSC
Prog
CTL
earlyEry
ProMono
ProB
Plasma
lateEry
FC
600
400
200
0
G
DE analysis from True Expression
DE analysis from CLIMB Expression
DE analysis from BayesPrism Expression
log10(p–value)
log2(FC)
True Positives
False Positives
False Positives (Prog–like)
False Negatives
True Negatives
p–value = 0.0001
H
True Positive Rate
False Positive Rate
True Expression
CLIMB prediction
BayesPrism prediction
Random

[p!]


A
$R^2 = 0.78$
Bulk Viability – Measured
Bulk Viability – Reconstructed
drug
5Aza2d
Cytarabine
Doxorubicin
Imatinib
Mitomycin
Nilotinib
Oxaliplatin
Rapamycin
timepoint
48h
72h
96h
B
Predictive Accuracy of CLIFF as PCC in function of sample number
Pearson Correlation Coefficient (PCC)
Number of samples in simulation
CLIFF+TrueProp
CLIFF+CLIMB
CLIFF+BayesPrism
CLIFF+CiberSortX
C
Predictive Accuracy of CLIFF as RMSE in function of sample number
RMSE
Number of samples in simulation
CLIFF+TrueProp
CLIFF+CLIMB
CLIFF+BayesPrism
CLIFF+CiberSortX
D
method
CLIFF+CLIMB
CLIFF+MUSIC
LR+CLIMB
LR+MUSIC
RMSE
Number of Samples in Simulation

[p!]