Risk Stratification of Acute Myeloid Leukemia Using Ex Vivo Drug Sensitivity Profiling

by

Andrea M. Brodersen[1-3,7], Aram N. Andersen[1-3,7,8], Pilar Ayuda-Durán[1,2,7], Laure Piechaczyk[1,2], Dagim Shiferaw Tadele[1,2,4], Lizet Baken[1,2], Julia Fredriksen[1,2,5], Mia Stoksflod[5], Andrea Lenartova[5], Yngvar Fløisand[2,5,6], and Jorrit M. Enserink[1-3,8]

[1] Department of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

[2] Centre for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway

[3] Section for Biochemistry and Molecular Biology, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

[4] Current address: Cleveland Clinic, Department of Translational Hematology and Oncology Research, Lerner Research Institute, Cleveland Ohio

[5] Department of Hematology, Oslo University Hospital, Oslo, Norway

[6] Present address: Department of Haematology, The Clatterbridge Cancer Centre NHS Foundation Trust, Liverpool, UK.

[7] Equal contribution

[8] Contact information: aram.andersen@rr-research.no or jorrit.enserink@rr-research.no

Running title: Computational precision medicine in AML

Abstract

Acute Myeloid Leukemia (AML) is a heterogeneous malignancy involving the clonal expansion of myeloid progenitor cells (blasts) in the bone marrow and peripheral blood. Most AML patients are treated with intensive chemotherapy. Current risk stratification is mainly based on cytogenetic screening for structural chromosomal alterations and targeted sequencing of a selection of common mutations. However, a large fraction of patients, such as cytogenetically normal AML, remains difficult to assign to an appropriate therapy regimen, and new methods are required to better stratify patients into treatment groups. Here, we developed a strategy that uses *ex vivo* drug sensitivity profiling to forecast patient survival and stratify patients into high and low-risk groups. Using regularized Cox regression, we demonstrate that this approach is superior at predicting survival compared to methods currently used in the clinic. Using these findings, we defined a specific set of drugs that can be used in miniaturized drug profiling assays to stratify patients in risk groups with high confidence. In conclusion, *ex vivo* drug sensitivity data can be used to make accurate predictions for cancer prognostics and forecasting.

Introduction

Acute Myeloid Leukemia (AML) is a heterogeneous cancer where the clonal expansion of myeloid progenitor cells (blasts) in the bone marrow and peripheral blood limit the production of normal cells, generating malfunctioning leukocytes and resulting in a lack of platelets and red blood cells[1,2]. The current 5-year survival rate of patients over sixty years of age is estimated to be 10-15%[3]. Although the minority of patients qualifies for potentially curative stem cell transplantation, most treatment-eligible individuals receive standard treatment, which consists of 100-200 mg/m^2 Cytarabine (or Ara-C) intravenously for seven days followed by a three-day treatment of 60 mg/m^2 Daunorubicin (or Idarubicin)[4]. Survival rates of older patients have not substantially improved over the past decades, underscoring the need for better clinical assessment and more accurate prognostic approaches. Current risk stratification methods are mainly based on parameters such as age, cytogenetics, and predictive genetic variants, such as the European LeukemiaNet (ELN) guidelines for AML stratification[5]. However, a substantial number of patients remains difficult to stratify, such as cytogenetically normal AML, and approximately 50% of patients are stratified as intermediate-risk patients for whom selection of the appropriate treatment regimen remains a major challenge[6,7].

Recently, *ex vivo* drug profiling has been used as a precision medicine approach to identify potential compounds that may be repurposed for treatment of various types of cancer, including AML[8]. Here, leukemic cells are obtained from bone marrow biopsies or from peripheral blood and incubated in the presence of a library of drugs. Drug responses can then be determined using a variety of different read-outs, such as CellTiter-Glo, which measures relative ATP levels in cell lysates, which correlates with relative cell numbers[9]. This approach allows for rapid screening of a high number of compounds and has identified potential novel treatment avenues for AML[8]. *Ex vivo* drug screening of cancer cells using chemotherapeutics that are used in the clinic has also been used to make predictions about treatment outcome[10,11,12,13]. However, correlations between *ex vivo* responses to more expansive drug sets and patient risk have not been fully explored.

Machine learning variations of the Cox proportional hazard model are rapidly gaining traction in precision medicine for improving clinical forecasting and patient diagnostics[14,15,16,17,18]. A major challenge with high-dimensional biological datasets is that the training of complex models requires large patient cohorts to avoid overfitting. Linear Cox regression with regularization and cross-validation is a robust approach for survival prediction compared with

more advanced algorithms and has the advantage of high interpretability through its learned risk associations [19,20,21].

In this study we used regularized Cox regression to analyze *ex vivo* drug profiling data of AML patients with the goal of (i) forecasting the response to standard treatment; and (ii) to improve patient stratification. We show that this approach allows for stratification of AML patients with high accuracy.

Methods

Patient cohort

Bone marrow and/or peripheral blood samples were collected from 220 adult patients diagnosed with AML and treated at the Department of Hematology at Oslo University Hospital in Norway between 2014-2017. Patients were included consecutively in this observational study. The study was performed in accordance with the Declaration of Helsinki and samples were collected following written informed consent. The study was approved by the Regional Committee for Medical Research Ethics South-East Norway (REK 2015/2012). Patient data that were collected included survival, gender, age and risk assessments, routine diagnostic workup consisting of flowcytometry and genetic biomarkers investigated by G-banding, RT-qPCR, FISH and fragment analysis of FLT3 and NPM1, as well as cytogenetic classification using the WHO and FAB systems (see Suppl. Table S1 for a complete overview). The median age was 62 years and 35 male and 39 female patients were included in the cohort. All patients were subjected to standard treatment, which consisted of a 30 min infusion of anthracycline [daunorubicin (60mg/m^2) or idarubicin (10-12 mg/m^2)] for three days in combination with 24 hour infusion of cytarabine (Ara-C) for seven days, although the length and dosage of each treatment was in some cases altered depending on the patient's age and general condition. Of note, one of the patients received Sorafenib in combination with standard treatment to combat renal failure, whereas another patient deceased before treatment could be initiated.

Ex vivo drug screens

Ex vivo drug sensitivity screens were performed on freshly isolated blasts cells from bone marrow- or peripheral blood from 74 patients purified by LymphoprepTMgradient centrifugation (Stemcell) and cultured in Mononuclear Cell Medium (PromoCell C-28030) supplemented with 1% Penicillin + Streptomycin (PS) (Gibco,15140-122). A total of 10,000 cells (25 µL) was added to each well in pre-drugged 384-well plates (Greiner Bio-One) using a Multi Drop Combi peristaltic dispenser (Thermo Scientific). The number of cells was adjusted for specific samples with lower cell counts without a noticeable change in the outcome of the experiments.

The Selleck Anti-Cancer Compound Library L3000 was used for drug screening, which consists of 349 anti-cancer drugs dissolved in dimethyl sulfoxide (DMSO) (see Suppl. Table S2 for the complete list). Compounds were pre-aliquoted in seven 384-well plates using five 10-fold dilution steps, reaching a concentration range from 1 nM to 10 µM. Eight positive

(benzethonium chloride, BzCl) and eight negative (DMSO only) controls were added to each plate. The drug handling was done at the Biotechnology Center of Oslo using an Echo 550 (Labcyte Inc.). After incubation (72 hours at 37°C in a 5% CO2, humidified environment) relative cell viability was quantified using the CellTiter-Glo Luminescent viability assay (Promega) and an EnVision 2104 Multilable plate reader (Perkin Elmer). The luminescence readout was measured in counts per second (CPS).

Data preprocessing

Computational analyses were performed in the statistical programming environment R (4.1.2 GUI 1.77 High Sierra build (8007)) and RStudio (version 2021.09.1 build 372). To measure the drug responses as a relative change in cell growth, all CPS values were normalized to the median CPS of the positive control (BzCL) and the negative control (DMSO) by subtracting the mean CPS of BzCl and dividing on the difference between median DMSO and median BzCl. All medians were calculated per plate per person. All negative response values were adjusted to the lowest positive response value in each plate. The drug sensitivity was measured as an area under the curve (AUC) of the dose responses for every patient-drug combination using the formula below.

$$AUC = \sum_{d=2}^{m} \frac{Response_d(log_{10}(C_d) - log_{10}(C_{d-1}))}{log_{10}(C_m) - log_{10}(C_1)}$$

C_d indicates a given concentration of a given drug, ranging from the lowest concentration C_1 to the highest concentration C_m. The AUCs were subsequently log_2-transformed to counter skewness of the relative growth scale. These AUCs were based on the measured dose responses (raw AUCs) and compared to the curve-fit derived metrics EC50, toxicity EC50 (TEC50), the drug sensitivity scores DSS1, DSS2 and DSS2, and the DSS AUC, computed using routines adapted from BREEZE [22,23]. All drug sensitivity metrics were standardized over each patient to counter potential screen batch effects, and/or for each drug to counter unevenness in penalization of unequally scaled model variables.

Clinical data were prepared by binarizing all mutations and genomic abnormalities, and creating dummy variables for all AML and patient classification categories. All missing data were set to zero. Age was computed between year of birth and year of diagnosis. The survival times were computed as date of diagnosis until the registered date of death, and the last recorded visiting date was used for censoring surviving patients.

Model training and cross-validation

Regularized Cox models were trained on the different drug sensitivity metrics and the binarized clinical variables using the glmnet package[24]. Ridge models were trained using only the L2 norm penalties by setting the mixture factor (α) to zero, and screening over a sequence of λ penalties using leave-one-out cross-validation (due to the low sample size). The model with the best average deviance score was selected for testing, or further analysis of model coefficients. For comparison with lasso or elastic net models, the α was set to one for L1 norm penalty, or screened through a range of values (between zero and one), respectively. Pre-selection of variables prior to model training was based on thresholds for standard deviations in drug sensitivity (AUC).

Model testing

For testing the models, a random sample of 10 patients were withheld from the training procedure, keeping a 4/10-ratio of deceased patients to maintain the approximate proportions in the original dataset. The predictivity on test data was scored with Harrell's concordance index (C-index), and the training and testing procedure was repeated 50 or 200 times[25].

Variable importance

The importance of individual drugs was assessed by their contribution to predictivity through a variable withdrawal test, and the statistical significance of the model coefficients. The contribution to predictivity was measured by training and testing models leaving out individual drugs on-by-one and computing the change in C-index from the model containing all drugs. Statistical significance of model coefficients was estimated by training 200 models on bootstrapped data and computing the mean and confidence interval for the fitted coefficients.

Model reduction

An algorithm cycling between variable withdrawal testing and backward selection was designed to find the minimal set of drugs with maximal out-of-sample predictivity. Following a variable withdrawal test, the withdrawn drugs negatively impacting test predictivity (C-index increase in response to drug withdrawal) were removed iteratively based on their rank. For each step the shrinking model was compared to its previous iteration until a drug removal resulted in a decrease in predictivity. Then a new variable withdrawal test was performed on the reduced model, followed by a new round of backward selection. This routine was repeated until a variable withdrawal test only yielded drugs positively impacting test predictivity.

Enrichment test and drug set mapping

A Fisher's exact test was used to analyze overrepresentation of annotated drug classes in the drug set from the reduced model. The drugs were clustered with UMAP using correlations in drug target information from Drug Target Commons (https://drugtargetcommons.fimm.fi/; June 2020) and colored with annotated drug classes[26,27].

Patient risk stratification

A regularized Cox model was fitted using the standardized $\log_2$-transformed AUCs, and the estimated coefficients were multiplied with their respective drug sensitivities. The resulting risk association-weighted matrix was then subjected to hierarchical clustering using maximum distance between patients, euclidean distance between drugs, and ward.D clustering.

Results

Study approach

We hypothesized that blast cells from AML patients display *ex vivo* drug susceptibility profiles that correlate with survival outcome of standard 7+3 therapy. To test this hypothesis, we analyzed data from *ex vivo* drug screens that were performed on treatment-naïve bone marrow and peripheral blood samples on a cohort of newly diagnosed AML patients. For an overview of the project outline see Figure 1A. In brief, samples were collected between 2014 and 2017 and patient data was recorded until spring 2022 or until loss of life, and all patients received standard treatment of daunorubicin and cytarabine. We obtained patient data for 220 AML patients in total. We then performed *ex vivo* drug screens on blast cells derived from 74 out of these 220 patients. Hereafter, we will refer to these groups as '*C220*' and '*C74*', respectively. The clinical dataset consisted of 70 variables (Suppl. Table S1). Drug screens were performed on blast cells isolated from fresh bone marrow biopsies or peripheral blood using a drug library of 349 anticancer drugs from the Selleck Anti-Cancer Compound Library (Fig. 1B and Suppl. Table S2). We used these data to determine to which extent *ex vivo* drug sensitivity profiling can predict patient survival. As a metric for drug sensitivity we calculated the area under the curve (AUC). See Figure 1C for how the AUC compares to other metrics, such as the drug sensitivity score (DSS) and EC50, and see Figure S1 for correlations between the different metrics.

Pattern conservation and association of clinical variables.

To ensure that our drug-screened patient cohort (C74) was representative of the larger clinical population (C220) we plotted the estimated overall survival for both groups. This revealed good similarity between the two groups, with the main notable difference that the median survival time of C220 was approximately 944 days (Fig. 2A), whereas the median survival time for C74 was approximately 1200 days (Fig. 2B).

Next, to investigate the risk association of the clinical variables we performed a Cox proportional hazard analysis of each genetic marker or category for both cohorts (Table 1). The analysis of C220 only indicated three statistically significant variables (Fig. 2C); i.e. WHO class 4 and deletions in chromosome 20 were associated with negative outcome in patients subjected to the standard treatment, whereas WHO class 0 was associated with positive outcome. Cohort C74 only exhibited a statistically significant association of CEBPA, which was associated with negative outcome (Table 1).

Assessing predictivity of drug sensitivity metrics

To find the optimal usage of *ex vivo* drug responses for clinical prognostics, we compared the accuracy of several multivariate Cox models based on different drug sensitivity metrics in predicting patient survival (Fig. 2D). In addition to the EC50, TEC50 and DSS scoring metrics, which are all derived from curve fitting on an inverted growth density scale representing growth inhibition, we also computed the area under the raw growth density curve (raw AUC) to preserve the empirical variation in the data. Furthermore, since the relative growth densities are naturally non-symmetric values we introduced a $\log_2$-transformation step of the raw AUCs. Due to the relatively low sample size compared to the number of drugs, a Ridge (L2-norm) penalty was introduced to counter overfitting of the models. This showed that the raw AUCs and DSS metrics outperformed EC50 and TEC50 in predicting patient survival. Moreover, when standardizing the patient distributions, there were substantial improvements for the raw AUCs with the $\log_2$-transformed AUC yielding the best overall prediction accuracy with a median C-index of 72%. In contrast, standardizing the drug sensitivity distributions, and thus equalizing penalty unevenness due to differences in drug sensitivity variation, resulted in a reduction in prediction accuracy for all metrics. This, together with the observation of substantial overfitting (Fig. S2A), prompted us to test the effects of pre-selecting drugs with standard deviations in AUCs above a set threshold to reduce the model size. In addition, we tested changing the regularization type to elastic net and lasso, which, due to an L1-norm penalty, not only penalize multicollinearity more strongly but also perform a variable selection (Fig. S2B). These operations, particularly variable pre-selection, improved the survival prediction from patient-standardized $\log_2$-transformed AUCs even further to C-index scores above 80% (Fig. S2C-D and Fig. 2E). For comparison, the predictive accuracy of the clinical data alone only resulted in a C-index of 57% (Fig. 2E). Since binary and categorical data are more sensitive to low sample sizes, we also tested models trained on the larger cohort, which yielded similar results. Furthermore, combining the clinical data with drug sensitivity data did not increase predictivity of the drug sensitivity data alone (Fig. 2E). This suggests that there is little prognostic information in the set of genetic markers and currently available diagnostic tools that is not already captured by the drug sensitivity data. Finally, a Ridge model trained on pre-selected drug sensitivities combined with clinical data resulted in the best overall predictive accuracy with a median C-index of 84% (Fig. 2E).

Drug associations

To identify specific drug responses that were associated significantly with survival we performed a bootstrap analysis of model coefficients trained on the entire drug library (with patient-standardized $\log_2$-transformed AUCs) under a Ridge penalty. Importantly, we found that *ex vivo* sensitivity of patient cells to several compounds was strongly associated with outcome of 7+3 treatment (Fig. 3A). These compounds included daunorubicin, which is a major component of 7+3 chemotherapy and which was strongly associated with survival (Fig. 3A,B), reflecting previous reports that *ex vivo* drug sensitivity can predict clinical efficacy of certain chemotherapeutics[10,11,12,13], thus validating our approach. Mitoxantrone, a chemotherapeutic agent that is currently in phase II clinical trials as treatment for relapsed or refractory AML[28], also showed a strong positive association with survival (Fig. 3A), whereas sensitivity to cytarabine was not associated significantly with survival (Fig. 3A,C). Interestingly, we found that *ex vivo* sensitivity to the Bcl2/Bcl-XL/Bcl-w inhibitor ABT-263 (navitoclax) was clearly associated with increased risk (Fig. 3A). ABT-263 is an experimental compound with a mechanism of action similar to the more selective Bcl2 inhibitor venetoclax, which has been reported to significantly improve survival of AML patients who are ineligible for standard chemotherapy[29].

Next, we determined to which extent each of the compounds contributed to predictivity by measuring the change in C-index upon removal of a single drug from the model (Fig. 3D). This revealed that ABT-263 was the drug that affected predictivity the most, along with mitoxantrone, the putative survivin inhibitor YM155, the HSP90 inhibitor geldanamycin and the JAK inhibitor Cyt387 (Fig. 3D). This drug withdrawal test also indicated that a substantial number of variables contributed negatively to the prediction, likely by causing overfitting of models trained on the entire drug library. Thus, we set out to find the minimal set of drugs that could maximize the out-of-sample prediction by performing a backward selection based on the predictive accuracy (test C-index; see Methods). This resulted in a model reduction to 86 drugs, with a median C-index of 96% (Fig. 4A,B). When assessing the statistical significance of the model coefficients, a greater number of drugs showed a statistically significant association with survival, suggesting that a more robust inference could be made from this model (Fig. 4C). By performing an enrichment test for drug classes within this reduced drug set, we found that only HSP inhibitors had a statistically significant enrichment (Fig. 4D). Moreover, upon clustering of the drugs based on their reported binding targets, we found two distinct groups of kinase inhibitors, in addition two more diverse groups enriched for hormone regulators clustering together with other drug classes (Fig. 4E).

Stratifying patients into risk groups

Finally, to investigate whether the clinical associations of *ex vivo* drug sensitivities could be used to stratify AML patients into risk groups, we performed hierarchical clustering based on the risk associations of the individual drug sensitivities within the reduced drug set (Fig. 5A). The analysis revealed distinct high-risk and low-risk patient groups that were otherwise not distinguishable based on genetic or clinical classification data alone. For instance, top responders to mitoxantrone or YM155 represented patient groups with favorable prognosis under the standard treatment (Fig. 5B,C), while top responders to ABT-263 or Mesna (Uromitexan) represented high-risk groups with a poor prognosis (Fig. 5D,E). Thus, *ex vivo* drug screening data can reveal novel insight that cannot be obtained through other means.

Discussion

In this study we have demonstrated that *ex vivo* drug sensitivity screening can be used to predict survival of AML patients subjected to standard chemotherapy with very high accuracy. The advantage of our study over previous precision medicine studies is that our approach eliminates the need for tedious and often expensive genetic analyses that require advanced equipment and highly skilled personnel. Furthermore, by using a drug library that not only includes chemotherapeutics used for AML treatment, but also a wide variety of other compounds, we were able to identify a 'drug fingerprint' that predicts treatment outcome with very high accuracy. An interesting observation is that merging the clinical variables dataset with the drug screening dataset did not improve predictivity compared to the drug dataset alone. This strongly suggests that all predictive information in the clinical variables is already present within the drug responses, thus making the clinical variables redundant. Given that we were able to reduce the number of compounds to a small set of 86 drugs with a strikingly high predictive power, we envision that this reduced drug set is the first step towards development of a rapid, small-scale *ex vivo* drug test that can be used in the clinical decision-making process to stratify patients into treatment groups.

In our cohort of 220 patients, only three clinical variables were statistically significantly associated with survival after standard treatment. Moreover, cytogenetics and general health based on the ELN guidelines were hardly predictive for this patient cohort. The ELN guidelines were defined based on the analysis of thousands of AML patients in multiple countries, and one reason for the low degree of predictivity in our cohort could be its relatively small size. Furthermore, our cohort of Norwegian patients is likely much more homogeneous than the group of patients that was analyzed by ELN, which includes patients from several European countries with relatively high genetic and cultural diversity. While our method needs to be further validated using a larger patient cohort and preferably also a more expansive and diverse drug library, we argue that our observations emphasize the need for novel stratification methods that are better tailored to the local population to provide more accurate treatment. The advantage of our method is that predictive drug fingerprints can be identified in local patient cohorts with relatively little effort and without the need for advanced and expensive equipment.

On a more technical note, we found that several commonly used drug sensitivity metrics were outperformed by AUCs. This is likely due to the fact that calculation of AUCs does not involve Hill curve fitting, which is typically used in many other drug response metrics. We believe that

the discrete integration under the empirical response-curve between the lowest and highest dose in $\log_{10}$-scale avoids curve-fitting biases due to artifacts of non-Hill type responses or due to insufficient data points to support an accurate curve fit. While the DSS has the advantage of normalizing drug sensitivities between studies using different dose ranges, our data suggests information may be lost through this procedure, and that careful standardization of patient sample processing, drug screening practices and study designs may be more critical for forecasting[30,23] .

While not the primary focus of this study, one interesting observation that may have more direct clinical relevance is that the response of cancer cells to navitoclax showed a major negative correlation with predictivity of patient survival; i.e. a strong response to navitoclax was associated with poor outcome of 7+3 treatment. We are currently exploring whether the related compound venetoclax, which is already approved for treatment of AML, has a similar effect on predictivity. If this is indeed the case, one might argue that this subgroup of patients should forgo standard 7+3 therapy and rather be subjected to first-line venetoclax treatment. Prospective clinical trials will be necessary to further investigate this possibility.

Taken together, we have developed an approach that is superior at predicting patient survival than current clinical decision-making methods.

Acknowledgments

We would like to thank all patients who participated in this study, as well as patient representatives from the Norwegian Blood Cancer Society (Blodkreftforingen), members of the Enserink and Knævelsrud laboratories for fruitful discussions, and the PERCATHE and PINpOINT environments for feedback on the study. We also thank Dr. M. Zucknick for suggesting the drug withdrawal test. This study was supported by grants from the Norwegian Health Authority South-East, grant numbers 2017064, 2018012 and 2019096; the Norwegian Cancer Society, grant numbers 182524 and 208012; and the Research Council of Norway through its Centers of Excellence funding scheme (262652) and through grants and 261936, 294916 and 314811.

Author contributions

Conceptualization: ANA and JME. Project supervision: ANA and JME. Coding and data analysis: AMB and ANA. Created figures: AMB and ANA. Performed drug screens: PAD with assistance from LP, DST, LB, JF. Sampled patients: YF and AL. Wrote the manuscript: AMB, ANA, JME. Extracted patient data: MS, PAD, AMB.

References

1. Blodkreft: Akutt myelogen leukemi. NHI.no. Accessed August 11, 2022. https://nhi.no/sykdommer/barn/kreftsykdommer/akutt-myelogen-leukemi-oversikt/
2. Lowenberg B, Downing JR, Burnett A. Acute Myeloid Leukemia. *N Engl J Med*. 1999;341(14):1051-1062. doi:10.1056/NEJM199909303411407
3. Survival | Acute myeloid leukaemia | Cancer Research UK. Accessed August 11, 2022. https://www.cancerresearchuk.org/about-cancer/acute-myeloid-leukaemia-aml/survival
4. Dombret H, Gardin C. An update of current treatments for adult acute myeloid leukemia. *Blood*. 2016;127(1):53-61. doi:10.1182/blood-2015-08-604520
5. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447. doi:10.1182/blood-2016-08-733196
6. Falini B, Nicoletti I, Martelli MF, Mecucci C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood*. 2006;109(3):874-885. doi:10.1182/blood-2006-07-012252
7. Mrózek K, Heinonen K, Bloomfield CD. Clinical importance of cytogenetics in acute myeloid leukaemia. *Best Pract Res Clin Haematol*. 2001;14(1):19-47. doi:10.1053/beha.2000.0114
8. Spinner MA, Aleshin A, Santaguida MT, et al. *Ex vivo* drug screening defines novel drug sensitivity patterns for informing personalized therapy in myeloid neoplasms. *Blood Adv*. 2020;4(12):2768-2778. doi:10.1182/bloodadvances.2020001934
9. CellTiter-Glo® 3D Cell Viability Assay | 3D Cell Culture. Accessed August 11, 2022. https://no.promega.com/products/cell-health-assays/cell-viability-and-cytotoxicity-assays/celltiter-glo-3d-cell-viability-assay/
10. Staib P, Staltmeier E, Neurohr K, Cornely O, Reiser M, Schinköthe T. Prediction of individual response to chemotherapy in patients with acute myeloid leukaemia using the chemosensitivity index Ci. Br J Haematol. 2005;128(6):783-791. doi:10.1111/j.1365-2141.2005.05402.x
11. Onecha E, Ruiz-Heredia Y, Martínez-Cuadrón D, et al. Improving the prediction of acute myeloid leukaemia outcomes by complementing mutational profiling with ex vivo chemosensitivity. Br J Haematol. 2020;189(4):672-683. doi:10.1111/bjh.16432
12. Zhang Y, Ji M, Zhao JY, et al. Ex Vivo Chemosensitivity Profiling of Acute Myeloid Leukemia and Its Correlation With Clinical Response and Outcome to Chemotherapy. Front Oncol. 2021;11:793773. doi:10.3389/fonc.2021.793773
13. Martínez-Cuadrón D, Gil C, Serrano J, et al. A precision medicine test predicts clinical response after idarubicin and cytarabine induction therapy in AML patients. Leuk Res. 2019;76:1-10. doi:10.1016/j.leukres.2018.11.006
14. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent

progress. *Npj Precis Oncol*. 2020;4(1):1-10. doi:10.1038/s41698-020-0122-1
15. Carpenter KA, Cohen DS, Jarrell JT, Huang X. Deep learning and virtual drug screening. *Future Med Chem*. 2018;10(21):2557-2567. doi:10.4155/fmc-2018-0314
16. Schork NJ. ARTIFICIAL INTELLIGENCE AND PERSONALIZED MEDICINE. *Cancer Treat Res*. 2019;178:265-283. doi:10.1007/978-3-030-16391-4_11
17. Javaid M, Haleem A, Pratap Singh R, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. *Int J Intell Netw*. 2022;3:58-73. doi:10.1016/j.ijin.2022.05.002
18. MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome*. 2021;64(4):416-425. doi:10.1139/gen-2020-0131
19. Hughey JJ, Rhoades SD, Fu DY, Bastarache L, Denny JC, Chen Q. Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record. *BMC Genomics*. 2019;20:805. doi:10.1186/s12864-019-6192-1
20. Spooner A, Chen E, Sowmya A, et al. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Sci Rep*. 2020;10(1):20410. doi:10.1038/s41598-020-77220-w
21. Zhang Z, Li J, He T, et al. Two predictive precision medicine tools for hepatocellular carcinoma. *Cancer Cell Int*. 2019;19(1):290. doi:10.1186/s12935-019-1002-z
22. Potdar S, Ianevski A, Mpindi JP, et al. Breeze: an integrated quality control and data analysis application for high-throughput drug screening. *Bioinformatics*. 2020;36(11):3602-3604. doi:10.1093/bioinformatics/btaa138
23. Yadav B, Pemovska T, Szwajda A, et al. Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci Rep*. 2014;4(1):5193. doi:10.1038/srep05193
24. Lasso and Elastic-Net Regularized Generalized Linear Models. Accessed September 29, 2022. https://glmnet.stanford.edu/
25. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the Yield of Medical Tests. *JAMA*. 1982;247(18):2543-2546. doi:10.1001/jama.1982.03320430047030
26. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Published online September 17, 2020. doi:10.48550/arXiv.1802.03426
27. Drug Target Commons (DTC). Accessed June 8, 2020. https://drugtargetcommons.fimm.fi/
28. Anderson R, Miller LD, Isom S, et al. Phase II trial of cytarabine and mitoxantrone with devimistat in acute myeloid leukemia. *Nat Commun*. 2022;13(1):1673. doi:10.1038/s41467-022-29039-4
29. DiNardo CD, Jonas BA, Pullarkat V, et al. Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia. *N Engl J Med*. 2020;383(7):617-629. doi:10.1056/NEJMoa2012971
30. Pemovska T, Kontro M, Yadav B, et al. Individualized Systems Medicine Strategy to Tailor Treatments for Patients with Chemorefractory Acute Myeloid Leukemia. *Cancer Discov*. 2013;3(12):1416-1429. doi:10.1158/2159-8290.CD-13-0350

Figure legends

Figure 1. Study overview. ***A,*** Peripheral blood or bone marrow samples were collected at the time of diagnosis. All patients received some form of standard 7+3 treatment and in select cases a stem cell transplantation, and clinical data was collected throughout the course of the disease. Drug screens were performed on mononuclear cells purified from the patient samples. Cox regression was used to associate the clinical variables and drug sensitivity data to patient survival. ***B,*** Overview of the main drug classes. Bubble size indicates relative representation in the library. ***C,*** Illustration comparing DSS (in green) to AUC (in orange).

Figure 2. **Predictivity of clinical variables and *ex vivo* drug screening data.** ***A, B,*** Survival curves of the C220 and C74 cohorts, respectively. ***C,*** Cox proportional hazard analysis of 34 clinical variables with sufficient coverage for C220. Significant associations are indicated with an asterisk. ***D,*** Test C-index results (n=200) for regularized Cox models based on various drug sensitivity metrics and standardization procedures using C74. ***E,*** Test C-index (top panel; n=200) and training C-index (bottom panel; n=200) for regularized Cox models trained on various dataset compositions. Pre-selection was done with a standard deviation threshold of 0.15, and elastic net models were trained with α=0.4.

Figure 3. Analysis of drug associations with survival prediction. ***A,*** Ranked model coefficients (log-hazard changes) with 95% confidence intervals for 349 drugs. Significant associations are indicated in black, standard treatments for AML are indicated in red and alternative treatments for AML are indicated in blue. ***B, C,*** Kaplan-Meier plots based on the top 50% of *ex vivo* responders to Daunorubicin and Cyterabine, respectively. Censoring is marked by ticks. ***D,*** Effects of drug withdrawal on C-index. Ranked changes in C-index with 95% confidence intervals for 349 drugs. Color-coding as in (A).

Figure 4. Backward selection of predictive drugs. ***A,*** Iterative removal of drugs to maximize predictivity. Test C-index results (n=50) for reduced models every 20 iterations of withdrawal testing and backward selection. Last model represents 134 iterations. ***B,*** Effects of drug withdrawal from the reduced model. Ranked changes in C-idex with 95% confidence intervals for 86 drugs. ***C,*** Ranked model coefficients (log-hazard changes) with 95% confidence intervals for 86 drugs. Significant associations are indicated in black, and alternative treatments for AML are indicated in blue. ***D,*** Enrichment analysis of drug classes within the selected drug set. The color scale indicates Fisher's exact test p-values in -$\log_{10}$-scale and the x-axis depicts the enrichment score in $\log_2$-scale. ***E,*** UMAP clustering of the selected drugs based on their drug target information. Colors indicate drug class.

Figure 5. **Stratification of patient risk groups.** ***A,*** Hierarchical clustering of risk associated drug sensitivities Patients are clustered along the columns and drugs are clustered along the rows. Patients are color annotated for clinical variables, gender, survival status, and predicted log hazard ratio (standardized) per patient. The heatmap color scale indicates log hazard ratio contribution of each drug sensitivity (drug sensitivity scores multiplied by their respective model coefficients). Rows with standard deviations over 0.2 are labeled. ***B-E,*** Kaplan-Meier plots based on the top responders to Mitoxantrone (*B*), YM155 (*C*), ABT-263 (Navitoclax; *D*) and Mesna (Uromitexan/Mesnex; *E*), respectively. The split of the population is indicated by the legends above the plots. Censoring is marked by ticks.

Supplemental figure legends

Figure S1. **Correlations between different drug sensitivity metrics.** Blue and yellow dots indicate dose response curve-fits with high and low confidence in the BREEZE routine, respectively.

Figure S2. ***A,*** Training set C-index results (n=200) for regularized Cox models based on various drug sensitivity metrics and standardization procedures using C74. ***B,*** Overview of routine for variable pre-selection and regularization-type testing. ***C, D,*** Mean and median test C-index (n=50) for models trained with different variable pre-selection thresholds and regularization types.

Supplementary table legends

Table S1. Clinical variable overview. A list over the variables included in the clinical dataset. Patient-sensitive information is not shown.

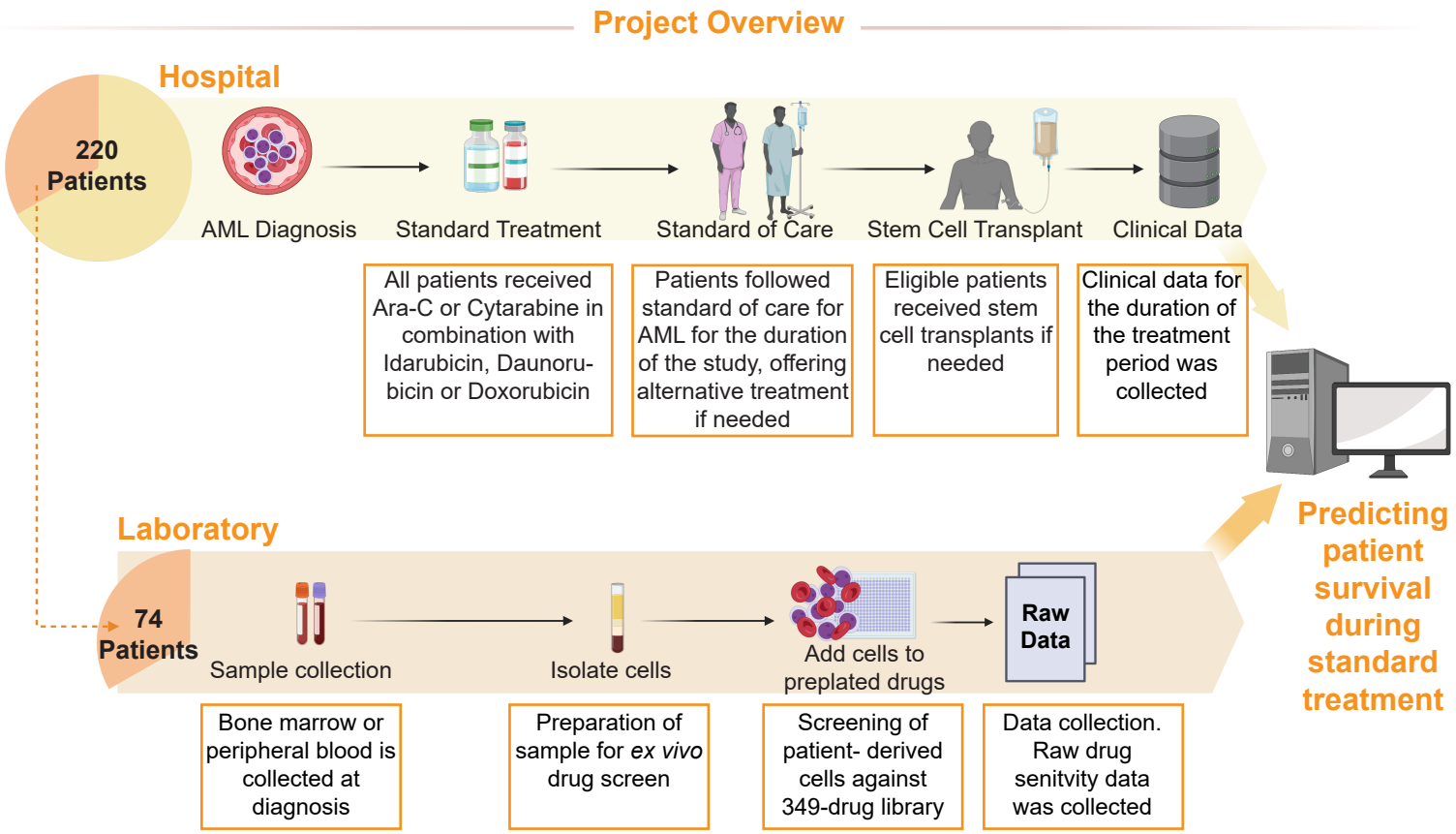
Table S2. Selleck anti-cancer library. An overview for the 349 drugs included in the Selleck anti-cancer library (version L3000). The left side indicates the name of the drugs with alternate names in parentheses. The second column indicates the primary drug target as described by the producers. Colum 3 shows the drug class that the drug targets.

Table 1. Clinical associations. An overview over the clinical variables' association with Hazard separated by cohort. The variable name is indicated on the left-hand side. The Hazard ratio (HR) is indicated in column 1 for C220 and 4 for C74. The 95% confidence intervals (CI) are indicated in columns 2 and 5 for C220 and C74, respectively. The p-values are indicated in column 3 for C220 and 6 for C74. If the model was unable to calculate values due to lack of coverage they are listed as NA.

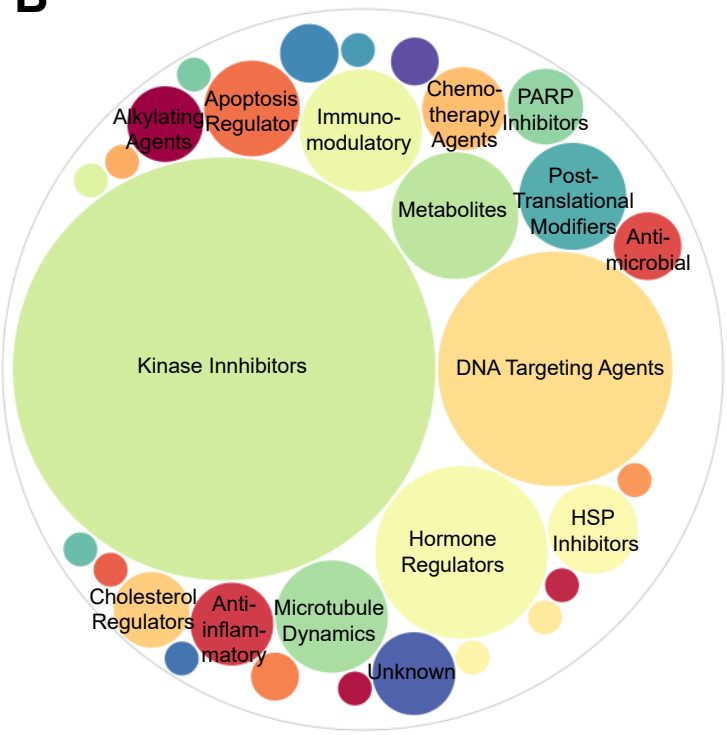
	C220			C74		
	HR	95% CI	p-value	HR	95% CI	p-value
WHO Class 4	5.41	2.34-12.51	0.000	NA	NA	NA
WHO Class 2	1.23	0.59-2.57	0.577	0.95	0.22-4.07	0.950
WHO Class 1	1.02	0.61-1.71	0.932	1.45	0.63-3.36	0.386
WHO Class 0	0.37	0.19-0.72	0.004	1.86	0.73-4.78	0.196
t911	1.31	0.32-5.37	0.704	NA	NA	NA
t821	1.20	0.16-8.77	0.859	NA	NA	NA
t1517	0.22	0.03-1.61	0.137	NA	NA	NA
STAG2	1.81	0.25-13.17	0.557	NA	NA	NA
SRSF2	2.49	0.60-10.26	0.207	NA	NA	NA
RUNX1	1.80	0.56-5.81	0.323	2.51	0.57-10.98	0.220
PMLRARA	0.17	0.02-1.24	0.080	0.00	0-inf	1.00
NRAS	2.54	0.35-18.66	0.359	NA	NA	NA
NPM1	0.88	0.50-1.55	0.658	0.77	0.29-2.04	0.598
MLL	0.98	0.24-4.00	0.973	NA	NA	NA
KMT2A	0.52	0.07-3.76	0.518	NA	NA	NA
inv16	0.54	0.17-1.70	0.291	0.57	0.07-4.20	0.580
Gender (Female)	1.20	0.76-1.90	0.427	1.23	0.99-1.03	0.585
FLT3TKD	0.44	0.14-1.40	0.164	0.00	0-inf	0.997
FLT3ITD	1.06	0.60-1.88	0.832	1.27	0.51-3.15	0.600
EVI1	1.07	0.43-2.66	0.882	0.00	0-inf	0.998
del20	2.82	1.03-7.77	0.044	NA	NA	NA
Cytogenetics Prognosis: Good	0.48	0.23-1.00	0.052	0.38	0.09-1.61	0.189
Cytogenetics Prognosis: Bad	1.11	0.68-1.80	0.672	0.73	0.30-1.80	0.50
Cytogenetics Prognosis: Aveereage	1.18	0.73-1.93	0.499	2.11	0.98-4.56	0.057
CEBPA	0.90	0.33-2.47	0.499	4.63	1.32-16.32	0.017
CBFMYH11	0.50	0.16-1.60	0.245	0.48	0.07-4.20	0.477
bcrabl	0.00	0.00-inf	0.996	0.00	0-inf	0.996
ASXL1	1.32	0.41-4.20	0.639	1.96	0.46-8.38	0.363
AMLFAB Class 0	1.05	0.33-3.34	0.931	1.43	0.33-6.10	0.196
AMLFAB Class 3	0.16	0.02-1.12	0.065	0.00	0-inf	0.998
AMLFAB Class 6	1.40	0.34-5.73	0.640	0.00	0-inf	0.998
AMLFAB Class 1-2	1.05	0.64-1.71	0.849	0.95	0.43-2.12	0.910
AMLFAB Class 4-5	0.96	0.58-1.60	0.890	1.73	0.78-3.81	0.176
Age	1.00	0.99-1.02	0.478	1.01	0.99-1.04	0.240

Figure 1

A


Project Overview
Hospital
220 Patients
AML Diagnosis
Standard Treatment
Standard of Care
Stem Cell Transplant
Clinical Data
All patients received Ara-C or Cytarabine in combination with Idarubicin, Daunoru-bicin or Doxorubicin
Patients followed standard of care for AML for the duration of the study, offering alternative treatment if needed
Eligible patients received stem cell transplants if needed
Clinical data for the duration of the treatment period was collected
Laboratory
74 Patients
Sample collection
Isolate cells
Add cells to preplated drugs
Raw Data
Bone marrow or peripheral blood is collected at diagnosis
Preparation of sample for ex vivo drug screen
Screening of patient- derived cells against 349-drug library
Data collection. Raw drug senitvity data was collected
Predicting patient survival during standard treatment

B


Alkylating Agents
Apoptosis Regulator
Immuno-modulatory
Chemo-therapy Agents
PARP Inhibitors
Metabolites
Post-Translational Modifiers
Anti-microbial
Kinase Innhibitors
DNA Targeting Agents
Hormone Regulators
HSP Inhibitors
Cholesterol Regulators
Anti-inflam-matory
Microtubule Dynamics
Unknown

C

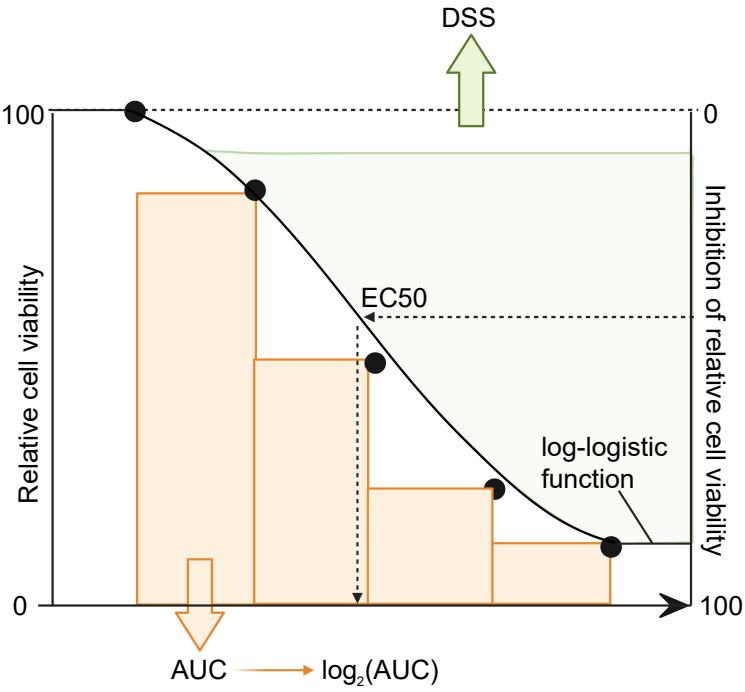

DSS
100
0
Relative cell viability
Inhibition of relative cell viability
EC50
log-logistic function
0
100
AUC
log2(AUC)

Figure 2

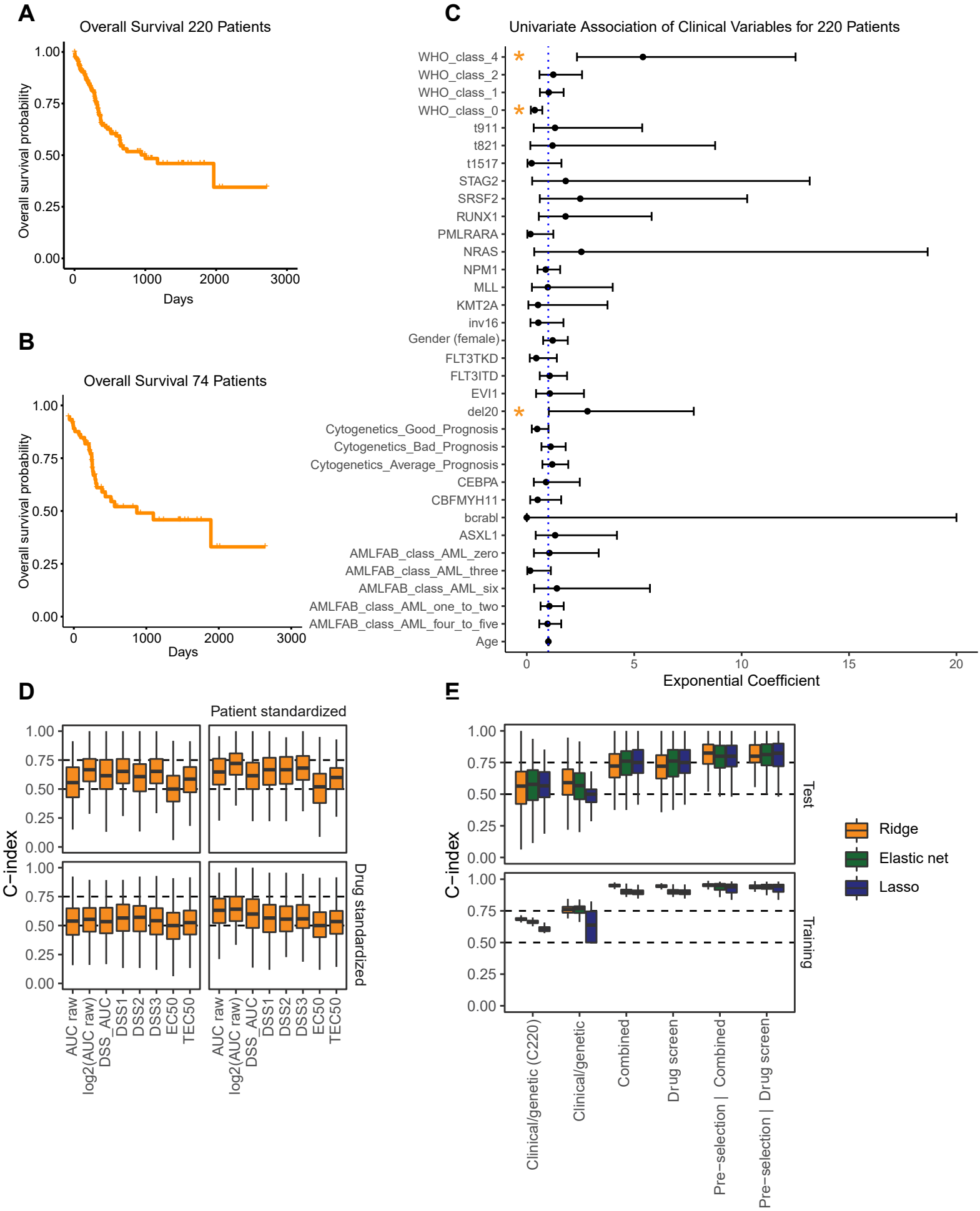

A
Overall Survival 220 Patients
Overall survival probability
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
Days
B
Overall Survival 74 Patients
Overall survival probability
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
Days
C
Univariate Association of Clinical Variables for 220 Patients
WHO_class_4
WHO_class_2
WHO_class_1
WHO_class_0
t911
t821
t1517
STAG2
SRSF2
RUNX1
PMLRARA
NRAS
NPM1
MLL
KMT2A
inv16
Gender (female)
FLT3TKD
FLT3ITD
EVI1
del20
Cytogenetics_Good_Prognosis
Cytogenetics_Bad_Prognosis
Cytogenetics_Average_Prognosis
CEBPA
CBFMYH11
bcrabl
ASXL1
AMLFAB_class_AML_zero
AMLFAB_class_AML_three
AMLFAB_class_AML_six
AMLFAB_class_AML_one_to_two
AMLFAB_class_AML_four_to_five
Age
0
5
10
15
20
Exponential Coefficient
D
Patient standardized
Drug standardized
C−index
1.00
0.75
0.50
0.25
0.00
AUC raw
log2(AUC raw)
DSS_AUC
DSS1
DSS2
DSS3
EC50
TEC50
E
Test
Training
C−index
Ridge
Elastic net
Lasso
Clinical/genetic (C220)
Clinical/genetic
Combined
Drug screen
Pre−selection | Combined
Pre−selection | Drug screen

Figure 3

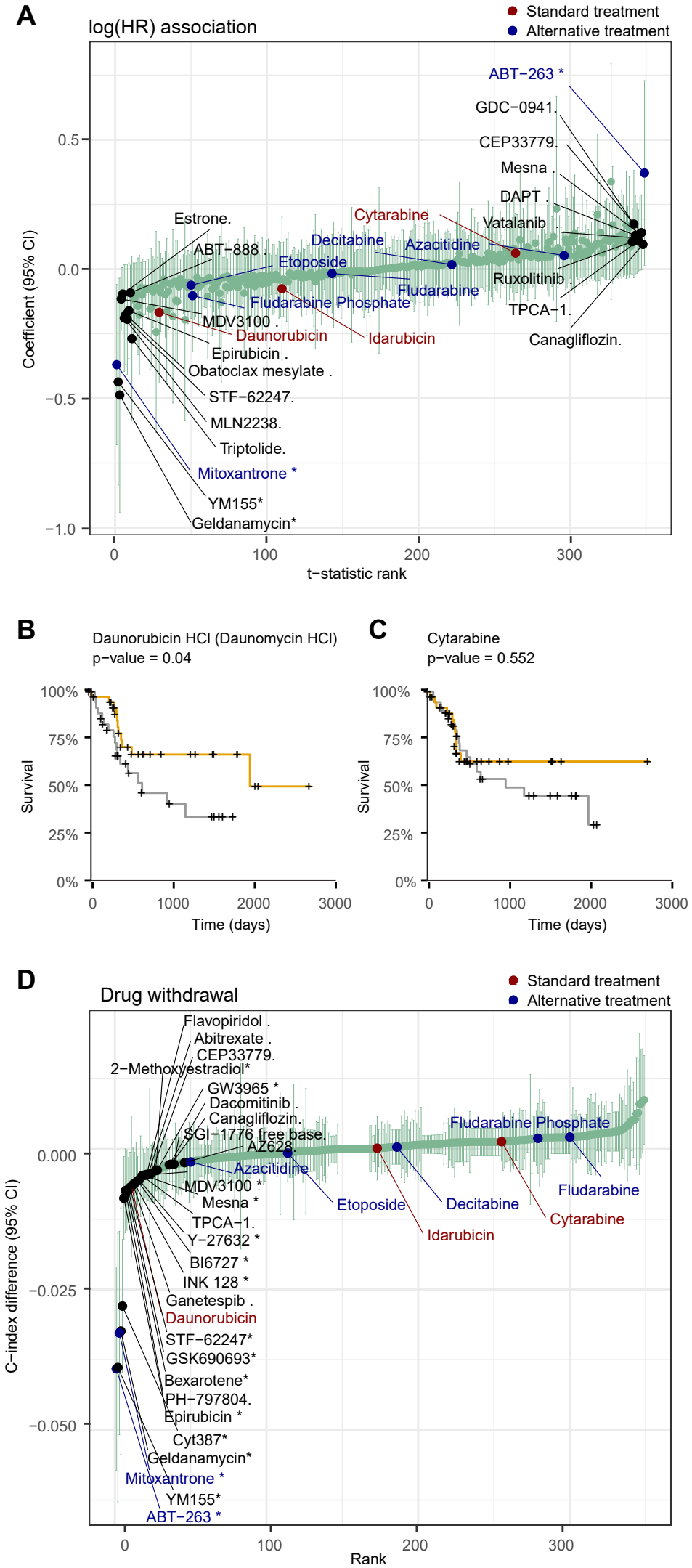

A
log(HR) association
Standard treatment
Alternative treatment
Coefficient (95% CI)
t−statistic rank
ABT−263 *
GDC−0941.
CEP33779.
Mesna .
DAPT .
Vatalanib .
Cytarabine
Decitabine
Azacitidine
Estrone.
ABT−888 .
Etoposide
Ruxolitinib .
Fludarabine
Fludarabine Phosphate
TPCA−1.
MDV3100 .
Daunorubicin
Idarubicin
Canagliflozin.
Epirubicin .
Obatoclax mesylate .
STF−62247.
MLN2238.
Triptolide.
Mitoxantrone *
YM155*
Geldanamycin*
B
Daunorubicin HCl (Daunomycin HCl)
p−value = 0.04
Survival
Time (days)
C
Cytarabine
p−value = 0.552
Survival
Time (days)
D
Drug withdrawal
Standard treatment
Alternative treatment
C−index difference (95% CI)
Rank
Flavopiridol .
Abitrexate .
CEP33779.
2−Methoxyestradiol*
GW3965 *
Dacomitinib .
Canagliflozin.
SGI−1776 free base.
AZ628.
Azacitidine
MDV3100 *
Mesna *
TPCA−1.
Y−27632 *
BI6727 *
INK 128 *
Ganetespib .
Daunorubicin
STF−62247*
GSK690693*
Bexarotene*
PH−797804.
Epirubicin *
Cyt387*
Geldanamycin*
Mitoxantrone *
YM155*
ABT−263 *
Fludarabine Phosphate
Fludarabine
Etoposide
Decitabine
Cytarabine
Idarubicin

Figure 4

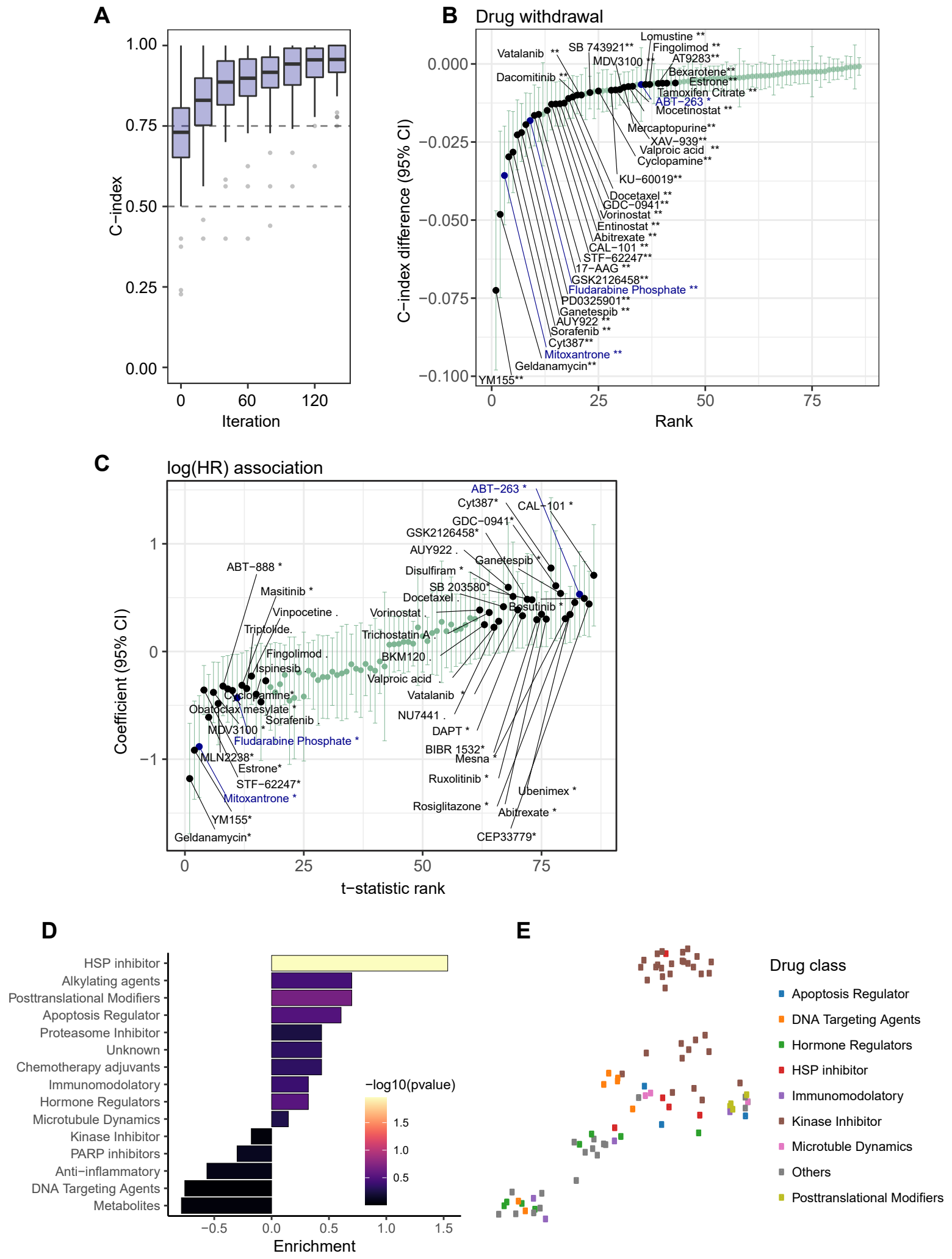
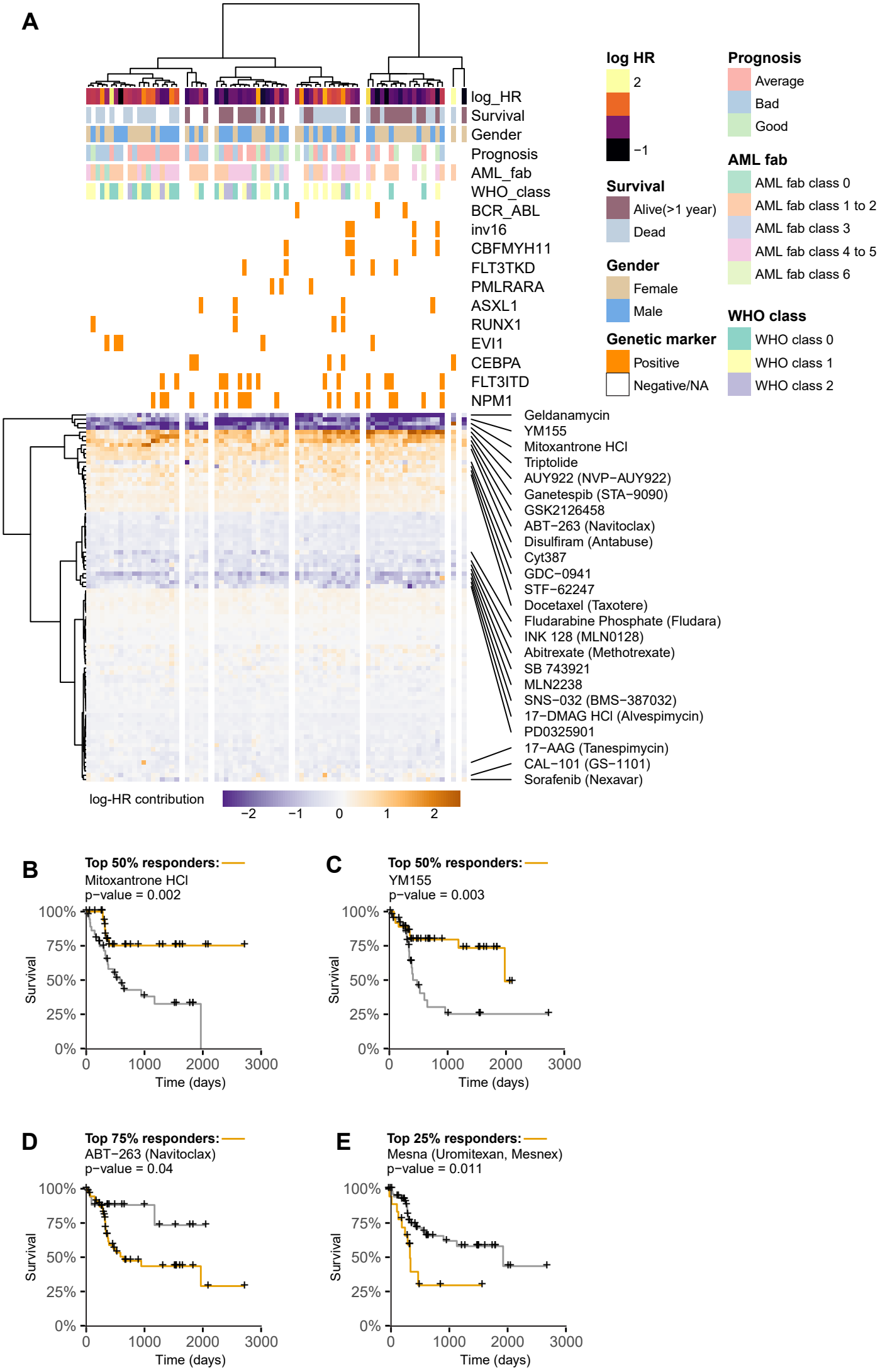

A
C−index
Iteration
B
Drug withdrawal
C−index difference (95% CI)
Rank
Lomustine **
Fingolimod **
SB 743921**
Vatalanib **
MDV3100 **
AT9283**
Bexarotene**
Estrone**
Dacomitinib **
Tamoxifen Citrate **
ABT−263 *
Mocetinostat **
Mercaptopurine**
XAV−939**
Valproic acid **
Cyclopamine**
KU−60019**
Docetaxel **
GDC−0941**
Vorinostat **
Entinostat **
Abitrexate **
CAL−101 **
STF−62247**
17−AAG **
GSK2126458**
Fludarabine Phosphate **
PD0325901**
Ganetespib **
AUY922 **
Sorafenib **
Cyt387**
Mitoxantrone **
Geldanamycin**
YM155**
C
log(HR) association
Coefficient (95% CI)
t−statistic rank
ABT−263 *
Cyt387*
CAL−101 *
GDC−0941*
GSK2126458*
AUY922 .
Ganetespib *
Disulfiram *
SB 203580*
Docetaxel .
Bosutinib *
Vorinostat .
Trichostatin A .
BKM120 .
Valproic acid .
Vatalanib *
NU7441 .
DAPT *
BIBR 1532*
Mesna *
Ruxolitinib *
Ubenimex *
Rosiglitazone *
Abitrexate *
CEP33779*
ABT−888 *
Masitinib *
Vinpocetine .
Triptolide.
Fingolimod .
Ispinesib .
Cyclopamine*
Obatoclax mesylate *
Sorafenib .
MDV3100 *
Fludarabine Phosphate *
MLN2238*
Estrone*
STF−62247*
Mitoxantrone *
YM155*
Geldanamycin*
D
HSP inhibitor
Alkylating agents
Posttranslational Modifiers
Apoptosis Regulator
Proteasome Inhibitor
Unknown
Chemotherapy adjuvants
Immunomodolatory
Hormone Regulators
Microtubule Dynamics
Kinase Inhibitor
PARP inhibitors
Anti−inflammatory
DNA Targeting Agents
Metabolites
−log10(pvalue)
Enrichment
E
Drug class
Apoptosis Regulator
DNA Targeting Agents
Hormone Regulators
HSP inhibitor
Immunomodolatory
Kinase Inhibitor
Microtuble Dynamics
Others
Posttranslational Modifiers

Figure 5


A
log_HR
Survival
Gender
Prognosis
AML_fab
WHO_class
BCR_ABL
inv16
CBFMYH11
FLT3TKD
PMLRARA
ASXL1
RUNX1
EVI1
CEBPA
FLT3ITD
NPM1
Geldanamycin
YM155
Mitoxantrone HCl
Triptolide
AUY922 (NVP−AUY922)
Ganetespib (STA−9090)
GSK2126458
ABT−263 (Navitoclax)
Disulfiram (Antabuse)
Cyt387
GDC−0941
STF−62247
Docetaxel (Taxotere)
Fludarabine Phosphate (Fludara)
INK 128 (MLN0128)
Abitrexate (Methotrexate)
SB 743921
MLN2238
SNS−032 (BMS−387032)
17−DMAG HCl (Alvespimycin)
PD0325901
17−AAG (Tanespimycin)
CAL−101 (GS−1101)
Sorafenib (Nexavar)
log-HR contribution
−2
−1
0
1
2
log HR
2
−1
Survival
Alive(>1 year)
Dead
Gender
Female
Male
Genetic marker
Positive
Negative/NA
Prognosis
Average
Bad
Good
AML fab
AML fab class 0
AML fab class 1 to 2
AML fab class 3
AML fab class 4 to 5
AML fab class 6
WHO class
WHO class 0
WHO class 1
WHO class 2
B
Top 50% responders:
Mitoxantrone HCl
p−value = 0.002
C
Top 50% responders:
YM155
p−value = 0.003
D
Top 75% responders:
ABT−263 (Navitoclax)
p−value = 0.04
E
Top 25% responders:
Mesna (Uromitexan, Mesnex)
p−value = 0.011
Survival
100%
75%
50%
25%
0%
0
1000
2000
3000
Time (days)