Robust scoring of selective drug responses for patient-tailored therapy selection

Yingjia Chen[1], Liye He[1], Aleksandr Ianevski[1], Pilar Ayuda-Durán[2,3], Swapnil Potdar[1], Jani Saarela[1], Juho J. Miettinen[1], Sari Kytölä[4], Susanna Miettinen[5,6], Mikko Manninen[7], Caroline A. Heckman[1,8], Jorrit M. Enserink[2,3,9], Krister Wennerberg[10], Tero Aittokallio[1,8,11,12,*]

1. Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, FI-20520 Helsinki, Finland
2. Department of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital, N-0310 Oslo, Norway
3. Centre for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway
4. Department of Hematology, Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland
5. Adult Stem Cell Group, Faculty of Medicine and Health Technology, Tampere University, FI-33520 Tampere, Finland
6. Research, Development and Innovation Centre, Tampere University Hospital, FI-33520 Tampere, Finland
7. Orton Orthopaedic Hospital, FI-00280 Helsinki, Finland
8. iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Hospital, Finland
9. Section for Biochemistry and Molecular Biology, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway
10. Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark
11. Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, N-0310 Oslo, Norway
12. Centre for Biostatistics and Epidemiology (OCBE), Faculty of Medicine, University of Oslo, N-0317 Oslo, Norway

* Correspondence: tero.aittokallio@helsinki.fi

Abstract

Most patients with advanced malignancies are treated with severely toxic, first-line chemotherapies. Personalized treatment strategies have led to improved patient outcomes and could replace one-size-fits-all therapies yet need to be tailored by testing a range of targeted drugs in primary patient cells. However, most functional precision medicine studies use simple drug response metrics, which cannot quantify the selective effects of drugs, that is, the differential responses of cancer and normal cells. We developed a computational method for selective drug sensitivity scoring (DSS), which enables the normalization of the individual patient's responses against normal cell responses. The selective response scoring uses the inhibition of non-cancerous cells as a proxy for potential drug toxicity, which can in turn be used to identify effective and safer treatment options. Here, we explain how to apply the selective DSS calculation for guiding precision medicine in patients with leukemia and treated across three cancer centers in Europe and the US; the generic methods are widely applicable also to other malignancies that are amenable to drug testing. The open-source and extendable R-codes provide a robust means to tailor personalized treatment strategies based on increasingly available *ex vivo* drug testing data from patients in real-world and clinical trial settings. We also make available drug response profiles to 527 anticancer compounds tested in 10 healthy bone marrow samples as reference data for selective scoring and de-prioritization of drugs that show broadly toxic effects. The procedure takes less than 60 minutes and requires basic skills in R.

Introduction

Functional testing of molecularly-targeted drugs in primary patient cells has shown clinical benefits in terms of patient stratification and patient-centric treatment optimization both in hematological malignancies[1,2] and in solid tumors[3,4]. Functional drug testing in patient cells offers clinically actionable treatment recommendations for clinical tumor boards beyond genetic testing alone[5], and *ex vivo* drug testing has faster turnaround time than sequencing data analysis (a typical functional assay takes 2-3 days, and analysis can be done in less than a day). While such direct testing of patient cells to hundreds of anticancer compounds provides an exciting approach to personalized treatment recommendations, and drug repurposing[6], critical experimental and analytical challenges remain, concerning the cancer-selectivity, reproducibility and translatability of the *ex vivo* drug testing results that needs to be considered when using these data in patient management[3,4,7,8].

To address these challenges in cell-based drug testing, we developed a model-based drug sensitivity score (DSS), which enables accurate quantification of the sensitivity of cancer cells to both standard chemotherapies and molecularly-targeted drugs[9]. The DSS calculation is applicable both to *ex vivo* drug testing of primary patient cells and to cell line drug screening *in vitro*. Using pan-cancer screening data from two independent *in vitro* studies[10,11], we showed that it is possible to achieve an improved consistency between different laboratories for drug response measurements by paying careful attention to harmonization of both laboratory assays and computational procedures[12]. For patient-centric clinical translation of *ex vivo* drug testing results, we have used the DSS calculation to tailor treatment regimens for relapsed/refractory (r/r) patients with acute myeloid leukemia (AML)[2,13], and to identify novel treatment vulnerabilities for r/r patients with chronic lymphocytic leukemia (CLL)[14].

In this Protocol, we describe several case studies and real-life examples to showcase the best use of the selective DSS calculation when tailoring targeted treatments for leukemia patients based on personalized *ex vivo* drug testing data from ongoing functional precision medicine studies carried out in three cancer centers in Finland, Norway and the US. We have implemented new features in the DSS calculation that enable better quantification of the cancer-selective effects of drugs by normalizing the responses measured in cancer patients against those from healthy controls; we have made the normal cell responses to 527 drugs available to the community to help predict and avoid potentially toxic effects when identifying safe and effective treatment options. Using the open-source DSS codes and data, we demonstrate an enhanced robustness of the selective DSS calculation in terms of improved correlation between technical replicates in two AML patient screening studies.

AML sample cohorts and *ex vivo* drug response data used in the protocol

We detail how to best use the selective DSS calculation in *ex vivo* functional drug profiling to guide treatment selection for patients with leukemia, using drug testing data from two cancer studies in Europe and in the US that showcase the wide applicability of the methods (**Table 1**).

The FIMM-AML patient cohort contains drug profiling data from a prospective series of 252 samples from 186 patients with acute myeloid leukemia (AML) collected in Helsinki, Finland[2]. Due to the evolving compound libraries used over the years in the ongoing study, the patient samples were tested with varied collections of drugs, ranging from about 300 to 527 chemical compounds. Here, we used the *ex vivo* drug testing data from 125 patient samples, tested with the two newest compound libraries, either FO4B (525 drugs) or FO5A (527 drugs). To have a standardized set of reference control data for selective drug response scoring, we profiled bone marrow samples from 10 healthy donors using the newest FO5A panel with compounds tested in 5 concentrations (see

Materials). To study reproducibility of drug response assays and metrics, we also made five new replicate screens of an AML patient tested in Oslo[15], using the FIMM FO5A compound collection.

The newest data release from the BeatAML study contains a cumulative cohort of 942 samples from 805 AML patients collected over a span of 10 years at the Oregon Health & Science University (OHSU)[16]. The patient samples were tested with 166 compounds in 7 increasing doses (see **Materials**). Since the BeatAML cohort does not include healthy control samples, we used the FIMM control data to score selective responses in the BeatAML patients. The BeatAML data cohort contains replicate measurements from a subset of the samples and compounds, which were used to investigate the robustness of the drug testing assay and the selective DSS calculation, compared to the standard response metrics, when using FIMM normal cell control data for the selective response scoring. The BeatAML drug testing is based on a relatively different assay and compound set, compared to those used in Helsinki or Oslo, demonstrating the wide applicability of the methods.

Table 1. Summary of the AML patient cohorts used in the protocol

Cohort (total number of patient samples)	Samples with drug response profiles	Disease stage	Sample types	Number of drugs tested
FIMM (n = 252)	n = 164	Diagnosis = 79 Relapse = 53 Refractory = 32	Bone marrow = 150 Peripheral blood = 14	Up to 527*
Beat AML (n = 942)	n = 631	Diagnosis = 397 Relapse = 45 Remission = 19 Residual = 138 NA = 32	Bone marrow = 321 Peripheral blood = 292 Leukapheresis = 18	166

*In this protocol, we used data from n=79 bone marrow (BM) samples tested with FIMM FO4B collection (525 drugs), and n=46 BM samples tested with FO5A collection (527 drugs). The overlap between these collections is 431 drugs. The earlier FIMM compound libraries contain less number of drugs.[2]

Drug sensitivity scoring pipeline for patient-centric treatment selection

We have previously developed a quantitative modeling approach, named drug sensitivity score (DSS), which extracts and integrates multiple dose-response parameters into a single response metric (selective DSS, sDSS) to identify differential drug response patterns between cancer and healthy control cells, rather than scoring drug activity in cancer cells alone[9]. Our initial DSS calculation was based on closed-form integration of the area under the dose-response curve (AUC), using the 4-parameter logistic model (4P-LM) as the dose-response model[9], but the generic modeling approach can also be used in the context of other parametric models, such as sigmoidal or Hill slope response functions, using either analytical or numerical integration procedures. More complex dose-response models require larger numbers of dose-measurements, which may limit the number of tested drugs in scarce patient cells, while the four parameters of 4P-LM can be accurately estimated with 5 dose-response pairs per each drug and sample.

The calculation of standard activity metrics, such as DSS, AUC or half-maximal inhibitory concentration (IC$_{50}$), is implemented in an interactive analysis application, Breeze, which is available both as a web-portal and an open-source R-package[17]. After quality control (QC) of the multi-well

plate drug response assay data, based on the preferred positive and negative controls (**Fig. 1a**), the user can select the dose-response model (e.g. 4P-LM), and then perform curve-fitting (**Fig. 1b**), and calculate sDSS as the area between the dose-response curves in patient cells and control cells (**Fig. 1c**). Such selective drug response scoring is a critical part of any translational application, since inhibition of non-cancerous control cells can be used as a preclinical proxy for potential toxic effects of the drugs. DSS-based response metrics enable direct calculation of the selective responses, whereas differential IC_{50} is not so straightforward to interpret.

Based on the healthy control samples, one can estimate standard deviation of compound-specific reference distribution using the z-score statistic (see **Box 1**). While sDSS normalizes the patient responses by the average value of the controls (i.e., location normalization; **Fig. 1e**), calculation of z-score (or robust z-score) for each compound further normalizes the patient responses in terms of variability over the controls (i.e., scale normalization; **Fig. 1f**). This additional scaling improves both the comparison of drug responses within a patient sample (i.e., identification of optimal treatment option for each individual patient; **Fig. 1g**), and comparison of the response of a particular drug across patients and controls (e.g., identification of selective responses for drug repurposing; **Fig. 1h**). The statistical significance of an observed zDSS level can be calculated with *z*-test or using one-sample permutation test for non-normal response data (**Fig. 1f**).

Box 1 | Calculation of selective DSS for each patient and drug

Let us use **R**(x) to model the control-well normalized drug response at a concentration x. The area under the dose-response curve (AUC) over the dose range (x_1, x_2), where the responses exceed a user-specified minimum activity level (A_{min}) can be calculated either using analytical or numerical integration. The original DSS article shows how to calculate closed-form solution when using the 4P-LM function the dose-response model **R**(x)[9].

Non-selective DSS is a modified AUC calculation in which the area of the rectangle below the minimum activity level (i.e., a user-defined noise level for a high-throughput screening assay) is subtracted from the integrated AUC, and the difference is then divided by the maximal response area to take into account potential differences in the minimal and maximal tested concentrations between the compounds (C_{min}, C_{max}):

$$\mathrm{DSS}_1 = \frac{\mathrm{AUC} - A_{min}(x_2 - x_1)}{(100 - A_{min})(C_{max} - C_{min})}$$

To penalize those compounds that are effective at higher tested concentrations only (i.e., show potential off-target toxic responses), the DSS_1 summary score is further normalized by the logarithm of the top asymptote (R_{max}) of the estimated dose-response model that corresponds to the maximal estimated response of the drug:

$$\mathrm{DSS}_2 = \frac{\mathrm{DSS}_1}{\log_{10} R_{max}}$$

To further favor those compounds that show potency across a relative wide concentration window in a given patient case, instead at the higher-end of the drug concentration range only, the DSS_2 score is further divided by the relative dose range (x_1, x_2) over which the drug response exceeds the minimum activity threshold A_{min}:

$$\mathrm{DSS}_3 = \mathrm{DSS}_2 \frac{x_2 - x_1}{C_{max} - C_{min}}$$

Using any of the above DSS versions (DSS_2 was used in the results of this protocol), the selective DSS (sDSS) for a particular patient-drug pair is calculated as the difference between the drug response in the patient cells (DSS_p) and the average drug response of the same drug in the normal control cells [mean(DSS_c)]:

$$sDSS = DSS_p - mean(DSS_c)$$

When there are enough (≥10) standardized dose-response measurements from the control cells, one can estimate also standard deviation (SD) of the drug responses in the control cells (σ_c), hence leading to a z-score:

$$zDSS = \frac{DSS_p - mean(DSS_c)}{\sigma_c + 1}$$

where the addition of one into the denominator makes the calculation feasible also for those cases, where the SD of controls becomes zero or close to zero (e.g. when DSS is close to zero in all the healthy control samples for a targeted agent).

To make the drug selectivity calculation less sensitive to outliers in the control samples, a robust (or modified) z-score can be calculated by replacing the mean and SD of controls by their median [median(DSS_m)] and median absolute deviation (MAD_c):

$$rDSS = \frac{DSS_p - median(DSS_m)}{MAD_c + 1}$$

where the addition of one is to again accommodate the case with a MAD of zero or close to zero.

Normalization of patient responses with control responses in various drug classes

The z-score normalization of drug responses (zDSS) is expected to better account for differences in responses between drug and target classes, and therefore lead to better comparability of the drug testing results both between patients and drugs, which is important for biomarker discovery and treatment tailoring, respectively. In particular, conventional chemotherapies often lead to higher DSS responses (and wider distributions), both in patients and healthy controls (**Fig. 2a**). While sDSS enables effective normalization of average responses between patients and controls, and hence better comparability between targeted and non-targeted drugs (**Fig. 2b**), zDSS further normalizes variances and scales of the distributions, hence improving identification of both cancer-selective and patient-specific responses (**Fig. 2c**). In the example patient case, the list of top-compounds changed from pan-PI3K/HDAC inhibitor fimepinostat and protein translation inhibitor omacetaxine to targeted BCL-2 inhibitor venetoclax when using different metrics (**Figs. 2a-c**).

A further investigation of the two targeted compounds, navitoclax and venetoclax, demonstrates the benefits of zDSS and robust z-score (rDSS), even for drugs with very similar modes of action (BCL-2 inhibition). When using the original DSS values, navitoclax and venetoclax appear to elicit relatively similar response distributions among most patients with AML (**Fig. 2d,g**); however, venetoclax shows more cancer-selective responses, i.e., lower responses in the controls while subset of the patient have high DSSs, which makes zDSS distributions and scales different between the two drugs (**Fig. 2e,h**). For targeted drugs, the distribution of control responses may be very narrow and centered close to zero, requiring the addition of "one" in the denominator of z-score when calculating zDSS (**Box 1**). As there may also appear outliers among the control responses, the use of robust

rDSS may become beneficial when comparing selective responses for subsets of patients (**Fig. 2f,i**). In the example patient (the red dot), venetoclax shows a higher rDSS, even if the DSS of navitoclax was much higher.

Improved reproducibility of selective drug sensitivity and robust z-scores

The dose-response model estimation is expected to make the DSS calculations robust against sources of technical variability. For instance, the dose-response curve fitting enables interpolation of potential missing or poor-quality dose-responses within the tested concentration range. Further, in contrast to single-point response metrics, such as IC_{50} and EC_{50}, which are dependent on the concentration ranges being tested for each compound, summary response metrics, such as AUC and DSS, provide more comparable results for compounds tested under variable concentration windows; for instance, we have shown that DSS metric enables a more standardized response calculation between studies by focusing on the overlapping concentration ranges when calculating DSS (modified AUC), leading to more consistent *in vitro* drug testing results between drug screening studies in overlapping pan-cancer cell lines[12]. We have also shown earlier that DSS leads to more robust responses than IC_{50} between replicated AML cell line screens *in vitro*[9].

To investigate reproducibility of *ex vivo* assays, we made 5 replicate drug screens of a single AML patient sample tested in Oslo. In general, IC_{50} had the lowest correlation among the replicates, whereas the selective DSS metrics improved correlations (**Fig. 3a**). While both sDSS and rDSS led to generally robust profiles, some variability remains, which resulted in deviation from the diagonal line (**Fig. 3b,c**). In a further analysis of four outlier drugs, we noted that the selective response of dasatinib was captured by both metrics (**Fig. 3f**), while GSK-2636771 gave mixed results (**Fig. 3e**); rDSS identified in a robust manner the two compounds (ridaforolimus and omipalisib) for which controls showed dose-responses similar to that of patients (**Fig. 3h,i**). There were no apparent differences in the correlations across the drug classes (**Fig. 3b,c**). We observed similar results in the replicated measurements of the BeatAML dataset for a subset of patients and compounds (**Fig. 3d**), where rDSS again resulted in robust response profiles (**Fig. 3g**).

Batch effect correction when combining drug testing data from multiple sources

When using drug response profiles from different cohorts or compound collections, e.g., from a multi-year functional precision medicine study, there may occur batch effects that need to be taken into account. As an example, response profiles of the BeatAML patient samples are rather different from those of the FIMM controls, due to both biological and experimental factors (**Fig. 4a**, left); the technical variability should be corrected, e.g., using the ComBat batch effect correction method[18], before the healthy control profiles can be used to normalize the response profiles of the BeatAML samples. When there are no clearly separable sample subgroups visible in the 2D projection of the high-dimensional DSS profiles, using e.g. principal component analysis (PCA), one can conclude that the batch effect correction was successful (**Fig. 4a**, right panel), or not needed at all.

However, it is equally important not to normalize out true biological variability or over-normalize in the case of no obvious batch effects. For example, there are true biological differences in drug response profiles between FIMM patients and healthy controls that are expected to remain in the DSS data, regardless of whether the old (FO4B) or new (FO5A) compound collection was used (**Fig. 4b**). However, there appeared a clear batch effect between the two compound collections, when

analyzing the healthy controls only, which needs to be corrected if all the healthy controls will be used for calculating the selective drug responses (**Fig. 4c**). *Note*: in the analyses shown elsewhere in this protocol, only the FO5A drug response data from the healthy controls were used to have a standardized set of control data for the selective DSS calculations.

Alternative methods

Half-maximal inhibitory concentration (IC_{50}) and half-maximal effective concentration (EC_{50}) are popular measures of the potency of a compound in inhibiting a cellular target such as protein function or cell viability. However, such single-point potency measures are not designed for personalized therapy optimization in high-throughput screening, where the compounds are often tested in different concentration windows. In contrast, model-based drug sensitivity quantitation with DSS effectively captures various types of activity information from complex dose-response measurements through combining IC_{50}, slope and the other model parameters for drug sensitivity scoring. Importantly, the area-based metrics, such as DSS and AUC, enable straightforward calculation of differential responses, relative to that of the control cells, hence identifying selective responses[9]. We note that the differential pIC_{50} (negative logarithm of IC_{50}, see **Fig. 3**) corresponds to the standard selectivity index, calculated as the average of the IC_{50} value on the normal cells divided by the IC_{50} value on the cancer cells.

In this protocol, using the new selective drug response calculations, we demonstrated that location and scale normalization of the patient *ex vivo* responses using the healthy control profiles (**Fig. 1**) leads to several benefits. When compared to existing drug response metrics, such as IC_{50}, AUC and DSS, the new zDSS and rDSS calculations improved both the comparability between the targeted and chemotherapy drugs when selecting the best treatment regimen for a given patient (**Fig. 2**), and the robustness of the patient *ex vivo* responses in terms of enhanced correlation between technical replicates in the screening data from two AML studies (**Fig. 3**). Even if there were differences in the results between the various selective responses metrics, depending on the application use case, the observed differences between zDSS and rDSS metrics were not so striking, when compared with the results of the other metrics. In the current results, we used the DSS_2 version in the sDSS, zDSS and rDSS calculations, since it has been widely used in the FIMM-AML programme[2,13].

Advantages and future directions
The current implementation of precision cancer medicine is mainly driven by genomics, and the current molecular-clinical tumor boards rely extensively on genomic characterization to identify personalized therapeutic interventions. However, genomic biomarkers can guide treatment decisions only for a fraction of cancer patients, and more often genomic data are neither actionable nor predictive[1,2,3]; therefore, there is an increasing interest in using functional assays to guide cancer treatment by capturing information from direct perturbations of patient-derived cells, such as using *ex vivo* drug sensitivity screening, with the aim to extend both the biomarker space and patient populations suitable for precision oncology treatment selection[3,4,5]. Such a functional precision medicine approach has been spearheaded in hematological malignancies, where *ex vivo* drug testing has been shown to contribute to significant additional and clinically actionable therapeutic insights for individual patients[13-16]. Drug testing profiles can be generated in a few days (two or three-day assay, and <1 day for the data analysis), hence enabling rapid clinical translation[1,2].

In the past, it has been challenging to predict patient responses to conventional chemotherapeutics using *ex vivo* testing, both in hematological malignancies and in solid tumors[21,22], because sensitivity to chemotherapy drugs depends on the cell proliferation rate of the *ex vivo* models. We note that the selective DSS calculation can also be extended to account for differences in cell growth rates to better capture drug-induced effects[23,24]. Since bone marrow toxicity is a common restrictive adverse effect among anticancer drugs, the healthy control bone marrow samples provide meaningful control responses also for other cancer types than AML[25], toward identification of cancer-selective drug responses and to de-prioritize drugs in preclinical testing that show cytotoxicity in the normal bone marrow cells. We therefore expect that the robust and selective DSS calculations will lead to more effective and safe therapy selections, especially when combined in the future with cell population-level drug testing assays, such as those based on imaging[26,27] and flow cytometry[28,29]. Once tested at multiple sites and in large-scale screening, these methods will provide improved means to identify patient-specific treatments that selectively inhibit cancer cells while avoiding inhibition of non-cancer cells.

Beyond monotherapy recommendations, an *ex vivo* drug testing platform was recently used to guide drug combination treatment in patients with r/r non-Hodgkin lymphoma (NHL)[30]. We have also shown earlier how DSS calculation enables prediction of cancer-selective and patient-specific drug combinations for patients with T-cell prolymphocytic leukemia (T-PLL)[31] and high-grade serous ovarian carcinoma using single-cell imaging assay[32]. The open-source R codes allow others to modify the drug scoring methodology and implement it as part of their in-house drug screening pipelines. Selective DSS calculation can be easily extended to include also other modules of drug screening data analysis, including various QC parameters and interactive visualization options, such as those implemented in *Breeze* for monotherapies[17], and *screenwerk* for combination therapies[33], as well as curve-fitting options that can model various dose-response relationships[34]. Finally, ComBat and other methods for batch effects correction enable meta-analyses between multiple patient cohorts that may help speeding-up and de-risking of clinical studies by identifying optimal patient sub-groups for new or existing targeted treatments or their combinations.

Regulatory approvals

In this computational protocol, we assume that all the necessary regulatory and ethical approvals are in place before one start screening and scoring compound sensitivities in patient samples or cell lines. Such approvals may require the use of Material Transfer Agreements (MTAs) for the transfer of cell lines or pharmaceuticals and other chemicals between institutions for use in research, as well as Ethical Approvals for the use of compound testing in human samples. The time required to obtain such permissions, as well as their respective costs, depend on the country and institution where the research is conducted, and whether the external parties, such as pharmaceutical companies or academic institutions, have any usage restrictions related to intellectual property (IP) rights of the compounds or cell lines. In case there are no relevant IP issues, and the research is well-justified from the ethical point of view, obtaining such permissions is typically straightforward, but it may take several months. When working on human material, appropriate national laws and institutional regulatory board guidelines must be followed, and informed consent obtained from the human subjects. The compound testing data used in this protocol originates from ongoing studies that have been approved by the local ethics committees, as described in the original studies[2,13, 15, 16].

Materials

AML patient cohorts and drug testing assays

The FIMM-AML and BeatAML cohorts contained adult patients with AML (median age at diagnosis of the patient cohorts were 62 and 61 years, respectively), from different disease stages (diagnosis, residual, remission and relapse/refractory), carrying various cytogenetic and molecular abnormalities. The patients had been treated with various treatment regimens at different stages, including conventional chemotherapies that combine anthracyclines with nucleoside analogs, cumulative treatment, induction treatment, and allogeneic hematopoietic stem cell transplantation (alloHSCT). The FIMM healthy donor cohort contained eight elderly individuals undergoing hip replacement surgery, and two healthy young adults (4 male and 6 female donors). The median age of the healthy donors was 65.5 years (eldest 78 years, youngest 19 years).

CAUTION Bone marrow aspirates from the elderly healthy donors were obtained with signed informed consent under approval of the Tampere University Hospital Ethics Committee, Tampere, Finland (R15174). Helsinki samples from adult patients with AML and young healthy donors were collected with signed informed consent with protocols in accordance with the Declaration of Helsinki (Ethical Committee Statement 303/13/03/01/201, latest amendment 7 dated June 15, 2016. Latest HUS study permit HUS/395/2018 dated February 13, 2018).

In the FIMM-AML and healthy donor cohort[2], mononuclear cells (MNC) were isolated by Ficoll-Paque centrifugation from freshly collected bone marrow aspirates and peripheral blood specimens. Comprehensive set of >500 anti-cancer compounds, consisting of conventional chemotherapeutics and a broad range of targeted oncology compounds was tested in 5 doses over a 10,000-fold concentration range, allowing for the establishment of accurate dose–response curves for each drug in each patient and control sample. The chemical compounds, DMSO (negative control) and benzethonium chloride (positive control) were added to 384-well plates using an acoustic liquid dispensing system Echo 550 (Labcyte). Freshly isolated MNCs were resuspended in conditioned media (CM), constituted of 77.5% RPMI 1640, 10% FCS, 12.5% human HS-5 bone marrow stromal cell line–derived CM, and 1% penicillin and streptomycin. For the AML samples and three of the healthy samples 5-µL cell-free medium was added to dissolve compounds, followed by 20 µL cell suspension containing 5,000 to 10,000 cells to each well using a Multidrop Combi (Thermo Fisher) or a Multiflo FX (BioTek) dispenser. For seven of the healthy samples 25 µL cell suspension containing 5,000 cells was added to each well without first dissolving the compounds in 5-µL cell-free medium. The plates were incubated at 37°C in 5% CO2 for 72 hours. Subsequently, CellTiter-Glo 2.0 (Promega) reagent was added to all wells, and cell viability as luminescence generated by total cellular ATP was measured using a PHERAstar FS (BMG Labtech).

A single AML patient treated in Oslo[15] was tested in 5 replicates with the FIMM FO5A compound collection after cryopreservation. MNCs were freshly isolated from bone marrow after lymphoprep density gradient centrifugation, and were placed in freezing media (50% MCM, 40% FBS, 10% DMSO) and placed at -80°C in a Corning® CoolCell® container until temperature equilibration, before transferring the vials to a nitrogen tank. For the screen, the cells were recovered in MCM (PromoCell) with 1% Penicillin + Streptomycin (PS) (Gibco,15140-122) overnight. After recovery, the cells were used in a similar way as the freshly isolated cells. 3800 cells per well were seeded into the 384-well plates using a CERTUS Flex liquid dispenser (Fritz Gyger AG) and incubated at 37°C in 5% CO2 for 72 hours. The cell viability was measured with CTG 2.0, using an EnVision 2105 multilabel plate reader (Perkin Elmer) with a stacker.

In the BeatAML cohort[16], mononuclear cells were isolated by Ficoll gradient centrifugation from freshly obtained bone marrow aspirates or peripheral blood draws. Freshly isolated MNCs were

plated into an ex vivo drug sensitivity assays within 24h. 10,000 cells per well were arrayed into three, 384-well plates containing 166 small-molecule inhibitors. Drug plates were created using inhibitors purchased from LC Laboratories and Selleck Chemicals and master stocks were reconstituted in dimethyl sulfoxide (DMSO) and stored at −80 °C. Master plates were created by distributing a single agent per well in a seven-point concentration series, created from threefold dilutions of the most concentrated stock resulting in a range of 10 μM to 0.0137 μM for each drug (except dasatinib, ponatinib, sunitinib and YM-155, which were plated at a concentration range of 1 μM to 0.00137 μM). DMSO-control wells and positive-control wells containing a drug combination of flavopiridol, staurosporine and bortezomib were placed on each plate, with the final concentration of DMSO ≤0.1% in all wells. Cells were seeded into 384-well assay plates at 10,000 cells per well in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with fetal bovine serum (FBS) (10%), L-glutamine, penicillin–streptomycin, and β-mercaptoethanol (100 μM). After three days of culture at 37 °C in 5% CO_2 MTS reagent (CellTiter96 AQueous One; Promega) was added, the optical density was measured at 490 nm, and absorbance values were adjusted to a reference blank value and then used to determine cell viability (normalized to untreated control wells).

Software

Instructions for setting up the DSS calculation in an R environment:
- R v.3.5.1 or newer (https://www.r-project.org)
- RStudio (https://www.rstudio.com/)
- Download the data and R scripts with all requisitions from the GitHub directory https://github.com/yingjchen/DSS-v2.0
- Code to run Procedure 1 with examples: https://github.com/yingjchen/DSS-v2.0/blob/main/procedure1.R
- Code to run Procedure 2 with examples: https://github.com/yingjchen/DSS-v2.0/blob/main/procedure2.R

Required hardware for the R environment on a desktop machine:
- Computer with ≥2 GB RAM memory
- Hard drive with ≥0.5 GB free space
- A stable broadband internet connection

Procedure 1

CRITICAL We illustrate the utility of the computational methods in a use case of AML patients and healthy controls from FIMM. We provide below code and data for selective drug sensitivity scoring using DSS, sDSS, zDSS and rDSS calculations (see **Box 1**) for normalization of patient responses (see **Fig. 2**). The links used for setting up the DSS computations are listed in the 'Materials' section. Fig. 5 summarizes the key steps of the procedures, with their input-output relationships.

Setup

1. Install the required packages and load these packages to the R environment:

```
packages.required <- packages.required <-
c("matrixStats","dplyr","reshape","reshape2", "scales", "drc", "caTools",
"ggplot2", "data.table", "stringr","MESS", "BiocManager","svMisc", "egg",
"pheatmap")

packages.bio <- c("sva", "pcaMethods")

packages.new <- packages.required[!(packages.required %in%
installed.packages()[,"Package"])]

if(length(packages.new)) install.packages(packages.new)

if (!requireNamespace(packages.bio, quietly = TRUE))

    BiocManager::install(packages.bio)

lapply(packages.required, library, character.only = T)

lapply(packages.bio, library, character.only = T)
```

?TROUBLESHOOTING

2. Set the working directory using *setwd* function (replace `'/path/to/working/directory/'` with the desired path):

```
path_to_working_directory <- '/path/to/working/directory/'

setwd(dir = path_to_working_directory)
```

3. Download the example data and R scripts from Github and set the working directory to the folder DSS-v2.0-main using *setwd* function:

```
download.file(url = 'https://github.com/yingjchen/DSS-
v2.0/archive/refs/heads/main.zip', destfile = 'DSS-v2.0-main.zip')

unzip('DSS-v2.0-main.zip')

setwd(dir = file.path(path_to_working_directory, 'DSS-v2.0-main'))
```

DSS computation

4. Import DSS.R script adapted from the Breeze R-package[17] to compute different versions of DSS (DSS_1, DSS_2 and DSS_3) and import helper functions from HelperFunctions.R script using *source* function:

```
source('./DSS.R')

source('./HelperFunctions.R')
```

5. Load the example *ex vivo* dose-response profiles (cell viability at five drug concentrations) of 3 FIMM-AML patient samples tested with up to 527 anticancer drugs.

```
path_to_exampledata <- './exampleData_procedure1.csv'

df_dose.responses <- read.csv(path_to_exampledata, header = T,sep =
',',check.names = F)
```

6. Use the function *DOSE_RESPONSE_PROCESS* from the HelperFunctions.R script to calculate the percentage of growth inhibition from the cell viability data. Set the viability argument to 'FALSE' if input data is cell growth inhibition data. After that, prepare the grouped dataset as input for quantifying drug sensitivity:

```
df_dose.responses.list <- DOSE_RESPONSE_PROCESS(df_dose.responses, viability =
TRUE)
```

7. Calculate different versions of DSS and other standard drug response metrics (e.g. AUC and IC_{50}) using the *CALC_METRICS* function in the HelperFunctions.R script.Users can find the fitted dose-response growth inhibition curves in the folder ./IC50 inside the current working directory when the graph argument is set 'TRUE'.

```
df.metrics <- CALC_METRICS(dose_responses = df_dose.responses.list[[1]],
dose_responses_grouped = df_dose.responses.list[[2]], graph = FALSE)
```

?TROUBLESHOOTING

8. Set DSS_2 as the drug response metric (as an example). By utilizing the function HEATMAP_SD from the HelperFunctions.R script, users can plot a heatmap showing DSS_2 of the drugs with the highest standard deviations across the samples. The argument proportion can be adjusted to specify the proportion of most variable drugs (e.g. 10% drugs selected with the code provided below). The heatmap is saved as a pdf file and customized file names can be assigned to the argument filename, where needed. If the argument filename is left empty, the file is named as 'procedure1_step8_Breeze_DSS_53drugs_heatmap.pdf'.

```
patients.dss <- as.data.frame(acast(df.metrics,df.metrics$Patient.num ~
df.metrics$drug, value.var = 'DSS2'))

HEATMAP_SD(patients.dss, proportion = 0.1, filename = "")
```

?TROUBLESHOOTING

Importing the control sample DSS profiles

9. Load drug response profiles of the 10 healthy controls (DSS values of 527 drugs per control):

```
path_to_controldss <-'./controls/File_1_Drugname_response_DSS_10Healthy.txt'

controls.dss <- read.csv(path_to_controldss, header = T, sep = '\t', row.names =
1,stringsAsFactors = F, check.names = F)
```

Searching for the drugs of interest from the drug library

10. Upload the FIMM FO5A compound library of 527 drugs. Users can find the corresponding control DSS values for their drugs of interest by using the drug synonyms or standard InChIKey (International Chemical Identifier) in the drug library. In an example below, the function *DRUG_FILTER_SYNONYMS* from the HelperFunctions.R script is implemented to search for the drugs with drug synonyms and to filter out unmatched drugs.

```
path_to_druglibrary <- './controls/File_2_Drugname_library_527D.txt'
```

```
df_drug.library <- read.csv(path_to_druglibrary, header = T, sep = '\t',
row.names = NULL,stringsAsFactors = F, check.names = F)

patients.dss <- DRUG_FILTER_SYNONYMS(patients.dss, df_drug.library)
```

Calculating the selective drug sensitivity scores

11. Compute the descriptive statistics (i.e. mean, standard deviation, median, and median absolute deviation) of DSSs for each drug over the controls:

```
controls.summary <-
as.data.frame(rbind(colMeans(as.matrix(controls.dss)),colSds(as.matrix(controls.
dss)),colMedians(as.matrix(controls.dss)), colMads(as.matrix(controls.dss))))

rownames(controls.summary ) <- c('mean', 'sd', 'median', 'mad')
```

12. Normalize and scale the patient-specific responses to each drug against the responses to the same drug in the healthy controls. By running the following commands in R line-by-line, calculate the preferred selective drug response score (i.e. sDSS, zDSS, and rDSS):

```
patients.sdss <- patients.dss - slice(controls.summary['mean',
colnames(patients.dss)],rep(1:n(), each = nrow(patients.dss)))

patients.zdss <- (patients.dss - slice(controls.summary['mean',
colnames(patients.dss)],rep(1:n(), each =
nrow(patients.dss))))/(slice(controls.summary['sd',
colnames(patients.dss)],rep(1:n(), each = nrow(patients.dss))) + 1)

patients.rdss <- (patients.dss - slice(controls.summary['median',
colnames(patients.dss)],rep(1:n(), each =
nrow(patients.dss))))/(slice(controls.summary['mad',
colnames(patients.dss)],rep(1:n(), each = nrow(patients.dss))) + 1)
```

Plotting the drug response distributions

CRITICAL The comparison of drug responses between targeted drugs and chemotherapy drugs for an example AML patient (AML_013_01) is shown in **Fig. 2a**. The *SAMPLE_DSS_CONCAT* function from the HelperFunctions.R script was used to merge the response metrics, where the sample_id variable can be modified to select other samples. The drug response distributions for patient AML_013_01 can be visualized using the function *CHEMO_TARGETED_PLOT*. In the example code below, DSS was used in the visualization.

13. To select the response metric (i.e. sDSS, zDSS, and rDSS), find the corresponding column name of the metric of interest in the sample_dss dataframe, and then set the argument metric to that column name. The plot is saved as a pdf file and customized file names can be assigned to the argument filename, where needed. If the argument filename is left empty, the file is named as 'procedure1_step13_chemotherapy_targeted_drugs_DSS_distribution.pdf'.

```
sample_id <- 'AML_013_01'

sample_dss <- SAMPLE_DSS_CONCAT(patients.dss, patients.sdss, patients.zdss,
patients.rdss, sample_id = sample_id)

CHEMO_TARGETED_PLOT(sample_dss, metric = 'DSS', filename = "")
```

14. The function *SELECTIVE_SCORE_PLOT* supports users to compare the data distributions of the response metric (i.e. DSS, sDSS, zDSS, and rDSS) using additional density plots. The

density plot is saved as a pdf file and customized file names can be assigned to the argument filename, where needed. If the argument filename is left empty, the file is named as 'procedure1_step13_density_distribution_response_scores.pdf'.

```
SELECTIVE_SCORE_PLOT(sample_dss, filename = "")
```

?TROUBLESHOOTING

Procedure 2

Batch effect correction

CRITICAL To combine drug response profiles from multiple sources, we applied the ComBat algorithm to adjust for the known batch effects. **Suppl. Fig. 1** shows the visual exploration of high-dimensional DSS profiles from BeatAML and FIMM-AML cohorts. Probabilistic PCA (PPCA)[19] was used to visualize the example dataset of DSS_2 profiles with missing data for 147 BeatAML patient samples, 125 FIMM-AML patient samples and 10 healthy controls. Given the differences between two cohorts, we used the *ComBat* function from SVA package[18] to correct the batch effects (**Suppl. Fig. 1**). ComBat uses either parametric or non-parametric empirical Bayes frameworks for adjusting data for batch effects. We recommend the use of the nonparametric version for DSS data, where no prior assumptions about DSS data distributions are required.

1. Run steps 1-3 of Procedure 1 to set up the R environment, if needed, before running the following codes:

```
source('./HelperFunctions.R')

path_to_exampledata <- './exampleData_procedure2.csv'

df.dss <- read.csv(path_to_exampledata, header = T,sep = ',',  row.names = 1,
check.names = F)
```

2. Utilize the function PCA_FUNC from the HelperFunctions.R script to identify the missing values in the example dataset. In case of missing DSS values, make a PPCA plot of the DSS profiles to check for potential batch effects (**Fig. 5**, Procedure 2). In case of no missing values, one can make a standard PCA plot instead. When there are no clearly separable sample subgroups visible in the PCA or PPCA projections, one can conclude that no batch effect correction is needed.

```
df.dss.1 <- df.dss[, 1 : (ncol(df.dss) - 3)]

score_pca <- PCA_FUNC(df.dss.1)

score_pca$group <- paste(df.dss$cohort,  df.dss$status,  sep = ' ')

ggplot(score_pca, aes(x = PC1, y = PC2, color = group)) +

  geom_point() + labs(title = "DSS",  x = "PC1", y = "PC2") +

  theme_classic()

ggsave("./procedure2_DSS_ppca.pdf", height = 10, width = 10)
```

3. Perform ComBat correction to adjust for all the identified batch effects:

```
df.dss.combat <- ComBat(dat = t(df.dss.1), batch = as.factor(df.dss$cohort), mod
= NULL, par.prior = F, prior.plots = F)
```

?TROUBLESHOOTING

4. Utilize the function PCA_FUNC to identify the missing values in ComBat-corrected DSS
   profiles and then make a PPCA plot to confirm the successful removal of the batch effects. One
   can make a PCA plot if no missing data are present in the ComBat-corrected DSS profiles.

```
score_pca_combat <- PCA_FUNC(t(df.dss.combat))

score_pca_combat$group <- paste(df.dss$cohort,  df.dss$status,  sep = ' ')

ggplot(score_pca_combat, aes(x = PC1, y = PC2, color = group)) +

  geom_point() + labs(title = "ComBat DSS",  x = "PC1", y = "PC2") +

  theme_classic()

ggsave("./procedure2_ComBat_DSS_ppca.pdf", height = 10, width = 10)
```

Heatmap visualization

Below is an example code to plot a combined heatmap for patient samples from different sources.

5. In case of obvious batch effects are visible based on the PCA or PPCA plot, correct the DSS
   data with ComBat (see **Procedure 2** above):

```
df.dss.2 <- as.data.frame(df.dss[df.dss$status != 'controls', ])

df.dss.2$datasource <- paste(df.dss.2$cohort, df.dss.2$plate,  sep = ' ')

r_ <- data.frame(datasource = as.factor(df.dss.2$datasource))

rownames(r_) <- rownames(df.dss.2)
```

6. Make a heatmap plot of combined DSS profiles as visualized in **Suppl. Fig. 2**:

```
p1 <- pheatmap(t(df.dss.2[, 1:(ncol(df.dss.2) - 4)]), annotation_col = r_,
show_colnames = F, show_rownames = F, clustering_distance_cols = "minkowski")

ggsave("./procedure2_DSS_heatmap.pdf", p1, height = 10,width = 10)
```

7. Make a heatmap of ComBat-corrected DSS profiles as visualized in **Suppl. Fig. 3**:

```
df.dss.3 <- as.data.frame(t(df.dss.combat)[df.dss$status != 'controls', ])

p2 <- pheatmap(rescale(t(df.dss.3), c(0, 50)), annotation_col = r_,
show_colnames = F, show_rownames = F, clustering_distance_cols = "minkowski")

ggsave("./procedure2_ComBat_DSS_heatmap.pdf", p2, height = 10,width = 10)
```

The below codes compute rDSS profiles from BeatAML and FIMM-AML studies and produce the
integrated heatmap and hierarchical clustering of rDSS profiles as visualized in **Suppl. Fig. 4.** The

user can also run the step 12 of **Procedure 1** to compute other selective drug responses for heatmap and clustering. The distance function for hierarchical clustering can be freely selected based on the user preferences.

8. Calculate ComBat-corrected rDSS profiles:

```
controls.dss <- t(df.dss.combat)[df.dss$status == 'controls', ]

controls.summary <-
as.data.frame(rbind(colMeans(as.matrix(controls.dss)),colSds(as.matrix(controls.
dss)),colMedians(as.matrix(controls.dss)), colMads(as.matrix(controls.dss))))

rownames(controls.summary ) <- c('mean', 'sd', 'median', 'mad')

patients.rdss <- (df.dss.3 - slice(controls.summary['median',
colnames(df.dss.3)],rep(1:n(), each = nrow(df.dss.3
))))/(slice(controls.summary['mad', colnames(df.dss.3 )],rep(1:n(), each =
nrow(df.dss.3 ))) + 1)
```

9. Make a heatmap of ComBat-corrected rDSS profiles:

```
p3 <- pheatmap(t(patients.rdss), annotation_col = r_, show_colnames = F,
show_rownames = F, clustering_distance_cols = "minkowski")

ggsave("./procedure2_ComBat_rDSS_heatmap.pdf", p3, height = 10,width = 10)
```

Troubleshooting

Troubleshooting advice for R-codes of 'Procedure 1' and 'Procedure 2' can be found in Table 2.

Table 2. Troubleshooting advice for R-codes of Procedure 1 and Procedure 2.

Step	Function	Problem	Possible cause	Solution
1 of Procedure 1	lapply(packages.required, library, character.only = T)	Error: package or namespace load failed for 'drc' in loadNamespace(i, c(lib.loc, .libPaths()), versionCheck = vI[[i]]): there is no package called 'mvtnorm'	The required package "mvtnorm" is not installed successfully	Install the package "mvtnorm" separately with the command : install.packages("mvtnorm", repos="http://R-Forge.R-project.org", dependencies = T)
7 of Procedure 1	CALC_METRICS	All DSS values equal to zero	The cell viability/growth inhibition data in the input file are percentages rather than decimals	Verify that the input data is correctly formatted by comparing it to the example data file provided

8 of Procedure 1	HEATMAP_SD	Error in HEATMAP_SD(patients.dss, proportion = 2) : The value of the argument proportion should be larger than 0 and no larger than 1	Non-numeric data is assigned to the argument proportion; The argument's value is less than or equal to 0 or larger than 1	Pass numeric values into the argument proportion; the value should be greater than 0 and not exceeding 1
14 of Procedure 1	CHEMO_TARGETED_PLOT	Error in CHEMO_TARGETED_PLOT(sample_dss, metric = "") : The argument metric should be one of 'DSS', 'sDSS', 'zDSS', or 'rDSS'	The input value is not a valid argument metric	Pass a string of either'DSS', 'sDSS', 'zDSS', or 'rDSS' into the argument metric
3 of Procedure 2	ComBat	Error in solve.default(crossprod(des), crossprod(des, y1)) : Lapack routine dgesv: system is exactly singular: U[1,1] = 0	Variable dat: in the input dataset, the DSS values for a drug are missing across all the samples in a batch	Filter out these drugs
		Unexpected missing values in output	Variable dat: in the input dataset, there is a single available DSS value for a drug across all the samples in a batch	Filter out these drugs or apply the mean.only=TRUE option accordingly
		Error in tcrossprod(x, y): requires numeric/complex matrix/vector arguments In addition: Warning message: in var(x) : NAs introduced by coercion	Variable dat: the input dataset contains non-numeric data	To confirm that the data is recognized as numeric by R and not as a factor or character, one can apply the *class* function to check the data type and the *as.numeric* function to convert to a numeric data type.

Timing

Initial testing of the R-codes in GitHub was performed on a desktop with Linux operating system, Intel Core i5-7300HQ CPU @ 2.50GHz machine, 4 processors and 8GB of memory. Code for the

example runs and timing calculator are available in the GitHub repository (https://github.com/yingjchen/DSS-v2.0; see **Table 3** for the most time-consuming functions).

Independent code testing by a naive user was performed on a Windows 10 Home 64-bit operating system, with an AMD Ryzen 7 4700U with Radeon GraphicsCPU @ 2.00GHz processor and 16GB memory, using a 64-bit R version 4.3.0. Downloading and installing the R and RStudio environment, downloading the files, setting up the working directory and installing the missing packages took 25min. Running the procedures 1 and 2 took 5min.

Table 3. Timing of the top-6 most time-consuming functions

Function	Description	Timing
CALC_METRICS(dose_responses, dose_responses_grouped, graph = FALSE)	Compute DSS_1, DSS_2, DSS_3, AUC and relative IC_{50} with cell viability data for the 3 example samples tested with up to 527 anticancer drugs	1.57min
CHEMO_TARGETED_PLOT	Visualize the distribution of DSS responses to 473 molecularly-targeted and 54 conventional chemotherapy (non-targeted) drugs in one sample	2.21s
HEATMAP_SD	Make a heatmap plot showing DSS_2 of the drugs with the highest standard deviations across the 3 example samples.	1.15s
ComBat	Adjust the batch effects in the DSS_2 profiles of 272 patient samples and 10 controls tested with up to 82 drugs	0.58s
SELECTIVE_SCORE_PLOT	Make a density plot for the response metric (i.e. DSS, sDSS, zDSS, and rDSS) of 527 drugs in one sample	0.46s

PCA_FUNC	Check for missing data and perform PCA/PPCA on the DSS_2 profiles of 272 patient samples and 10 controls tested with up to 82 drugs	0.18s

Anticipated results

We anticipate that selective scoring of *ex vivo* drug responses provides opportunities for many scientific and translational applications, ranging from (i) improved robustness and translatability of the preclinical ex vivo response profiles, to (ii) better understanding of the mechanisms of action of selective, non-toxic treatment responses, and (iii) improved identification of effective and safe, individualized treatment options for guiding clinical decision making. Recent reports have shown that *ex vivo* drug sensitivity profiles correlate with patient clinical responses and can hence guide personalized therapy selection in pediatric AML/ALL[35] and in relapsed/refractory (r/r) adult AML[36]. Similar findings have been made also in aggressive NHL[1], r/r CLL[14] and multiple myeloma[37], where functional testing of molecularly-targeted drugs has shown clinical benefits in terms of improved patient outcomes and patient-tailored treatment options. Our recent findings also demonstrated how DSS stratifies drug response profiles across ovarian cancer subtypes and optimizes patient-specific treatment regimens using *ex vivo* drug testing in solid tumors[25].

In this protocol, we demonstrated that the selective DSS calculation is widely-applicable to various drug screening studies, using *ex-vivo* drug testing data from ongoing functional precision medicine programmes carried out in three cancer centers in Finland, Norway and the US. However, high-throughput drug screens may result in technical variability, requiring stringent QC, which leads to missing data for those drugs that do not pass the QC criteria. Therefore, the downstream analysis methods need to be able to deal with missing data in the drug response matrices. The missing data becomes even more frequent, when analyzing data from longitudinal studies or from multiple studies that use different, partially overlapping compound collections. Missing values may also make batch effect correction more challenging, and we recommend the use of PPCA for visual analysis of batch effects in the presence of missing data (**Fig. 4**). Furthermore, since z-score may be sensitive to calculation of SD over control samples for zDSS, we recommend the use of rDSS to normalize patient responses after the ComBat batch effect correction.

To demonstrate how to combine *ex vivo* drug response data from multiple sources, we carried out an integrated analysis of patient samples from the FIMM-AML and BeatAML studies, using the FIMM healthy controls to normalize the patient responses in both cohorts. ComBat allows users to adjust for known batch covariates in large-scale datasets; here, the sample source (BeatAML patients vs. FIMM patients and controls) was used as the batch variable, and the non-parametric empirical Bayes version of ComBat was used for the batch effect adjustment[18]. ComBat outputs a new DSS matrix that has been corrected for the user-defined batch variables (see **Procedure 2**). It is generally recommended to make adjustments for all the known batch factors simultaneously (**Suppl. Fig. 1**), rather than correcting one batch effect at a time (**Fig. 4**). When we used rDSS to normalize patient responses against the healthy controls, the patients from BeatAML and FIMM-AML cohorts clustered together (**Fig. 6)**, and the two FIMM compound collections did not lead to separate larger clusters, even if the collection label was not used as a batch variable (**Suppl. Fig. 4**).

Besides using ComBat or other methods for batch effect correction, another option is to model and adjust for known batch covariates in a standard linear model, that is, one can use sample sources or compound libraries as confounding variables, and then subtract those out when performing differential sensitivity analyses and selective DSS calculations. However, in addition to differences in experimental assays that may lead to differences in the scales and missing value patterns in the drug response profiles, there may also exist additional experimental factors that may be more challenging to correct with the batch adjustment methods. For instance, the FIMM-AML and BeatAML studies often use different concentration ranges in their compound testing, which leads to differences in drug responses, even for the overlapping drugs. These factors may be difficult to correct using ComBat, which mainly removes effects that result in difference in the means and variances across the batches (i.e., making it easier to use the z-score in selective response calculation). Use of overlapping dose windows in the DSS calculation has been shown to further improve the correlation between studies[12].

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

Data availability
The original clinical summary and drug response data from the FIMM-AML and BeatAML cohorts are available at https://zenodo.org/record/7274740 and https://biodev.github.io/BeatAML2/. The drug response data file consisting of 10 healthy control responses and AML patient data for the Procedures 1 and 2 are available at: https://github.com/yingjchen/DSS-v2.0/tree/main/controls

Code availability

The R-codes can be used by anyone with basic skills in R and with basic knowledge of dose-response testing data. The codes come with readme files, user instructions, example data from FIMM-AML patients and healthy controls, along with the expected visualization outcomes (a zip file from GitHub, 2.5 MB), freely available at: https://github.com/yingjchen/DSS-v2.0

Competing interests

Author contributions

Y.C. Implemented the new drug sensitivity scores, analyzed the data, drafted the figures, wrote the protocols and troubleshooting advice. L.H. Designed the experiments, supervised the code writing, and tested the codes. A.I. Made the final figures, reviewed the text, and tested the codes. P.A.D. Screened the Oslo patient samples and wrote the Oslo materials section. S.P. Contributed to the drug response analysis codes and tested the Procedures. J.S. Supervised the drug response data analyses, reviewed the text, and completed Table 1 for the FIMM cohort. J.J.M. wrote the FIMM materials section and reviewed the text. S.K., S.M. M.M. Obtained and prepared the healthy donor samples for screening. C.A.H. Supervised the FIMM drug screening, reviewed and edited the text. J.M.E. Supervised the Oslo drug screening, reviewed and edited the text. K.W. Designed the drug sensitivity scores, reviewed and edited the text. T.A. Designed the drug sensitivity scores and analyses, supervised the work, wrote the first version of the manuscript.

References

1. Kornauth, C. et al. Functional Precision Medicine Provides Clinical Benefit in Advanced Aggressive Hematologic Cancers and Identifies Exceptional Responders. Cancer Discov. **12**, 372-387 (2022).
2. Malani, D. et al. Implementing a Functional Precision Medicine Tumor Board for Acute Myeloid Leukemia. Cancer Discov. **12**, 388-401 (2022).
3. Letai, A., Bhola, P. & Welm, A.L. Functional precision oncology: Testing tumors with drugs to identify vulnerabilities and novel combinations. Cancer Cell **40**, 26-35 (2022).
4. Tognon, C.E., Sears, R.C., Mills, G.B., Gray, J.W. & Tyner, J.W. Ex Vivo Analysis of Primary Tumor Specimens for Evaluation of Cancer Therapeutics. Annu. Rev. Cancer Biol. **5**, 39-57 (2021).
5. Flobak, Å., Skånland, S.S., Hovig, E., Taskén, K. & Russnes, H.G. Functional precision cancer medicine: drug sensitivity screening enabled by cell culture models. Trends Pharmacol. Sci. **43**, 973-985 (2022).
6. Pemovska, T. et al. Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. Nature **519**, 102-105 (2015).
7. Hatzis, C. et al. Enhancing reproducibility in cancer drug screening: how do we move forward? Cancer Res. **74**, 4016-4023 (2014).
8. Haibe-Kains, B. et al. Inconsistency in large pharmacogenomic studies. Nature **504**, 389-393 (2013).
9. Yadav, B. et al. Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. Sci Rep **4**, 5193 (2014).

10. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature **483**, 603-607 (2012).
11. Garnett, M.J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature **483**, 570-575 (2012).
12. Mpindi, J.P. et al. Consistency in drug response profiling. Nature **540**, E5-E6 (2016).
13. Pemovska, T. et al. Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. Cancer Discov. **3**, 1416-1429 (2013).
14. Yin, Y. et al. Functional Testing to Characterize and Stratify PI3K Inhibitor Responses in Chronic Lymphocytic Leukemia. Clin. Cancer Res. **28**, 4444-4455 (2022).
15. Andersen, A.N. et al. Clinical Forecasting using Ex Vivo Drug Sensitivity Profiling of Acute Myeloid Leukemia. Preprint at https://www.biorxiv.org/content/10.1101/2022.10.11.509866v2 (2023).
16. Bottomly, D. et al. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. Cancer Cell **40**, 850-864 e859 (2022).
17. Potdar, S. et al. Breeze 2.0: an interactive web-tool for visual analysis and comparison of drug response data. Nucleic Acids Research (in press). doi: 10.1093/nar/gkad390 (2023).
18. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics **28**, 882-883 (2012).
19. Tipping, M.E. & Bishop, C.M. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society. Series B (Statistical Methodology) **61**, 611-622 (1999).
20. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science **2**, 559-572 (1901).
21. Yamada S, Hongo T, Okada S, Watanabe C, Fujii Y & Ohzeki T. Clinical relevance of in vitro chemoresistance in childhood acute myeloid leukemia. Leukemia **15**, 1892-7 (2001).
22. Volm, M. & Efferth, T. Prediction of Cancer Drug Resistance and Implications for Personalized Medicine. Front. Oncol. **5**, 282 (2015).
23. Gupta, A., Gautam, P., Wennerberg, K. & Aittokallio, T. A normalized drug response metric improves accuracy and consistency of anticancer drug sensitivity quantification in cell-based screening. Commun. Biol. **3**, 42 (2020).
24. Hafner, M., Niepel, M., Chung, M. & Sorger, P.K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods **13**, 521-527 (2016).
25. Murumagi, A. et al. Drug response profiles in patient-derived cancer cells across histological subtypes of ovarian cancer: real-time therapy tailoring for a patient with low-grade serous carcinoma. Br. J. Cancer **128**, 678-690 (2023).
26. Heinemann, T. et al. Deep Morphology Learning Enhances Ex Vivo Drug Profiling-Based Precision Medicine. Blood Cancer Discov. **3**, 502-515 (2022).
27. Kropivsek, K. et al. Ex vivo drug response heterogeneity reveals personalized therapeutic strategies for patients with multiple myeloma. Nat Cancer. **4**, 734-753 (2023).
28. Kuusanmäki, H. et al. Phenotype-based drug screening reveals association between venetoclax response and differentiation stage in acute myeloid leukemia. Haematologica **105**, 708-720 (2020).
29. Ianevski, A. et al. Patient-tailored design for selective co-inhibition of leukemic cell subpopulations. Sci. Adv. **7**, eabe4038 (2021).
30. Goh, J. et al. An ex vivo platform to guide drug combination treatment in relapsed/refractory lymphoma. Sci. Transl. Med. **14**, eabn7824 (2022).
31. He, L. et al. Patient-Customized Drug Combination Prediction and Testing for T-cell Prolymphocytic Leukemia Patients. Cancer Res. **78**, 2407-2418 (2018).
32. He, L. et al. Network-guided identification of cancer-selective combinatorial therapies in ovarian cancer. Brief. Bioinform. **22**, bbab272 (2021).
33. Hanes, R. et al. screenwerk: a modular tool for the design and analysis of drug combination screens. Bioinformatics **39** (2023).
34. Ritz, C., Baty, F., Streibig, J.C. & Gerhard, D. Dose-Response Analysis Using R. PLoS One 10, e0146021 (2015).
35. Lee, S.H.R. et al. Pharmacotypes across the genomic landscape of pediatric acute lymphoblastic leukemia and impact on treatment response. Nat. Med. **29**, 170-179 (2023).
36. Kuusanmäki, H. et al. Ex vivo venetoclax sensitivity testing predicts treatment response in acute myeloid leukemia. Haematologica 108, 1768-1781 (2023).
37. Majumder, MM. et al. Identification of precision treatment strategies for relapsed/refractory multiple myeloma by functional drug sensitivity testing. Oncotarget **8**, 56338-56350 (2017).

Figure legends

Fig. 1 | Drug sensitivity scoring pipeline demonstrated in the FIMM-AML patient cohort. *First row:* design and processing of dose-response measurements from high-throughput drug screening. (**a**) Quality control using assay-specific positive and negative control wells. (**b**) Dose-response curve fitting using the user-selected dose-response model (here, log-logistic function). (**c**) Calculation of selective DSS (sDSS) as the area between the drug-response curves in patient and control cells (see **Box 1**). *Second row:* the new healthy control samples enable a better estimation of compound-specific reference control distribution (see **Box 1**). (**d**) Example responses of compound GSK-2334470 in the 10 healthy controls and 46 patients with AML tested with the FO5A compound collection. (**e**) sDSS normalizes the patient response by the average value of the controls (i.e., location normalization). Compound-specific sensitivity cut-off can be defined based on extreme values of sDSS (e.g. 95% quantile). (**f**) z-scored sDSS (zDSS) normalizes the patient responses also by variability over the controls (i.e., scale normalization), and enables significance assessment of an observed zDSS based on normal distribution approximation (z-test). *Bottom row:* application of the location and scale normalizations improves comparison of drug responses both within a patient sample and across patients. (**g**) Waterfall plot of the top-10 most effective and resistant drugs in an individual AML patient enables identification of optimal treatment options for the patient. (**h**) Volcano plot shows the statistical significance of the drug responses in the patient cohort (n=46; Wilcoxon test with false discovery rate [FDR]), which enables statistical association of the functional drug sensitivities with genetic dependencies and other biomarkers.

Fig. 2 | Comparison of responses across drug classes and patients in the FIMM-AML cohort. (**a-c**) Distributions of DSS, sDSS and zDSS responses to 473 molecularly-targeted and 54 conventional chemotherapy (non-targeted) drugs in an example patient case (AML_013_01). The top-effective compounds with each metric are labeled in the boxplots. The response distribution of a randomly selected healthy control is shown for comparison. (**d-i**) Responses to navitoclax (BCL-2 and BCL-XL inhibitor, upper panels **d-f**) and venetoclax (BCL-2 antagonist, bottom panels **g-i**) across 46 patients with AML and 10 healthy controls. The dotted vertical lines indicate either the 95% quantile (DSS metric), or the significance cutoff based on a normal distribution approximation (zDSS/rDSS metrics). The red dot indicates an example patient case (AML_016_01), with observed DSS, zDSS and rDSS values marked in the boxplots. In the boxplots, the horizontal lines indicate the median, the boxes the interquartile range (IQR), and the whiskers are Q1 - 1.5*IQR and Q3 + 1.5*IQR, where Q1 and Q3 are the first and third quartile, respectively. *Note:* the differences between the x-axis scales are due to varying scales of the metrics.

Fig. 3 | Robustness of drug response profiles in Oslo and BeatAML patients. (**a**) Correlation of drug response profiles among 5 technical replicates of an Oslo AML patient sample tested with FO5A collection. *p<0.05, **p<0.01; Kruskal-Wallis test with Bonferroni-Dunn correction for multiple testing. Correlation of 527 drug responses between two replicates of the Oslo AML sample when using (**b**) sDSS metric and (**c**) rDSS metric. Four example drugs that show sDSS differences between replicates are highlighted. (**d**) Correlation of drug response profiles between two technical replicates among 42 BeatAML samples (highest number of 16 overlapping drugs with FIMM FO5A collection with 2 replicates). **p<0.01, ***p<0.001; repeated measures ANOVA with Bonferroni correction for multiple testing. (**e-i**) Dose-response curves of the four selected drugs in the two replicates of the AML sample tested in Oslo. Responses of the 10 FIMM healthy controls are shown for comparison. (**g**) Correlation of 16 rDSS responses between two BeatAML replicates in a selected sample. The 10 healthy controls profiled in Helsinki with FO5A collection were used for selective response scoring both for the Oslo and BeatAML samples to demonstrate a wide applicability of the control data. In the boxplots (panels a and d), horizontal lines mark the median, the boxes the interquartile range

(IQR), and the whiskers are Q1 - 1.5*IQR and Q3 + 1.5*IQR, where Q1 and Q3 are the first and third quartile, respectively.

Fig. 4 | Batch effect correction in BeatAML and FIMM-AML cohorts. Drug response profiles of samples mapped to a 2D plane spanned by the first two principal components (PC1 and PC2), with and without ComBat[18] batch effect correction (left and right panels, respectively). The coefficient of determination (R^2) shows the proportion of variance explained by each PC. The dashed line represents the decision boundary of a support vector machine (SVM) classifier using a linear kernel, optimized independently for each sample group to distinguish either between sample sources, patients and controls, or compound collections. (a) 82 overlapping drugs between BeatAML patient samples (n=147) and FIMM controls (n=10). (b) 201 overlapping drugs between FIMM-AML patient samples (n=79 for FO4B compound collection; n=46 for FO5A collection) and FIMM controls (n=4 for FO4B compound collection; n=10 for FO5A collection). (c) 431 overlapping drugs tested in FIMM controls (n=4 for FO4B compound collection; n=10 for FO5A collection). The drug response profiles in panel a contain missing values (7.37%), due to rigid QC process implemented for drug responses in the BeatAML cohort; these data were analyzed using probabilistic principal component analysis (PPCA)[19], whereas the complete data in the other panels were analyzed with the standard PCA[20].

Fig. 5 | The key steps of Procedures 1 and 2, and their input-output relationships. The images show example outcomes of the visualization steps in a dummy data; see the GitHub folder for the expected outcomes of the R codes in the real-world data (https://github.com/yingjchen/DSS-v2.0/tree/main/expected_outcomes).

Fig. 6 | Integrated heatmap of selective drug responses in BeatAML and FIMM-AML cohorts. Robust rDSS response profiles of 147 BeatAML samples (n = 147) and 125 FIMM-AML samples (n=79 for FO4B compound collection; n=46 for FO5A collection), where the rDSS normalization was done against the FIMM healthy control samples (n = 10, FO5A compound collection). The columns correspond to the patient samples and the rows to the 82 overlapping drugs between the BeatAML and FIMM compound collections; grey color indicates missing values. Batch effects from the sample sources were adjusted with ComBat. ComBat uses either parametric or non-parametric empirical Bayes frameworks for adjusting data for batch effects. We recommend the use of the nonparametric version for DSS data, where no prior assumptions about DSS data distributions are required. Example visualizations from Procedure 2 before and after ComBat correction using PPCA plots and heatmap clusterings are shown in **Suppl. Figs. 1-4**.