

# Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies

Nikita Pozdeyev<sup>1</sup>, Minjae Yoo<sup>1</sup>, Ryan Mackie<sup>1</sup>, Rebecca E. Schweppe<sup>1</sup>, Aik Choon Tan<sup>1,\*</sup>, Bryan R. Haugen<sup>1,\*</sup>

<sup>1</sup>Department of Medicine, University of Colorado Cancer Center, University of Colorado School of Medicine, Aurora, CO, USA

\*Two senior authors contributed equally to this work

Correspondence to: Nikita Pozdeyev, email: nikita.pozdeyev@ucdenver.edu

Keywords: database integration, drug sensitivity, pharmacogenomics, cancer, cell line

Received: April 08, 2016

Accepted: May 29, 2016

Published: June 14, 2016

## ABSTRACT

The consistency of *in vitro* drug sensitivity data is of key importance for cancer pharmacogenomics. Previous attempts to correlate drug sensitivities from the large pharmacogenomics databases, such as the Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer (GDSC), have produced discordant results. We developed a new drug sensitivity metric, the area under the dose response curve adjusted for the range of tested drug concentrations, which allows integration of heterogeneous drug sensitivity data from the CCLE, the GDSC, and the Cancer Therapeutics Response Portal (CTRP). We show that there is moderate to good agreement of drug sensitivity data for many targeted therapies, particularly kinase inhibitors. The results of this largest cancer cell line drug sensitivity data analysis to date are accessible through the online portal, which serves as a platform for high power pharmacogenomics analysis.

## INTRODUCTION

Systematic pharmacogenomics studies of large panels of cancer cell lines have proven to be a useful pre-clinical resource in identifying clinically relevant drugs and putative predictive biomarkers in cancer research. For example, the Cancer Cell Line Encyclopedia (CCLE) [1], the Genomics of Drug Sensitivity in Cancer (GDSC) [2] and the Cancer Therapeutics Response Portal (CTRP) [3] projects have systematically collected drug sensitivity profiles and genomic information across hundreds of compounds and cancer cell lines. These studies have recapitulated known drugs and cancer-dependency relationships as well as revealed novel biomarkers associated with the drug sensitivity [1–4]. As large-scale pharmacogenomics studies are becoming more available, integrating these various data sets will provide a rich resource for the robust novel hypothesis generation and discovery in cancer therapeutics development. However, integrating these diverse pharmacogenomics studies poses a methodological and analytical challenge due to the differences in the types of assays, maximum tested drug concentration, the range of tested drug concentrations and

drug sensitivity metrics employed by different studies. These challenges were highlighted, when Haibe-Kains et al. found inconsistencies in drug sensitivity data from CCLE and GDSC [5]. In response, the authors of CCLE and GDSC re-analyzed drug sensitivity data accounting for the differences in the analytical approach employed by the two studies and reported a significantly better agreement [6].

In this study, we developed a new metric, area under the dose-response curve adjusted for the range of tested concentrations (adjusted AUC), to combine heterogeneous drug sensitivity data generated by multiple pharmacogenomics studies. To illustrate the utility of this new metric, we performed an unbiased comparison of the drug screening data from CCLE, GDSC and CTRP, resolved analytical challenges of the previous comparative analyses, and defined the variables associated with the better or worse agreement of the drug sensitivities between studies. We also developed a companion online portal, the Quantitative Analysis of Pharmacogenomics in Cancer (QAPC, <http://tanlab.ucdenver.edu/QAPC>), to explore and download the summarized data for high-power pharmacogenomic analysis by the scientific community.

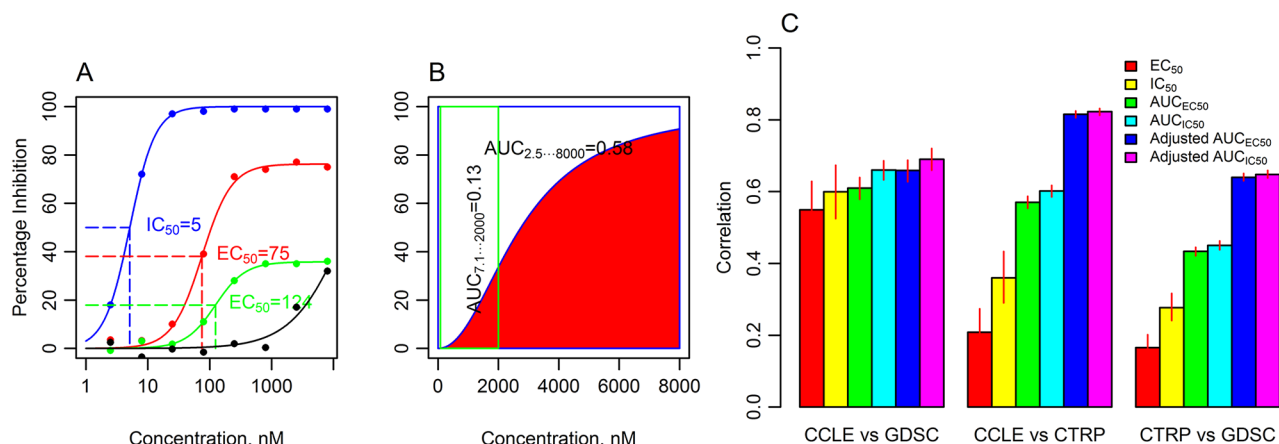
## RESULTS AND DISCUSSION

The CCLE and GDSC projects used half maximal inhibitory concentration ( $IC_{50}$ ), defined as a drug concentration producing absolute 50% inhibition of growth in the proliferation assay (Figure 1A, blue). By definition, this metric relies on the assumption, that at a high concentration of the drug, 100% effect is achieved (all cells die in a proliferation assay). This is true for potent cytotoxic drugs such as doxorubicin and paclitaxel (see examples in the QAPC portal, <http://tanlab.ucdenver.edu/QAPC>). However, cancer pharmacogenomics increasingly focuses on the targeted therapies, and many of these drugs, such as MEK inhibitors, are cytostatic [7]. Dose response curves produced for cytostatic drugs, such as MEK inhibitor selumetinib, plateau at a percent inhibition (maximal drug effect,  $A_{max}$ ) of less than 100 % (Figure 1A, red). 4-point logistic regression can estimate lower and upper asymptotes of the sigmoid dose-response curve that differ from 0 and 100 % of an  $IC_{50}$  curve and calculate half maximal effective concentration ( $EC_{50}$ , also called relative  $IC_{50}$ ), which is defined as a concentration of a drug causing an effect equal to the 50% of the  $A_{max}$ . More flexible,  $EC_{50}$  curves model experimental data better (median residuals standard error is 7.0 and 9.2 for  $EC_{50}$  and  $IC_{50}$  curve fits, respectively, Mann-Whitney-Wilcoxon,  $p < 0.001$ ,  $n = 478086$ ) and may provide more accurate measure of the drug sensitivity for cytostatic drugs. However,  $EC_{50}$  is ambiguous for the low amplitude curves (Figure 1A, green). Such a curve may represent a true low amplitude drug response to a cytostatic drug, a beginning

part of the high amplitude dose response curve with an  $EC_{50}$  exceeding maximal tested concentration, or a signal drift due to the technical imperfections of the large scale screening.

Incomplete dose response curves (Figure 1A, black) present a major challenge in interpreting drug sensitivity data. Neither  $IC_{50}$  nor  $EC_{50}$  can be accurately estimated from these drug responses. Other proposed drug sensitivity metrics, such as Hill slope and  $A_{max}$  [8] cannot be used as well. The three pharmacogenomic projects handled this problem differently. The CCLE assigned an  $IC_{50}$  of 8  $\mu$ M (maximal tested concentration) to incomplete dose response curves. This approach provides a finite number for the downstream analysis, but capped  $IC_{50}$  values are not always supported by the experimental data. The GDSC extrapolated data from incomplete curves outside of the range of tested concentrations that produced high and inaccurate  $IC_{50}$  values. The CTRP used AUC as a measure of drug sensitivity, but not  $IC_{50}$  or  $EC_{50}$ .

Haibe-Kains et al. compared CCLE and GDSC  $IC_{50}$  data as published without considering the differences in methodology [5]. This resulted in the expected poor agreement between the two datasets. To account for the analytical differences, Stransky et al. capped  $IC_{50}$  at a maximal drug concentration tested by the GDSC study (Supplementary Table S1) [6]. This produced a large number of capped  $IC_{50}$  values, which are not supported by the experimental data for the 15 drugs shared by the two databases (Supplementary Table S2), and overestimated the correlation between  $IC_{50}$  values. For



**Figure 1: The six drug sensitivity metrics and the agreement of pooled pharmacologic data between CCLE, GDSC and CTRP databases.** **A.** An  $IC_{50}$  dose-response curve has minimal ( $A_{min}$ ) and maximal ( $A_{max}$ ) asymptotes set to 0 and 100, respectively (blue line, CCLE, paclitaxel, cell line REH);  $EC_{50}$  is estimated from the dose response data with flexible  $A_{min}$  and  $A_{max}$  (red line, CCLE, selumetinib, cell line A375), but  $EC_{50}$  is ambiguous for low amplitude curves (green line, CCLE, selumetinib, cell line 639V, see detailed explanation in the text). The  $IC_{50}$  or  $EC_{50}$  cannot be estimated from incomplete dose response curve (black line, CCLE, selumetinib, cell line EFO27). **B.** AUCs calculated from the dose response curve (CCLE, crizotinib, cell line KMS26) using the range of drug concentrations from CCLE (2.5-8000 nM, blue box) and GDSC (7.125-2000 nM, green box) differ (0.58 and 0.13, respectively). Adjusted AUC is calculated for the range of concentrations shared by the databases (green box). **C.** Pearson correlation coefficients ( $r$ ) for the comparison of drug sensitivity data in 3 databases using six drug sensitivity metrics. Pearson correlation was calculated for a subset of  $EC_{50}$  and  $IC_{50}$  with finite values (estimated to be within the range of tested concentrations). Adjusted AUC provides the best consistency among databases. Error bars illustrate 95% confidence intervals (random permutations).

the drugs reported to have the greatest  $IC_{50}$  correlation between CCLE and GDSC, such as crizotinib, lapatinib, and nilotinib (Pearson correlation  $r = 0.74, 0.78, 0.89$ , respectively), up to 98% of  $IC_{50}$  values were capped (Supplementary Table S2)). Since the GDSC used lower maximal tested concentrations for 14 out of 15 drugs shared with CCLE (Supplementary Table S1), a portion of reliable CCLE  $IC_{50}$  estimates ( $IC_{50}$  values within the range of tested drug concentrations) were capped, eliminating potentially useful drug sensitivity information. For the drugs with the least number of capped  $IC_{50}$  values (17-AAG and paclitaxel, 19 % and 29 % capped  $IC_{50}$  values, respectively), no improvement in the correlation was found by Stransky et al. (0.57 and 0.15 for 17-AAG and paclitaxel, respectively) when compared to Haibe-Kains et al. analysis (0.61 and 0.16 for 17-AAG and paclitaxel, respectively). In addition, four drugs analyzed by both studies have more than one maximal tested concentration in the GDSC dataset, adding ambiguity to this artificial  $IC_{50}$  cap (Supplementary Table S1). The above considerations suggest that  $IC_{50}$  calculations (and  $EC_{50}$  calculations, following the same logic) are not suitable drug sensitivity metrics to accurately compare and reconcile drug sensitivities from the large pharmacogenomic studies, which contain many incomplete dose response curves.

The AUC, on the other hand, (Figure 1B, red area) is an attractive drug sensitivity metric, as it can be calculated for any dose response curve. AUC is unambiguous, combines information about the potency ( $EC_{50}, IC_{50}$ ) and efficacy ( $A_{max}$ ) of the drug into a single measure, and has been shown to be a robust metric for comparing a single drug across cell lines [8], as well as a better measure of cell line selectivity, when compared to  $IC_{50}$  [9]. However, AUC depends on the range of tested drug concentrations, which varies between studies. Figure 1B illustrates that AUC calculated from the same dose-response data for the range of concentrations used by GDSC (7.8-2000 nM) and CCLE (2.5-8000 nM) differs by more than 4-fold (0.13 and 0.58 respectively). The influence of the range of concentration on AUC estimate is particularly obvious, when maximal tested concentration differs significantly between databases, such as for crizotinib. For example, the median AUC calculated from the  $IC_{50}$  model of crizotinib responses in our analysis were 0.02, 0.14 and 0.84 for GDSC, CCLE and CTRP, respectively, reflecting the marked differences in maximal tested concentrations (2, 8 and 66  $\mu M$ , respectively), and making meaningful comparison of these data difficult (the distribution of AUC estimates for crizotinib and other drugs can be visualized with the QAPC portal, *Summary* tab). To solve this problem, we applied adjusted AUC, our new metric that takes into account the differences in the range of tested drug concentrations. The adjusted AUC uses sigmoid curve parameters estimated with a standard logistic regression ( $IC_{50}$  or  $EC_{50}$  models for adjusted  $AUC_{IC_{50}}$  and

adjusted  $AUC_{EC_{50}}$ , respectively), however, it is calculated only for the range of concentrations that is shared by the dose-response curves being compared (Figure 1B, green box). In contrast to  $IC_{50}$  capping, no data is discarded (non-overlapping high drug concentration data points are used for the sigmoid curve modeling).

To evaluate the performance of adjusted  $AUC_{IC_{50}}$  and adjusted  $AUC_{EC_{50}}$ , and compare it to the traditional drug sensitivity metrics ( $IC_{50}, EC_{50}$ , and unadjusted  $AUC_{IC_{50}}$  and  $AUC_{EC_{50}}$ ), we correlated drug sensitivity data obtained from CCLE, GDSC and CTRP (Figure 1C). There is a significant overlap in the drugs and cell lines analyzed by CCLE, GDSC and CTRP (Supplementary Figure S1, Supplementary Tables S3 and S4), permitting such analysis. Specifically, twelve drugs and 264 cell lines are represented in all 3 databases and pairwise intersection is even larger (Supplementary Figure S1).

First, we compared combined data for all compounds for the purpose of identifying the sensitivity metric that provides the best agreement between databases and, therefore, is the most reproducible quantitative assessment of the drug sensitivity for the pooled pharmacogenomic analysis.  $IC_{50}, EC_{50}$  and unadjusted AUC drug sensitivity metrics produced mild to moderate agreement between the pharmacogenomic databases (Figure 1C, Supplementary Table S5, Supplementary Figures S2-S4, QAPC portal). When adjusted for the range of drug concentrations tested, CCLE drug sensitivity data agreed very well with that of CTRP (Pearson correlation,  $r$ , for adjusted  $AUC_{IC_{50}} = 0.82$ ), and moderately well with the pharmacologic data from GDSC ( $r$  for adjusted  $AUC_{IC_{50}} = 0.69$ ). An improvement in the agreement between drug sensitivities measured with adjusted AUC is particularly noticeable when CTRP data is included, which is expected, because the CTRP tested highest maximal drug concentration among 3 studies (Supplementary Table S1) that skews the CTRP unadjusted AUC data towards the higher values. Both CCLE and CTRP projects were performed by the Broad Institute and used the same proliferation assay (CellTiterGlo), which likely contributed to the good reproducibility between these two studies. The correlation between CTRP and GDSC data was moderate ( $r$  for adjusted  $AUC_{IC_{50}} = 0.65$ ). Flexible curve modeling ( $EC_{50}, AUC_{EC_{50}}$ ) did not provide an additional advantage (Figure 1C).

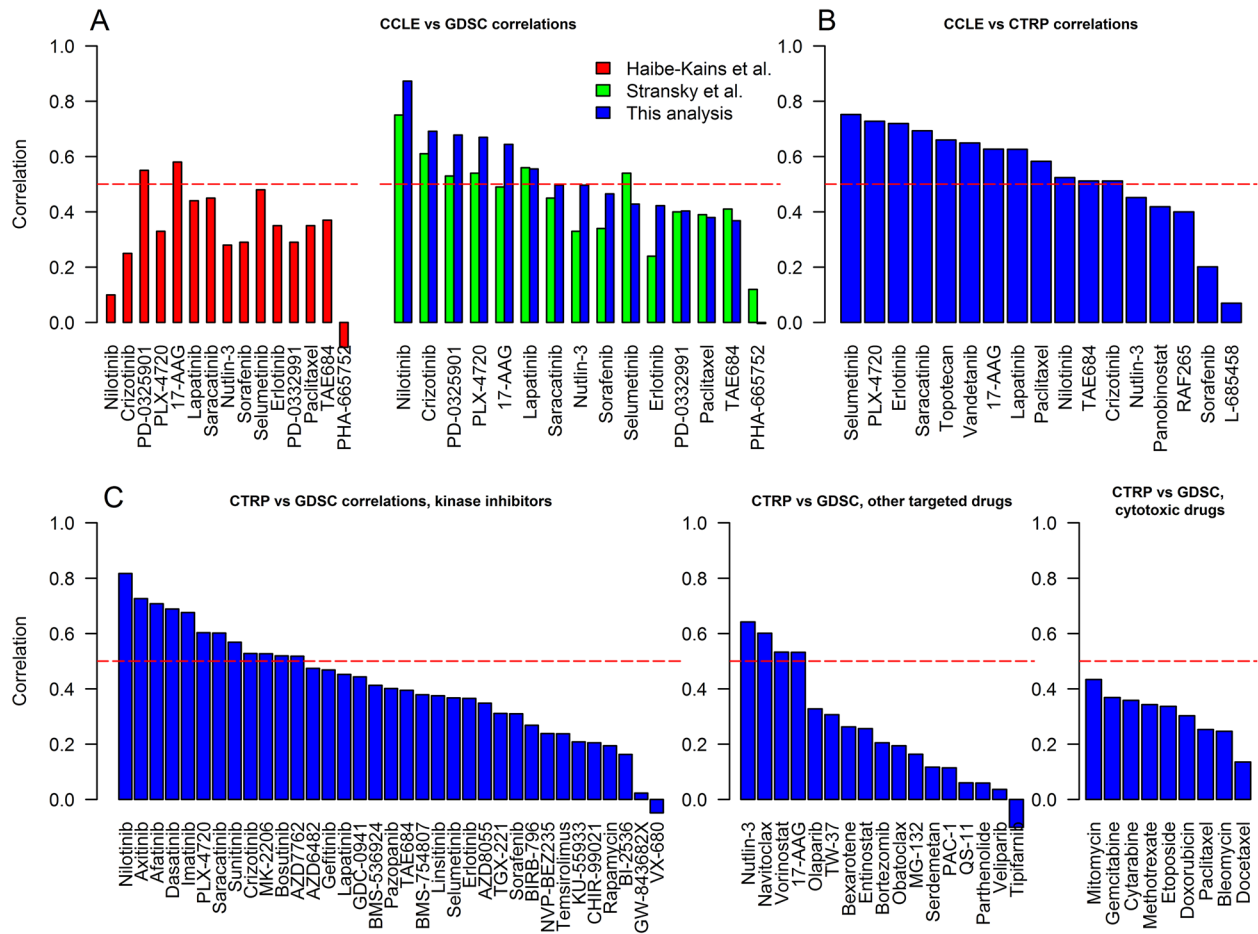
The main purpose of pharmacogenomics is to associate molecular features with the sensitivity to a particular drug, therefore we studied whether the correlations hold when data is stratified by an individual compound (Figure 2, Supplementary Table S5, QAPC Portal). For most drugs the highest correlation was achieved with the adjusted  $AUC_{IC_{50}}$  (Supplementary Table S5, QAPC Portal), supporting the superiority of our drug sensitivity metric. The  $r$  value for the adjusted  $AUC_{IC_{50}}$  was significantly higher than *any* other unadjusted drug sensitivity metric for lapatinib (Supplementary Figure S5)

and selumetinib in the CCLE-CTRP comparison, and for afatinib, axitinib, gefitinib, and nilotinib in the CTRP-GDSC comparison (higher  $r$  with non-overlapping 95% confidence intervals obtained with random permutations, Supplementary Table S5). Because adjusted  $AUC_{IC50}$  outperformed other drug sensitivity metrics, we chose it to assess the agreement between the databases and to compare our analysis with previously published analyses.

Adjusted  $AUC_{IC50}$  resulted in an improved correlation of the drug sensitivity data in CCLE and GDSC, when compared to the similar analyses (that used unadjusted AUC) performed by Haibe-Kains et al. [5] and Stransky et al. [6] (Kruskal-Wallis,  $p = 0.02$ ; paired Mann-Whitney-Wilcoxon (this analysis vs Stransky et al. analysis),  $p = 0.05$ ). The correlation has improved by  $> 0.1$  for 6 out of 15 drugs: nilotinib, PD-0325901, PLX-4720, nutlin-3, sorafenib, and erlotinib, (Figure 2A, Supplementary Table S5). As expected, CCLE and CTRP drug sensitivities agreed well (Figure 2B), with 12

out of 17 drugs demonstrating at least a moderate degree of correlation ( $r > 0.5$ ). Sixty one drugs were analyzed by both GDSC and CTRP allowing us to subdivide compounds based on the primary mechanism of action. The drug responses of kinase inhibitors showed the best agreement with 12 out 35 (34%) drugs with  $r > 0.5$ . Only 4 out of 17 (24%) other targeted therapies demonstrated moderate correlation. Surprisingly, drug sensitivity data for cytotoxic drugs showed minimal or poor correlation. ( $r < 0.5$ ).

Close examination of the scatterplots and raw dose response curves for the targeted therapies with the help of the QAPC portal demonstrated that most drugs with very low  $r$  show no or low activity in the proliferation assay, explaining the lack of correlation. The variability of drug responses for inactive compounds is likely due to the intrinsic noise of the high-throughput screening and has no biologic meaning. This is true for PHA-665752 (Supplementary Figure S6) and sorafenib (CCLE and



**Figure 2: The correlations of cell line responses to individual drugs from CCLE, GDSC and CTRP.** A. CCLE and GDSC correlations. Pearson correlation ( $r$ ) coefficients of adjusted  $AUC_{IC50}$  calculated by this study (blue bars) were matched to the similar analyses performed by Haibe-Kains et al. [5] (red bars, unadjusted AUC, Spearman correlation ( $\rho$ )), and Stransky et al. [6] (green bars, unadjusted AUC,  $r$ ). B. CCLE and CTRP correlations ( $r$ ; adjusted  $AUC_{IC50}$ ). C. CTRP and GDSC correlations (drugs are subdivided into kinase inhibitors, other targeted therapies, and cytotoxic chemotherapies;  $r$ ; adjusted  $AUC_{IC50}$ ). Horizontal dashed line shows a threshold for moderate correlation ( $> 0.5$ ).

GDSC comparison), L-685458 and sorafenib (CCLE and CTRP comparison), bexarotene, BIRB-796, CHIR-99021, erlotinib, KU-55933, olaparib, PAC-1, parthenolide, QS-11, serdemetan, sorafenib, TGX-221, veliparib, and VX-680 (CTRP and GDSC comparison). The scatterplots for the individual drugs are not included in this paper due to the space limitations but can be examined using the QAPC Portal (*Correlation Plots* tab) or downloaded together with the with R bioinformatics pipeline (“fig” folder, Supplementary Figures 2-4; [http://tanlab.ucdenver.edu/QAPC/Downloads/Drug\\_Sensitivity\\_Metrics.zip](http://tanlab.ucdenver.edu/QAPC/Downloads/Drug_Sensitivity_Metrics.zip)).

To improve the accessibility of the drug sensitivity data, we re-calculated 6 drug sensitivity metrics for all drugs and cell lines in CCLE, GDSC and CTRP (including adjusted AUC for the drugs analyzed by more than one database) and have made this data available for downloading through the QAPC portal. Alternatively, AUC can be calculated for the specified range of drug concentrations using functions from the recently published PharmacGx package in R [10] (at this time, the PharmacGx analysis is limited to the data from CCLE and GDSC databases) or by using the code from our downloadable bioinformatics pipeline.

In conclusion, we developed a new drug sensitivity metric, AUC adjusted for the range of tested concentrations, which allows reconciliation of publicly available heterogeneous pharmacologic data enabling pooled analysis for drug/gene associations. Using this methodology we have achieved the best correlation to date between overlapping data in CCLE, GDSC and CTRP databases, and we demonstrated that there is a moderate-to-good agreement between the drug sensitivity data for many targeted therapies, particularly for kinase inhibitors with activity in an *in vitro* proliferation assay. We predict that the adjusted AUC can be used to match *in vitro* drug responses generated by the high-throughput testing of patient-derived primary tumor cells with an existing pool of data to rapidly identify outlier drug sensitivities with potential therapeutic implications.

While our method accounts for several differences in the data analysis and experimental design, we acknowledge, that the variability introduced by the intrinsic features of each study (proliferation assay, cell line passage, growth media, plating density and others) outlined in the supplementary data for Haibe-Kains et al. [5] is unlikely to be alleviated with bioinformatics analysis alone.

## MATERIALS AND METHODS

### Data

Raw dose-response data was downloaded from CCLE portal (<http://www.broadinstitute.org/ccle>) [1], GDSC website (<http://www.cancerrxgene.org/downloads/>) [2], and The National Cancer Institute’s CTD2 Network (CTRP2, <https://ctd2.nci.nih.gov/dataPortal/>) [3].

## Calculations of drug sensitivity metrics

Log-logistic regression was used to model sigmoid dose response curve using the following formula:

$$y = A_{\max} + \frac{A_{\min} - A_{\max}}{1 + \left(\frac{x}{EC_{50}}\right)^{Hill}} \quad (1)$$

where  $A_{\min}$  and  $A_{\max}$  are lower and upper asymptotes of the sigmoid curve (no response and the maximal response to the drug, respectively),  $EC_{50}$  is drug concentration causing the effect equal to the 50% of the  $A_{\max}$ , and *Hill* is a Hill slope of the dose response curve. R package *drc* [11], [www.bioassay.dk](http://www.bioassay.dk), was used for modeling. 4-point regression analysis was performed first, and minimal and maximal asymptotes were allowed within the following ranges:  $R_{\min} \leq A_{\min} \leq 0$  and  $\min(0, R_{\min}) \leq A_{\max} \leq \max(100, R_{\max})$ , respectively.  $R_{\min}$  and  $R_{\max}$  are minimal and maximal measured drug responses, respectively. The coordinate of the upper bend point of the dose response curve was calculated using the formula [12]:

$$x_{bend} = \left(\frac{A_{\min} - A_{\max}}{1 + 4.6805}\right) + A_{\max} \quad (2)$$

If two or more data points were present with the drug concentration  $x > x_{bend}$ , the curve was assumed to be complete and the upper asymptote estimated accurately [13]. Otherwise, regression analysis was repeated with the  $A_{\max} = 100$  (for the lack of a better upper asymptote estimate for incomplete curves).  $A_{\max} = 100$  was also set for very low amplitude curves ( $A_{\max} - A_{\min} < 30$ ) to avoid the noise being misinterpreted as a true drug response. To estimate  $IC_{50}$ ,  $A_{\min}$  and  $A_{\max}$  were set to 0 and 100, respectively (by the definition of  $IC_{50}$ ). The parameters of the sigmoid curve were used for AUC calculations.

AUC was estimated in between minimal ( $x_{\min}$ ) and maximal ( $x_{\max}$ ) tested concentrations (nM) for the curve. For the purpose of comparing two dose response curves, an adjusted AUC was calculated for the range of concentrations shared by both curves (Figure 2B, green box). AUC was calculated using the following formula:

$$AUC = \frac{\sum_{x=x_{\min}}^{x_{\max}} (f(x) - A_{\min})}{\sum_{x=x_{\min}}^{x_{\max}} (\max(A_{\max}, 100) - A_{\min})} \quad (3)$$

where  $f(x)$  = formula (1)

To perform an unbiased comparison of  $EC_{50}$  and  $IC_{50}$ , the estimated values that exceed maximal tested concentration were set to infinity acknowledging the fact, that  $EC_{50}$  and  $IC_{50}$  cannot be accurately estimated from incomplete dose response curves.

## Comparisons of drug sensitivity metrics

Pearson and Spearman correlations were used to compare drug sensitivity data between studies. Pearson correlation for pooled  $EC_{50}$  and  $IC_{50}$  data was calculated for a subset of finite values. p-values were computed by performing random permutations and bootstrapping using published algorithm [6]. P-values and confidence intervals calculated using random permutations are cited in the text.

## Implementation and source code

All data manipulations were performed in R (<https://www.r-project.org/>). To ensure reproducibility, all figures and tables of this manuscript were generated programmatically from the raw data. R scripts and data can be downloaded from: [http://tanlab.ucdenver.edu/QAPC/Downloads/Drug\\_Sensitivity\\_Metrics.zip](http://tanlab.ucdenver.edu/QAPC/Downloads/Drug_Sensitivity_Metrics.zip)

## QAPC portal

All aspects of this analysis can be explored interactively using companion portal the Quantitative Analysis of Pharmacogenomics in Cancer (QAPC, <http://tanlab.ucdenver.edu/QAPC/>).

QAPC features include:

1. Interactive graphic interface to access drug sensitivity data for 585 drugs/drug combinations and 1201 cell lines.

2. Choice of 6 drug sensitivity metrics:

$EC_{50}$  – half maximal effective concentration

$IC_{50}$  – half maximal inhibitory concentration

AUC  $EC_{50}$  – area under the dose response curve calculated from  $EC_{50}$  model

AUC  $IC_{50}$  – area under the dose response curve calculated from  $IC_{50}$  model

Adjusted AUC  $EC_{50}$  – area under the dose response curve calculated from  $EC_{50}$  model adjusted for the range of tested drug concentrations

Adjusted AUC  $IC_{50}$  – area under the dose response curve calculated from  $IC_{50}$  model adjusted for the range of tested drug concentrations

3. Option to reconcile data from up to three databases using adjusted AUC  $EC_{50}$  or adjusted AUC  $IC_{50}$ . Adjusted AUC for the cell line/drug pairs analyzed by more than one study are averaged. Adjusted AUC are calculated even for the cell lines analyzed by one database only, if the drug is represented in more than one database, thus allowing a fair comparison of drug sensitivities across studies and enabling high-power pharmacogenomics analysis.

4. Functionality to visualize dose-response curves for each drug/cell line combination to evaluate the quality of the raw data and the accuracy of the dose response model.

5. Graphic representation of the drug sensitivity data agreement between CCLE, GDSC and CTRP databases with 2D and 3D scatterplots.

6. Functionality to download raw data,  $EC_{50}$  and  $IC_{50}$  regression model parameters, adjusted AUC calculations, and correlation statistics for the drug.

Detailed description of the QAPC controls and interface can be accessed through the *Help* tab of the online portal.

## CONFLICTS OF INTEREST

None of the authors report conflicts of interest.

## GRANT SUPPORT

This work was supported by the Paul R. Ohara II Seed Grant Program and University of Colorado Cancer Center ACS-IRG Grant to NP.

## REFERENCES

1. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012; 483: 603–607.
2. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*. 2012; 483: 570–575.
3. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price E V, Coletti ME, Jones V, Bodycombe NE, Soule CK, Gould J, Alexander B, Li A, Montgomery P et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov*. 2015; 5: 1210–1223.
4. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, Ebright RY, Stewart ML, Ito D, Wang S, Bracha AL, Liefeld T, Wawer M et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*. 2013; 154: 1151–1161.
5. Haibe-Kains B, El-Hachem N, Birbak NJ, Jin AC, Beck AH, Aerts HJ, Quackenbush J. Inconsistency in large pharmacogenomic studies. *Nature*. 2013; 504: 389–393.
6. Stransky N, Ghandi M, Kryukov GV, Garraway LA, Lehár J, Liu M, Sonkin D, Kauffmann A, Venkatesan K, Edelman EJ, Riester M, Barretina J, Caponigro G et al. Pharmacogenomic agreement between two cancer cell line data sets. *Nature*. 2015; 528: 84–87.

7. Corcoran RB, Cheng KA, Hata AN, Faber AC, Ebi H, Coffee EM, Greninger P, Brown RD, Godfrey JT, Cohoon TJ, Song Y, Lifshits E, Hung KE et al. Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models. *Cancer Cell*. 2013; 23: 121–128.
8. Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW, Sorger PK. Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat Chem Biol*. 2013; 9: 708–714.
9. Shimada K, Hayano M, Pagano NC, Stockwell BR. Cell-Line Selectivity Improves the Predictive Power of Pharmacogenomic Analyses and Helps Identify NADPH as Biomarker for Ferroptosis Sensitivity. *Cell Chem Biol*. 2016; 23: 225–235.
10. Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, Freeman M, Selby H, Gendoo DM, Grossmann P, Beck AH, Aerts HJ, Lupien M et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics*. 2016; 32: 1244–1246.
11. Ritz C, Streibig JC. Bioassay Analysis using R. *J. Stat. Softw*. 2005; 12: 1–21.
12. Sebaugh JL, McCray PD. Defining the linear portion of a sigmoidshaped curve: bend points. *Pharm. Stat*. 2003; 2: 167–174.
13. Sebaugh JL. Guidelines for accurate EC50/IC50 estimation. *Pharm. Stat*. 2011; 10: 128–134.