# Consistency in drug response profiling

ARISING FROM B. Haibe-Kains *et al. Nature* **504,** 389–393 (2013); doi:10.1038/nature12831

The comparative analysis by Haibe-Kains *et al.*[1] concluded that data from two large-scale studies of cancer cell lines[2,3] showed highly discordant results for drug sensitivity measurements, whereas gene expression data were reasonably concordant. Here, we cross-compared the two original datasets[2,3] against our own data of drug response profiles in overlapping cancer cell line panels. Our results indicate that it is possible to achieve concordance between different laboratories for drug response measurements by paying attention to the harmonization of assays and experimental procedures. There is a Reply to this Comment by Safikhani, Z. *et al. Nature* **540,** http://dx.doi.org/10.1038/nature20172 (2016).

Haibe-Kains *et al.*[1] reported on a comparative evaluation of two drug sensitivity and molecular profiling datasets, one from the Cancer Genome Project (CGP)[2] and the other from the Cancer Cell Line Encyclopedia (CCLE)[3]. In their analyses, gene expression profiles between hundreds of common cancer cell lines across all genes showed high consistency between the two studies (median rank correlation (MRC) = 0.85), whereas the drug response data for 15 common compounds were highly discordant (MRC = 0.28 for half-maximum inhibitory concentration ($IC_{50}$) values). This report[1] and the accompanying commentary[4] suggested that differences in laboratory protocols, compounds and their tested concentration ranges, and computational methods may account for the differences, but these reports did not elaborate which of these factors are important and whether they can be controlled for.

Here, we reanalysed the dose–response data from both CGP and CCLE using a standardized area under the curve (AUC) response metric, which we call the drug sensitivity score (DSS)[5]. We then compared the CGP and CCLE data with a new dataset of drug responses profiled using the Institute for Molecular Medicine Finland (FIMM) compound testing assay[6], covering 308 drugs across 106 cancer cell lines. The FIMM data included 45 compounds in common with CGP and 14 with the CCLE in 50 cell lines (Supplementary Data 1). In the AUC calculation, we unified the drug concentration ranges across the CGP, CCLE and FIMM assays. We observed a significantly higher level of consistency ($P = 4.2 \times 10^{-5}$), especially between the CCLE and FIMM drug response data (MRC = 0.74), as compared to the consistency between FIMM and CGP data (MRC = 0.54) (Fig. 1a).

Similar experimental protocols were applied at FIMM and CCLE, including the same readout (CellTiter-Glo, Promega), similar controls (vehicle as negative control and positive controls of toxic compounds 100 μM benzethonium chloride or 1 μM MG132). However, there were also differences, such as the plate format used (1,536 versus 384 wells). Importantly, there was no effort made to standardize cell numbers used or any other parameters between the three laboratories, such as the source, passage number and media used for cells, nor the origin and handling of drugs. Therefore, this observed level of drug response agreement could be substantially improved by further standardization of the laboratory protocols. The CGP experimental protocol differed from the two others in terms of the readout (fluorescent nucleic acid stain Syto 60, Life Technologies), in the use of controls (drug-free cells as negative and no cells as positive controls), and the plate format used (96- or 384-well plates).

We compared the drug response profiles between the same cell lines from different laboratories, in line with the approach of Haibe-Kains *et al.*[1], in which they showed consistency in gene expression profiles from CGP and CCLE (MRC = 0.85)[1]. The Haibe-Kains *et al.*[1] approach, in which the correlation is calculated for each drug separately across the cell lines, showed more variability (Fig. 1b), owing to the fact that some drugs show minimal efficacy in all the tested cell lines. Analogously, gene expression correlations vary more widely when analysed at the level of genes across cell lines (MRC = 0.58 between CGP and CCLE), as certain genes are not expressed above technical noise. Although both ways to compare the data are relevant to the overall goal of personalized

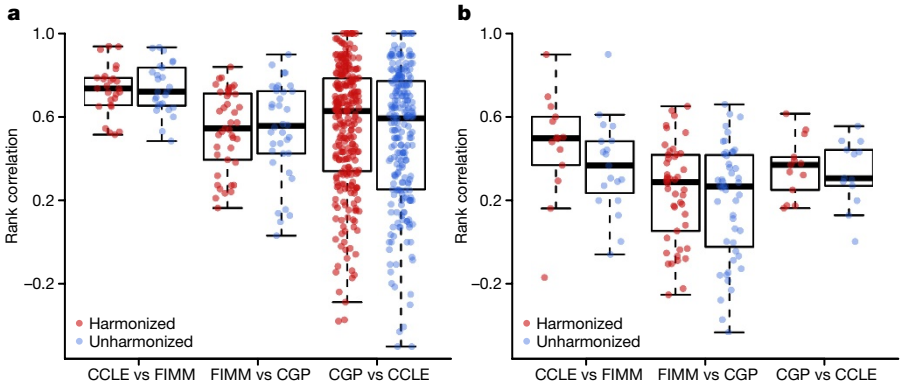


**Figure 1 | Consistency between drug response profiles across FIMM, CCLE and CGP. a**, In the 'between' cell line comparison, the Spearman rank correlation coefficient was calculated for the pairwise overlapping cell lines and over shared sets of compounds between FIMM–CCLE (26 cell lines, 14 compounds), FIMM–CGP (41 cell lines, 45 compounds), and CCLE–CGP (268 cell lines, 13 compounds). The drug sensitivity score (DSS) profiles between FIMM and CCLE showed improved correlation compared with the correlation between FIMM versus CGP or between CGP versus CCLE ($P = 4.2 \times 10^{-5}$ or $P = 0.0045$, respectively, two-sided Wilcoxon rank-sum test). **b**, In the 'across' cell line comparison[1], the correlation was calculated for each drug across all of the cell lines. This led to decreased consistency, yet still showed the same inter-laboratory differences after standardized data analysis ($P = 0.00034$, two-sided Wilcoxon rank-sum test, for comparisons both between FIMM versus CCLE and FIMM versus CGP or CGP versus CCLE correlations). The coloured points indicate whether the comparison made use of the harmonized data analysis (see Methods), or the AUC values provided in the original studies[1,2]. The box and horizontal bar represent the interquartile range and median of the correlation coefficients, respectively, and the whiskers denote the most extreme data points.

therapy, emphasized in the original publications[2,3], the same evaluation approach should ideally be used when comparing the consistency of gene expression and drug response measurements.

In summary, we show that standardization of assay methods and laboratory conditions will help to improve the inter-laboratory agreements in drug response profiling. Global standards, similar to the minimum information about a microarray experiment (MIAME) standard for the microarray data[7], should be developed.

## Methods

We scaled the drug response readouts using the available positive and negative controls from CGP, CCLE and FIMM screens (FIMM dose–response data points are provided in Supplementary Data 2). Curve fitting for the scaled dose–response curves was based on the original dose ranges, using the four-parameter logistic model and the Levenberg–Marquardt algorithm[8,9]. We used the median of the triplicate measurements from CCLE (http://www.broadinstitute.org/ccle/home), and the single dose–response measurements from CGP (provided by M. Garnett). In the harmonized data analysis, DSS was calculated based on the fitted dose–response models (R-package available at https://bitbucket.org/BhagwanYadav/drug-sensitivity-score-dss-calculation)[5]. The DSS integration was restricted to the common concentration window shared between CGP, CCLE and FIMM, which was limited by the narrower dose range of the CGP data. The compounds were identified and matched using InChIKeys[10]. Cell line matching was based on the named mapping file provided by Haibe-Kains *et al.*[1], followed by manual curation to guarantee that we compared exactly the same cell lines and chemical compounds.

This work was financially supported by Academy of Finland (grants 269862 and 279163, T.A.; 277293, K.W.; Center of Excellence for Translational Cancer Biology, O.K.), Cancer Society of Finland (K.W., O.K., T.A.), JSPS Bilateral programs (A.H.), Sigrid Juselius Foundation (A.H., K.W., O.K.), European Union's Seventh Framework Programme (FP7, grant 258068, EU-FP7-Systems Microscopy, O.K.), and Doctoral Programme in Biomedicine (J.P.M. and B.Y.).

**John Patrick Mpindi[1], Bhagwan Yadav[1], Päivi Östling[1,2], Prson Gautam[1], Disha Malani[1], Astrid Murumägi[1], Akira Hirasawa[1,3], Sara Kangaspeska[1], Krister Wennerberg[1], Olli Kallioniemi[1,2] & Tero Aittokallio[1,4]**

[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00290 Helsinki, Finland.
email: tero.aittokallio@fimm.fi
[2]Science for Life Laboratory, Department of Oncology and Pathology, Karolinska Institutet, 17165 Solna, Sweden.
[3]Department of Obstetrics & Gynecology, Keio University School of Medicine, 160-8582, Tokyo, Japan.
[4]Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland.

**Received 4 April 2014; accepted 12 October 2016.**

1. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504,** 389–393 (2013).
2. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483,** 570–575 (2012).
3. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
4. Weinstein, J. N. & Lorenzi, P. L. Cancer: Discrepancies in drug sensitivity. *Nature* **504,** 381–383 (2013).
5. Yadav, B. *et al.* Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci. Rep.* **4,** 5193 (2014).
6. Pemovska, T. *et al.* Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov.* **3,** 1416–1429 (2013).
7. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29,** 365–371 (2001).
8. Levenberg, K. A method for the solution of certain problems in least squares. *Q. Appl. Math.* **2,** 164–168 (1944).
9. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11,** 431–441 (1963).
10. Nicola, G., Liu, T. & Gilson, M. K. Public domain databases for medicinal chemistry. *J. Med. Chem.* **55,** 6987–7002 (2012).

**Supplementary Information** is available in the online version of the paper.

**Author Contributions** J.P.M. and B.Y. performed computational experiments and analysed data; these authors contributed equally to this work. P.G., D.M., A.M., A.H. and S.K. designed and performed drug response experiments. J.P.M., B.Y., P.Ö., O.K. and T.A. designed the study, interpreted data and wrote the manuscript. K.W. provided critical review and comments on the study and manuscript. O.K. and T.A. supervised the study.

**Competing Financial Interests** Declared none.

**doi:**10.1038/nature20171

# Safikhani *et al.* reply

**REPLYING TO** J. P. Mpindi *et al. Nature* **540,** http://dx.doi.org/10.1038/nature20171 (2016)

The accompanying Comment[1] by Mpindi *et al.* is an important contribution to the discussion of pharmacogenomic consistency for several reasons. Mpindi *et al.*[1] were able to reproduce our initial finding[2] of a substantial inconsistency between the pharmacological profiles generated within the Cancer Genome Project (CGP)[3] and the Cancer Cell Line Encyclopedia (CCLE)[4], and explored potential reasons behind the problem and possible solutions. To do this, they compared the CGP and CCLE to a new dataset generated by the Institute for Molecular Medicine Finland (FIMM) that includes 308 drugs that were tested across 106 cancer cell lines[5]. The authors shared the subset of the FIMM data overlapping with CGP and CCLE, including 52 drugs tested in up to 50 cell lines (drug dose–response curves and their comparison with CGP and CCLE curves are available in Supplementary Data). Overall, their comparative analysis[1] of this newly released dataset supports our published finding[2] of greater consistency between studies in which there is similarity in experimental methods. We agree with Mpindi *et al.*[1] that harmonizing the readout, drug concentration range, and statistical estimator makes it possible to achieve greater consistency across pharmacogenomic studies. Here we provide specific responses to the main results reported by Mpindi *et al.*[1]

The FIMM and CCLE studies used a similar experimental protocol that included the CellTiter-Glo pharmacological assay, as opposed to the Syto60 assay used in CGP. As pointed out by the authors[1], there were also parts of the experimental protocols that were different between all the three studies, effectively preventing perfect replication of *in vitro* molecular and pharmacological profiles. In our initial analysis[2], we showed that drug sensitivity measures for paclitaxel and lapatinib were more consistent between the GlaxoSmithKline (GSK) dataset and CCLE, than between GSK and CGP or CCLE and CGP. Given that both GSK and CCLE used the CellTiter-Glo assay, we concluded that the use of different pharmacological readouts has a substantial effect on the consistency of drug sensitivity measurements. The results from Mpindi *et al.*[1] further confirm this observation. However, a recent study from Genentech suggested that laboratory-specific effects might induce even greater biases than the use of different readouts[6]. Indeed, Haverty *et al.*[6] showed that, although their CellTiter-Glo screen was more concordant

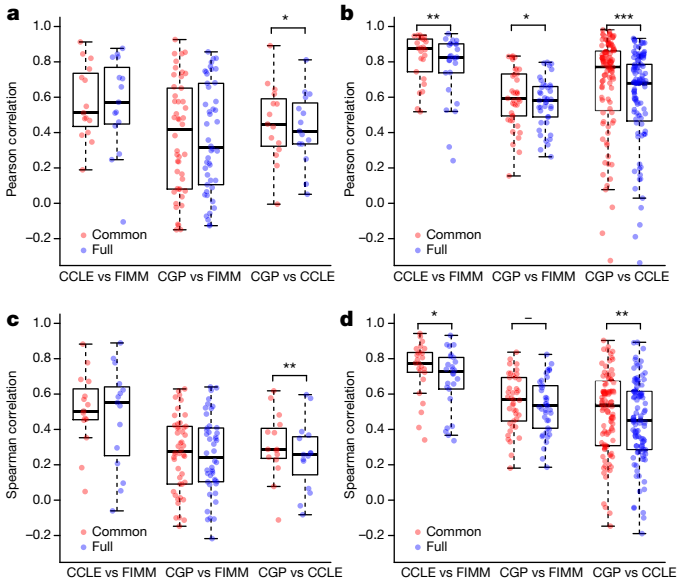


**Figure 1 | Consistency between drug sensitivity measures (AUC) across FIMM, CGP and CCLE. a–d**, AUC values were computed with the PharmacoGx package using the full concentration range (blue) or the concentration range shared between two studies (red). **a**, **b**, Boxplots of Pearson correlations between AUC values across (**a**) and between (**b**) cell lines. **c**, **d**, Boxplots of Spearman rank correlations between AUC values across (**c**) and between (**d**) cell lines. The box and horizontal bar represent the interquartile range and median of the correlation coefficients, respectively, and the whiskers denote the most extreme data points. Significance of the difference between AUC correlations computed using the full or common concentration range is calculated using a one-sided Wilcoxon signed-rank test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; '–' denotes $0.05 \leq P < 0.10$.

with CCLE, their new drug sensitivity data generated using the Syto60 assay were more consistent to their previous screen than CGP, despite the use of the same pharmacological assay[6]. It therefore remains unclear whether there are other experimental factors that drive the observed inconsistencies between large-scale pharmacogenomic studies and further argues for a detailed analysis of experimental protocols.

Similar to Pozdeyev *et al.*[7] and the Comment by Bouhaddou *et al.*[8], the authors investigated whether sensitivity metrics computed from the drug concentration range shared between studies yield higher consistency than the published metrics computed on the full (only partially overlapping) concentration range using different curve fitting algorithms. Concurring with previous results, Mpindi *et al.*[1] showed that the modified area under the curve (AUC) statistic (referred to as the drug sensitivity score (DSS)) computed on the shared concentration range (harmonized) was better correlated between studies than published AUC values (unharmonized). To test whether this marginal but significant improvement was due to the use of the same drug dose–response curve modelling or the choice of concentration range, we reproduced the authors' analysis using our PharmacoGx package[9] and used the same curve-fitting algorithm for the FIMM CGP and CCLE studies. We observed significantly higher correlations, across and between cell lines, for AUC values computed using a shared concentration range for the CGP and CCLE comparison ($P < 0.05$, Wilcoxon signed-rank test; Fig. 1). Although this observation held true for the correlations between cell lines for all comparisons, restricting AUC computation to the common concentration range did not yield significantly higher correlation across cell lines for CCLE versus FIMM and CGP versus. FIMM (Fig. 1). Our results confirm that restricting the analysis to the common concentration range improves consistency between CGP and CCLE, and to a lesser extent with the FIMM dataset.

In our original report[2], we computed correlation coefficients for each individual drug across cell lines, which was relevant to the overall goal of the CCLE and CGP studies to discover new genomic biomarkers of drug responses in order to increase the emergence of 'personalized' treatment regimens. Mpindi *et al.*[1] also computed the correlation of cell line sensitivity data across drugs (referred to as between cell lines in their presentation; see Supplementary Information). While the application of the across cell line correlation analysis is more relevant for identifying biomarkers predictive of response to individual drugs, we agree with Mpindi *et al.*[1] that correlations across and between cell lines should be compared in a consistent manner, as was done in our published in-depth reanalysis of CCLE and CGP[10]. Consistent with Mpindi *et al.*[1], our results clearly demonstrate that the overall correlation across cell lines is lower than the correlation between cell lines (Supplementary Fig. 1). However, gene expression data are significantly more concordant between studies than the drug response summary statistics (half-maximum inhibitory concentration ($IC_{50}$) and AUC) values in all comparisons ($P < 0.002$, Wilcoxon rank-sum test). Consequently, our observation that gene expression data are significantly more correlated than pharmacological response still holds.

As we argued previously[2], we agree with Mpindi *et al.*[1] that there is a need for harmonization of experimental protocols and cross-validation of large pharmacogenomic studies, and that doing so will improve robustness and reproducibility of the associated data.

Author A. C. Jin was a student in A.H.B.'s laboratory and left shortly after publication of the initial study, and did not participate in in the writing of this Reply. Authors Z.S., P.S. and M.F. developed the PharmacoGx software package, which enabled the analyses presented here; A.G. helped with the comparison of the different drug sensitivity metrics, and participated in the interpretation of the results and writing of this Reply.

## Methods

The methods are described in detail in our published reanalysis[10] and in the Supplementary Information. The code and associated files required to reproduce this analysis are publicly available on the cdrug-rebuttals GitHub repository (https://github.com/bhklab/cdrug-rebuttals). The procedure to set up the software environment and run our analysis pipeline is provided in the Supplementary Information. This work complies with the guidelines proposed previously[11] in terms of code availability and replicability of results.

**Zhaleh Safikhani[1,2], Nehme El-Hachem[3], Petr Smirnov[1], Mark Freeman[1], Anna Goldenberg[4,5], Nicolai J. Birkbak[6,7], Andrew H. Beck[8,9], Hugo J. W. L. Aerts[9,10,11], John Quackenbush[10,12] & Benjamin Haibe-Kains[1,2,5,13]**

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada.
email: bhaibeka@uhnresearch.ca
[2]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada.
[3]Institut de recherches cliniques de Montréal, Montreal, Quebec H2W 1R7, Canada.
[4]Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada.
[5]Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada.
[6]The Francis Crick Institute, London NW1 1AT, UK.
[7]University College London Cancer Institute, London WC1E 6BT, UK.
[8]Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA.
[9]Harvard Medical School, Boston, Massachusetts 02115, USA.
[10]Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

[11]Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.
[12]Harvard TH Chan School of Public Health, Boston, Massachusetts 02115, USA.
[13]Ontario Institute of Cancer Research, Toronto, Ontario M5G 1L7, Canada.

1. Mpindi, J. P. *et al.* Consistency in drug response profiling. *Nature* **540,** http://dx.doi.org/10.1038/nature20171 (2016).
2. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504,** 389–393 (2013).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483,** 570–575 (2012).
4. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
5. Pemovska, T. *et al.* Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov.* **3,** 1416–1429 (2013).
6. Haverty, P. M. *et al.* Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533,** 333–337 (2016).
7. Pozdeyev, N. *et al.* Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget* http://dx.doi.org/10.18632/oncotarget.10010 (2016).
8. Bouhaddou, M. *et al.* Drug response consistency in CCLE and CGP. *Nature* **540,** http://dx.doi.org/10.1038/nature20580 (2016).
9. Smirnov, P. *et al.* PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics* **32,** 1244–1246 (2016).
10. Safikhani, Z. *et al.* Revisiting inconsistency in large pharmacogenomic studies. *F1000Research* http://dx.doi.org/10.12688/f1000research.9611.1 (2016).
11. Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* **9,** e1003285 (2013).

**Supplementary Information** is available in the online version of the paper.

**doi:**10.1038/nature20172