Check for updates

RESEARCH NOTE

Assessment of pharmacogenomic agreement [version 1; peer review: 3 approved]

Zhaleh Safikhani[1,2], Nehme El-Hachem[3], Rene Quevedo[1,2], Petr Smirnov[1], Anna Goldenberg[4,5], Nicolai Juul Birkbak[6], Christopher Mason[7-9], Christos Hatzis [ID][10,11], Leming Shi[12,13], Hugo JWL Aerts[14,15], John Quackenbush[14,16], Benjamin Haibe-Kains [ID][1,2,5]
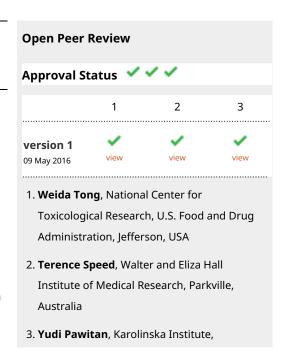
[1]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, M5G 1L7, Canada
[2]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 1L7, Canada
[3]Institut de recherches cliniques de Montréal, Montreal, Quebec, H2W 1R7, Canada
[4]Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada
[5]Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 2E4, Canada
[6]University College London, London, WC1E 6BT, UK
[7]Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY, 10065, USA
[8]The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, NY, 10021, USA
[9]The Feil Family Brain and Mind Research Institute (BMRI), New York, NY, 10065, USA
[10]Section of Medical Oncology, Yale School of Medicine, New Haven, CT, 06520, USA
[11]Yale Cancer Center, Yale University, New Haven, CT, 06510, USA
[12]Fudan University, Shanghai City, 200135, China
[13]University of Arkansas for Medical Sciences, Little Rock, AR, 72205, USA
[14]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, 02215, USA
[15]Department of Radiation Oncology and Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02215, USA
[16]Department of Biostatistics and Computational Biology and Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA, 02215, USA

Abstract

In 2013 we published an analysis demonstrating that drug response data and gene-drug associations reported in two independent large-scale pharmacogenomic screens, Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer Cell Line Encyclopedia (CCLE), were inconsistent. The GDSC and CCLE investigators recently reported that their respective studies exhibit reasonable agreement and yield similar molecular predictors of drug response, seemingly contradicting our previous findings. Reanalyzing the authors' published methods and results, we found that their analysis failed to account for variability in the genomic data and more importantly compared different drug sensitivity measures from each study, which substantially deviate from our more stringent consistency assessment. Our comparison of the most updated genomic and pharmacological data from the GDSC and CCLE confirms our

Open Peer Review

Approval Status ✔ ✔ ✔

	1	2	3
version 1 09 May 2016	✔ view	✔ view	✔ view

1. **Weida Tong**, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, USA

2. **Terence Speed**, Walter and Eliza Hall Institute of Medical Research, Parkville, Australia

3. **Yudi Pawitan**, Karolinska Institute,

published findings that the measures of drug response reported by these two groups are not consistent. We believe that a principled approach to assess the reproducibility of drug sensitivity predictors is necessary before envisioning their translation into clinical settings.

Keywords

Cancer Cell Lines , Pharmacogenomics , High-Throughput Screening , Biomarkers , Drug Response , Experimental Design , Bioinformatics , Statistics

Stockholm, Sweden

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Preclinical Reproducibility and Robustness gateway.

Corresponding author: Benjamin Haibe-Kains (bhaibeka@uhnresearch.ca)

Introduction

Pharmacogenomic studies correlate genomic profiles and sensitivity to drug exposure in a collection of samples to identify molecular predictors of drug response. The success of validation of such predictors depends on the level of noise both in the pharmacological and genomic data. The groundbreaking release of the Genomics of Drug Sensitivity in Cancer[1] (GDSC) and Cancer Cell Line Encyclopedia[2] (CCLE) datasets enables the assessment of pharmacogenomic data consistency, a necessary requirement for developing robust drug sensitivity predictors. Below we briefly describe the fundamental analytical differences between our initial comparative study[3] and the recent assessment of pharmacogenomic agreement published by the GDSC and CCLE investigators[4].

Which pharmacological drug response data should one use?

The first GDSC and CCLE studies were published in 2012 and the investigators of both studies have continued to generate data and to release them publicly. One would imagine that any comparative study would use the most current versions of the data. However, the authors of the reanalysis used an old release of the GDSC (July 2012) and CCLE (February 2012) pharmacological data, resulting in the use of outdated IC_{50} values, as well as missing approximately 400 new drug sensitivity measurements for the 15 drugs screened both in GDSC and CCLE. Assessing data that are three years old and which have been replaced by the very same authors with more recent data seems to be a substantial missed opportunity. It raises the question as to whether the current data would be considered to be in agreement and which data should be used for further analysis.

Comparison of drug sensitivity predictors

Given the complexity and high dimensionality of pharmacogenomic data, the development of drug sensitivity predictors is prone to overfitting and requires careful validation. In this context, one would expect the most significant predictors derived in GDSC to accurately predict drug response in CCLE and *vice versa*. This will be the case if both studies independently produce consistent measures of both genomic profiles and drug response for each cell line. In our comparative study[3], we made direct comparison of the same measurements generated independently in both studies by taking into account the noise in both the genomic and pharmacological data (Figure 1a). By investigating the authors' code and methods, we identified key shortcomings in their analysis protocol, which have contributed to the authors' assertion of consistency between drug sensitivity predictors derived from GDSC and CCLE.

For their ANOVA analyses, the authors used drug activity area (1-AUC) values independently generated in GDSC and CCLE, but used the same GDSC mutation data across the two different datasets (Figure 1b; see Methods). By using the same mutation calls for both GDSC and CCLE, the authors have disregarded the noise in the molecular profiles, while creating an information leak between the two studies. For their ElasticNet analysis, the authors followed

a similar design by reusing the CCLE genomic data across the two datasets, but comparing different drug sensitivity measures that are IC_{50} in GDSC vs. AUC in CCLE (Figure 1c; see Methods).

We are puzzled by the seemingly arbitrary choices of analytical design made by the authors, which raises the question as to whether the use of different genomic data and drug sensitivity measures would yield the same level of agreement. Moreover, by ignoring the (inevitable) noise and biological variation in the genomic data, the authors' analyses is likely to yield over-optimistic estimates of data consistency, as opposed to our more stringent analysis design[3].

What constitutes agreement?

In examining correlation, there is no universally accepted standard for what constitutes agreement. However, the FDA/MAQC consortium guidelines define good correlation for inter-laboratory reproducibility[5–8] to be ≥0.8. The authors of the present study used two measures of correlation, Pearson correlation (ρ) and Cohen's kappa (κ) coefficients, but never clearly defined *a priori* thresholds for consistency, instead referring to ρ>0.5 as "reasonable consistency" in their discussion. Of the 15 drugs that were compared, their analysis found only two (13%) with ρ>0.6 for AUC and three (20%) above that threshold for IC_{50}. This raises the question whether ρ~0.5–0.6 for one third of the compared drugs should be considered as "good agreement." If one applies the FDA/MAQC criterion, only one drug (nilotinib) passes the threshold for consistency.

Similarly, the authors referred to the results of their new Waterfall analysis as reflective of "high consistency," even though only 40% of drugs had a κ≥0.4, with five drugs yielding moderate agreement and only one drug (lapatinib) yielding substantial agreement according to the accepted standards[9]. Based on these results, the authors concluded that 67% of the evaluable compounds showed reasonable pharmacological agreement, which is misleading as only 8/15 (53%) and 6/15 (40%) drugs yielded ρ>0.5 for IC_{50} and AUC, respectively. Taking the union of consistency tests is bad practice; adding more sensitivity measures (even at random) would ultimately bring the union to 100% without providing objective evidence of actual data agreement.

Consistency in pharmacological data

The authors acknowledged that the consistency of pharmacological data is not perfect due to the methodological differences between protocols used by CCLE and GDSC, further stating that standardization will certainly improve correlation metrics. To test this important assertion, the authors could have analyzed the replicated experiments performed by the GDSC using identical protocols to screen camptothecin and AZD6482 against the same panel of cell lines at the Wellcome Trust Sanger Institute and the Massachusetts General Hospital.

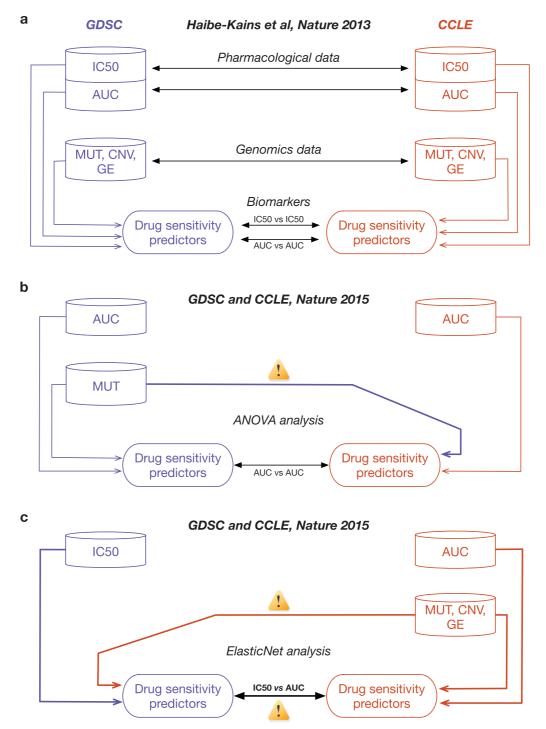Our re-analyses[3,10] of drug sensitivity data from these drugs found a correlation between GDSC sites on par with the

Figure 1. Analysis designs used to compare pharmacogenomic studies. (**a**) Analysis design used in our comparative study (Haibe-kains *et al.*, Nature 2013) where each data generated by GDSC and CCLE are independently compared to avoid information leak and biased assessment of consistency. (**b**) Analysis design used by the GDSC and CCLE investigators for their ANOVA analysis where the mutation data generated with GDSC were duplicated for use in the CCLE study. (**c**) Analysis design for the ElasticNet analysis where the molecular profiles from CCLE were duplicated in the GDSC study and the GDSC IC$_{50}$ were compared to CCLE AUC data. Differences between our analysis design and those used by the GDSC and CCLE investigators are indicated by yellow signs with exclamation mark symbol.

correlations observed between GDSC and CCLE (ρ=0.57 and 0.39 for camptothecin and AZD6482, respectively; Figure 2 a,b). These results suggest that intrinsic technical and biological noise of pharmacological assays is likely to play a major role in the lack of reproducibility observed in high-throughput pharmacogenomic studies, which cannot be attributed solely to the use of different experimental protocols.

Consistency in genomic data

In their comparative study, the authors did not assess the consistency of genomic data between GDSC and CCLE[4]. Consistency of gene copy number and expression data were significantly higher than for drug sensitivity data (one-sided Wilcoxon rank sum test p-value=3×10^{-5}; Figure 3), while mutation data exhibited poor consistency as reported previously[11]. The very high consistency
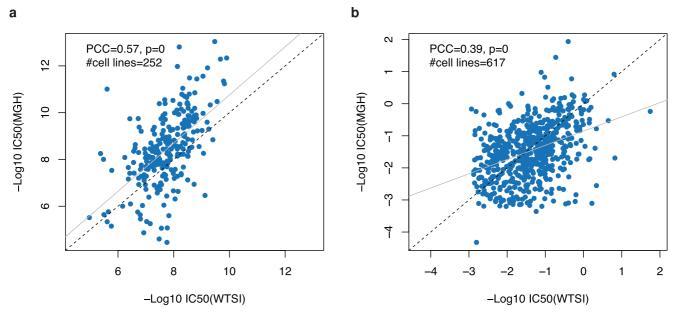
a

b

Figure 2. Consistency of sensitivity profiles between replicated experiments across GDSC sites. (**a**) Camptothecin and (**b**) AZD6482. PCC: Pearson correlation coefficient; MGH: Massachusetts General Hospital (Boston, MA, USA); WTSI: Wellcome Trust Sanger Institute (Hinxton, UK).
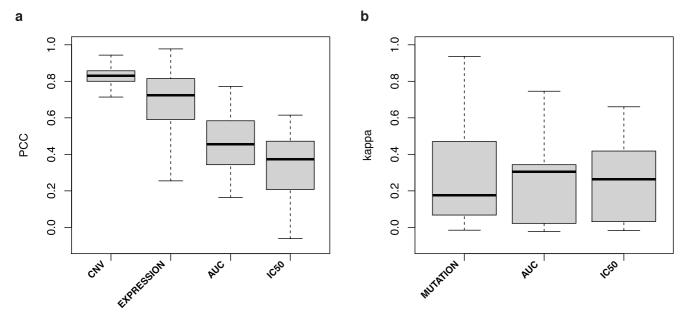
a

b

Figure 3. Consistency of molecular profiles between GDSC and CCLE. (**a**) Continuous values for gene copy number ratio (CNV), gene expression (EXPRESSION), AUC and IC_{50} and (**b**) for binary values for presence/absence of mutations (MUTATION) and insensitive/sensitive calls based on AUC >= 0.2 and IC_{50} > 1 microMolar values. PCC: Pearson correlation coefficient; Kappa: Cohen's Kappa coefficient.

of copy number data is quite remarkable (Figure 3a) and could be partly attributed to the fact that CCLE investigators used their SNP array data to compare cell line fingerprints with those of the GDSC project prior to publication and removed the discordant cases from their dataset[2].

Conclusions

We agree with the authors that their and our observations "[…] *raise important questions for the field about how best to perform comparisons of large-scale data sets, evaluate the robustness of such studies, and interpret their analytical outputs.*" We believe that a principled approach using objective measures of consistency and an appropriate analysis strategy for assessing the independent datasets is essential. An investigation of both the methods described in the manuscript and the software code used by the authors to perform their analysis[4] identified fundamental differences in analysis design compared to our previous published study[3]. By taking into account variations in both the pharmacological and genomic data, our assessment of pharmacogenomic agreement is more stringent and closer to the translation of drug sensitivity predictors in preclinical and clinical settings, where zero-noise genomic information cannot be expected.

Our stringent re-analysis of the most updated data from the GDSC and CCLE confirms our 2013 finding that the measures of drug response reported by these two groups are not consistent and have not improved substantially as the groups have continued generating data since 2012[10]. While the authors make arguments suggesting consistency, it is difficult to imagine using these post hoc methods to drive discovery or precision medicine applications.

The observed inconsistency between early microarray gene expression studies served as a rallying cry for the field, leading to an improvement and standardization of experimental and analytical protocols, resulting in the agreement we see between studies published today. We are looking forward to the establishment of new standards for large-scale pharmacogenomic studies to realize the full potential of these valuable data for precision medicine.

Methods

The authors' software source code. As the authors' source code, we refer to the 'CCLE.GDSC.compare' (version 1.0.4 from December 18, 2015) and DRANOVA (version 1.0 from October 21, 2014) R packages available from http://www.broadinstitute.org/ccle/Rpackage/.

Pharmacogenomic data

As evidenced in the authors' code (lines 20 and 29 of CCLE.GDSC.compare::PreprocessData.R), they used GDSC and CCLE pharmacological data released on July 2012 and February 2012, respectively. However the GDSC released updated sets of pharmacological data (release 5) on June 2014, gene expression arrays (E-MTAB-3610) and SNP arrays (EGAD00001001039) on July 2015. CCLE released updated pharmacological data on February 2015, the mutation and SNP array on October 2012, and the gene expression data, on March 2013. These updates

substantially increased the overlap in genomic features between the two studies, thus providing new opportunities to investigate the consistency between GDSC and CCLE[10].

ANOVA analysis

In the authors' ANOVA analyses, identical mutation data were used for both GDSC and CCLE studies as can be seen in the authors' analysis code in lines 20, 25–35 of CCLE.GDSC.compare:: plotFig2A_biomarkers.R.

ElasticNet (EN) analysis

In their EN analyses, the authors compared different drug sensitivity measures, using IC_{50} in GDSC and AUC in CCLE, as described in the Supplementary Data 5 and stated in the Methods section of their published study:

> "*Since the IC50 is not reported in CCLE when it exceeds the tested range of 8 μM, we used the activity area for the regression as in the original CCLE publication. We also used the values considered to be the best in the original GDSC study: the interpolated log(IC50) values.*"

This was confirmed by looking at the authors' analysis code, lines 83 and 102 of CCLE.GDSC.compare::ENcode/prepData.R. Moreover, identical genomic data were used for both GDSC and CCLE studies, as described the Methods section of the published study:

> "*In order to compare features between the two studies, we used the same genomic data set (CCLE).*"

This was confirmed by looking at the authors' analysis code, lines 17, 38, 51, and 70 of CCLE.GDSC.compare::ENcode/ genomic.data.R, and lines 10–11 of CCLE.GDSC.compare:: plotFigS6_ENFeatureVsExpected.R.

Statistical analysis

All analyses were performed using the most updated version of the GDSC and CCLE pharmacogenomic data based on our *PharmacoGx* package[12] (version 1.1.4).

Research replicability

All analyses were performed using the most updated version of the GDSC and CCLE pharmacogenomic data based on our PharmacoGx package[12] (version 1.1.4). PharmacoGx provides intuitive function to download, intersect and compare large pharmacogenomics datasets. The PharmacoSet for the GDSC and CCLE datasets are available from pmgenomics.ca/bhklab/sites/ default/files/downloads/ using the downloadPSet() function. The code and the data used to generate all the results and figures are available as Data Files 1 and 2. The code is also available on GitHub: github.com/bhklab/cdrug-rebuttal.

The Waterfall approach

In the Methods, the authors use all cell lines to optimally identify the inflection point in the response distribution curves. The authors

stated that "*This is a major difference to the Haibe-Kains et al. analysis, as that analysis only considered the cell-lines in common between the studies when generating response distribution curves.*" This is not correct. As can be seen in our publicly available R code, we performed the sensitivity calling (using the Waterfall approach as published in the CCLE study[2] before restricting our analysis to the common cell lines, for the obvious reasons that the authors mentioned in their manuscript. See lines 308 and 424 in https://github.com/bhklab/cdrug/blob/master/CDRUG_format.R.

Data and software availability

Open Science Framework: Dataset: Assessment of pharmacogenomic agreement, doi 10.17605/osf.io/47rfh[13]

Author contributions

Z Safikhani, N El-Hachem R Quevedo, and P Smirnov were responsible for downloading and curating the pharmacogenomic data. Z Safikhnao wrote most of the analysis code with the help of N El-Hachem R Quevedo, and P Smirnov. Z Safikhani, J Quackenbush and B Haibe-Kains designed the study. B Haibe-Kains supervised the study. All authors participated in the interpretation of the results. Z Safikhani, A Goldenberg, N Juul Birkbak, C Mason, C Hatzis, L Shi, H Aerts, J Quackenbush and B Haibe-Kains participated in the manuscript writing.

Competing interests

No competing interests were disclosed.

References

1. Garnett MJ, Edelman EJ, Heidorn SJ, *et al.*: **Systematic identification of genomic markers of drug sensitivity in cancer cells.** *Nature.* 2012; **483**(7391): 570–575.
   PubMed Abstract | Publisher Full Text | Free Full Text

2. Barretina J, Caponigro G, Stransky N, *et al.*: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.** *Nature.* 2012; **483**(7391): 603–607.
   PubMed Abstract | Publisher Full Text | Free Full Text

3. Haibe-Kains B, El-Hachem N, Birkbak NJ, *et al.*: **Inconsistency in large pharmacogenomic studies.** *Nature.* 2013; **504**(7480): 389–393.
   PubMed Abstract | Publisher Full Text | Free Full Text

4. Cancer Cell Line Encyclopedia Consortium, Genomics of Drug Sensitivity in Cancer Consortium: **Pharmacogenomic agreement between two cancer cell line data sets.** *Nature.* 2015; **528**(7580): 84–87.
   PubMed Abstract | Publisher Full Text

5. SEQC/MAQC-III Consortium: **A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium.** *Nat Biotechnol.* 2014; **32**(9): 903–914.
   PubMed Abstract | Publisher Full Text | Free Full Text

6. MAQC Consortium, Shi L, Reid LH, *et al.*: **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *Nat Biotechnol.* 2006; **24**(9): 1151–1161.
   PubMed Abstract | Publisher Full Text | Free Full Text

7. Shi L, Campbell G, Jones WD, *et al.*: **The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models.** *Nat Biotechnol.* 2010; **28**(8): 827–838.
   PubMed Abstract | Publisher Full Text | Free Full Text

8. Li S, Tighe SW, Nicolet CM, *et al.*: **Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study.** *Nat Biotechnol.* 2014; **32**(9): 915–925.
   PubMed Abstract | Publisher Full Text | Free Full Text

9. Sim J, Wright CC: **The kappa statistic in reliability studies: use, interpretation, and sample size requirements.** *Phys Ther.* 2005; **85**(3): 257–268.
   PubMed Abstract

10. Safikhani Z, Freeman M, Smirnov P, *et al.*: **Revisiting inconsistency in large pharmacogenomic studies.** *bioRxiv.* 2015; 026153.
    Publisher Full Text

11. Hudson AM, Yates T, Li Y, *et al.*: **Discrepancies in cancer genomic sequencing highlight opportunities for driver mutation discovery.** *Cancer Res.* 2014; **74**(22): 6390–6396.
    PubMed Abstract | Publisher Full Text | Free Full Text

12. Smirnov P, Safikhani Z, El-Hachem N, *et al.*: **PharmacoGx: An R package for analysis of large pharmacogenomic datasets.** *Bioinformatics.* 2015; **32**(8): 1244–1246. pii: btv723.
    PubMed Abstract | Publisher Full Text

13. Safikhani Z, El-Hachem N, Quevedo R, et al.: **Dataset: Assessment of pharmacogenomic agreement.** Open Science Framework. 2016.
    Data Source

F1000Research

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 1

Reviewer Report 29 June 2016

✔ **Yudi Pawitan**

Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

The paper highlights the curious lack of rigorous standards for what constitutes 'agreement', 'consistency' between genomic studies, or more generally, the fundamental issues of 'validation' and 'reproducibility', etc. The problem is even more serious of results based on high-throughput omics data as the potential for false positive is substantial.

The persistent lack of consensus or standards may partly indicate that these issues are not so straightforward. The main problem is that when we say we 'validate' a result, this can be done at different strengths. For example, consider the commonly performed method in statistical analyses, the so-called 'cross-validation', where we split our total sample into training and validation sets. If the split is done randomly, then we have only a 'soft validation', since it applies to the same sample (or same lab, same population, same measurement method, etc) so the 'validation' is internal and corresponds to statistical significance only. In contrast a scientist may wish for something stronger, for an external validation, for example, for the 'biological truth' to apply other populations; thus, one study may be performed in a European population, but the external validation is done in an Asian population. The latter is a stronger validation than the random-split validation, giving a more compelling and general biological story. What is relevant here is that both validations are commonly done in practice, and both are valid, but they carry different levels of information. I think what matters in practice is that the implication of the validation should always be clear (or clarified), so that the user of the information can judge its relevance.

The key point of Safikhani *et al* is that their 2013 validation study of the genomic predictors of drug-sensitivity was more stringent than the 2015 validation studies by the GDSC and CCLE investigators. This is clearly highlighted in Figure 1, where the latter used the same molecular data, so the 'validation' is only of the pharmacological data and perhaps (not clear to me) the method of analyses. Which level of validation is more relevant here? Let us imagine how the results (eg the genomic predictors) are to be used in patients. The molecular data are likely to be generated and analyzed in a diversity of labs, so the genomic predictors should really be robust to the actual heterogeneity in the molecular data. The results (the genomic predictors) may not

survive such stringent requirements, but that is what we need to know. So, overall, I agree with Safikhani *et al* that a more stringent validation allowing for variability in both molecular and pharmacological data is more relevant in this context of drug prediction.

(However, reading Haibe-Kains *et al*, there seemed to be an emphasis that the failure of agreement was due to the high variability in the pharmacological data. So it is possible that the later studies by the GDSC-CCLE investigators responded to this concern only.)

Regarding specific issues in the paper:
1. I do not consider the use of most recent data as a key issue.
2. I agree that the choice of IC_{50} in GDSC vs AUC in CCLE is puzzling and only raises a question mark regarding the results.
3. Arbitrary cutoffs in defining what constitutes an 'agreement' are unnecessary if authors can refrain from using judgmental words like 'high consistency' etc., especially when used as a summary statement across distinct drugs. It would be better to just report the actual performance for each drug or for each cancer type, since it is still not clear how these statistics would translate in terms of clinical cost-benefit balance.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 27 June 2016

✔ **Terence Speed**
Bioinformatics division, Walter and Eliza Hall Institute of Medical Research, Parkville, Vic, Australia

I found the title appropriate, and that the abstract represented a suitable summary of the work.

I believe that the design, methods and analysis of results are appropriate for the topic being studied, and that for the most part, they were clearly explained. A couple of perceived shortcomings are itemized here.

p.3, column 2, line 2. The "but" would be better replaced by "and".

p.5. Figure 2. The dotted and solid diagonal lines on these plots are not identified in either the caption or the text.

p.5, Figure 3. It is nowhere explained whose Pearson correlations (PCC) are summarized in these

box plots. I suppose that some number (to be stated) of cell lines were profiled in both GDSC and CCLE, and that in all cases, the PCC in the box plots are calculated from molecular data from pairs consisting of the data on the same cell line generated in GDSC and in CCLE. A clear statement along these lines would be helpful.

p.6, column 1, lines 1-4. This assertion would have more force if the authors told the reader how many cell lines *could* have contributed PCC to the box plot of Figure 3a, and how many *did* do so.

Further, I do believe that the conclusions are sensible, balanced and justified on the basis of the results of the study.

Finally, I understand that all the data used in this study is available, and this is also true for the code used to generate all the results and figures.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 10 June 2016

https://doi.org/10.5256/f1000research.9367.r14316

✔ **Weida Tong**
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA

It is a lot to take/digest the manuscript. I break this story into three parts:

- In 2012, both GDSC and CCLE released/published drug sensitivity data (both pharmacological and genomic). In 2013, the authors compared the two studies using the drugs in common between two. Their analysis was carried out in a direct fashion which account for variations of both genomic and pharmacological data from the same site (GDSC or CCLE) and found the results between two did not agree.

- Recently, GDSC/CCLE did an independent analysis and demonstrated that the agreement between two are actually higher (using ANOVA) than what the authors reported. They concluded that the results between GDSC and CCLE were consistent. However, the comparison was only focused on the pharmacological data because the genomic data used actually came from one site. That means their analysis did not include the noise introduced by both sites in this comparison.

- The authors, again, reanalyzed data by including pharmacological and genomic data from both sites and the conclusions remain as the same as they reported in 2013.

I have no problem with their analysis and support their conclusions. With that said, I did find the paper could flow better by moving two sections into Discussion. These are:

- "Which pharmacological drug response data should one use?" - It seems odd and smell bad that GDSC/CCLE used the data published in 2012 and totally ignored the most current data in their analysis. This could be due to many different reasons. Thus, speculation is not necessary considered as "results". I would say this will be better justified as "discussion".

- "What constitutes agreement" – Again, this is a difficult call. I believe there is no single baseline that can be used to justify consistency. Thus, most text in this section will sit better in "discussion".

Overall, I support its indexation with revision by focusing on the flow of the story and the structure of manuscript.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research