

Enhancing Reproducibility in Cancer Drug Screening: How Do We Move Forward?

Christos Hatzis^{1,3}, Philippe L. Bedard^{10,11}, Nicolai J. Birkbak¹³, Andrew H. Beck⁴, Hugo J.W.L. Aerts^{5,7}, David F. Stern^{2,3}, Leming Shi^{8,14,15}, Robert Clarke⁹, John Quackenbush^{5,6}, and Benjamin Haibe-Kains^{10,12}

Abstract

Large-scale pharmacogenomic high-throughput screening (HTS) studies hold great potential for generating robust genomic predictors of drug response. Two recent large-scale HTS studies have reported results of such screens, revealing several known and novel drug sensitivities and biomarkers. Subsequent evaluation, however, found only moderate interlaboratory concordance in the drug response phenotypes, possibly due to differences in the experimental protocols used in the two studies. This highlights the need for community-wide implementation of standardized assays for measuring drug response phenotypes so that the full potential of HTS is realized. We suggest that the path forward is to establish best practices and standardization of the critical steps in these assays through a collective effort to ensure that the data produced from large-scale screens would not only be of high intrastudy consistency, so that they could be replicated and compared successfully across multiple laboratories. *Cancer Res*; 74(15); 4016–23. ©2014 AACR.

Introduction

Pharmacogenomic high-throughput screening (HTS; see definitions box) offers tremendous promise in the rational development of targeted therapies, but its full potential will not be realized unless experimental protocols and analysis methods are standardized through a collective effort.

With dropping costs of genetic testing, cutting-edge sequencing technologies are now at the forefront of new cancer drug development. Yet, how to fully incorporate the presence of individual mutations and other specific genomic features to

inform personalized therapeutic decisions remains a major challenge. Pharmacogenomics has emerged as a promising strategy to rapidly elucidate the link between genes and drugs by systematically characterizing the effects of pharmacologic agents on whole biologic systems, allowing concurrent identification of therapeutic targets and discovery of drug candidates (1).

Advancements in molecular biology and the genomic technologies have had a profound impact on drug discovery, and together with the advent of automated HTS have dramatically changed the drug development endeavor (2). Pharmacogenomic HTS presents both opportunities and challenges. The goal is to measure the response of hundreds or thousands of cell lines to drugs or other perturbations and to associate the response with the genomic characteristics of each cell line. The extensive data generated from such efforts provide opportunities to search for biomarkers and to explore the mechanism associated with the underlying response. Two recent pharmacogenomic HTS studies, the Cancer Cell Line Encyclopedia (CCLE; ref. 3) and Cancer Genome Project (CGP; ref. 4) evaluated an impressive array of cell lines and drugs (1,036 cell lines and 24 drugs, and 727 cell lines and 138 drugs, respectively), generating gene-expression profiles and drug-sensitivity data for each combination. A subsequent comparative analysis of the CCLE and CGP found that although the gene-expression profiles were highly concordant between studies, the measured cell line drug sensitivities were inconsistent (5), which has been further confirmed by an independent research group (6). Although the apparent variability in drug response presents a serious barrier to the ultimate goal of such studies—to develop signatures predictive of response—the high degree of correlation in the gene-expression measures provides hope for a potential path forward. Here, we review potential explanations for these findings and extend recommendations

Authors' Affiliations: ¹Section of Medical Oncology, Yale University School of Medicine; ²Department of Pathology, Yale University School of Medicine; ³Yale Cancer Center, Yale University, New Haven, Connecticut; ⁴Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School; ⁵Department of Biostatistics and Computational Biology and Center for Cancer Computational Biology; ⁶Department of Cancer Biology, Dana-Farber Cancer Institute; ⁷Department of Radiation Oncology and Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; ⁸National Center for Toxicological Research, US Food and Drug Administration, Jefferson, Arkansas; ⁹Georgetown University Medical Center, Washington, DC; ¹⁰Princess Margaret Cancer Centre, University Health Network; ¹¹Faculty of Medicine; ¹²Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; ¹³Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; ¹⁴State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, Schools of Life Sciences and Pharmacy, Fudan University, Shanghai; and ¹⁵Zhanjiang Center for Translational Medicine, Shanghai, China

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: B. Haibe-Kains, Princess Margaret Cancer Centre, University Health Network, 101 College Street, Toronto, ON M5G 1L7, Canada. Phone/Fax: 14165818626; E-mail: bhaibeka@uhnresearch.ca

doi: 10.1158/0008-5472.CAN-14-0725

©2014 American Association for Cancer Research.

Definitions

- **Pharmacogenomics:** the use of genetic information from populations to inform drug design and development or from individual patients to inform the clinical management of pharmacotherapy, including drug selection, dosing, and analysis of drug toxicities. Here, this term is used more broadly to include experimentation on model systems.
- **(Conventional) High-Throughput Screening:** highly automated methodology that heavily relies on robotics, liquid handling devices, and automated detection to quickly conduct millions of pharmacologic tests, typically on specific protein targets
- **In vitro cell-based/pharmacogenomic HTS:** extension of conventional HTS to include large panel of cancer cell lines to assess *in vitro* drug response.

for improving the reproducibility of pharmacogenomic HTS studies.

Pharmacogenomics in Cancer Drug Development

With more than 900 new cancer drugs in clinical testing (7) and better technologic tools to characterize patient populations and modulation of drug targets, the paradigm of oncology drug development is shifting (8, 9). A new drug must now demonstrate proof of concept that it can be beneficial as early as possible in clinical development. Preclinical pharmacogenomic HTS could identify genomic predictors allowing investigators to enrich early-phase clinical studies for those patients most likely to receive benefit. An early signal of proof-of-concept activity, such as vemurafenib in BRAF V600 mutant melanoma (10) and crizotinib in ALK fusion non-small cell lung cancer (11), can rapidly reduce the time to clinical testing and subsequent market approval. With a growing repertoire of potential drug combination partners, HTS could help to prioritize genotype-selective drug combinations for clinical testing (12). Genomic predictors developed from HTS data can support this new paradigm when the number of false-positive leads is contained.

HTS has become one of the primary scientific tools used in the pharmaceutical industry to generate new leads (13). Very large compound libraries are screened for "hits" against a target (14). Hits are generally target-specific active small molecules, which are further characterized for dose-response effects in secondary screens. Functional characterization and validation in animal models leads to selection of candidates for clinical testing. Pharmacogenomic screening extends the conventional HTS screening by using high-throughput approaches to evaluate the effect of compounds on large panels of cancer cell lines; this approach enables to identify simultaneously drug-gable targets and biologically active compounds (1, 15). Effectiveness of a compound is assessed by functional assays that measure changes in such activities as cell proliferation, DNA

replication, or cell death. Associating these changes with drug-induced transcriptomic alterations can identify cellular processes that are affected by a given treatment. A key advantage of pharmacogenomic screening is that it can compress the timeline of conventional HTS. By using systems biology methods, at least partial functional characterization and initial validation of the specificity/selectivity of a given compound can be obtained.

Given the complexity of HTS assay, there are many sources of experimental noise that should be carefully controlled to provide sufficient power to detect reliable "hits" (16). Occasional false-positive hits passing through the primary screen will likely be caught in secondary screening steps. However, the biologic outcomes measured by pharmacogenomic HTS assays are often associated with increased variability, which reduces our ability to detect weaker target-drug-genotype associations in a single step. Several sources of variation affect these assays and may limit their power to reliably detect a broad range of biologic associations, especially where a compound is effective against multiple potential targets in one or more pathways, or when the overall cytotoxic effect reflects cross-talk among interconnected biologic pathways.

Lack of Concordance in Large Pharmacogenomic Studies

For pharmacogenomics to realize its full potential, the assessment of drug response needs to be robust, reliable, and reproducible. The CCLE and CGP drug screening studies are the largest public pharmacogenomic screens to date. To the credit of the investigators, they shared important details of their experimental protocols and all the pharmacogenomic data generated, allowing a careful reanalysis. Fifteen drugs and 471 cancer cell lines were screened in both studies (Fig. 1A). These studies provided a unique opportunity to assess the consistency of both the raw data and the set of significant associations between genomic features, defined by gene-expression measurements or mutation calls, and measures of drug response (5).

The experimental protocols used to profile the transcriptomes at baseline before drug treatments were very similar in both studies, with the exception of the different density of the Affymetrix chips used (Fig. 1A). Transcriptional profiles from replicate samples within the same study and between identical cell lines across studies were highly correlated, suggesting that the high level of standardization for gene-expression profiling and analysis enabled reproducible findings. Only moderate concordance was observed in mutations detected in 64 genes in the 471 common cell lines between CCLE and CGP, which could be due to differences in the platforms used for mutational profiling (Fig. 1A), or as a result of genetic divergence of the cell lines during passaging.

CCLE and CGP studies generated large amounts of drug-sensitivity data using experimental protocols that differed in many important aspects (Fig. 1B; Supplementary Table S1 and Supplementary Information), including the specific pharmacologic assay used [adenosine triphosphate (ATP)-based vs. reductase-based metabolic activity for CCLE and CGP,

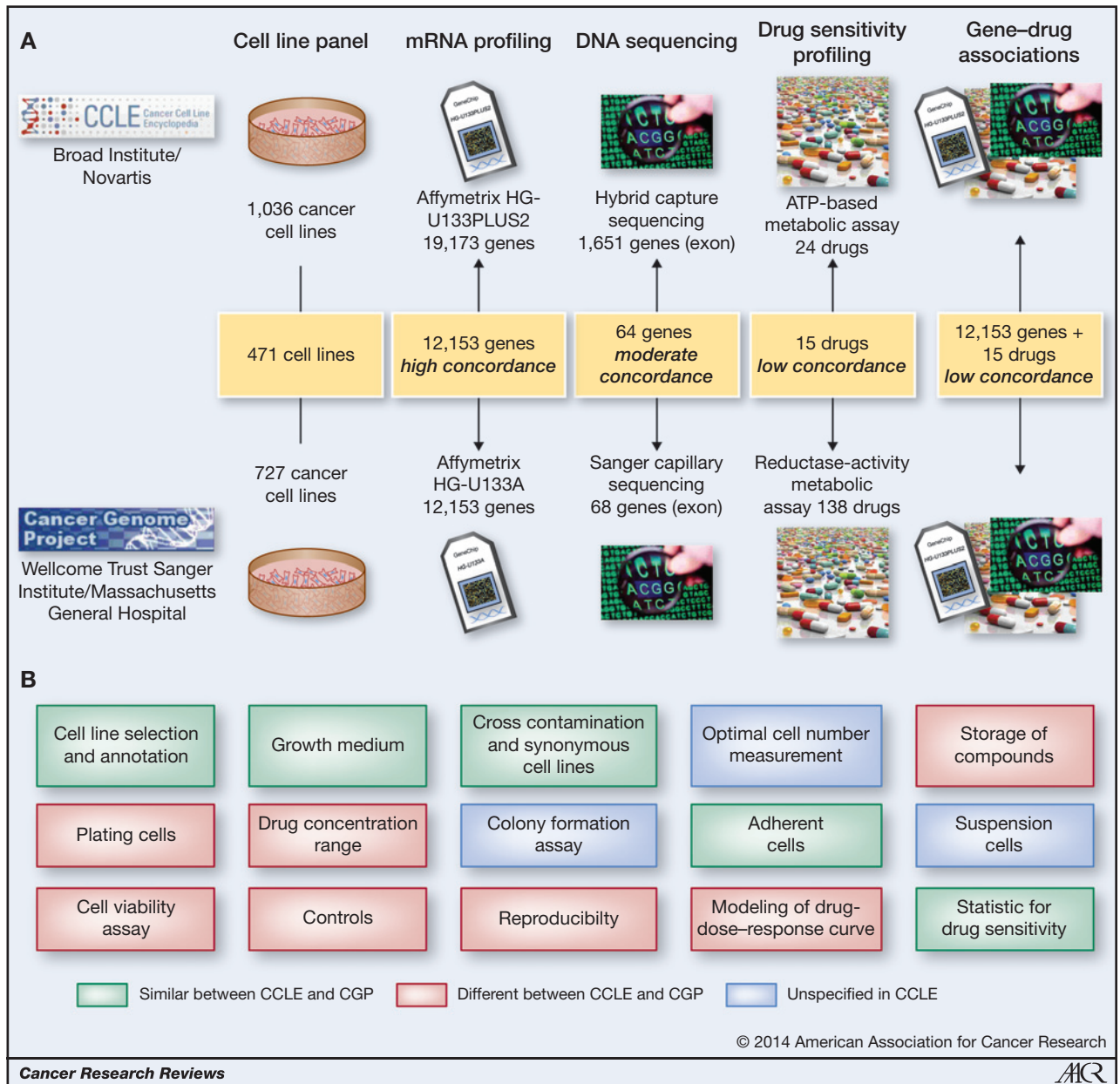


Figure 1. A, comparison of the Cancer Genome Project (4) and The Cancer Cell Line Encyclopedia (3) as published by Haibe-Kains and colleagues (5). Our comparative study revealed a high degree of concordance between the gene-expression datasets, moderate concordance for the mutation data, but low concordance between drug-sensitivity assays; such inconsistency propagates into the gene–drug associations found by the two studies. B, main steps in the experimental protocols used in the CCLE and CGP studies; each step is further described in Supplementary Table S1. The choice of experimental procedures vastly differs between the two studies, illustrating the lack of standardization in large pharmacogenomic studies.

respectively]. Although the two studies were not designed specifically to address interlaboratory reproducibility, the differences reflect a lack of consensus for pharmacogenomic HTS protocols. For 13 of the 15 drugs tested in common in CCLE and CGP (87%), a high level of inconsistency between drug-sensitivity measurements was observed (Spearman rank correlation <0.5), regardless of which characteristic of the drug-dose–response curve was used to summarize drug potency (Fig. 1A). Intriguingly, when the same protocol was used at two different sites in the CGP study, only a fair intersite

correlation for sensitivity to camptothecin was observed [Spearman rank correlation for the half maximal inhibitory concentration (IC_{50}) value of 0.57]. These results led us to conclude that the two experimental protocols are not equivalent and that measuring drug response is a complex and poorly reproducible process. Although compounds were assayed at least in duplicate for each cell line, unfortunately, the raw data from these technical replicates were not publicly released by either study, preventing a direct assessment of assay analytical reproducibility.

To confirm our findings, we compared the data for two drugs, lapatinib and paclitaxel, against a third-large pharmacogenomic dataset generated by GlaxoSmithKline (GSK) that used the same ATP-based assay as in the CCLE study (17). As expected, the correlation was higher between CCLE and GSK compared with CGP and GSK, suggesting that the choice of assay has a major impact on drug response measurements. However, the correlation between GSK and CCLE drug-sensitivity data was still only fair (Spearman correlation <0.51), implying that the drug-sensitivity assay used is not the only source of inconsistency (5).

The primary goal of CCLE and CGP studies was to discover new associations between genomic and transcriptomic features of cancer cells and drug response that could advance our understanding of drug sensitivity and allow development of robust genomic predictors of response. Thus, we tested whether these associations were consistent across studies. We observed that, although we were able to validate the strongest associations identified in both studies, coupling consistent transcriptomic data with inconsistent drug-sensitivity measurements resulted in inconsistent pharmacogenomic associations for most of the investigated drugs. Hence, any single study may not be able to identify robust predictors of drug response because a strong association observed in one study may not necessarily be evident in independent studies (5).

***In Vitro* Cell-Based Drug Screening: What Can Affect Assay Consistency?**

When considered independently, both the CCLE and CGP studies approached their experimental design, implementation, and data analysis with apparently adequate care and rigor; however, there were methodologic differences between the two studies (Supplementary Information). Several factors could introduce variability in the results of *in vitro* cell-based assays. It is important to establish best practices on how the chemical compound collection is managed, maintained, and delivered to the assay plates for screening (18). Critical factors include compound handling to ensure accuracy in the amount of the compound transferred to the master solution, maximal solubility in the solvent used (usually DMSO), optimal storage conditions to ensure compound integrity, and to minimize liquid loss through evaporation or leaching of the test compounds through the storage containers (19). Because of such potential artifacts, the purity, integrity, and concentration of the compound stock solutions should be verified before being used in the assay. Of the two studies, only CCLE reported preassay compound library validation. Furthermore, the type of the liquid transfer system can affect the accuracy of the amount of compound and cell culture stocks delivered to the recipient plates (20). Tip-based serial dilution and dispensing could lead to larger IC_{50} values for some compounds, resulting in more than 100-fold underestimation of their potency compared with acoustic dispensing methods (21).

Minor variations in cell culture conditions, including seeding density, plating efficiency, growth rate, and cell-cycle distribution, may affect cell metabolism and drug

responsiveness. The number of cells per well at the time of pharmacologic testing also has a major impact on the dynamic range and sensitivity of the assay, and this variable depends both on the seeding density and on the growth rate of the cells. Several of these variables are affected by cell culture conditions, including substrate, choice of cell culture medium (including relative concentrations of glucose, glutamine, and amino acids), and serum (final concentration, type, source, and batch effects). Other factors affecting the cell lines include their reference source, passage number, and the quality of the stock cultures used to seed the experiments. Even cell lines obtained from the same reference source may diverge enormously unless similar passages from the reference culture are used. Both the CCLE and CGP groups appropriately attempted to reduce the contribution of several of these variables, which would help control intrastudy variability. However, these actions do not necessarily improve interstudy comparability.

Adaptive responses to the stress of drug treatment can occur within the transcriptome, proteome, metabolome, kinome, and methylome; some of these changes may be detected in hours or less of the initiation of exposure to a drug. Where rapid drug-induced changes in cells reflect common and/or evolutionarily conserved stress response pathways, key features that drive profile changes may seem qualitatively more similar than different across studies. Thus, it is likely that, provided the cell lines are of the same origin, the genomic data will appear broadly similar. If the same cell lines respond in the same way to the same drugs, selective features of the transcriptomes might be expected to exhibit at least qualitatively similar changes. These observations are broadly consistent with the findings of our comparative study (5).

The type of cytotoxicity assay used in assessing drug sensitivity could significantly affect the measured phenotypic response. Among the several assays compatible with HTS (22, 23), bioluminescent detection of ATP is one of the most sensitive and reliable methods (24) and has been used extensively in HTS studies (Table 1). CCLE used an ATP detection cell proliferation assay, but CGP used two different cell viability assays, a fluorescent nucleic acids stain and a redox indicator dye to assess drug inhibition (for details see Supplementary Information). A recent study by Chan and colleagues directly compared cell viability assays based on quantifying total amount of nucleic acid using fluorescent DNA-binding dyes (similar to the SYTO 60 assay used in the CGP study) versus ATP-dependent luminescence (CellTiter-Glo assay also used in the CCLE study; ref. 25). The study shows that the ATP-dependent luminescence assay is prone to underestimation of drug potency and efficacy, particularly for DNA synthesis-targeting agents (25). The ATP-dependent luminescence and fluorescent DNA-binding assays are measuring different aspects of the drug response phenotype, and, therefore, it is not surprising that the assays show only moderate correlation in our comparative analysis (5). Given the limitations of each assay it has been suggested that multiparameter testing, incorporating complementary cell-viability assays, yields the most robust and informative phenotypic measures (23).

Table 1. Comparison of available pharmacological assays for measuring cell viability

Cell viability assay	Type of assay	Pros	Cons	Suitable for HTS	Used in ^a
CellTiter-Glo (Promega)	Viability, membrane integrity, ATP	Nontoxic, does not require fixing or washing steps	ATP could be reduced by metabolic interferences, leading to false-positive results; underestimation of drug effect by DNA-targeting agents, in which cells may be incapacitated but intact	Yes	CCLC, GSK, FIMM, SU2C, NTP
Syto60 (Invitrogen)	Proliferation; fluorescent DNA stain	Simple	Requires fixation, washing; could stain DNA of intact but not-viable cells	Yes	CGP
Bromodeoxyuridine-ELISA (CytoSelect, Cell BioLabs)	Proliferation, DNA synthesis	Only live and replicating cells can incorporate bromodeoxyuridine	Assay is time consuming and cost prohibitive in high-throughput format	No	
DAPI	Proliferation, nucleic acids stain	Simple	Could stain nonviable cells if DNA is intact	Yes	GSK
MTT, MTS, XTT	Viability, metabolic reduction of tetrazolium dye	Simple colorimetric assay	Formazan product insoluble in aqueous media; secondary step required to solubilize it before optical detection	No	NCI60
SRB	Proliferation, anionic general biomass stain	Simple colorimetric assay; stable for extended periods; differentiates cell kill from growth inhibition	Requires fixation, washing steps	Yes	NCI60
Resazurin (CellTiter-Bue; Promega)	Viability, cell permeable redox indicator	Simple and inexpensive; more sensitive than tetrazolium assays	Potentially cytotoxic; requires long incubation 1–4 h; could underestimate toxic effect	Yes	
GF-AFC (CellTiter-Fluor; Promega)	Viability; penetrates live cells where cytoplasmic cytopeptidase activity releases AFC that fluoresces; protease becomes inactive upon cell death	Nontoxic, does not require fixing or washing steps; short incubation time <1 h	Potential underestimation of drug effect by DNA-targeting agents, in which cells may be incapacitated but intact	Yes	
Ruthenium dye	Proliferation; the fluorescence of the dye is quenched by oxygen; proliferating cells reduce external oxygen and increase fluorescence	Nontoxic, simple, rapid assay	Narrow dynamic range; signal depends on metabolic activity	Yes	
LDH (Cytotox-ONE; Promega)	Cytotoxicity; released LDH is measured by enzymatically coupled reagent chemistry	Reagents compatible with viable cells; quick incubation; cost effective	Susceptible to background signal from serum sources of LDH in growth media or from compounds inhibiting LDH activity	Yes	

^aCCLC (3); CGP (4); FIMM, Institute for Molecular Medicine, Finland (38); GSK, GlaxoSmithKline Study (17); NCI60, National Cancer Institute (39); NTP: National Toxicology Program (40); SU2C, the Lawrence Berkeley National Laboratory (41).

Primary phenotypic responses measured in pharmacogenomic studies include dose–response curves and two related measures of drug potency: the IC_{50} value and the area under the dose–response curve (AUC). These measures can be estimated using many different laboratory assays and analytic methods—no standard “best practices” have been shown to be either highly reproducible or generally reflective of the response in humans (26). The shape of the dose–response curve, and, thus, the estimates for both AUC and IC_{50} , depends not only on the drug–cell line combination but also on the range of drug concentrations tested, which was not identical in the two studies. More broadly, the physiologic definitions of AUC and IC_{50} , originally intended to describe the mass-action law of a single-substrate competitive inhibition enzyme kinetics (27), might not provide similarly meaningful measures of growth inhibition kinetics in whole-cell systems. Cell population heterogeneity and cell-to-cell variability in drug response may affect the shape of the dose–response curve, yielding shallow curves with reduced steepness or Hill slope (26). An independent reanalysis of the two studies found that in the CCLE study a drug could be assessed both as effective or ineffective against a cell line depending on whether the assessment of drug sensitivity was based on an IC_{50} or AUC (6). Thus, it may not be a simple task to obtain meaningful quantitative, and perhaps even qualitative, comparisons of responsiveness to many drugs across studies.

Standardization of *In Vitro* Cell-Based HTS

Reproducibility is a fundamental prerequisite for any scientific study, and this requirement becomes even more critical when experimental results affect medical decision-making. As large panels of genomically annotated cancer cell lines (such as those in the CCLE and CGP studies) are being used to screen cancer drugs and guide treatment selection for patients whose cancers show similar genomic alterations, it is important to ensure the reproducibility of cell line–based drug-sensitivity data from HTS assays. The unexpectedly low concordance in the results of the CCLE and CGP studies also illustrates a broader challenge. Scientifically appealing, novel high-throughput analytic methods often become widely used before the factors that affect assay performance are well understood. Because of the substantial experimental costs, limited funding, and lower priority assigned to this type of technology assessment, parameters that influence the results may not be established and standardized until after widespread use of the method.

When DNA microarray analysis was first introduced in the 1990s, there was great excitement about the potential of genome-wide expression profiling. The initial euphoria soon evaporated as many studies failed to be replicated. In 1999, the Microarray Gene Expression Data Society (MGED) was formed to address the need for standardization in microarray experiments and the reporting of experimental data and metadata. The group produced the Minimal Information about a Microarray Experiment standard (28, 29) and, through a series of meetings and workshops, helped to catalyze development of experimental and data analysis

standards. Data preprocessing tools such as dChip (30) and RMA (31) provided the means to compare large-scale assays, methods for removing batch effects (32, 33) helped to facilitate meta-analysis across experimental datasets, and approaches to creating robust biomarkers (34–37) have provided a framework to use expression data to make more reliable predictions of disease phenotypes.

It is clear that there is a need for a similar standardization of drug response measurements in addition to the development of new, robust drug-sensitivity assays that can both be replicated across studies and that are physiologically relevant for the treatment of human disease. Lack of standardization may limit the potential of high-throughput drug screening studies for discovering novel drug sensitivities. Although gene-expression and genome sequencing data may be reliable measures of the individual cell lines being screened, the lack of a robust phenotypic anchor for developing and validating predictive models can limit the use of data from HTS studies. A community-based effort, similarly to that undertaken by MGED, and an investment of resources to support the development of common standards to facilitate comparisons, would pay a huge dividend in establishing HTS as a valuable tool in drug discovery and could enable a host of robust and reproducible clinical and translational applications.

Community-wide Assessment for Standardization

Given the complexity of HTS assays, a community-wide consortium effort with multiple stakeholders from industry, academia, and government working together may be needed to reach consensus on best practices and strategies for objective quality assessment and intersite reproducibility of drug-sensitivity data. HTS assays are complex, multistep processes with many options; however, the respective advantages, limitations, and intersite reproducibility of the alternative protocols have not been adequately evaluated. Noise introduced at any step could seriously limit the robustness of an entire experiment. For the objective assessment of quality control of cell-based HTS assays, a community-wide consortium effort (“HTSQC”) similar to the MAQC-I (37) could be invaluable.

We offer some initial ideas for further consideration. A common panel of cancer cell lines whose identity would be verified genomically, and a collection of standardized anti-cancer compounds could be used as common “reference materials” for assessing the performance of each participating HTS platform and testing laboratory, providing a common ground to support meaningful comparisons. The reference materials (cell lines and compounds) could be maintained under strict quality control and could be used with defined protocols at a centralized location before being distributed to testing laboratories. Specifying defined cultured media and conditions, number of passages, seeding densities, and duration of assays could help establish common assay protocols. Similarly, reagents required for a given HTS platform could be kept as consistent as possible among the testing laboratories.

The same HTS platform could be tested in at least three independent laboratories using the same reference materials and the same lots of reagents with multiple replicates in the same laboratory, allowing both inter- and intra-laboratory reproducibility, and cross-platform comparability to be assessed. All study protocols and experimental details would be recorded and submitted along with raw experimental data to designated data analysis teams for independent analysis, allowing variability among data analysis methods to be assessed.

Finally, members of the consortium should share and discuss the results transparently, and publish the results in a peer-reviewed journal as a consortium consensus. Raw data and experimental details should be deposited in a public repository such as PubChem, to allow the community to replicate, further evaluate, and improve HTS assays and data analysis approaches. Such a consortium effort would provide an estimate of the overall reproducibility of current HTS assays, develop best practices, and provide an upper limit of the robustness of measured drug response phenotypes. If executed properly, this project could provide critical information for industry, academia, and regulatory agencies to objectively assess the implications of the limits of consistency of HTS assay results in drug discovery and development.

Conclusion

Although our evaluation suggests caution when considering the data generated from large-scale pharmacogenomic HTS studies, our results do not undermine the value of such studies. Indeed, large-scale pharmacogenomic HTS studies hold great potential for generating robust genomic predictors of drug response. However, this potential will not be fully realized until

improvements are made in the design, application, and community-wide implementation of robust, standardized assays for measuring drug response phenotypes. Efforts should be made to promote transparency and availability of raw assay data for HTS studies. These data will be critical to identify the sources and magnitudes of experimental uncertainties, and for developing methods to reduce their impact. One proven path forward is the establishment of best practices and standardization of the critical steps in these assays through a collective effort. Such an approach could ensure that the data produced from large-scale efforts can be replicated over a broad range of conditions and compared successfully across multiple laboratories.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors would like to thank the investigators of the Cancer Genome Project, the Cancer Cell Line Encyclopedia, and the GlaxoSmithKline cell line study who have made their invaluable data available to the scientific community.

Grant Support

C. Hatzis was supported by Yale University. P.L. Bedard was supported by a grant from the United States National Institute of Health (UMI CA186644). N. Juul Birkbak was funded by The Villum Kann Rasmussen Foundation. A.H. Beck was supported by an award from the Klarman Family Foundation and by support from US National Institutes of Health (NIH) grant CA087969. D.F. Stern was supported by the NIH National Cancer Institute (NCI) under award number R01CA80065. R. Clarke was supported by Public Health Service grants (U54 CA149147 and R01 CA131465) from the NIH. J. Quackenbush was supported by grants from the NCI GAME-ON Cancer Post-GWAS initiative (5U19 CA148065) and the NHLBI (5R01HL111759). B. Haibe-Kains was supported by the Gattuso Slight Personalized Cancer Medicine Fund at Princess Margaret Cancer Centre.

Received March 10, 2014; revised April 29, 2014; accepted April 30, 2014; published OnlineFirst July 11, 2014.

References

- Bredel M, Jacoby E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat Rev Genet* 2004;5:262–75.
- Drews J. Drug discovery: a historical perspective. *Science* 2000;287:1960–4.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. *Nature* 2012;483:603–7.
- Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;483:570–5.
- Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ, et al. Inconsistency in large pharmacogenomic studies. *Nature* 2013;504:389–93.
- Jang IS, Neto EC, Guinney J, Friend SH, Margolin AA. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac Symp Biocomput* 2014;19:63–74.
- Pharmaceutical Research and Manufacturers of America. More than 900 medicines and vaccines in clinical testing offer new hope in the fight against cancer. *Medicines in development for cancer 2012*. Available from: <http://www.phrma.org/research/new-medicines>.
- Tan DS, Thomas GV, Garrett MD, Banerji U, de Bono JS, Kaye SB, et al. Biomarker-driven early clinical trials in oncology: a paradigm shift in drug development. *Cancer J* 2009;15:406–20.
- Yap TA, Sandhu SK, Workman P, de Bono JS. Envisioning the future of early anticancer drug development. *Nat Rev Cancer* 2010;10:514–23.
- Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* 2010;363:809–19.
- Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small cell lung cancer. *N Engl J Med* 2010;363:1693–703.
- Held MA, Langdon CG, Platt JT, Graham-Steed T, Liu Z, Chakraborty A, et al. Genotype-selective combination therapies for melanoma identified by high-throughput drug screening. *Cancer Discov* 2013;3:52–67.
- Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* 2011;10:188–95.
- Walters WP, Namchuk M. Designing screens: how to make your hits a hit. *Nat Rev Drug Discov* 2003;2:259–66.
- Melnick JS, Janes J, Kim S, Chang JY, Sipes DG, Gunderson D, et al. An efficient rapid system for profiling the cellular activities of molecular libraries. *Proc Natl Acad Sci U S A* 2006;103:3153–8.
- Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;24:167–75.
- Greshock J, Bachman KE, Degenhardt YY, Jing J, Wen YH, Eastman S, et al. Molecular target class is predictive of *in vitro* response profile. *Cancer Res* 2010;70:3677–86.
- Matson SL, Chatterjee M, Stock DA, Leet JE, Dumas EA, Ferrante CD, et al. Best practices in compound management for preserving

- compound integrity and accurately providing samples for assays. *J Biomol Screen* 2009;14:476–84.
19. McDonald GR, Hudson AL, Dunn SM, You H, Baker GB, Whittall RM, et al. Bioactive contaminants leach from disposable laboratory plasticware. *Science* 2008;322:917.
 20. Marx V. Pouring over liquid handling. *Nat Methods* 2014;11:33–8.
 21. Ekins S, Olechno J, Williams AJ. Dispensing processes impact apparent biological activity as determined by computational and statistical analyses. *PLoS ONE* 2013;8:e62325.
 22. Slater K. Cytotoxicity tests for high-throughput drug discovery. *Curr Opin Biotechnol* 2001;12:70–4.
 23. Niles AL, Moravec RA, Riss TL. *In vitro* viability and cytotoxicity testing and same-well multi-parametric combinations for high-throughput screening. *Curr Chem Genomics* 2009;3:33–41.
 24. Crouch SP, Kozlowski R, Slater KJ, Fletcher J. The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *J Immunol Methods* 1993;160:81–8.
 25. Chan GK, Kleinheinz TL, Peterson D, Moffat JG. A simple high-content cell-cycle assay reveals frequent discrepancies between cell number and ATP and MTS proliferation assays. *PLoS ONE* 2013;8:e63583.
 26. 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;9:708–14.
 27. Chou T. Relationships between inhibition constants and fractional inhibition in enzyme-catalyzed reactions with different numbers of reactants, different reaction mechanisms, and different types and mechanisms of inhibition. *Mol Pharmacol* 1974;10:235–47.
 28. Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, et al. Standards for microarray data. *Science* 2002;298:539.
 29. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.
 30. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* 2001;2:RESEARCH0032.
 31. Bolstad B, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19:185–193.
 32. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;11:733–9.
 33. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–27.
 34. Prat A, Parker JS, Fan C, Perou CM. PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. *Breast Cancer Res Treat* 2012;135:301–6.
 35. Haibe-Kains B, Desmedt C, Loi S, Culhane AC, Bontempi G, Quackenbush J, et al. A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst* 2012;104:311–25.
 36. Beck AH, Knoblauch NW, Hefti MM, Kaplan J, Schnitt SJ, Culhane AC, et al. Significance analysis of prognostic signatures. *PLoS Comput Biol* 2013;9:e1002875.
 37. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, et al. The MicroArray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;24:1151–61.
 38. Pemovska T, Kontro M, Yadav B, Edgren H, Eldfors S, Szewajada A, et al. Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov* 2013;12:1416–29.
 39. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;10:813–23.
 40. Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect* 2008;3:284–91.
 41. Heiser LM, Wang NJ, Talcott CL, Laderoute KR, Knapp M, Guan Y, et al. Integrated analysis of breast cancer cell lines reveals unique signaling pathways. *Genome Biol* 2009;3:R31.

Correction: Enhancing Reproducibility in Cancer Drug Screening: How Do We Move Forward?

In this article (Cancer Res 2014;74:4016–23), which appeared in the August 1, 2014, issue of *Cancer Research* (1), an author's name was misspelled. The correct author listing is below. The publisher regrets this error.

The online version has been modified and no longer matches the print.

Christos Hatzis, Philippe L. Bedard, Nicolai J. Birkbak, Andrew H. Beck, Hugo J.W.L. Aerts, David F. Stern, Leming Shi, Robert Clarke, John Quackenbush, Benjamin Haibe-Kains.

Reference

1. Hatzis C, Bedard PL, Birkbak NJ, Beck AH, Aerts HJWL, Stern DF, et al. Enhancing reproducibility in cancer drug screening: how do we move forward? *Cancer Res* 2014;74:4016–23.

Published OnlineFirst August 29, 2014.

doi: 10.1158/0008-5472.CAN-14-2275

©2014 American Association for Cancer Research.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Enhancing Reproducibility in Cancer Drug Screening: How Do We Move Forward?

Christos Hatzis, Philippe L. Bedard, Nicolai J. Birkbak, et al.

Cancer Res 2014;74:4016-4023. Published OnlineFirst July 11, 2014.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-14-0725
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2014/07/23/0008-5472.CAN-14-0725.DC1

Cited articles	This article cites 40 articles, 6 of which you can access for free at: http://cancerres.aacrjournals.org/content/74/15/4016.full#ref-list-1
Citing articles	This article has been cited by 24 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/74/15/4016.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/74/15/4016 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.