**Comparison of experimental protocols used in CGP and CCLE**

A major potential source of variability in phenotypic measurements between CGP and CCLE phenotype is due to differences in procedures used for growing cells, storing compounds, treating cells with drugs, measuring cell viability, and assessing assay reproducibility (Supplementary Table 1). Based on the data provided in these two studies, it is not possible to determine which experimental procedure (CCLE or GCP) provides more accurate estimates of chemo-sensitivity, as there is no “gold-standard” set of phenotype measurements to use for comparison and benchmarking. Several published studies and reviews have evaluated relative strengths and weaknesses of experimental approaches for assessing drug sensitivity. Most of the literature has been focused on assays for measuring cell viability with relatively little published data on systematic comparisons of the methods for earlier steps in the protocols (e.g., media for growing cells, methods for storing compounds, procedures for treating cells) [1-18]. Each of these components may influence drug sensitivity results, and it would be ideal to standardize these steps, where possible.

We describe below some of the main differences between the CCLE and CGP experimental protocols.

Detection assay

The CCLE study used the CellTiter Glo cell viability assay that measures ATP using firefly luciferase activity [11].  Cells that loose membrane integrity also loose the ability to synthesize ATP and endogenous ATPases quickly deplete any APT remaining in the cytoplasm.  However non-lethal perturbations leading to reduced proliferation or inhibited mitochondrial respiration can reduce intracellular ATP and therefore ATP does not always correlate with cell viability.  Such metabolic interference could produce false positive results [16].

The CGP study used two different cell viability assays: 1) a fluorescent nucleic acid stain (Syto60) for adherent cell lines and 2) a redox indicator dye (resazurin) which is reduced to a fluorescent form by NADH in metabolically active viable cells for suspension cell lines.

The Syto60 assay determines the relative cell viability as the ratio of fluorescence from the treatment wells to the average fluorescence from the positive control wells (vehicle) on a plate.  Lower fluorescence ratio corresponds to greater inhibition.  The assay requires fixation and aspiration/washing steps to reduce non-specific fluorescence, and these steps can introduce considerable variation.  Furthermore, dead cells that are still intact will contribute a viable signal.

The resazurin assay is applied directly to the cell culture and thus avoids fixation/washing steps that could introduce additional variation.  However resazurin itself can have a direct cytotoxic effect and has been shown to reduce intracellular ATP levels [11].  Furthermore, the resazurin assay requires incubation of the substrate with viable cells at 37C for at least 1 hr, which can also increase the possibility of artifacts resulting from the chemical interaction of the assay chemistry with the compounds tested and with the biochemistry of the cell.

The manufacturer of a commercial resazurin kit (CellTiter Blue) has reported relatively good concordance and "similar" IC50 levels with the CellTiter Glo ATP assay (Promega Technical Bulletin no. 317).  However, we could reasonably expect low concordance between relative nucleic acid level and cellular metabolic activity.

We could not find any literature directly comparing performance of the Syto60 cell-permeant red fluorescent nucleic acid stain and intracellular ATP-based assays for estimating drug sensitivity; however, it is clear that these assays are measuring different aspects of the drug response phenotype, and therefore it is not surprising that the assays show only moderate correlation in the CGP/CCLE analysis. It has been suggested that multi-parameter testing, incorporating multiple, complementary cell-viability assays yields the most robust and informative phenotypic measures [7, 8].

Quality assurance of screening plates

One of the biggest challenges in HTS is to identify and subsequently to control HTS-generated data for systematic biases that can arise from mechanical causes (liquid handling, clogged transfer tips, temperature gradients), biological causes (faulty batch of cells, cell clumping, overestimation/underestimation of stock culture cell concentration, non-adherence on well) or experimental errors (treatment assignment to wrong well).  Anomalies can be isolated affecting individual wells or parts of plates (edge, region effects), or more pervasive affecting whole plates or entire runs.  Effective quality evaluation of assays helps reduce false positive and false negative hits that might results from systematic biases.

The CCLE study employed an ultra-high throughput screening system built by the Genomics Institute of the Novartis Research Foundation using 1536 well plates.  Details on plate layout or the number of positive and negative controls per plate were not provided.  Also details on procedures for evaluating the quality of plates were not provided.  The original publication where this system was used did not provide any QC details either [19].

The CGP study reported that all screening plates were subjected to stringent quality control using the Z'-factor to compare positive and negative control and their variation (median = 0.70, lower quartile = 0.46).  With the standard minimum quality cutoff being 0.5 [20], 25% of the plates in this study might have been of suboptimal quality.  Details of plate layout for the 96 or 384-well plates were not given and thus we cannot assess whether edge effects could be detected.

Controls

An additional area of protocol development and standardization that would likely aid in obtaining robust estimates of chemo-sensitivity would be a more thorough use of controls. CGP used the proteasome inhibitor MG132 as a control, as MG132 is known to be extremely cytotoxic. CCLE used drug-free positive controls and cell-free negative controls. While these controls may establish a bare minimum level of assay function, they are likely inadequate for ensuring accurate quantitative cell viability measurements. Development and distribution of a library of high-quality benchmarked drug-cell line “control” pairs and associated measurements, ranging from highly sensitive to highly resistant, would likely be useful for ensuring adequate assay function and for estimating accuracy and variability of measurements, compared with a “gold-standard” set of measurements. Similarly, more systematic use of technical replicates and reporting of raw data values would facilitate statistical estimates of assay reproducibility, which would enable modeling of experimental reproducibility in downstream analyses.

Measures of Drug Potency (IC50)

The two studies used different models to describe the dose-response relationship of the various drugs tested and different methods to estimate their parameters and the half-maximum inhibitory concentration (IC50).  The CCLE study used three different models depending on the statistical quality of the fits of duplicate assay data generated for each cell line on a single day: a 4-parameter logistic model (4PL), a constant model, or a non-parametric spline model.  For inactive compounds, the maximum tested concentration was used as the default IC50.  No details were provided on how the 4PL model was fit and therefore we assume standard nonlinear least squares fitting was used.

The CGP study used a 5-parameter logistic model (5PL) to describe the dose-response relationship under a Bayesian framework.  The IC50 was estimated using MCMC with a normal prior.  Non-informative priors were used for the other parameters.  This procedure provides uncertainty estimates for the IC50 from its associated posterior.

The low concordance in IC50 values between the two studies is not surprising given that different dose-response models and estimation procedures were used in the two studies.

Same comments apply to the AUC as this measure is also derived from the dose-response curves.  Furthermore, neither study specified whether the AUC of the dose-response was calculated on a log-concentration scale or on a linear scale.

**References**

1. Andreotti, P.E., et al., *Chemosensitivity testing of human tumors using a microplate adenosine triphosphate luminescence assay: clinical correlation for cisplatin resistance of ovarian carcinoma.* Cancer Res, 1995. **55**(22): p. 5276-82.

2. Bellamy, W.T., *Prediction of response to drug therapy of cancer. A review of in vitro assays.* Drugs, 1992. **44**(5): p. 690-708.

3. Burstein, H.J., et al., *American Society of Clinical Oncology clinical practice guideline update on the use of chemotherapy sensitivity and resistance assays.* J Clin Oncol, 2011. **29**(24): p. 3328-30.

4. Chan, G.K., et al., *A simple high-content cell cycle assay reveals frequent discrepancies between cell number and ATP and MTS proliferation assays.* PLoS One, 2013. **8**(5): p. e63583.

5. Cree, I.A., *Chemosensitivity and chemoresistance testing in ovarian cancer.* Curr Opin Obstet Gynecol, 2009. **21**(1): p. 39-43.

6. Miret, S., E.M. De Groene, and W. Klaffke, *Comparison of in vitro assays of cellular toxicity in the human hepatic cell line HepG2.* J Biomol Screen, 2006. **11**(2): p. 184-93.

7. Niles, A.L., R.A. Moravec, and T.L. Riss, *In vitro viability and cytotoxicity testing and same-well multi-parametric combinations for high throughput screening.* Curr Chem Genomics, 2009. **3**: p. 33-41.

8. Niles, A.L., R.A. Moravec, and T.L. Riss, *Update on in vitro cytotoxicity assays for drug development.* Expert Opin Drug Discov, 2008. **3**(6): p. 655-69.

9. Petty, R.D., et al., *Comparison of MTT and ATP-based assays for the measurement of viable cell number.* J Biolumin Chemilumin, 1995. **10**(1): p. 29-34.

10. Riss, T.L. and R.A. Moravec, *Use of multiple assay endpoints to investigate the effects of incubation time, dose of toxin, and plating density in cell-based cytotoxicity assays.* Assay Drug Dev Technol, 2004. **2**(1): p. 51-62.

11. Riss, T.L., R.A. Moravec, and A.L. Niles, *Cytotoxicity testing: measuring viable cells, dead cells, and detecting mechanism of cell death.* Methods Mol Biol, 2011. **740**: p. 103-14.

12. Sumantran, V.N., *Cellular chemosensitivity assays: an overview.* Methods Mol Biol, 2011. **731**: p. 219-36.

13. Weyermann, J., D. Lochmann, and A. Zimmer, *A practical note on the use of cytotoxicity assays.* Int J Pharm, 2005. **288**(2): p. 369-76.

14. Wlodkowic, D., et al., *Apoptosis and beyond: cytometry in studies of programmed cell death.* Methods Cell Biol, 2011. **103**: p. 55-98.

15. Wlodkowic, D., J. Skommer, and Z. Darzynkiewicz, *Cytometry of apoptosis. Historical perspective and new advances.* Exp Oncol, 2012. **34**(3): p. 255-62.

16. Kepp, O., et al., *Cell death assays for drug discovery.* Nat Rev Drug Discov, 2011. **10**(3): p. 221-37.

17. Klionsky, D.J., et al., *Guidelines for the use and interpretation of assays for monitoring autophagy.* Autophagy, 2012. **8**(4): p. 445-544.

18. Tolliday, N., *High-throughput assessment of Mammalian cell viability by determination of adenosine triphosphate levels.* Curr Protoc Chem Biol, 2010. **2**(3): p. 153-61.

19. Melnick, J.S., et al., *An efficient rapid system for profiling the cellular activities of molecular libraries.* Proc Natl Acad Sci U S A, 2006. **103**(9): p. 3153-8.

20. Zhang, J.H., T.D. Chung, and K.R. Oldenburg, *A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays.* J Biomol Screen, 1999. **4**(2): p. 67-73.