

Consistency in large pharmacogenomic studies

Paul Geeleher¹, Eric R. Gamazon¹, Hae Kyung Im², Cathal Seoighe³, Nancy J. Cox^{1*}, and R.

Stephanie Huang^{1*}

¹Department of Medicine, ²Department of Health Studies, The University of Chicago, Chicago, IL 60637, USA; and ³Department of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland

* Address for correspondence and reprints:

Nancy J. Cox, 900 E 57th street, KCBD room 3100, the University of Chicago, Chicago, IL 60637.
Telephone: (773) 702-3001; Fax: (773) 702-3000; E-mail: ncox@medicinebsd.uchicago.edu

R. Stephanie Huang, 900 E 57th street, KCBD room 7148, the University of Chicago, Chicago, IL 60637. Telephone: (773) 702-9363; Fax: (773) 702-0963; E-mail:
rhuang@medicinebsd.uchicago.edu

Haibe-Kains *et al.*¹ reported inconsistencies between two large-scale pharmacogenomic studies (the Cancer Cell Line Encyclopedia² (CCLE) and the Cancer Genome Project³ (CGP)). Upon careful analysis of these same data, we would draw quite different and much more positive conclusions. We highlight here the most obvious reasons for this different interpretation.

1. On close inspection, the authors reported correlation *between* cell lines for gene expression but, inconsistently, *across* cell lines for drug sensitivity (see Methods). On re-analysis, we found higher correlations between cell lines than across cell lines for both expression profiles and drug sensitivity measures (median $r_s = 0.88$ between cell lines, $r_s = 0.56$ across cell lines for expression; median $r_s = 0.62$ between cell lines and $r_s = 0.35$ across cell lines for AUC, a drug sensitivity measure). Thus, by correcting this inconsistency, the correlations for expression and drug sensitivity data were far more similar than was originally reported, severely undermining the authors' interpretation of the relative quality of expression and drug sensitivity data.
2. For gene-drug association analysis, the response data include severe outliers ($P < 3.9 \times 10^{-10}$ from Shapiro-Wilk test for all drug AUC in CGP) and the residuals in the linear regression models deviate dramatically from the normal distribution. Violation of the “normality assumption” causes unstable models⁴ rendering the concordance analysis premised on the fit of these models statistically invalid.
3. Lack of biological variability in either CGP or CCLE renders a correlation metric ineffective in the presence of any measurement error. This was often the case, because many of the drugs were highly-targeted agents requiring, by design, specific molecular targets for response. Consider nilotinib, which was not significantly correlated between CGP and CCLE ($r_s = 0.1$ for AUC in the original report). Nilotinib targets the BCR-ABL1 fusion-gene and in CGP, BCR-ABL1 status was predictive of drug sensitivity ($P = 2.54 \times 10^{-65}$), accurately reflecting known biology. BCR-ABL1 status was not reported in CCLE, however, two BCR-ABL1-positive cell lines were among the 188 nilotinib-treated cell lines that overlap CCLE where we also found them to be the two most sensitive samples ($P = 5.7 \times 10^{-5}$). Given that this canonical drug-specific biomarker was strongly associated with AUC in both studies, it seems contradictory that a correlation measure identifies nilotinib sensitivity as discordant. In fact, this highlights a key problem with the authors' approach. Of 577 cell lines screened in CGP (median AUC = 0.99; $\sigma^2 = 4 \times 10^{-3}$; Fig. 1(a)) only four were BCR-ABL1 positive. Clearly, given no expectation of cell death in almost all samples, as expected for a highly targeted-agent, there is little biological variability in drug response, resulting in low correlation. The fact that BCR-ABL1 status is in strong agreement with drug sensitivity in both studies further demonstrates that correlation is *not*, in this case, an appropriate concordance measure.

Given that most drugs compared were also targeted agents, this lack of variability was widely observed; indeed, for 9 of the 15 drugs, median AUC is greater than 0.95 in CGP. Valid and useful comparisons between CGP and CCLE must consider the pharmacology of the drugs screened (Fig. 1(b)).

As proof-of-concept, we leveraged all data in both studies to test whether findings informed by one study (source) can improve discovery of novel gene-drug associations in the other study (target). The false discovery rate (FDR) in the target-study was markedly improved by filtering based on increasingly strict significance thresholds in the source-study^{5–8} (Fig. 1(c)). The improvement in FDR demonstrates the power of combining large-scale pharmacogenomic studies for biological discovery. Finally, recent findings that expression and drug sensitivity data in CGP could reliably predict drug response in multiple clinical trials further support pharmacogenomic data integration⁹.

Methods

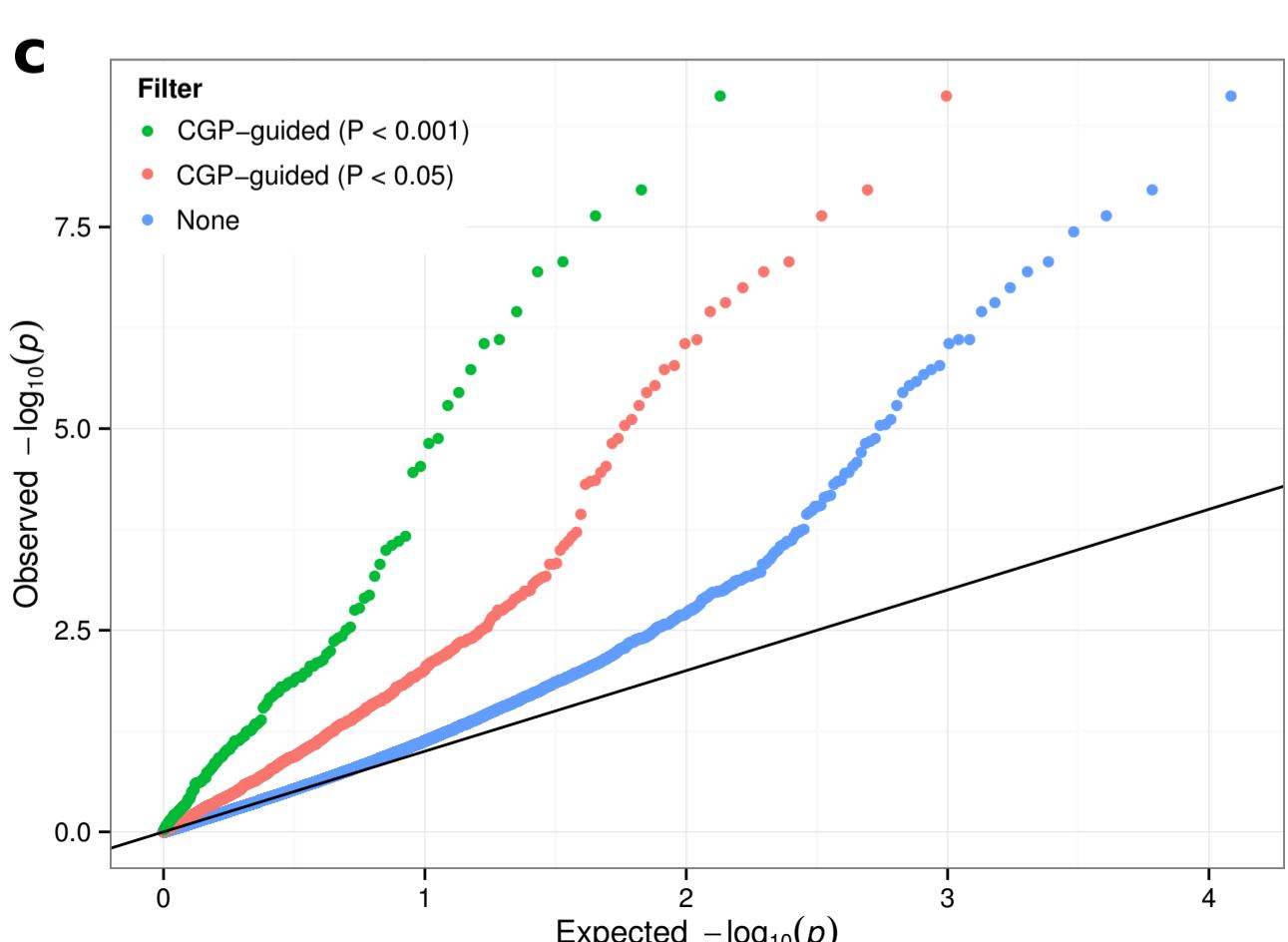
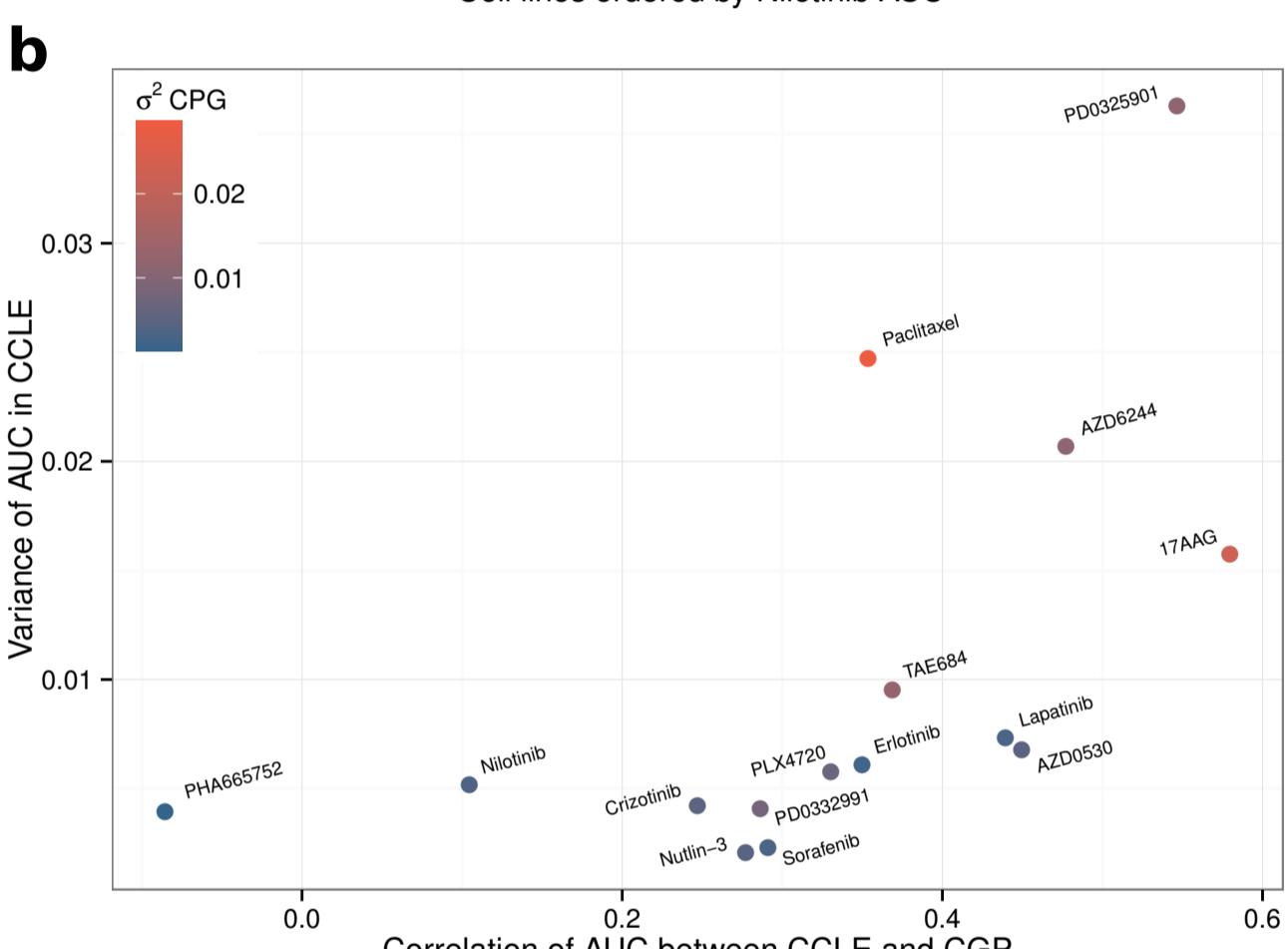
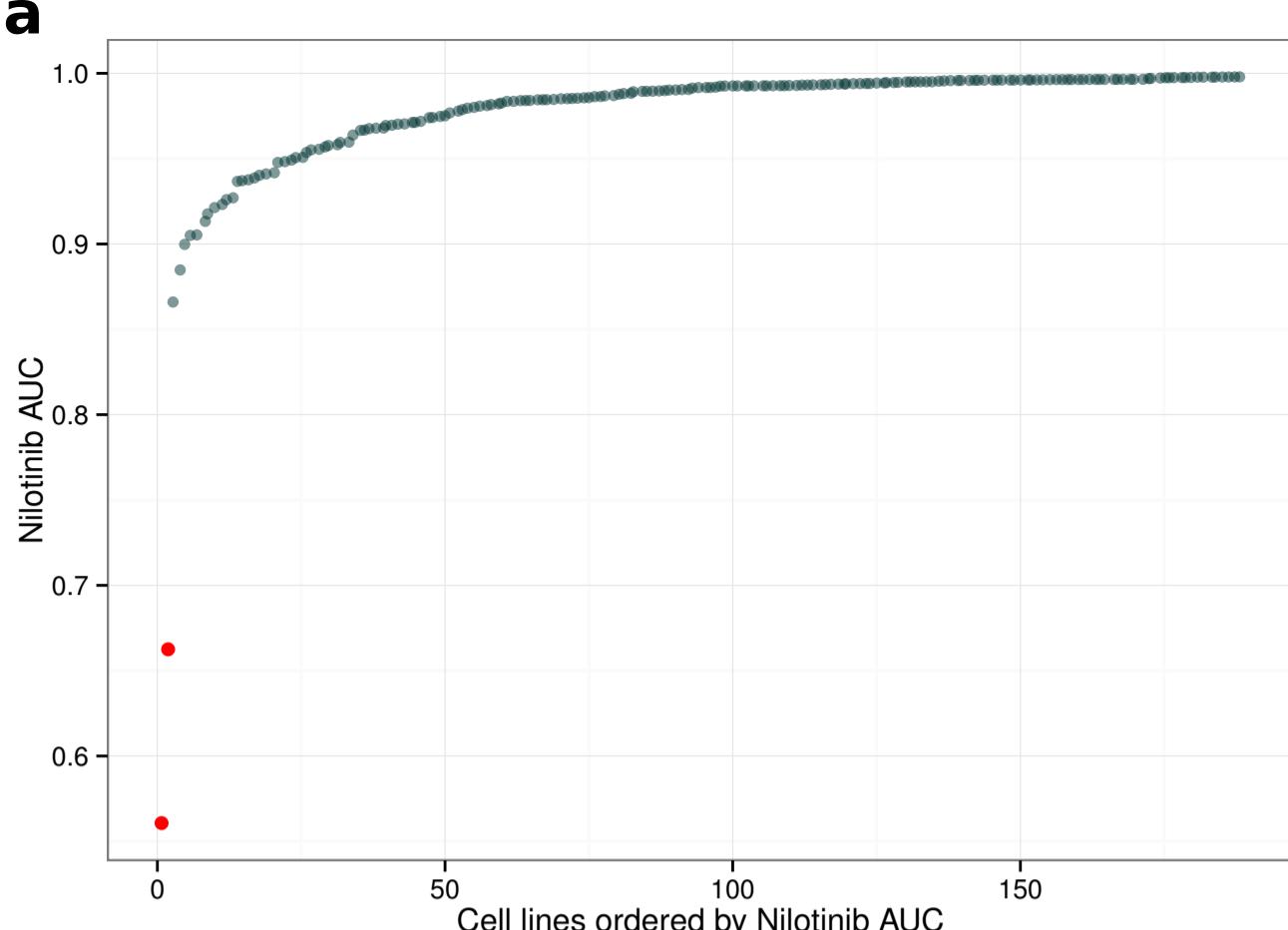
In CGP and CCLE, using ordered data common to both studies, gene expression and drug sensitivity (AUC) values can be arranged in $n_1 \times m$ and $n_2 \times m$ matrices respectively where m is the number of cell lines, n_1 is the number of genes and n_2 is the number of drugs common to both studies. Correlations “between” cell lines are calculated by the correlation of matching columns of CGP and CCLE matrices (vectors of length n_1 for expression or n_2 for AUC). Correlations “across” cell lines are the correlations of matching rows (vectors of length m for both data).

Figure Legends

- (a) Highly targeted agents (nilotinib) highlight a major limitation of the authors' test for concordance.** Scatterplot showing the nilotinib AUC values (in CGP) for the 188 cell lines that were screened by both CGP and CCLE. Only a very small proportion of cell lines achieve a response, e.g. the two BCR-ABL1 positive cell lines, which are highlighted in red.
- (b) The authors' test for concordance is confounded by variability in drug response.** Scatterplot showing the strong association between “Spearman’s correlation of AUC between CCLE and CGP” and “variance of AUC in CCLE”. Drugs whose AUC is more variable are more likely to be highly correlated between CCLE and CGP ($r_s = 0.83$, $P = 1.9 \times 10^{-4}$). The points have been color coded by their “variance of AUC in CGP”, which is also significantly associated with both “variance of AUC in CCLE” and “Spearman’s correlation of AUC between CCLE and CGP”.
- (c) Results from “source” study can improve discovery of novel gene-drug associations in “target” study.** For nilotinib, the false discovery rate (FDR) in the target-study (CCLE) was markedly improved by filtering based on increasingly strict significance thresholds in the source-study (CGP). The QQ-plots show the expected and observed p-values for the gene-drug Spearman correlations from the full CCLE dataset (blue), the CCLE Spearman correlations from those that meet $P < 0.05$ in CGP (red), and the CCLE correlations that satisfy $P < 0.001$ in CGP (green). As increasingly more significant thresholds in CGP are used to filter genes, the QQ-plot shifts leftward, indicating greater improvement in FDR^{5–8}. An earlier departure from the null line (shown as a more pronounced leftward shift in the QQ-plot) suggests a greater proportion of true associations, for a given nominal p-value. This implies that the fraction of non-null effects is dramatically greater with pre-selected top genes derived from the source study. Thus, an integrative analysis, rather than the authors’ problematic test of concordance, can substantially improve biological discovery in large-scale pharmacogenomic studies.

References:

1. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–93 (2013).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–7 (2012).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–5 (2012).
4. Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to linear regression analysis*. **821**, (John Wiley & Sons, 2012).
5. Li, C., Li, M., Lange, E. M. & Watanabe, R. M. Prioritized subset analysis: improving power in genome-wide association studies. *Hum. Hered.* **65**, 129–41 (2008).
6. Genovese, C. R., Roeder, K. & Wasserman, L. False discovery control with p-value weighting. *Biometrika* **93**, 509–524 (2006).
7. Schork, A. J. *et al.* All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* **9**, e1003449 (2013).
8. Sun, L., Craiu, R. V., Paterson, A. D. & Bull, S. B. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genet. Epidemiol.* **30**, 519–30 (2006).
9. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).



Response to “Consistency in large pharmacogenomic studies”

Benjamin Haibe-Kains, Nehme El-Hachem, Nicolai Juul Birkbak, Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush

We read the report by Geeleher *et al.* with great interest and we welcome the opportunity to respond. While we agree that there are anecdotal cases for some targeted agents, such as crizotinib and nilotinib, that demonstrate consistent identification of a few highly sensitive cell lines shared by the CGP [1] and CCLE [2], our initial conclusions about the broader inconsistency in reported phenotypes between the CGP and CCLE still holds [3].

1. In our initial publication, we computed the correlation between gene expression and mutation profiles of cell lines to assess whether large transcriptomic changes and/or genetic drift might be the cause of the observed inconsistency in drug sensitivity data, and subsequently the gene-drug associations [3]. We agree with the authors that correlations *across* and *between* cell lines should be thoroughly compared. Overall concordance across cell lines was lower than concordance between cell lines; however we still observed that gene expression and mutation data were significantly more concordant than IC₅₀ and AUC values in all comparisons (Wilcoxon rank sum test p<0.01), except for agreement of presence/absence of mutations and AUC sensitivity calls across cell lines (Cohen's kappa: 0.28 vs. 0.23 for mutations and AUC sensitivity calls, p=0.16). In light of these results, our conclusion that genomic data are significantly more concordant than pharmacological data still holds, although the difference under these assumptions is less pronounced than reported in our initial publication.

2. We agree with the authors that the normality assumption of linear regression is violated for

sensitivity data of highly targeted drugs. However our goal was to follow the approach used by the CGP and CCLE investigators to infer gene-drug associations (*elasticnet* linear regression). The authors of the current manuscript also used linear regression models (*ridge*) [4] to develop genomic predictors of response to cytotoxic drugs (such as paclitaxel in our publication), and presented an *ad hoc* logistic ridge regression model, using the 15 most sensitive vs. the 55 most resistant CGP cell lines, for targeted drugs. This suggests that discretizing AUC measurements into sensitivity calls would yield more robust gene-drug associations. However, we reported in our comparative study that the concordance of AUC sensitivity calls (computed using the Waterfall method [2]) is poor for all drugs [3]. Therefore, selecting extreme cases for model fitting is unlikely to result in robust predictors of drug response for the vast majority of cases.

3. We agree that lack of variability in drug sensitivity measurements may prevent biologically meaningful assessment of concordance between pharmacogenomic datasets. The authors focused their discussion on nilotinib for which there are three sensitive cell lines in both datasets. Even among these cell lines the AUC values are not concordant; the least sensitive cell line in CGP is actually the most sensitive one in CCLE. Therefore the only way to consider these results concordant is to classify these three cell lines as sensitive and the rest as resistant, which cannot easily be done using an the Waterfall approach [3] and as noted previously, limits sensitivity and robustness. We agree with the authors that the strong association between BCR-ABL1 gene fusion and sensitivity to nilotinib is reproduced in both studies. However this observation is anecdotal and is not generalizable to the other targeted and cytotoxic drugs investigated in our original report. For these drugs, there are multiple cell lines exhibiting high sensitivity in only one of the datasets. Moreover, for drugs with large variability in measured sensitivity, we observed only fair to poor correlation for these drugs ($r_s < 0.6$), which supports our report of inconsistencies between CGP and CCLE datasets [3].

Figure 1c. The authors attempted to leverage all data in both studies to test whether findings informed by one study can improve discovery of gene-drug associations in the other study. We are seriously concerned about multiple aspects of this analysis. First, the authors state that one should not use a correlation metric when there is lack of variability in the measurements, which is the case for nilotinib. However the authors appear to contradict themselves by using this metric to compute gene-drug associations. Second, the authors focused on nilotinib, for which the sensitivity data are somewhat reproducible between CGP and CCLE. We were not able to reproduce the authors' results; however we found that the Q-Q plot for nutlin3 is almost identical to Figure 1c, suggesting that drug names might have been switched, which cannot be confirmed as the author's software code was unavailable. Third, the Q-Q plot for nilotinib does not support the authors' conclusions. Indeed, we observed only a weak leftward shift compared to the unfiltered p-value distribution. Such a shift can easily be obtained by filtering genes based on variance to keep only the genes whose expression is part of the top 10% and top 1% most variant in the source dataset. Strikingly, we observed a stronger leftward shift for such a filtering process that is, by construction, fully independent from the drug sensitivity data. We made similar observations for 11 out of the common set of 15 drugs. These results demonstrate that the authors' analysis should be controlled for basic filtering procedures that are not based on drug sensitivity data. For the remaining 4 drugs, we observed a strong leftward shift for the filtered p-value distributions compared to the unfiltered distribution, suggesting that the source study is highly informative for the target one. We are not at all surprised that there is value in integrating the CGP and CCLE datasets; it is well known that increasing sample size is important in filtering signal from noise (if everything were perfectly consistent across the datasets, integrating the two datasets would not be necessary since they would contain the same information). In our original report, we found statistically significant non-zero correlations between phenotype measurements for almost all drugs, supporting the fact that there is relevant signal in these datasets; the main point of our study was that more work is

necessary to improve consistency of phenotypic measures to make these large-scale projects more useful for developing robust predictors of drug response.

References

1. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J *et al*: **Systematic identification of genomic markers of drug sensitivity in cancer cells.** *Nature* 2012, **483**(7391):570-575.
2. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D *et al*: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.** *Nature* 2012, **483**(7391):603-607.
3. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ, Quackenbush J: **Inconsistency in large pharmacogenomic studies.** *Nature* 2013, **504**(7480):389-393.
4. Geeleher P, Cox NJ, Huang RS: **Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines.** *Genome Biol* 2014, **15**(3):R47.

Extended Response to "Consistency in large pharmacogenomic studies"

Benjamin Halbe-Kains, Nehme El-Hachem, Nicolai Juul Birkbak, Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush

Contents

1	Extended Response	3
2	Supplementary Methods	7
2.1	Data retrieval and curation	7
2.1.1	CGP (release 4, March 2013)	7
2.1.2	CCLE (release March 2013)	7
2.2	Cell line annotations	7
2.3	Mutation data	8
2.4	Drug sensitivity calling	9
2.5	Concordance	10
2.5.1	Gene expression	10
2.5.2	Drug sensitivity	11
2.5.3	Mutation	11
2.5.4	Consistency between drug sensitivity calls	11
2.6	Gene-drug associations	11
3	Full Reproducibility of the Analysis Results	12
3.1	Set up the software environment	12
3.2	Run the R scripts	13
3.3	Additional parameters	14
3.4	Generate the Supplementary Information	14
4	List of Abbreviations	15
5	Supplementary Tables	16
6	Supplementary Figures	17

1 Extended Response

Point 1: In our original publication [?], we reported a poor correlation in drug response phenotypes in the CGP [2] and CCLE [1] studies. We computed the correlation between gene expression and mutation profiles of cell lines to assess whether large transcriptomic changes and/or genetic drift of the investigated cell lines might be the cause of the observed inconsistency in drug sensitivity data, and subsequently the gene-drug associations [?]. We agree with the authors that correlations *across* and *between* cell lines should be thoroughly compared. Therefore we extended our initial curation of the CGP and CCLE datasets to increase the number of cell lines overlapping between the two datasets (504 cancer cell lines, spanning 24 tissue types; Supplementary Methods; Supplementary Figure 1) and computed across and between cell lines concordance for gene expression and mutation data, IC₅₀ and AUC measurements and sensitivity calls (Supplementary Methods). Overall concordance across cell lines was lower than concordance between cell lines (Supplementary Figure 2); however we still observed than concordance for genomic data (gene expression and mutation) was statistically significantly higher than for pharmacological data (IC₅₀ and AUC) in all comparisons (Wilcoxon rank sum test $p < 0.01$), except for agreement of presence/absence of mutations and AUC across cell lines (Cohen's κ : 0.28 vs. 0.23 for mutations and AUC sensitivity calls, respectively; Wilcoxon rank sum test $p = 0.16$). In light of these results, our conclusion that genomic data are significantly more concordant than pharmacological data still holds, although the difference under these assumptions is less pronounced than reported in our initial publication.

Point 2: We agree with the authors that the normality assumption of linear regression is violated for sensitivity data of highly targeted drugs. However our goal was to follow the approach used by the authors of the CGP and CCLE studies to infer gene-drug associations from their pharmacogenomic data (linear regression via *elasticnet*). In their recent publication [3], the authors of the present manuscript used linear regression models (*ridge*) to develop genomic predictors of response to cytotoxic drugs (such as paclitaxel in our comparative study; Supplementary Table 1), and presented an *ad hoc* logistic regression model (logistic ridge regression models for the 15 most sensitive vs. the 55 most resistant CGP cell lines) for targeted agents. This suggests that discretizing AUC measurements into sensitivity calls would yield more robust gene-drug associations. However, we reported in our comparative study that the concordance of AUC sensitivity calls (computed using the Waterfall method introduced in [1]) is poor for all drugs. To test whether focusing on the extreme cases (highly sensitive and resistant cell lines) would improve the consistency of gene-drug associations between CGP and CCLE, we modified the Waterfall approach for extreme sensitivity calling (Supplementary Methods) to discard all cell lines with intermediate phenotypes and identify the highly sensitive cell lines for targeted drugs. In this setting the consistency of sensitivity calling remains poor (PD0325901 and 17AAG yielded the highest Cohen's κ coefficients with $\kappa = 0.26$; Supplementary Figure 3). When binary sensitive calls – highly sensitive vs. highly resistant, therefore ignoring the cell lines with intermediate sensitivity – are used to infer gene-drug associations using logistic regression (similar to the method used by the authors in their recent study; [3]), the consistency was still globally poor (Supplementary Figures 4 and 5 for gene-drug associations inferred from the shared set of 504 cell lines and all cell lines in CGP and CCLE datasets). Such inconsistency is probably due to the small sample size (few extreme cases) and the large proportion of cell lines identify as intermediate in one study but resistant or sensitive in the other study (Supplementary Figure 3). The consistency and significance of the gene-drug associations inferred from all available AUC measurements was much higher (Supplementary Figures 6 and 7 for gene-drug associations inferred from the shared set of 504 cell lines and all cell lines in CGP and CCLE datasets), which suggests that, despite the violation of the normality assumption of the linear regression analysis, more consistent results can be obtained by using the full drug sensitivity data.

Point 3: We agree that lack of variability in drug sensitivity measurements may prevent biologically meaningful assessment of concordance between pharmacogenomic datasets. Indeed, if

the vast majority of cell lines are resistant to a given drug, the concordance, as estimated using Spearman's correlation, will be low even if the noise present in the two datasets is well controlled (but still present). This is the case for highly targeted drugs, such as nilotinib, for which only a few cell lines are expected to be sensitive, while cytotoxic drugs exhibit more universal effects (Supplementary Figure 8). We compared the distribution of variability, estimated by the median absolute deviation (MAD), for targeted and cytotoxic drugs in CGP and CCLE (Supplementary Figure 9). As expected the variability of drug sensitivity data for cytotoxic drugs is significantly greater than for targeted drugs; a MAD cutoff of 0.10 enables to roughly discriminate between drugs with highly targeted or broader effect (Supplementary Figures 9 and 10). Based on this cutoff, paclitaxel, 17AAG and PD0325901 are classified as broad effect drugs in both datasets; similar observation for TAE684 and PD0332991 in CGP, and AZD6244 in CCLE. Highly targeted drugs include erlotinib, lapatinib, PHA665752, nilotinib, nutlin3, PLAX4720, and sorafenib (MAD of AUC < 0.10; Supplementary Figure 10). We concur with the authors that highly targeted drugs tend to yield lower spearman's correlation coefficients, likely due to the lack of variability of the corresponding sensitivity data (Supplementary Figure 11).

The authors focused their discussion on nilotinib. Given the prior knowledge that BCR-ABL1 fusion is the primary target of nilotinib and given that this gene fusion is rare in cancer cell lines, only few sensitive cell lines are expected. There are three such cell lines that have been screened with nilotinib both in CGP and CCLE:

Cell line	AUC in CGP	AUC in CCLE
EM-2	0.37	0.76
KU812	0.64	0.65
MEG-01	0.44	0.58

Among the three sensitive cell lines the AUC values are not quite concordant as the least sensitive cell line in CGP is actually the most sensitive one in CCLE. Therefore the only way to consider these results concordant is to classify these three cell lines as sensitive and the rest as resistant, yielding a perfect agreement between the two datasets. However, the use of an automated method for drug sensitivity calling, such as the waterfall approach introduced in the CCLE study, do not allow to reach high level of concordance (Supplementary Figure 3).

Of note there are other BCR-ABL1 positive cell lines screened in only one of the datasets and they do not systematically yield high sensitivity in CGP; see the table below.

Cell line	AUC in CGP	AUC in CCLE
K-562	0.10	NA
LAMA-84	0.67	NA
BV-173	0.30	NA
KCL-22	NA	0.70

It is worth pointing out that nilotinib is not a representative example of all targeted drugs investigated in CCLE and CGP. As can be seen in Supplementary Figure 1, there are several cell lines for erlotinib, PHA665752, AZD0530, PLX4720, and nutlin3 that exhibit high sensitivity in only one of the datasets. The authors state that given that most drugs investigated in our comparative study are targeted (Supplementary Table 1), such lack of variability in drug sensitivity data was widely observed. However, in addition to paclitaxel (the only cytotoxic drug in our study), we observed that several targeted drugs exhibited broad effect in the set of 504 cell lines screened both in CGP and CCLE (17AAG and PD0325901, as well as TAE684, PD0332991 and AZD6244, yielded MAD of AUC values > 0.10 in at least one dataset). For these drugs, one could argue that the variability of drug sensitivity measures warrants the use of a correlation metric for assessing the concordance between CGP and CCLE studies. However we observed only fair to poor correlation for these drugs ($r_s < 0.6$, Supplementary Figure 1). We agree with the authors that the strong association between BCR-ABL1 gene fusion and sensitivity to nilotinib is reproduced in both studies. However this observation is anecdotal and is not generalizable to the

other targeted and cytotoxic drugs investigated in our original report. Further, much of the motivation for the CGP and CCLE studies was the creation of large datasets through the screening of hundreds of cell lines that could be used to develop predictive models of drug response; if such studies are only useful in assessing targeted agents, it undermines the value of such large-scale pharmacogenomic studies.

Figure c: The authors attempted to leverage all data in both studies to test whether findings informed by one study (source) can improve discovery of gene-drug associations in the other study (target). They claimed that "*The false discovery rate (FDR) in the target-study was markedly improved by filtering based on increasingly strict significance thresholds in the source-study (Fig. 1(c))*" We are seriously concerned about three aspects of the authors' analysis. First, the authors did not take into account the *direction* of the gene-drug associations inferred from CGP and CCLE, in other words whether high expression of a given gene is associated with drug sensitivity or *vice versa*. As reported in Supplementary Figures 12-26, the proportion of filtered gene-drug associations with concordant direction could be as low as 15% for PH665752 and > 90% for PD0325901, AZD6244 and nutlin3. This aspect of the results should be considered because a gene-drug association significant in both datasets with opposite direction/sign cannot be considered as concordant. Second, the authors focused on a single drug, nilotinib, for which the sensitivity data are somewhat reproducible between CGP and CCLE. We were not able to reproduce the authors' results as illustrated in Supplementary Figure 12; however we found that the Q-Q plot for nutlin3 in Supplementary Figure 13A is almost identical to Figure 1c in the authors' report, suggesting that the drug names might have been inadvertently switched. Second, the Q-Q plot for nilotinib does not support the authors' conclusions. Indeed, we observed only a weak leftward shift compared to the unfiltered p-value distribution (Supplementary Figure 12A,B for associations filtered using CGP and CCLE data, respectively). Such a shift can easily be obtained by filtering genes based on other criteria than one based on drug sensitivity data. As an example we filtered the genes based on variance to keep only the genes whose expression is part of the top 10% and top 1% most variant in the source dataset; strikingly, we observed a stronger leftward shift for such a filtering process that is, by construction, fully independent from the drug sensitivity data (Supplementary Figure 12C,D for associations filtered using CGP and CCLE data, respectively). We made similar observations for most of the drugs (17AAG, AZD0530, crizotinib, erlotinib, lapatinib, paclitaxel, PD0332991, PHA665752, sorafenib, and TAE684; Supplementary Figures 14-23). These results demonstrate that the authors' analysis should be controlled for basic filtering procedures that are not based on drug sensitivity data. We also observed that in the case of nutlin3, AZD6244, PD0325901, and PLX4720, there is a strong leftward shift for the filtered p-value distributions compared to the unfiltered distribution, suggesting that the source study is highly informative for the target one for this small set of drugs. We are not at all surprised that there is value in integrating the CGP and CCLE datasets; it is well known that increasing sample size is important in filtering signal from noise (if everything were perfectly consistent across the datasets, integrating the two datasets would not be necessary since they would contain the same information). In our original report, we found statistically significant non-zero correlations between phenotype measurements for almost all drugs, supporting the fact that there is relevant signal in these datasets; the main point of our study was that more work is necessary to improve consistency of phenotypic measures to make these large-scale projects more useful for developing robust predictors of drug response.

Additional comments

Given that most drugs compared were also targeted agents, this lack of variability was widely observed; indeed, for 9 of the 15 drugs, median AUC is greater than 0.95 in CGP. Valid and useful comparisons between CGP and CCLE must consider the pharmacology of the drugs screened (Fig. 1(b)).

If one considers median AUC ≤ 0.05 (1-AUC value as reported in CGP) as the criterion to

define lack of variability in drug sensitivity, there are indeed 9 drugs in CGP that fulfill this criterion. However all the 15 drugs have a median AUC > 0.05 in CCLE (Supplementary Figure 8B), which contradicts the authors' claim. As shown in Supplementary Figures 9 and 10, a cutoff of MAD > 0.10 for AUC values is more relevant to discriminate between highly targeted drugs and those with broader effects. Applying this cutoff yields 10 and 11 highly targeted drugs in CGP and CCLE, respectively.

Lack of biological variability in either CGP or CCLE renders a correlation metric ineffective in the presence of any measurement error. This was often the case, because many of the drugs were highly-targeted agents requiring, by design, specific molecular targets for response. Consider nilotinib, which was not significantly correlated between CGP and CCLE ($rs = 0.1$ for AUC in the original report).

The authors state that one should not use a correlation metric when there is lack of variability in the measurements, which is apparently the case for nilotinib (median AUC of 0.008 and 0.07 for CGP and CCLE, respectively). However the authors appear to contradict themselves by using Spearman's correlation to compute gene-drug associations (Figure 1(c)). If there is not enough variability in the AUC values for nilotinib, it is not clear why the use of Spearman's correlation is appropriate in this case.

Finally, recent findings that expression and drug sensitivity data in CGP could reliably predict drug response in multiple clinical trials further support pharmacogenomic data integration.

In their recent paper, the authors did not assess the (in)consistency in large pharmacogenomic studies [3]; instead, they used CGP data to develop genomic predictors of response to three targeted and one cytotoxic drugs, and subsequently validated them in small sets of clinical samples. We suggest that, given our comparative analysis of CCLE and CGP, a more appropriate analysis would be a broad comparison to test the accuracy of their method in developing predictors on the shared 15 drugs and 471 cell lines in either the CGP or CCLE and testing them on the reported response measures in the other study. Alternatively, it may make sense to develop independent predictors using the CGP and CCLE datasets using data for the 15 drugs tested in both studies and then assess the concordance of these independent classifiers in predicting clinical response. On the one hand, if they produce highly accurate and concordant predictions of response, one could argue that the discrepancies we observed are not important. On the other hand, if the predictors are not performant in clinical studies, it would suggest that additional work is necessary to standardize phenotypic measurements before laboratory-based pharmacogenomic studies could be translated to clinical applications.

2 Supplementary Methods

2.1 Data retrieval and curation

2.1.1 CGP (release 4, March 2013)

Gene expression, mutation data and cell line annotations were downloaded from ArrayExpress. Drug sensitivity measurements and drug information were downloaded from the CGP website ([CGP](#)) and the Nature website ([Nature article](#)), respectively.

Minimum and maximum screening concentrations for each drug/cell line were extracted from `gdsc_compounds_conc_w2.csv` (μM).

The natural logarithm of IC_{50} measurements were retrieved from `gdsc_manova_input_w2.csv` in column " $*_{\text{IC_50}}$ " (referred to as x) and subsequently transformed using $-\log_{10}(\exp(x))$; high values are representative of cell line sensitivity to drugs.

The AUC measurements were retrieved from `gdsc_manova_input_w2.csv` in column " $*_{\text{AUC}}$ " (referred to as x); high values are representative of cell line sensitivity to drugs.

2.1.2 CCLE (release March 2013)

Gene expression, mutation data cell line annotations and drug information were downloaded from the CCLE website ([CCLE](#)). Drug sensitivity data were downloaded from the Nature website ([Nature letter addendum](#));

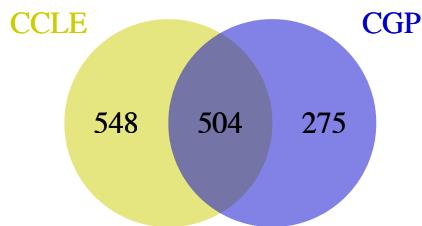
Screening concentrations for each drug/cell line were extracted from Supplementary Table 11 in column E (μM).

IC_{50} measurements were retrieved from Supplementary Table 11 in column J (" $\text{IC}_{50} \ \mu\text{M} \ (\text{norm})$ ") (referred to as x) and subsequently transformed into logarithmic scale, $-\log_{10}(x)$; high values are representative of cell line sensitivity to drugs.

AUC measurements were retrieved from Supplementary Table 11 in column L (" $\text{ActArea} \ (\text{norm})$ ") and subsequently divided by the number of drug concentrations tested (8); high values are representative of cell line sensitivity to drugs.

2.2 Cell line annotations

Cell line names were harmonized in both CGP and CCLE to match identical cell lines; this was done through manual search over alternative names of cell lines, as reported in CGP and CCLE cell line annotation files and online databases such as [hyperCLDB](#) and [BioInformationWeb](#). In our comparative analysis published in Nature [?], we focused on the set of 504 cell lines for which both gene expression and drug sensitivity were available. In the present work we extended our curation to all the cell lines for which at least one data type (gene expression, mutation or drug sensitivity) is available, increasing the shared set of cell lines to 504:



Tissue type nomenclature from CGP [2] was chosen throughout this study, CCLE tissue type information [1] was therefore updated to follow this nomenclature, which resulted in 24 tissue types:

Tissue type	Number of cell lines
lung	109
haematopoietic_and_lymphoid_tissue	77
breast	39
central_nervous_system	36
large_intestine	33
skin	30
oesophagus	21
urinary_tract	17
ovary	16
pancreas	16
stomach	16
autonomic_ganglia	12
soft_tissue	12
upper_aerodigestive_tract	12
liver	11
kidney	10
bone	8
endometrium	8
thyroid	8
pleura	6
prostate	4
biliary_tract	1
salivary_gland	1
small_intestine	1

All the curation steps have been documented in the scripts `cdrug2_normalization_cgpr.R`, `cdrug2_normalization_ccle.R`, and `cdrug2_format.R`.

2.3 Mutation data

We focused our analyses on missense mutations identified both in CGP and CCLE. For CGP, coding variants in 68 genes were extracted from `gdsc_manova_input_w2.csv`. For CCLE, coding variants in 1667 genes (column 'Protein_Change') measured using the Oncomap3 and hybrid capture platforms were extracted from `CCLE_Oncomap3_2012-04-09.maf` and

`CCLE_hybrid_capture1650_hg19_NoCommonSNPs_NoNeutralVariants_CDS_2012.05.07.maf.`, respectively.

For the 504 cell lines investigated in both CGP and CCLE, we identified 64 genes whose missense mutations in protein coding genes have been measured in both studies. We followed the approach used in the original studies by transforming the mutation data into binary matrices reporting the presence or absence of at least one mutation in a given gene and cell line.

2.4 Drug sensitivity calling

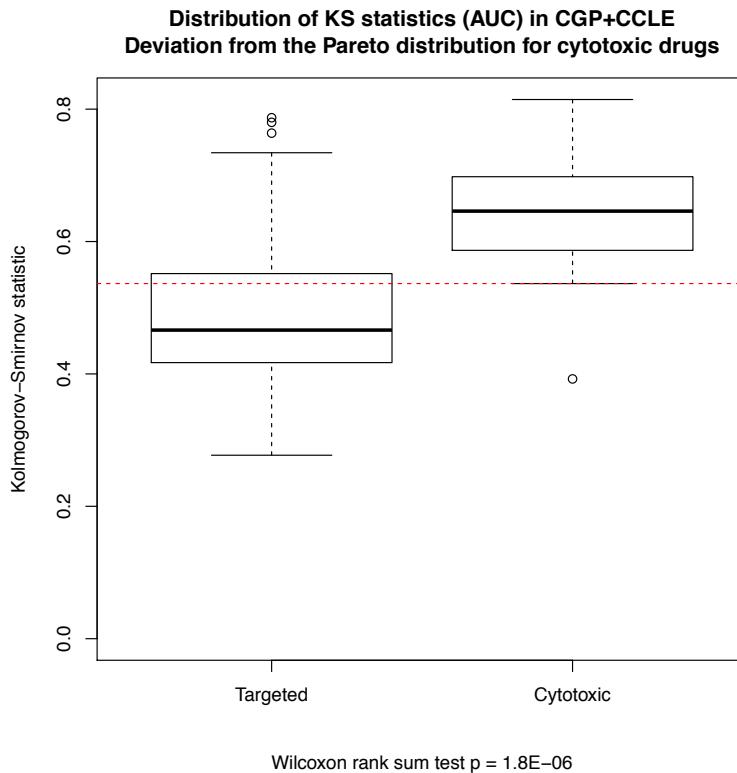
To categorize cell line sensitivity into three categories (resistant, intermediate and sensitive) we used the *Waterfall* approach described in the CCLE study [1]. The full procedure, as communicated by Dr. Kavitha Venkatesan (personal communication) is described below:

1. Extract the drug sensitivity measurements, either IC_{50} or AUC
2. Sort log IC_{50} values (or AUC) of the cell lines to generate a *waterfall* distribution.
3. If the waterfall distribution is non-linear (Pearson correlation coefficient to the linear fit ≤ 0.95), estimate the major inflection point of the log IC_{50} curve as the point on the curve with the maximal distance to a line drawn between the start and end points of the distribution.
4. If the waterfall distribution appears linear (Pearson correlation coefficient to the linear fit > 0.95), then use the median IC_{50} instead.
5. Cell lines within a 4-fold IC_{50} (or within a 1.2-fold AUC) difference centered around this inflection point are classified as being *intermediate*, cell lines with lower IC_{50} (or AUC) values than this range are defined as *sensitive*, and those with IC_{50} (or AUC) values higher than this range are called *resistant*.
6. Require at least $x = 5$ sensitive and $x = 5$ resistant cell lines after applying these criteria.

Modified Waterfall approach

In order to automatically identify the set of highly sensitive and resistant cell lines, we adapted the Waterfall approach described in the previous section. The main change resides in the use of a outlier detection method to identify highly sensitive cel lines for highly targeted drugs. The full list of modifications is provided below and were implemented in the new function `callingWaterfallAUC` in `cdrug2_foo.R`.

- We first discriminated between highly targeted and broad effect drugs by computing the Kolmogorov-Smirnov KS statistic to quantify the deviation of the AUC measurement distribution form the Pareto type I distribution [4]. We first computed such KS statistic in CGP and CCLE datasets and merged them to test whether this statistics can discriminate between targeted and cytotoxic drugs. As can be seen in the box plot below, the KS statistic is indeed highly discriminative and we determined the optimal cutoff for discriminating between targeted vs. cytotoxic drugs using Youden's index



- We choose large fold change to identify the cell lines exhibiting an intermediate drug sensitivity; this parameter was set to 1.2 and 4 for highly targeted and broader effect drugs, respectively.
- If a drug was classified as highly targeted using the KS statistic, then we identified the few highly sensitive cell lines based on the outlier detection approach implemented in the R package `extremevalues`, function `getOutliersI`, using the Pareto Type I distribution.

2.5 Concordance

We assessed the concordance of the gene expression, mutation and drug sensitivity of CGP and CCLE studies *across* and *between* cell lines. When data are compared across cell lines, we assess whether, for a given gene expression, mutation or drug, the cell line data were concordant (a given mutation is observed in the same set of cell lines for instance). When data are compared between cell lines, we assessed whether, for a given cell line, the genomic and pharmacological data were concordant in the two studies (a given cell line harbours similar mutations for instance).

2.5.1 Gene expression

We assessed consistency between gene expression measures in CGP and CCLE by computing Spearman rank-ordered correlations when ≥ 10 measures were available. Typically qualitative descriptions of correlation coefficients are associated with intervals: $r_s < 0.5$, poor correlation; $0.5 \leq r_s < 0.6$, fair correlation; $0.6 \leq r_s < 0.7$, moderate correlations; $0.7 \leq r_s < 0.8$, substantial correlation; and $r_s \geq 0.8$, almost perfect correlation.

2.5.2 Drug sensitivity

We assessed consistency between drug sensitivity measures (IC_{50} and AUC), in CGP and CCLE by computing Spearman rank-ordered correlations when ≥ 10 measures were available. We used the same qualitative descriptions of correlation coefficients as for gene expression data.

2.5.3 Mutation

We assessed the consistency between presence and absence of mutations measured in CGP and CCLE by using Cohen's κ statistics. We used the following qualitative descriptions of κ values associated with intervals: $\kappa < 0.5$, poor agreement; $0.5 \leq \kappa < 0.6$, fair agreement; $0.6 \leq \kappa < 0.7$, moderate agreement; $0.7 \leq \kappa < 0.8$, substantial agreement; and $\kappa \geq 0.8$, almost perfect agreement.

2.5.4 Consistency between drug sensitivity calls

We assessed the consistency of drug sensitivity calling, as computed using the waterfall method described in the CCLE study [1], using Cohen's Kappa (κ) coefficient. We used the same qualitative descriptions of κ coefficients as for mutation data.

2.6 Gene-drug associations

We assessed the association between gene expression and AUC values (referred to as gene-drug association), using either

- Spearman's correlation for continuous AUC values, as suggested by Geeleher *et al.*
- A linear regression model controlled for tissue type:

$$Y = \beta_0 + \beta_i G_i + \beta_t T$$

where Y denote the drug sensitivity variable, G_i and T denote the expression of gene i and the tissue type respectively and β s are the regression coefficients. The strength of gene-drug association is quantified by β_i , above and beyond the relationship between drug sensitivity and tissue type. The variables Y and G are scaled (standard deviation equals to 1) to estimate standardized coefficients from the linear model. Significance of the gene-drug association is estimated by the statistical significance of β_i (t statistic).

- A logistic regression model, similar to the one described above, except that Y represents the extreme sensitivity calls (highly sensitive vs. resistant) as computed by the modified Waterfall approach.

3 Full Reproducibility of the Analysis Results

We will describe how to fully reproduce the figures and tables reported in the main manuscript. We automated the analysis pipeline so that minimal manual interaction is required to reproduce our results. To do this, one must simply:

1. Set up the software environment
2. Run the R scripts
3. Generate the Supplementary Information

The code and associated files are publicly available on GitHub: <https://github.com/bhaibeka/cdrug2>.

3.1 Set up the software environment

We developed and tested our analysis pipeline using R running on linux and Mac OSX platforms.

To mimic our software environment the following R packages should be installed:

- R version 3.0.2 (2013-09-25), x86_64-unknown-linux-gnu
- Base packages: base, datasets, graphics, grDevices, grid, methods, parallel, splines, stats, utils
- Other packages: affxparser 1.34.0, affy 1.40.0, affyio 1.30.0, amap 0.8-7, AnnotationDbi 1.24.0, Biobase 2.22.0, BiocGenerics 0.8.0, biomaRt 2.18.0, bitops 1.0-6, cluster 1.14.4, corpcor 1.6.6, DBI 0.2-7, epibasix 1.3, extremevalues 2.2, Formula 1.1-1, frma 1.14.0, gdata 2.13.2, genefu 1.13.2, gPdtest 0.4, gplots 2.12.1, hgu133plus2cdf 2.13.0, hgu133plus2frmavecs 1.3.0, Hmisc 3.13-0, hthgu133acdf 2.13.0, hthgu133afrmavecs 1.1.0, igraph 0.6.6, inSilicoDb2 2.0.0, jetset 1.6.0, lattice 0.20-24, MASS 7.3-29, mclust 4.2, MetaGx 0.0.2, mgcv 1.7-27, mRMRe 2.0.4, nlme 3.1-113, OptimalCutpoints 1.1-1, org.Hs.eg.db 2.10.1, PharmacoGx 0.0.2, prodlm 1.3.7, RCurl 1.95-4.1, rjson 0.2.13, R.methodsS3 1.5.2, R.oo 1.15.8, RSQLite 0.11.4, R.utils 1.28.4, survcomp 1.5.5, survival 2.37-4, sva 3.8.0, vcd 1.3-1, VennDiagram 1.6.5, WriteXLS 3.2.2, xtable 1.7-1
- Loaded via a namespace (and not attached): BiocInstaller 1.12.0, Biostrings 2.30.1, bit 1.1-11, bootstrap 2012.04-1, caTools 1.16, codetools 0.2-8, colorspace 1.2-4, ff 2.2-12, foreach 1.4.1, GenomicRanges 1.14.4, gtools 3.1.1, IRanges 1.20.6, iterators 1.0.6, KernSmooth 2.23-10, Matrix 1.1-0, oligo 1.26.0, oligoClasses 1.24.0, plotrix 3.5-2, preprocessCore 1.24.0, rmeta 2.16, stats4 3.0.2, SuppDists 1.1-9.1, survivalROC 1.0.3, tools 3.0.2, XML 3.98-1.1, XVector 0.2.0, zlibbioc 1.8.0

All these packages are available on CRAN¹ or Bioconductor², except for jetset which is available on the CBS website³.

Run the following commands in a R session to install all the required packages:

```
source("http://bioconductor.org/biocLite.R")
biocLite(c("AnnotationDbi", "affy", "affyio", "hthgu133acdf",
          "hthgu133afrmavecs", "hgu133plus2cdf", "hgu133plus2frmavecs",
          "org.Hs.eg.db", "genefu", "biomaRt", "frma", "Hmisc", "vcd",
          "epibasix", "amap", "gdata", "WriteXLS", "xtable", "plotrix",
          "R.utils", "DBI", "GSA", "gplots"))
```

¹<http://cran.r-project.org>

²<http://www.bioconductor.org>

³<http://www.cbs.dtu.dk/biotools/jetset/>

Note that you may need to install Perl⁴ and its module Text::CSV_XS for the WriteXLS package to write xls file; once Perl is installed in your system, use the following command to install the Text::CSV_XS module through CPAN⁵:

```
cpan Text/CSV_XS.pm
```

Lastly, follow the instructions on the CBS website to properly install the jetset package or use the following commands in R:

```
download.file(url="http://www.cbs.dtu.dk/biotools/jetset/current/jetset_1.4.0.tar.gz",
  destfile="jetset_1.4.0.tar.gz")
install.packages("jetset_1.4.0.tar.gz", repos=NULL, type="source")
```

Once the packages are installed, uncompress the archive (cdrug2.zip) provided as **Supplementary File 2** accompanying the manuscript. This should create a directory on the file system containing the following files:

`cdrug2_foo.R` Script containing the definitions of all functions required for the analysis pipeline.
`cdrug2_normalization_cgp.R` Script to curate, annotate and normalize of CGP data.
`cdrug2_normalization_ccle.R` Script to curate, annotate and normalize of CCLE data.
`cdrug2_format.R` Script to identify common cell lines, tissue types and drugs investigated both in CGP and CCLE.
`cdrug2_analysis.R` Script generating all the figures and tables reported in the manuscript.
`cdrug2_pipeline.R` Master script running all the scripts listed above to generate the analysis results.
`matching_cell_line_CCLE_CGP.csv` Curation of cell line name to match CGP and CCLE nomenclatures.
`matching_tissue_type_CCLE_CGP.csv` Curation of tissue type name to match CGP and CCLE nomenclatures.
`HaibeKains_Nature_2013_common_cellines.csv` Set of 504 cell lines shared between CGP and CCLE as published in Haibe-Kains *et al.*, Nature, 2013.
`cdrug2_suppl_info.tex` The L^AT_EX file of the present supplementary information

All the files required to run the automated analysis pipeline are now in place. It is worth noting that raw gene expression and drug sensitivity data are voluminous, please ensure that at least 25GB of storage are available.

3.2 Run the R scripts

Open a terminal window and go to the `cdrug2` directory. You can easily run the analysis pipeline either in batch mode or in a R session. Before running the pipeline you can specify the number of CPU cores you want to allocate to the analysis (by default only 1 CPU core will be used). To do so, open the script `cdrug2_pipeline.R` and update line #41:

```
nbcore <- 4
```

⁴<http://www.perl.org/get.html>

⁵<http://www.cpan.org/modules/INSTALL.html>

to allocate four CPU cores for instance.

To run the full pipeline in batch mode, simply type the following command:

```
R CMD BATCH code/cdrug2_pipeline.R Rout &
```

The progress of the pipeline could be monitored using the following command:

```
tail -f Rout
```

To run the full analysis pipeline in an R session, simply type the following command:

```
source("code/cdrug2_pipeline.R")
```

Key messages will be displayed to monitor the progress of the analysis.

The analysis pipeline was developed so that all intermediate analysis results are saved in the directories `data` and `saveres`. Therefore, in case of interruption, the pipeline will restart where it stopped.

3.3 Additional parameters

download.method Method to download all the data from the CCLE and CGP websites; options are 'wget' (default), 'curl', 'lynx', or 'auto'.

cosmic.version Version of the COSMIC database ('v68' as for 2014-04-07; see Sanger's FTP)

saveres Path to the directory where all teh results should be stored. Default is ".`/saveres`"

max.cellfiles Maximum number of CEL files taht can be processed at once by `frma`.

minsample Minimum number of samples to compute the Spearman correlation. Default is 10.

genedrugm Method to estimate the association gene-drug, controlled for tissue type. Method "lm" = linear regression or logistic regression, depending on the output variable

concordance.method Estimator for concordance between gene-drug associations; possible options are "spearman" (default), "cosine", and "pearson" for Spearman's rank-ordered correlation coefficient, cosine similarity and Pearson's correlation coefficient, respectively.

3.4 Generate the Supplementary Information

After completion of the analysis pipeline a directory `saveres` will be created to contain all the intermediate results, tables and figures reported in the main manuscript and this Supplementary Information.

4 List of Abbreviations

AUC	Area under the drug sensitivity curve.
CGP	Cancer Genome Project initiated by the Wellcome Sanger Institute.
CCLE	The Cancer Cell Lines Encyclopedia initiated by Novartis and the Broad Institute.
IC ₅₀	Concentration in micro molar [μ M] at which the drug inhibited 50% of the cellular growth.
FDR	False Discovery Rate
R _s	Spearman correlation coefficient

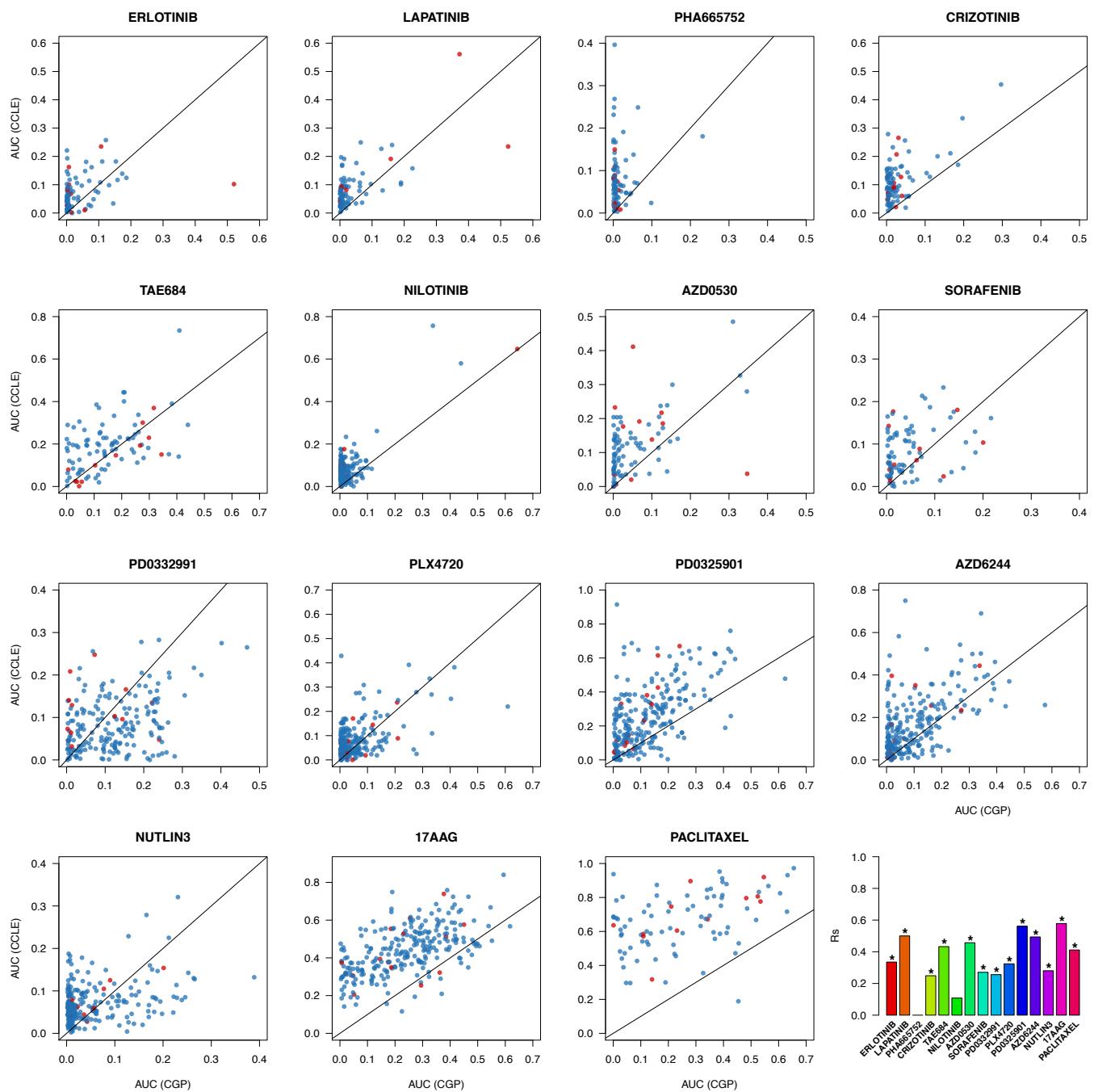
5 Supplementary Tables

Supplementary Table 1: Description of the 15 anticancer drugs screened both in CGP and CCLE studies.

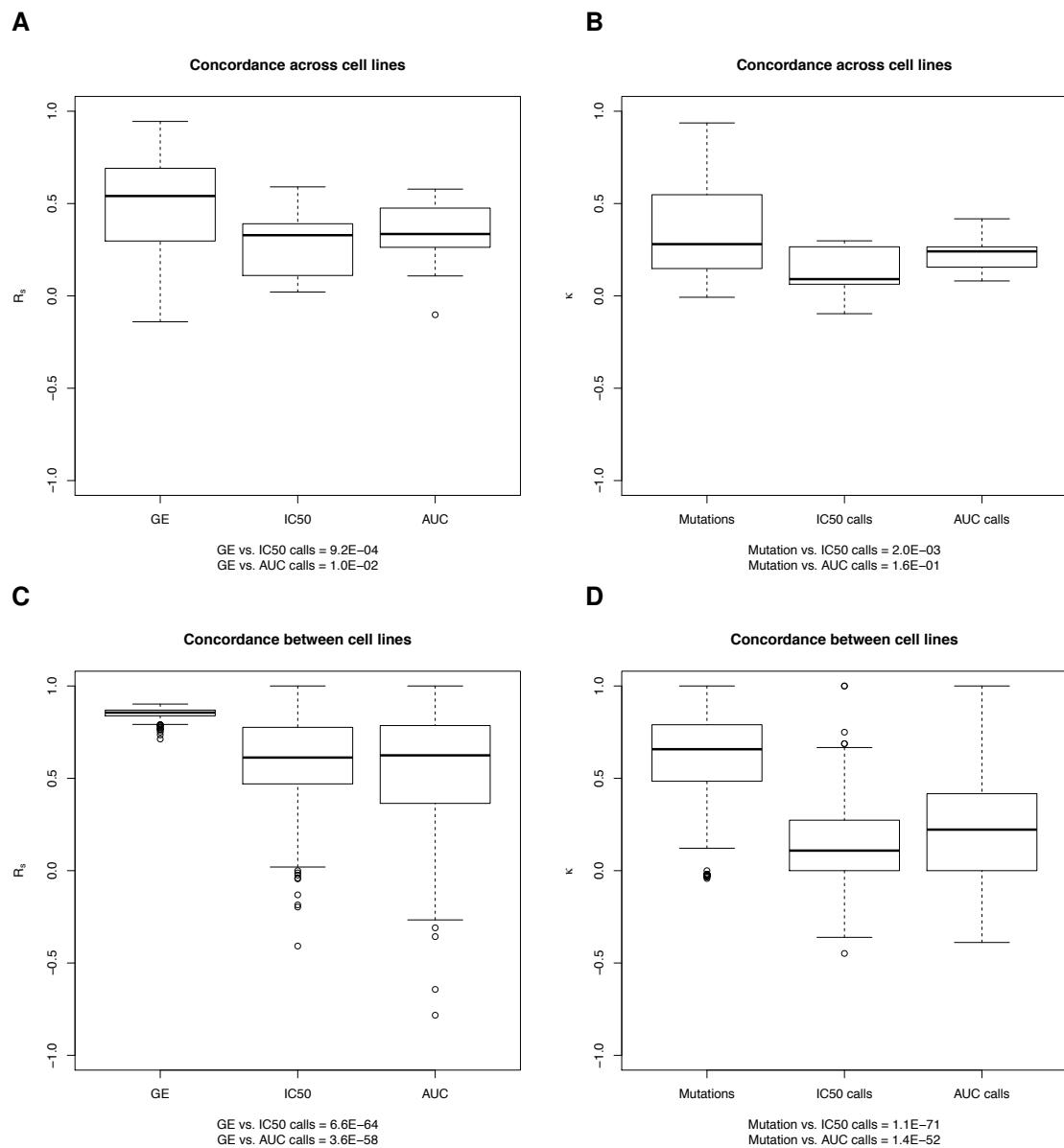
Compound	Class	Target(s)	Class	Organization
Erlotinib	Targeted	EGFR	Kinase inhibitor	Genentech
Lapatinib	Targeted	EGFR, HER2	Kinase inhibitor	GlaxoSmithKline
PHA-665752	Targeted	c-MET	Kinase inhibitor	Pfizer
Crizotinib	Targeted	c-MET, ALK	Kinase inhibitor	Pfizer
TAE684	Targeted	ALK	Kinase inhibitor	Novartis
Nilotinib	Targeted	Abl/Bcr-Abl	Kinase inhibitor	Novartis
AZD0530	Targeted	Src, Abl/Bcr-Abl, EGFR	Kinase inhibitor	AstraZeneca
Sorafenib	Targeted	Flt3, C-KIT, PDGFRbeta, RET, Raf kinase B, Raf kinase C, VEGFR-1, KDR, FLT4	Kinase inhibitor	Bayer
PD-0332991	Targeted	CDK4/6	Kinase inhibitor	Pfizer
PLX4720	Targeted	RAF	Kinase inhibitor	Plexxikon
PD-0325901	Targeted	MEK	Kinase inhibitor	Pfizer
AZD6244	Targeted	MEK	Kinase inhibitor	AstraZeneca
Nutlin-3	Targeted	MDM2	Other	Roche
17-AAG	Targeted	HSP90	Other	Bristol-Myers Squibb
Paclitaxel	Cytotoxic	beta-tubulin	Cytotoxic	Bristol-Myers Squibb

6 Supplementary Figures

Supplementary Figure 1: Consistency across cell lines of AUC values between CGP and CCLE for the 504 cell lines and 15 drugs investigated both in CGP and CCLE. The scatter plot represent each data point, where the 471 cell lines analyzed in our original study are represented by blue points, while the new cell lines identified after recreation of the data are presented in red. In case of perfect consistency, all the points should lie on the diagonal. The last bar plot (bottom right corner) reports Spearman's correlation coefficient (R_s) for each drug.



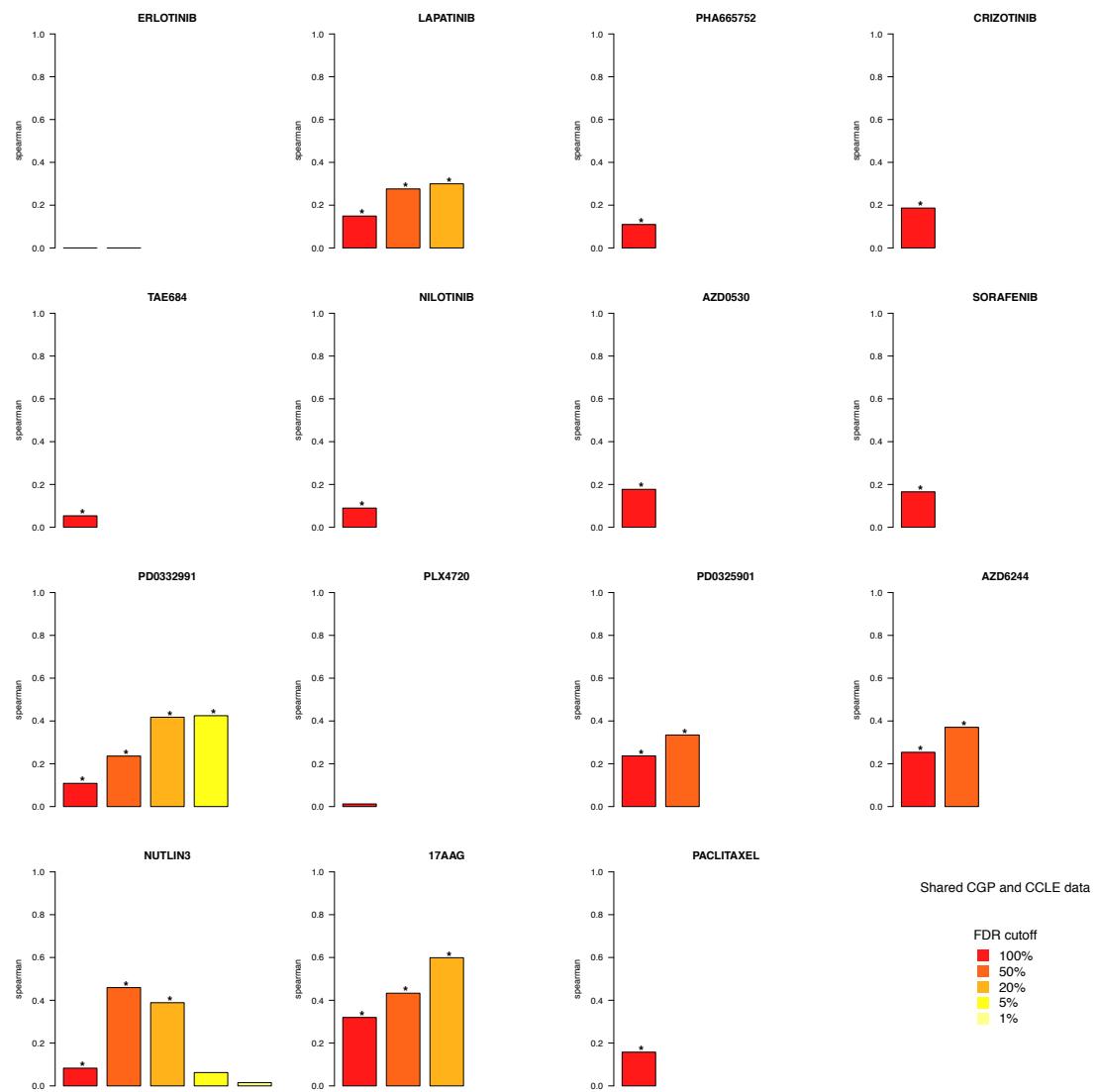
Supplementary Figure 2: **(A)** Box plot comparing Spearman's correlation coefficients estimating the concordance of gene expressions, IC_{50} and AUC measures across cell lines in CGP and CCLE. **(B)** Box plot comparing Cohen's κ coefficients estimating the concordance of presence/absence of mutations, IC_{50} sensitivity calls and AUC sensitivity calls across cell lines in CGP and CCLE. **(C)** Box plot comparing Spearman's correlation coefficients estimating the concordance of gene expressions, IC_{50} and AUC measures between cell lines in CGP and CCLE. **(D)** Box plot comparing Cohen's κ coefficients estimating the concordance of gene expressions, IC_{50} and AUC measures between cell lines in CGP and CCLE. Significance of the difference between concordance observed for genomic and pharmacological data, as computed using the Wilcoxon rank sum test, is provided under each plot. GE: gene expression; R_s : Spearman's rank-ordered correlation; κ : Cohen's κ coefficient.



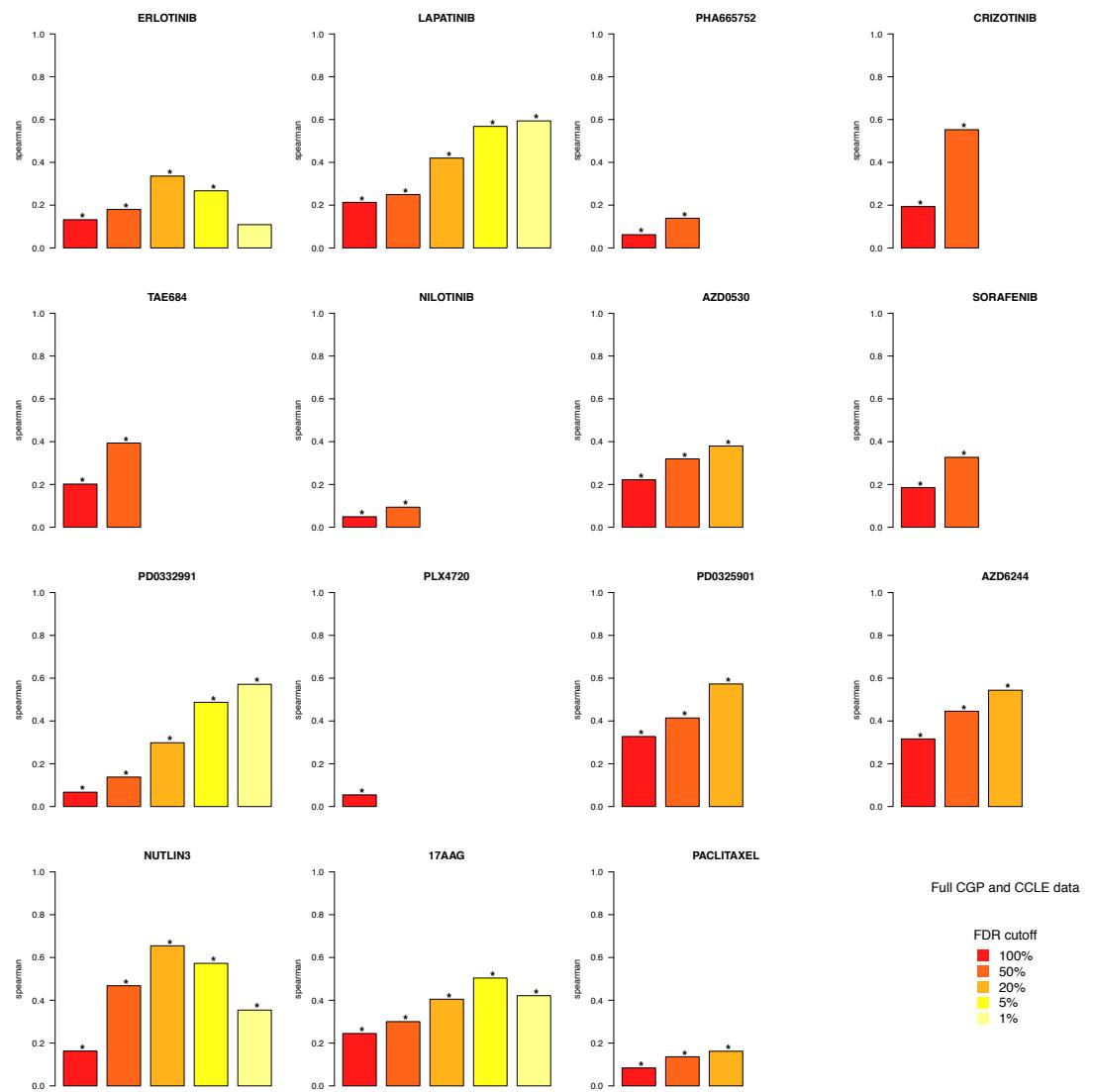
Supplementary Figure 3: Contingency tables comparing the (extreme) sensitivity calls (res, inter, and sens standing for resistant, intermediate and sensitive drug phenotype, respectively) computed from AUC measures for each of the 15 drugs screened both in CGP and CCLE. The Kappa coefficient, its confidence interval and its significance are reported below each contingency table. The intermediate drug sensitivity calls are made large to enable to identify the extreme sensitive and resistant cell lines. It is worth to note that the sensitivity calls have been computed using the modified Waterfall approach on the full CGP and CCLE datasets.

AUC sensitivity calling ERLOTINIB	AUC sensitivity calling LAPATINIB	AUC sensitivity calling PHA665752
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res 27 0 1 inter 41 3 6 sens 1 2 0	res 18 6 0 inter 34 16 6 sens 0 1 4	res 15 7 1 inter 42 13 5 sens 5 1 2
Kappa=0.052, 95%CI [-0.08,0.18], p=1.0E-02	Kappa=0.14, 95%CI [-0.036,0.31], p=2.9E-04	Kappa=-0.0036, 95%CI [-0.17,0.16], p=5.0E-01
AUC sensitivity calling CRIZOTINIB	AUC sensitivity calling TAE684	AUC sensitivity calling NILOTINIB
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res 10 3 0 inter 41 25 3 sens 1 5 4	res 24 47 3 inter 3 7 1 sens 0 5 2	res 40 17 2 inter 97 24 13 sens 3 2 4
Kappa=0.1, 95%CI [-0.059,0.27], p=1.3E-03	Kappa=0.06, 95%CI [-0.081,0.2], p=8.7E-02	Kappa=-0.017, 95%CI [-0.13,0.096], p=8.7E-03
AUC sensitivity calling AZD0530	AUC sensitivity calling SORAFENIB	AUC sensitivity calling PD0332991
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res 26 1 1 inter 46 10 1 sens 1 1 3	res 13 9 0 inter 25 22 5 sens 3 7 2	res 12 5 11 inter 72 29 63 sens 2 0 4
Kappa=0.14, 95%CI [-0.0034,0.29], p=9.8E-04	Kappa=0.05, 95%CI [-0.14,0.24], p=1.2E-01	Kappa=0.015, 95%CI [-0.067,0.097], p=5.2E-01
AUC sensitivity calling PLX4720	AUC sensitivity calling PD0325901	AUC sensitivity calling AZD6244
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res 49 22 1 inter 97 60 9 sens 1 5 9	res 25 10 0 inter 53 132 11 sens 3 10 6	res 42 23 0 inter 56 88 12 sens 3 6 6
Kappa=0.12, 95%CI [0.0086,0.23], p=2.8E-08	Kappa=0.26, 95%CI [0.15,0.38], p=6.2E-08	Kappa=0.23, 95%CI [0.11,0.35], p=4.5E-07
AUC sensitivity calling NUTLIN3	AUC sensitivity calling 17AAG	AUC sensitivity calling PACLITAXEL
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res 30 2 5 inter 134 16 45 sens 9 1 10	res 57 8 11 inter 32 24 61 sens 6 14 44	res 14 4 5 inter 20 6 24 sens 2 0 13
Kappa=0.054, 95%CI [-0.011,0.12], p=4.9E-02	Kappa=0.26, 95%CI [0.17,0.35], p=4.4E-16	Kappa=0.16, 95%CI [0.023,0.3], p=1.4E-03

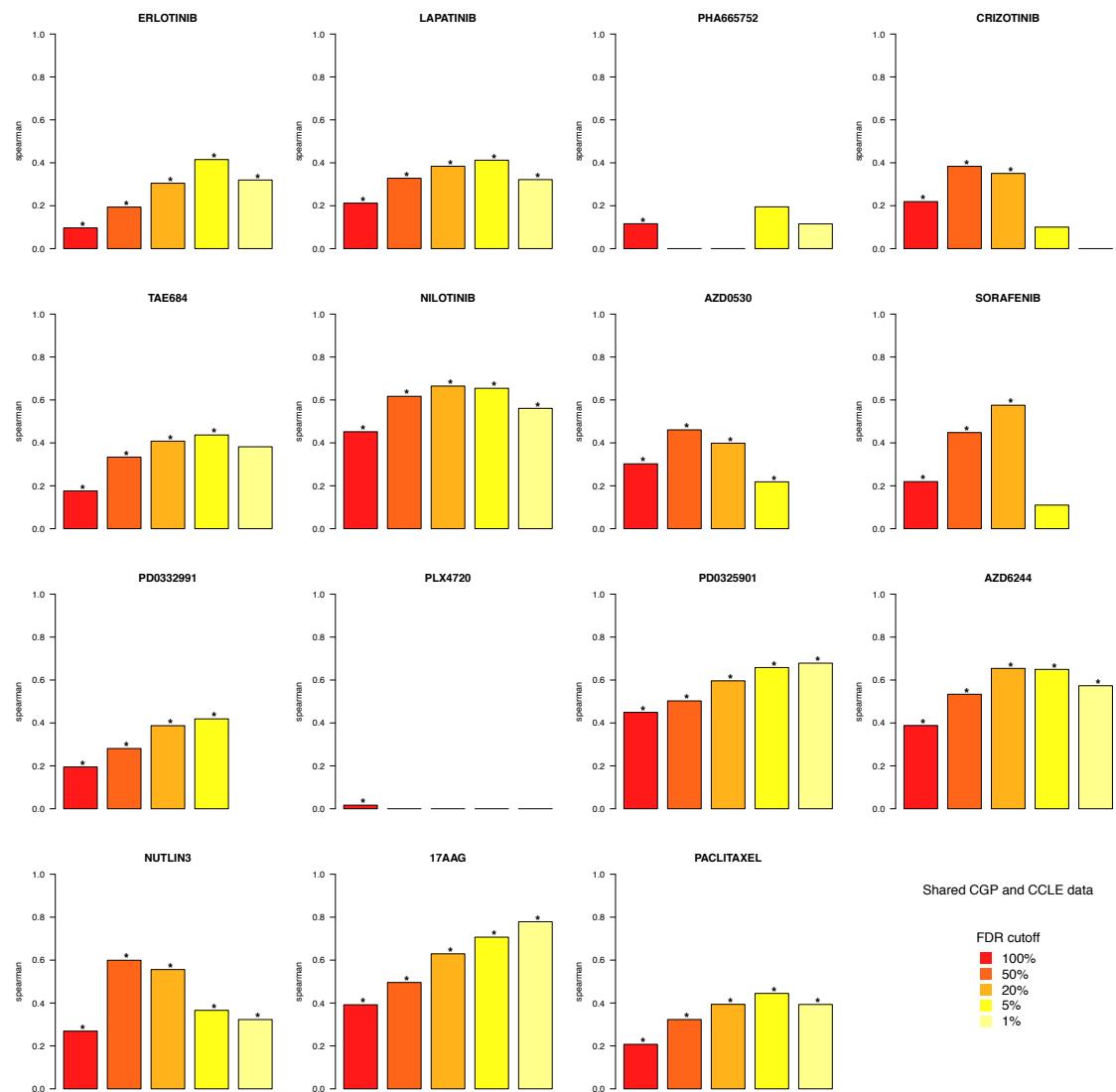
Supplementary Figure 4: Consistency across cell lines of gene-drug associations computed with logistic regression analysis based on extreme AUC sensitivity calls between CGP and CCLE for the 504 cell lines and 15 drugs investigated both in CGP and CCLE. The logistic regression model is adjusted for cell lines' tissue of origin. The consistency is estimated by Spearman's rank-ordered correlation for multiple false discovery rate (FDR) thresholds; the significance of Spearman's correlation coefficients is represented by ** for $p < 0.05$.



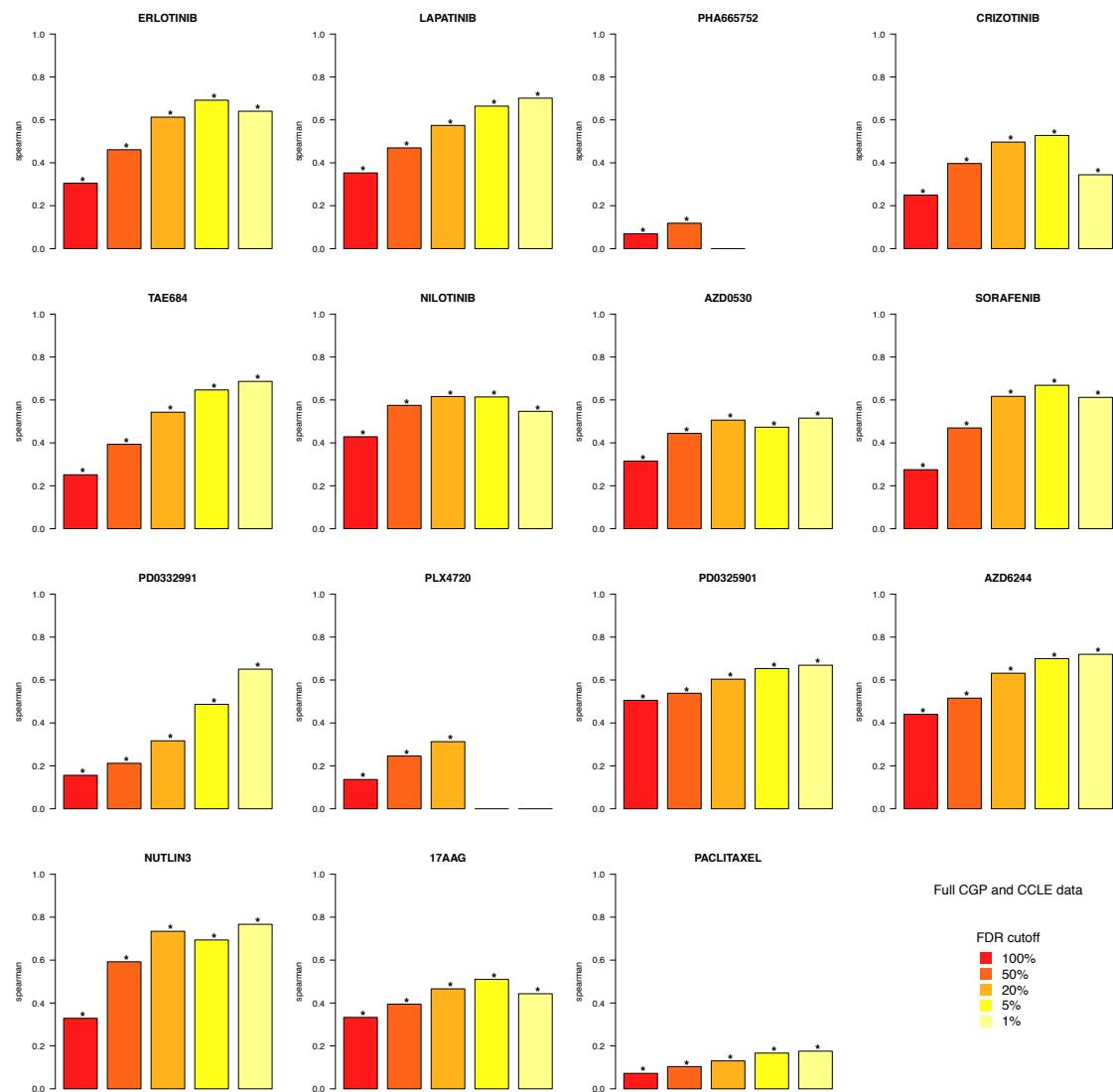
Supplementary Figure 5: Consistency across cell lines of gene-drug associations computed with logistic regression analysis based on extreme AUC sensitivity calls between CGP and CCLE for all the cell lines and the 15 drugs investigated both in CGP and CCLE. The model is adjusted for cell lines' tissue of origin. The consistency is estimated by Spearman's rank-ordered correlation for multiple false discovery rate (FDR) thresholds; the significance of Spearman's correlation coefficients is represented by '*' for $p < 0.05$



Supplementary Figure 6: Consistency across cell lines of gene-drug associations computed with linear regression analysis based on AUC values between CGP and CCLE for the 504 cell lines and the 15 drugs investigated both in CGP and CCLE. The model is adjusted for cell lines' tissue of origin. The consistency is estimated by Spearman's rank-ordered correlation for multiple false discovery rate (FDR) thresholds; the significance of Spearman's correlation coefficients is represented by '*' for $p < 0.05$

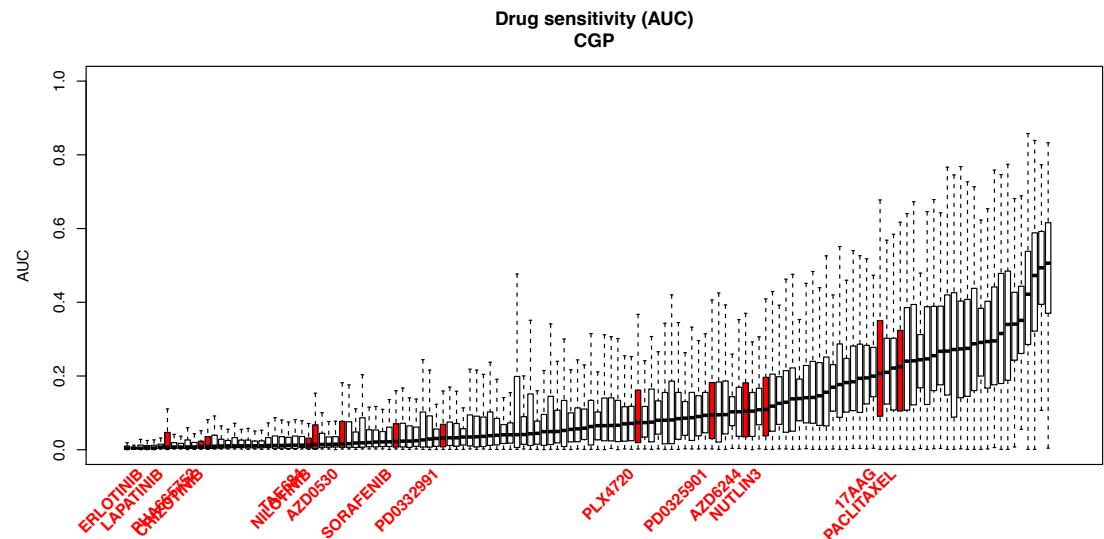


Supplementary Figure 7: Consistency across cell lines of gene-drug associations computed with linear regression analysis based on AUC values between CGP and CCLE for all the cell lines and the 15 drugs investigated both in CGP and CCLE. The model is adjusted for cell lines' tissue of origin. The consistency is estimated by Spearman's rank-ordered correlation for multiple false discovery rate (FDR) thresholds; the significance of Spearman's correlation coefficients is represented by '*' for $p < 0.05$

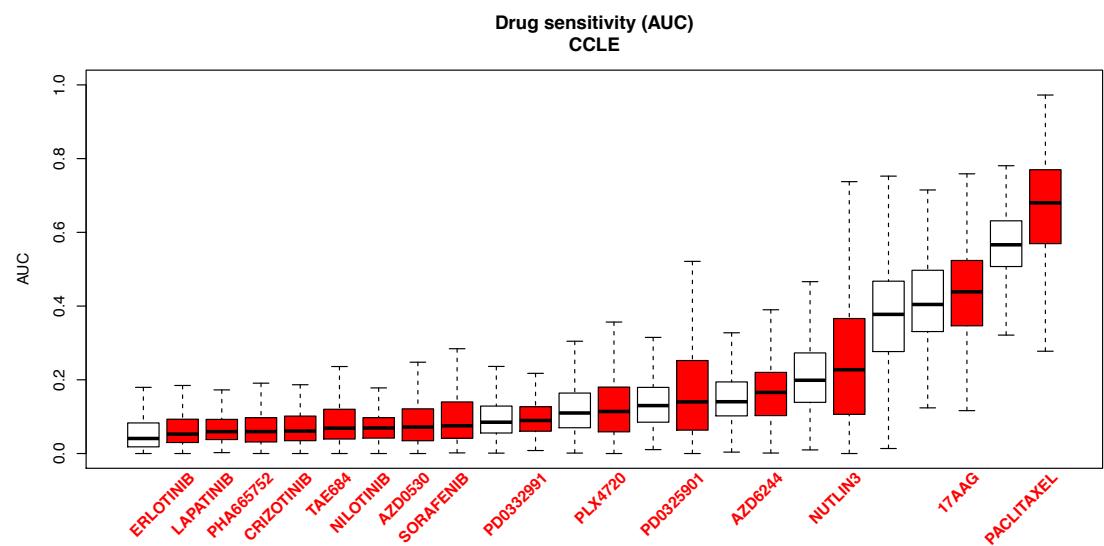


Supplementary Figure 8: Variability of drug sensitivity data. Box plot showing the distribution of AUC measures for all the drugs screened in **(A)** CGP and **(B)** CCLE for the shared set of 504 cell lines. In red are highlighted the drugs screened both in CGP and CCLE.

A

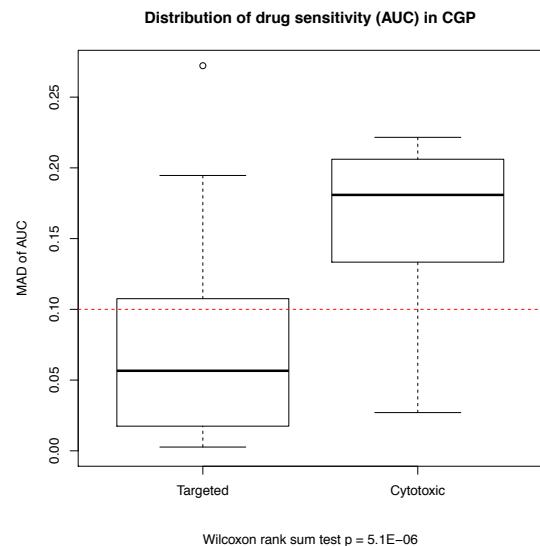


B

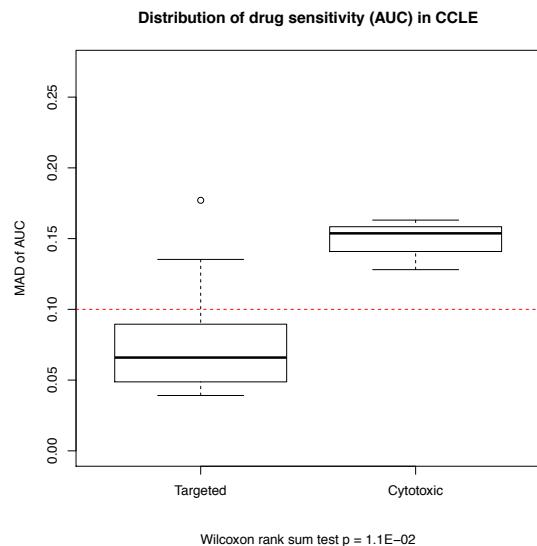


Supplementary Figure 9: Variability of drug sensitivity data, as computed by median absolute deviation (MAD), of cytotoxic and targeted drugs screened in CGP (**A**) and CCLE (**B**). The red horizontal dashed line is the cutoff of MAD used to discriminate between targeted drugs and those with broader effect.

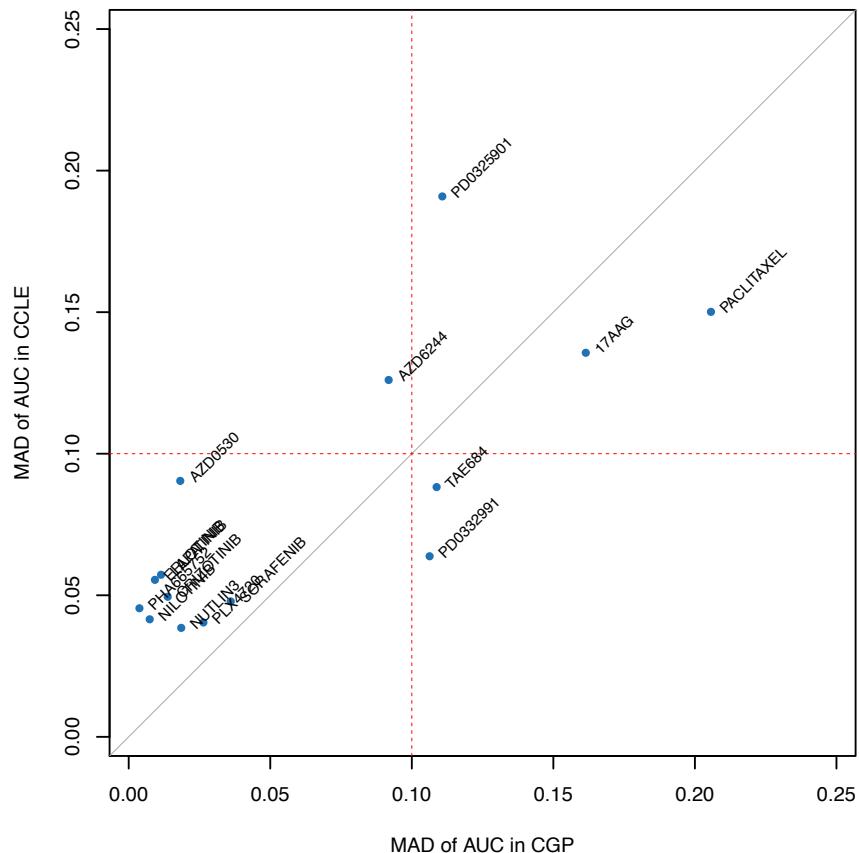
A



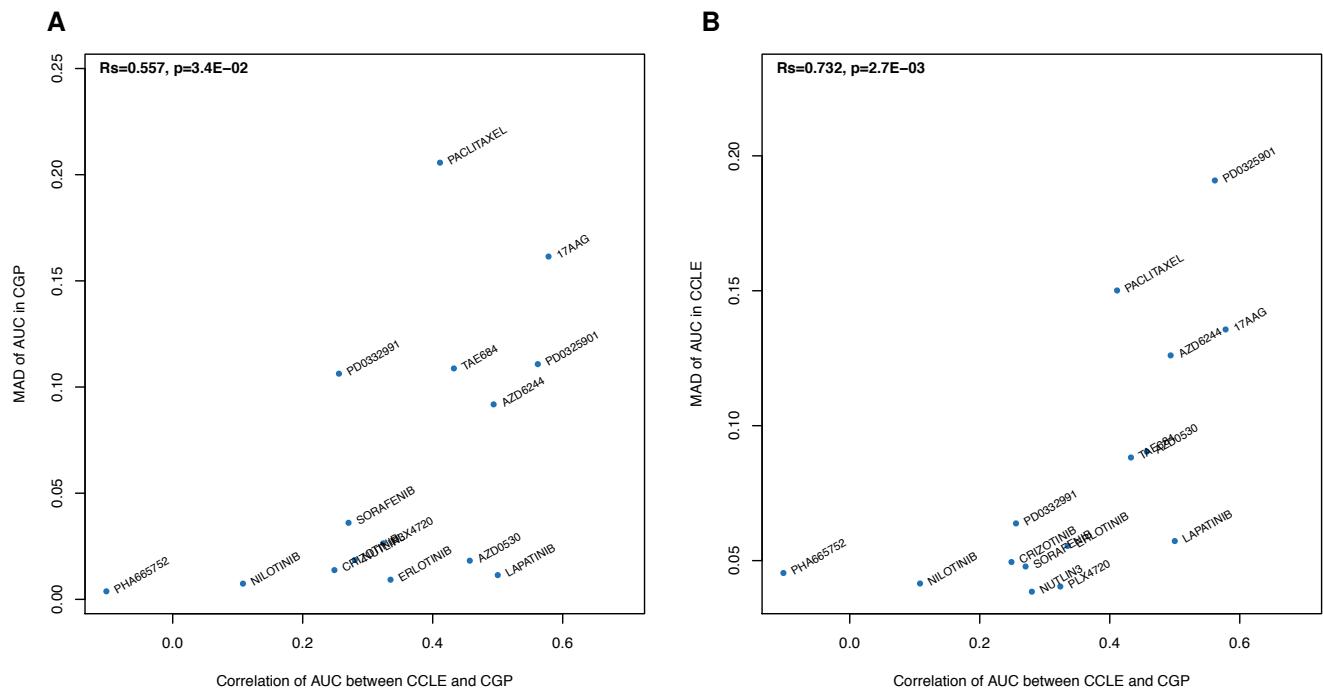
B



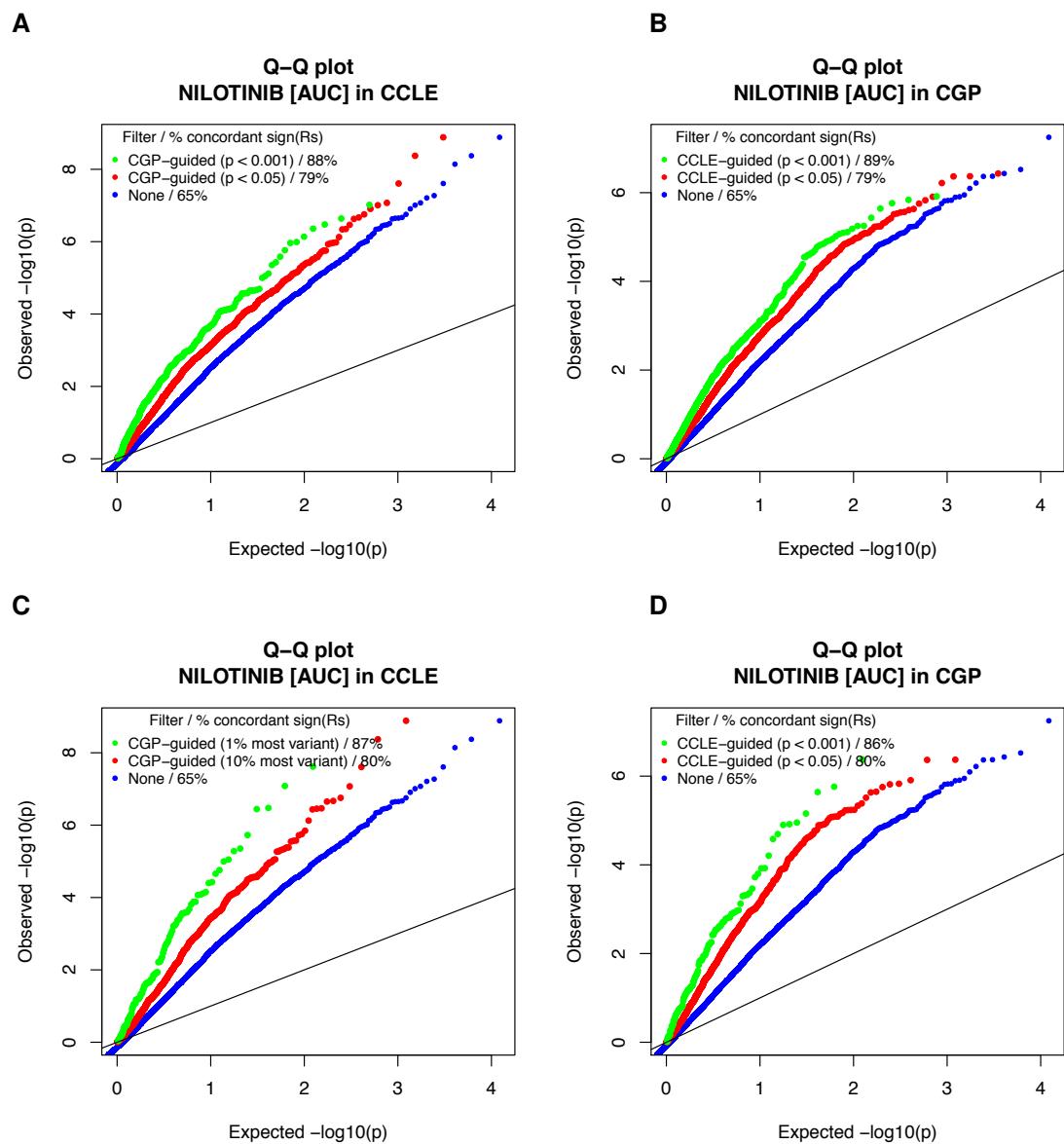
Supplementary Figure 10: Comparison of variability, as computed by the median absolute deviation (MAD), of sensitivity data of targeted vs. cytotoxic drugs in CGP and CCLE. The red horizontal dashed line is the cutoff of MAD used to discriminate between targeted drugs and those with broader effect.



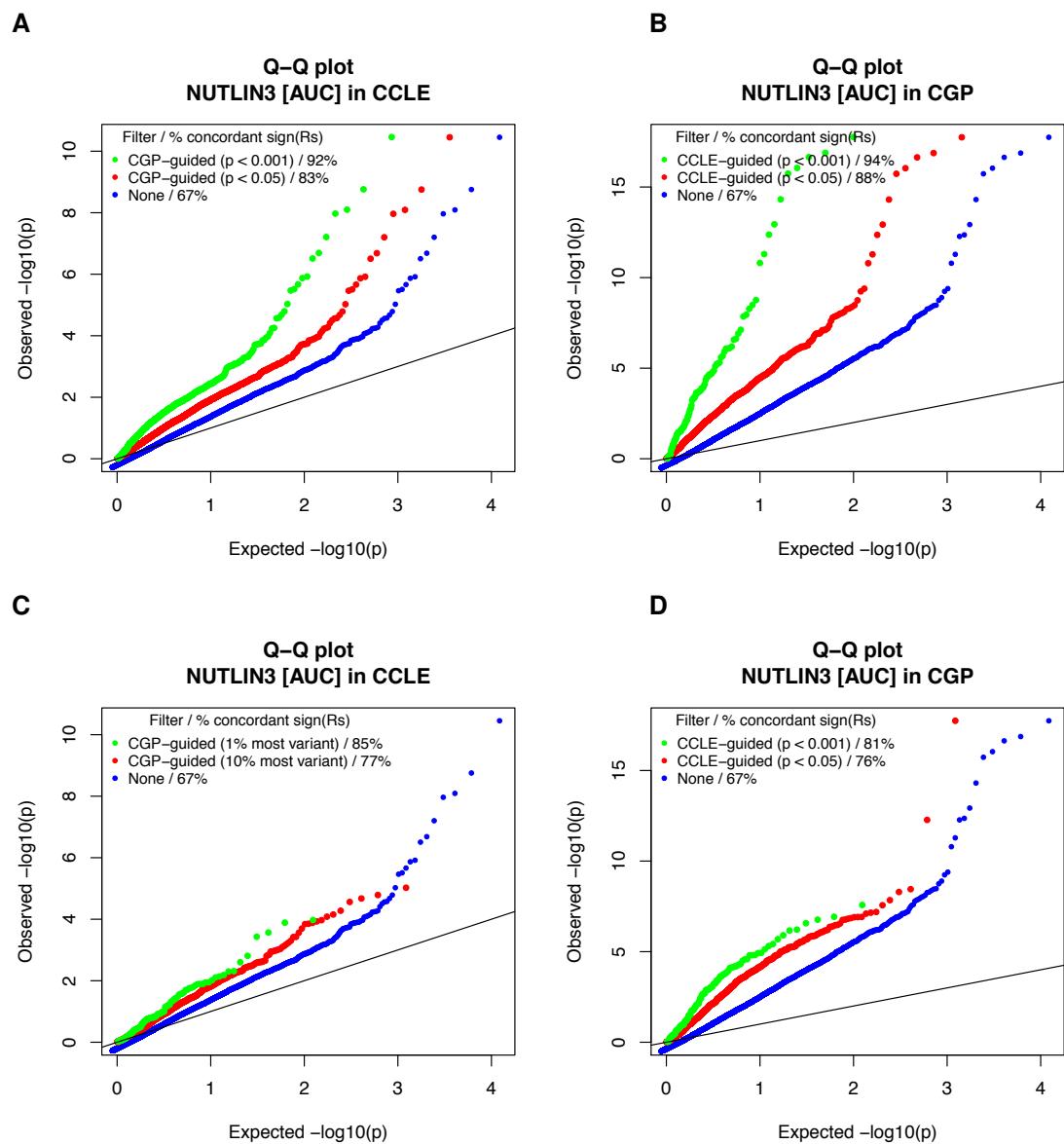
Supplementary Figure 11: Association between concordance of drug sensitivity, computed using Spearman's correlation coefficient, and variability of drug sensitivity data, computed using median absolute deviation (MAD).



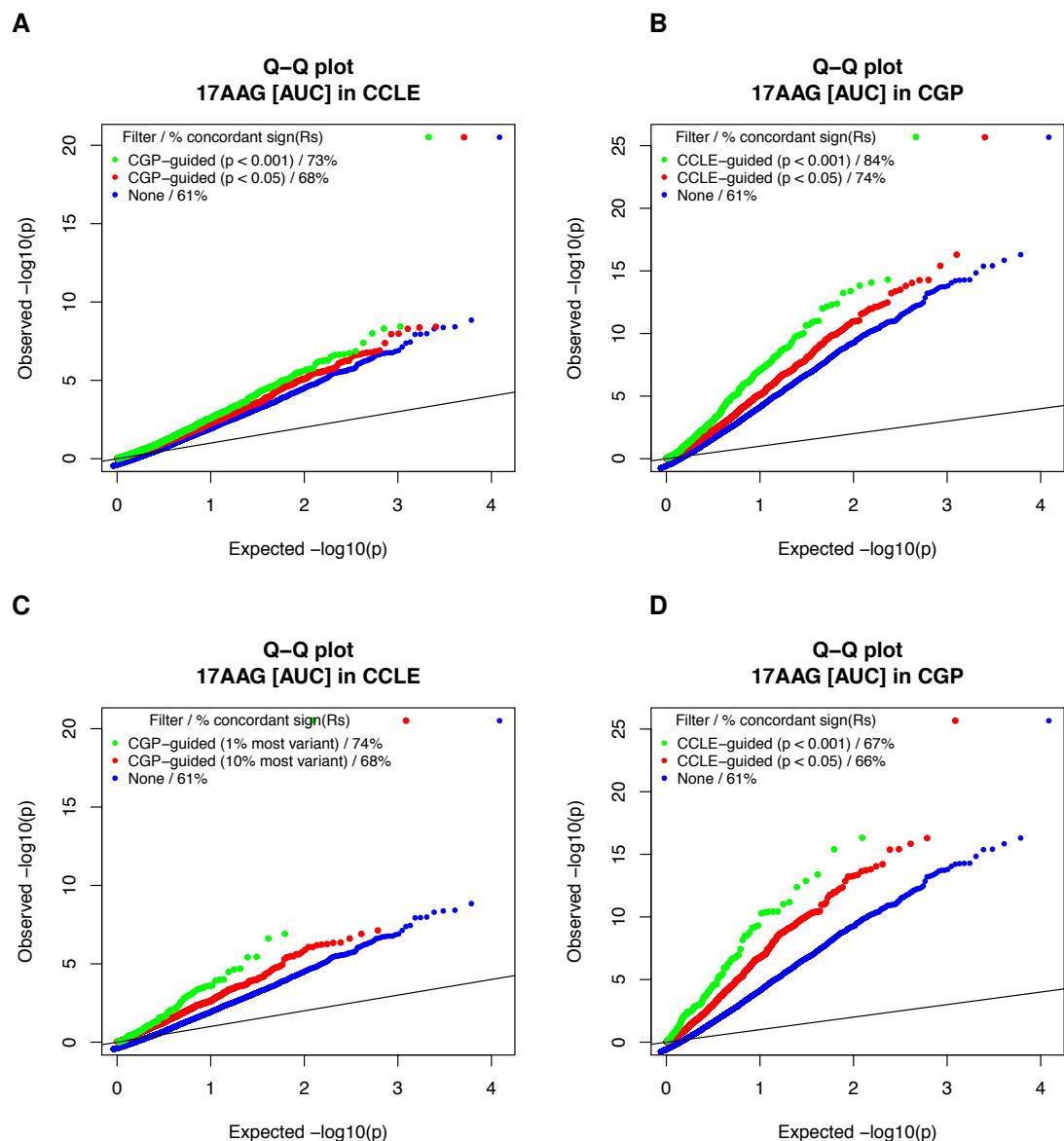
Supplementary Figure 12: Results from source study can improve discovery of novel gene-drug associations in target study for nilotinib. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



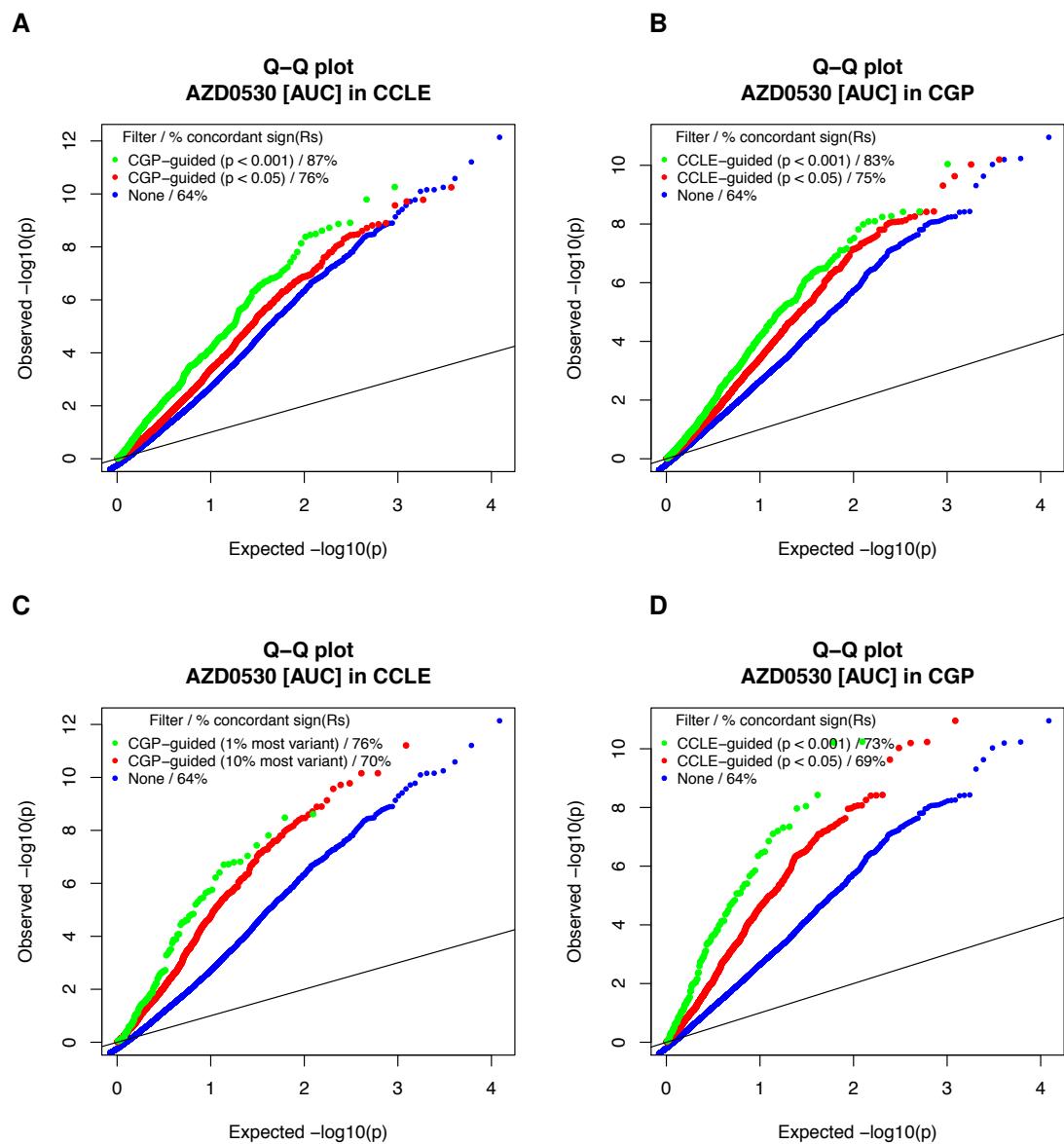
Supplementary Figure 13: Results from source study can improve discovery of novel gene-drug associations in target study for nutlin3. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



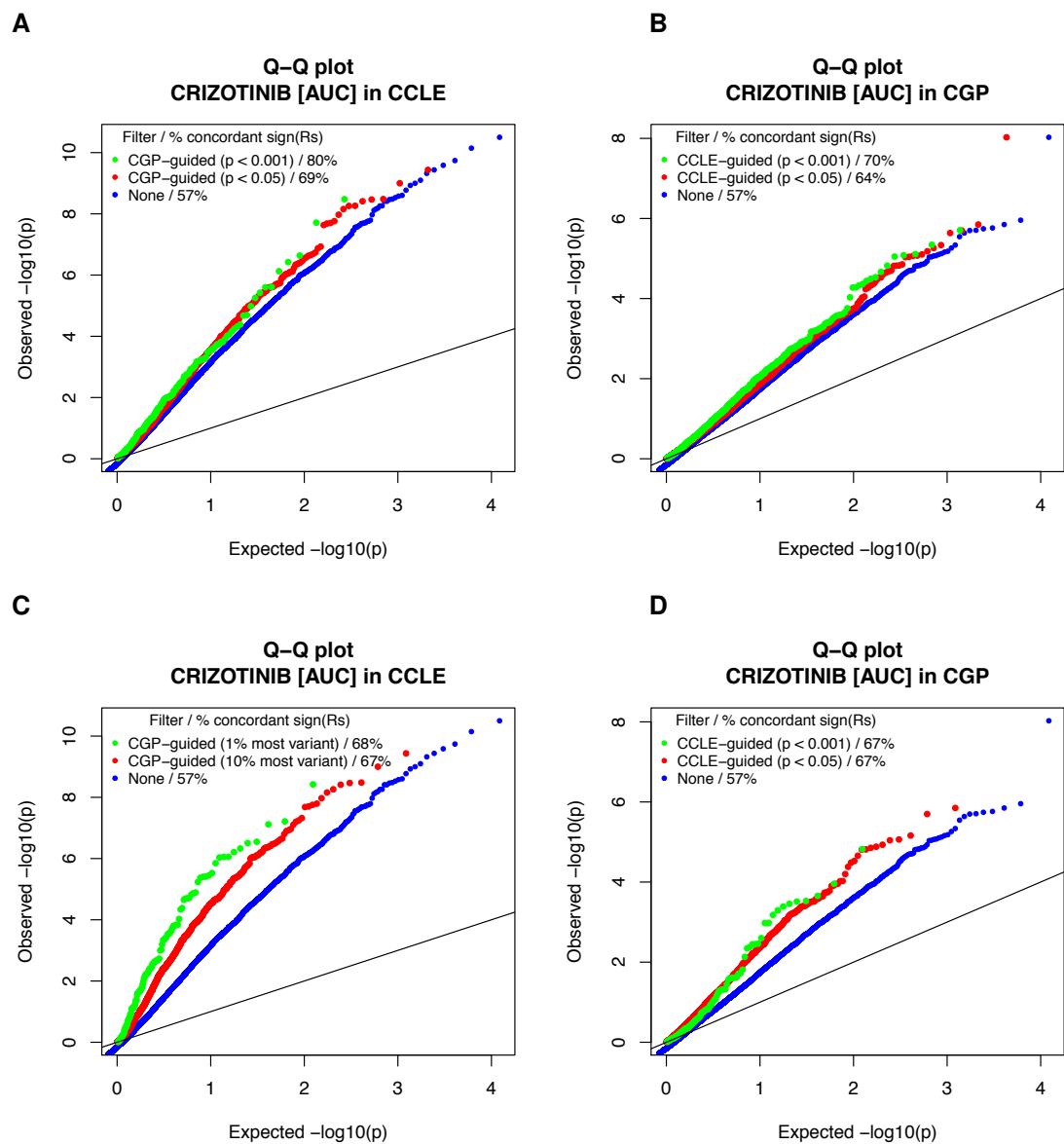
Supplementary Figure 14: Results from source study can improve discovery of novel gene-drug associations in target study for 17AAG. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



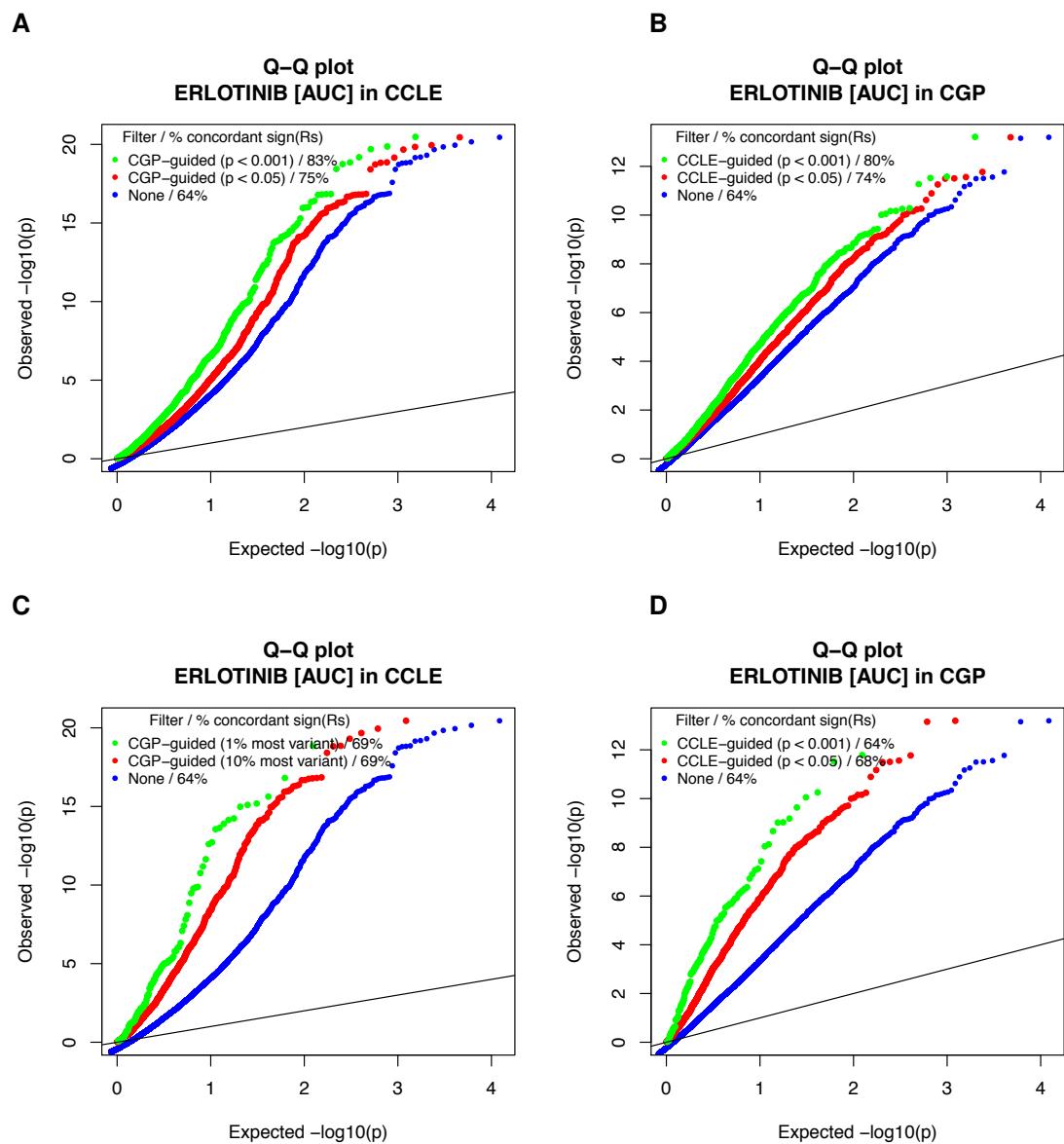
Supplementary Figure 15: Results from source study can improve discovery of novel gene-drug associations in target study for AZD0530. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



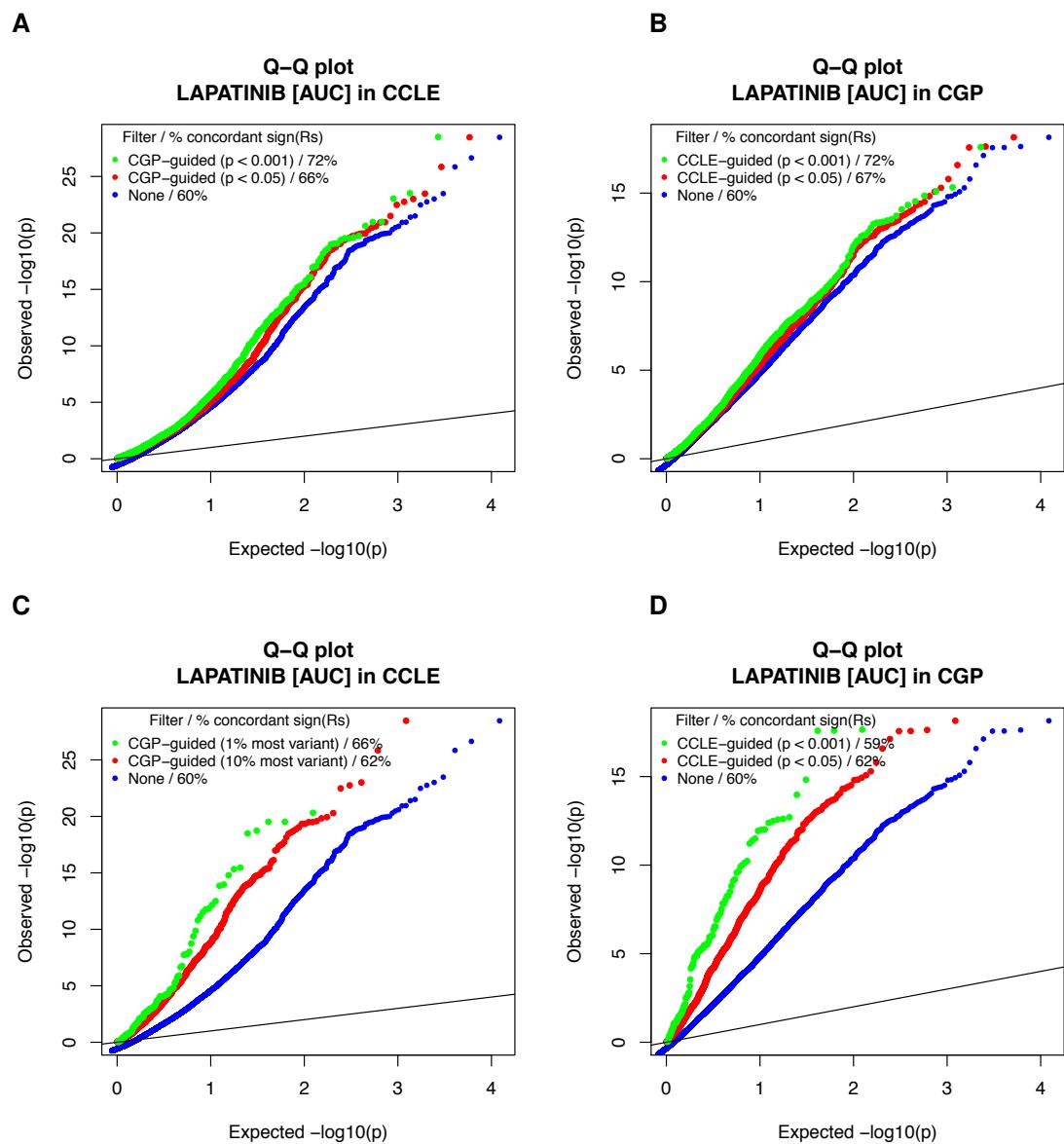
Supplementary Figure 16: Results from source study can improve discovery of novel gene-drug associations in target study for crizotinib. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



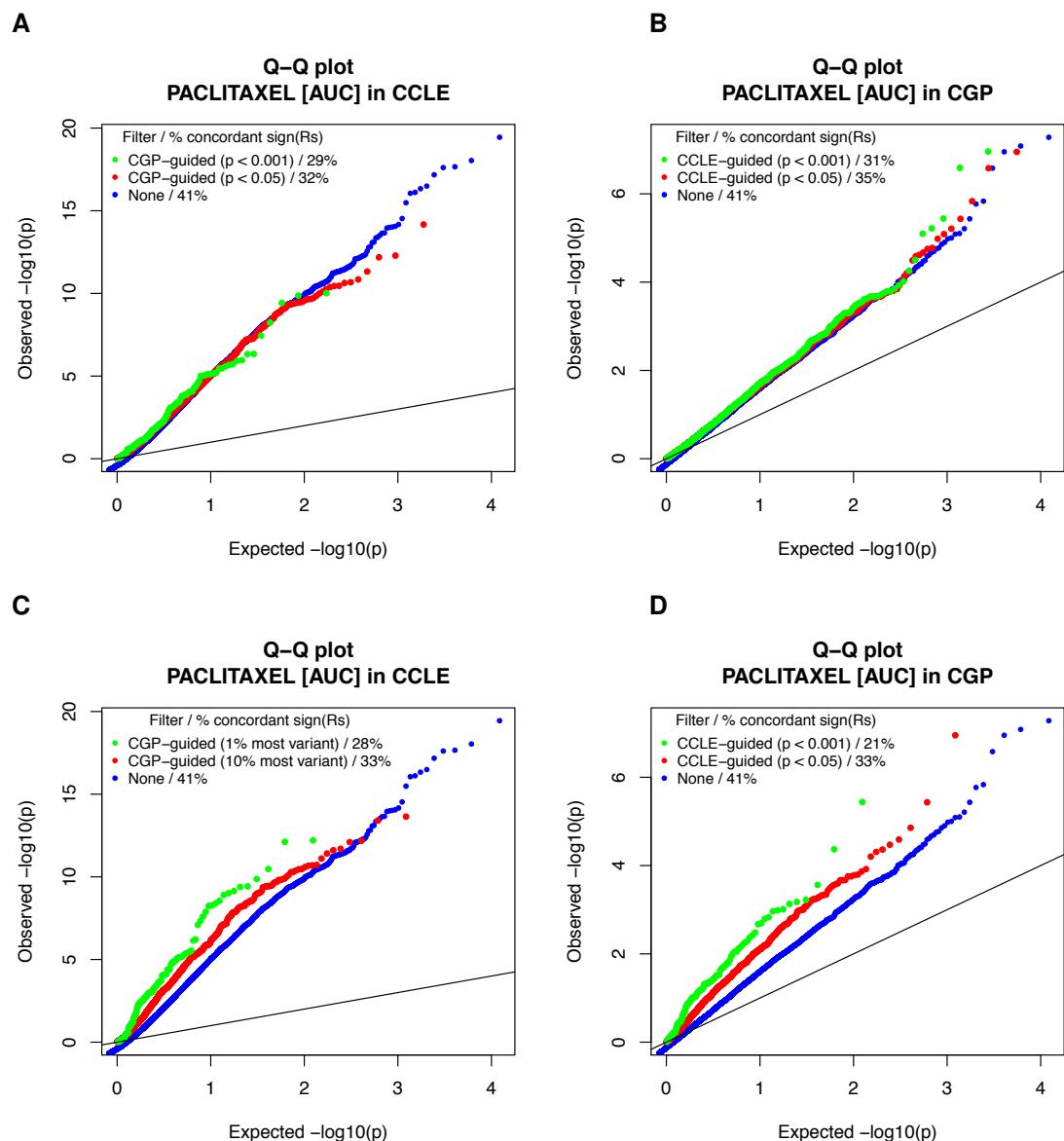
Supplementary Figure 17: Results from source study can improve discovery of novel gene-drug associations in target study for erlotinib. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



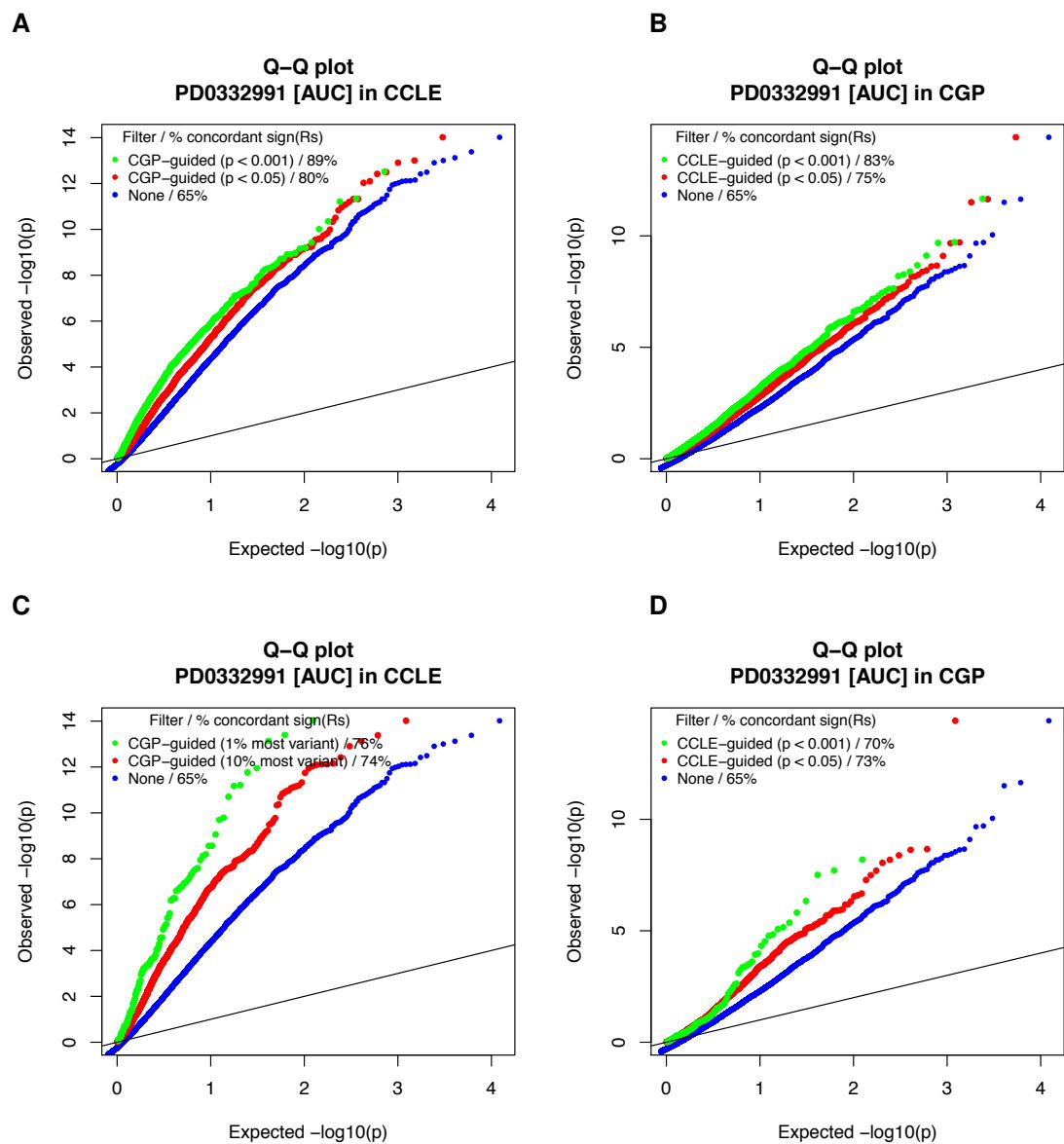
Supplementary Figure 18: Results from source study can improve discovery of novel gene-drug associations in target study for lapatinib. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



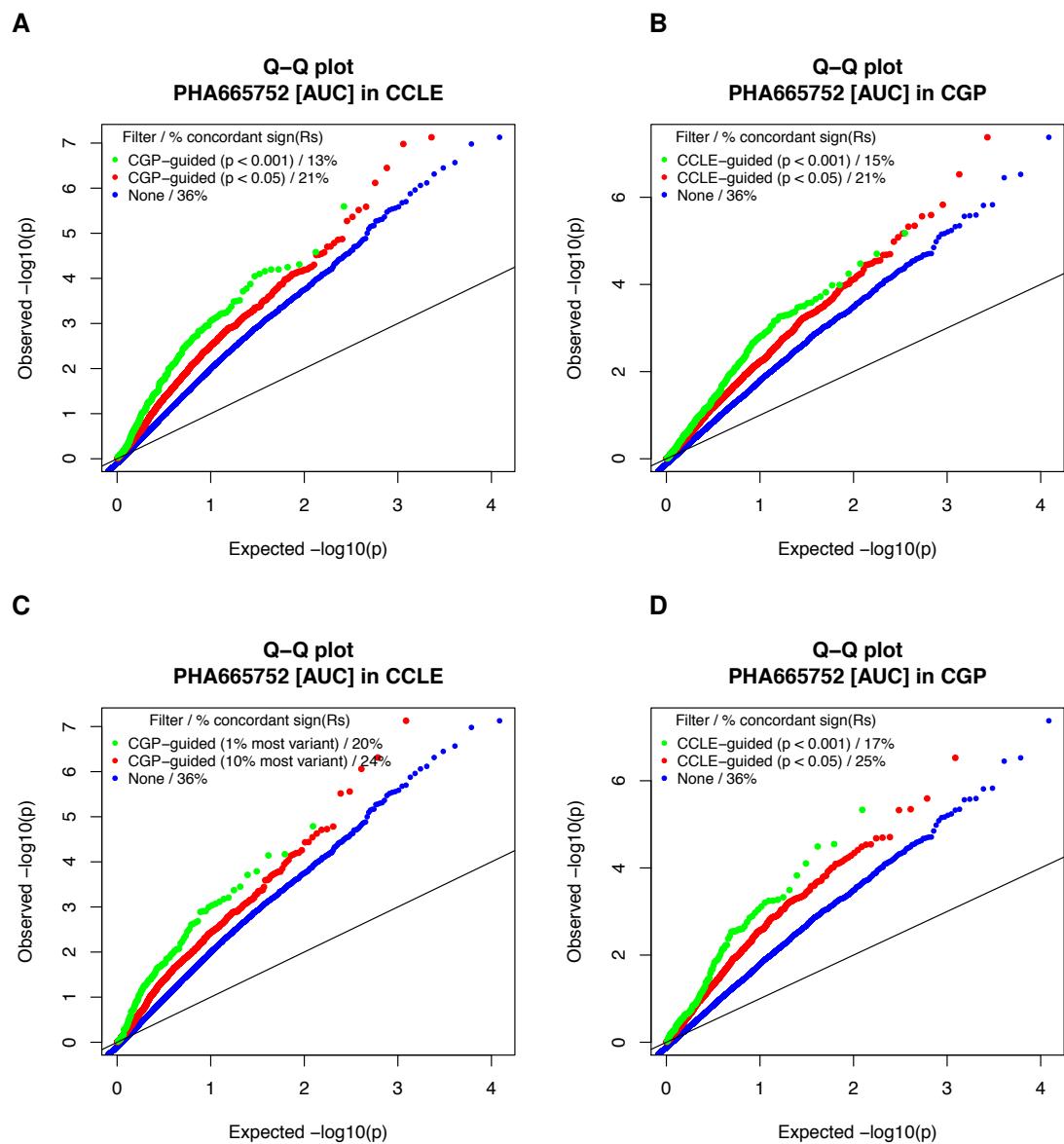
Supplementary Figure 19: Results from source study can improve discovery of novel gene-drug associations in target study for paclitaxel. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



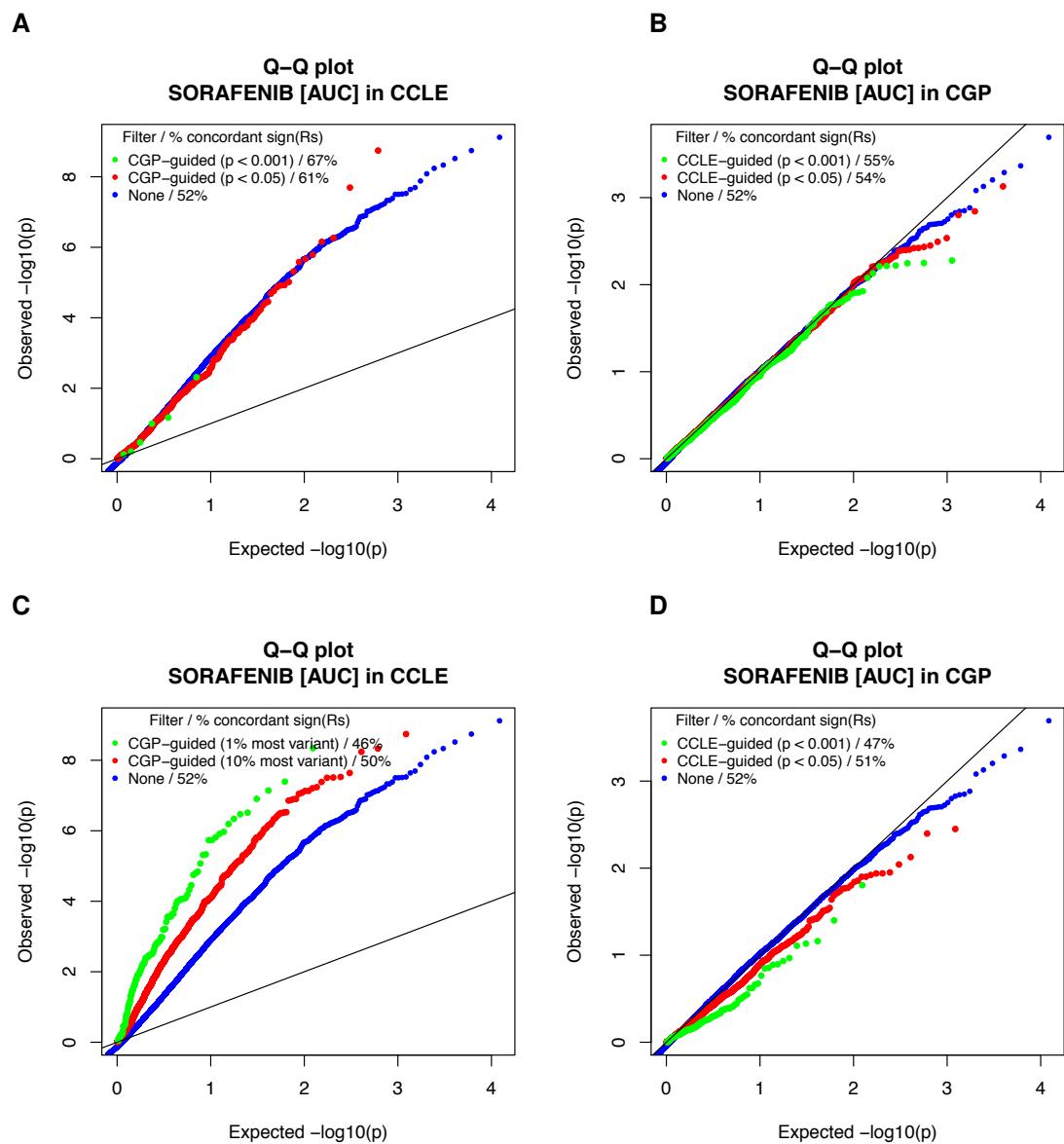
Supplementary Figure 20: Results from source study can improve discovery of novel gene-drug associations in target study for PD0332991. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



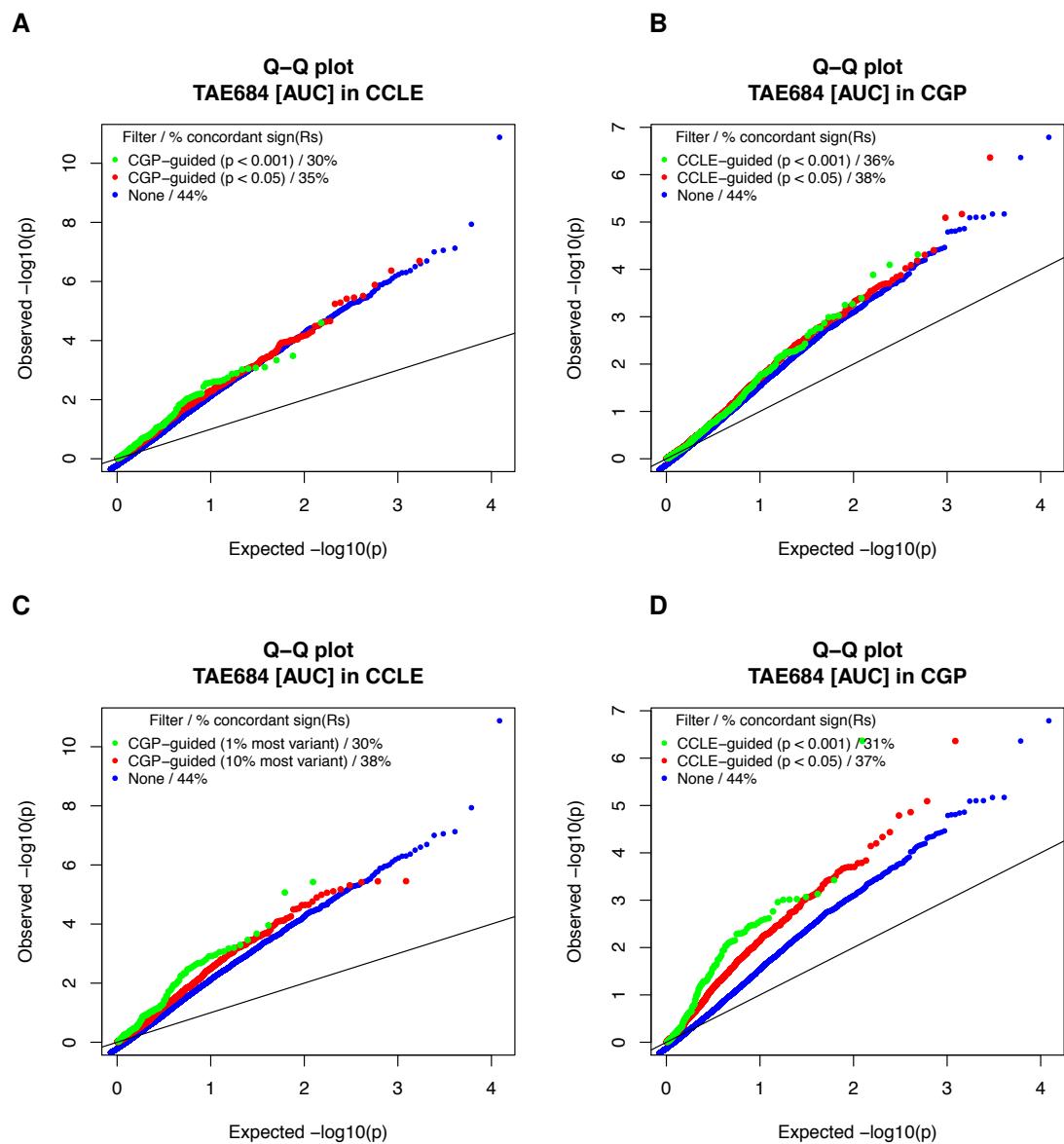
Supplementary Figure 21: Results from source study can improve discovery of novel gene-drug associations in target study for PHA665752. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



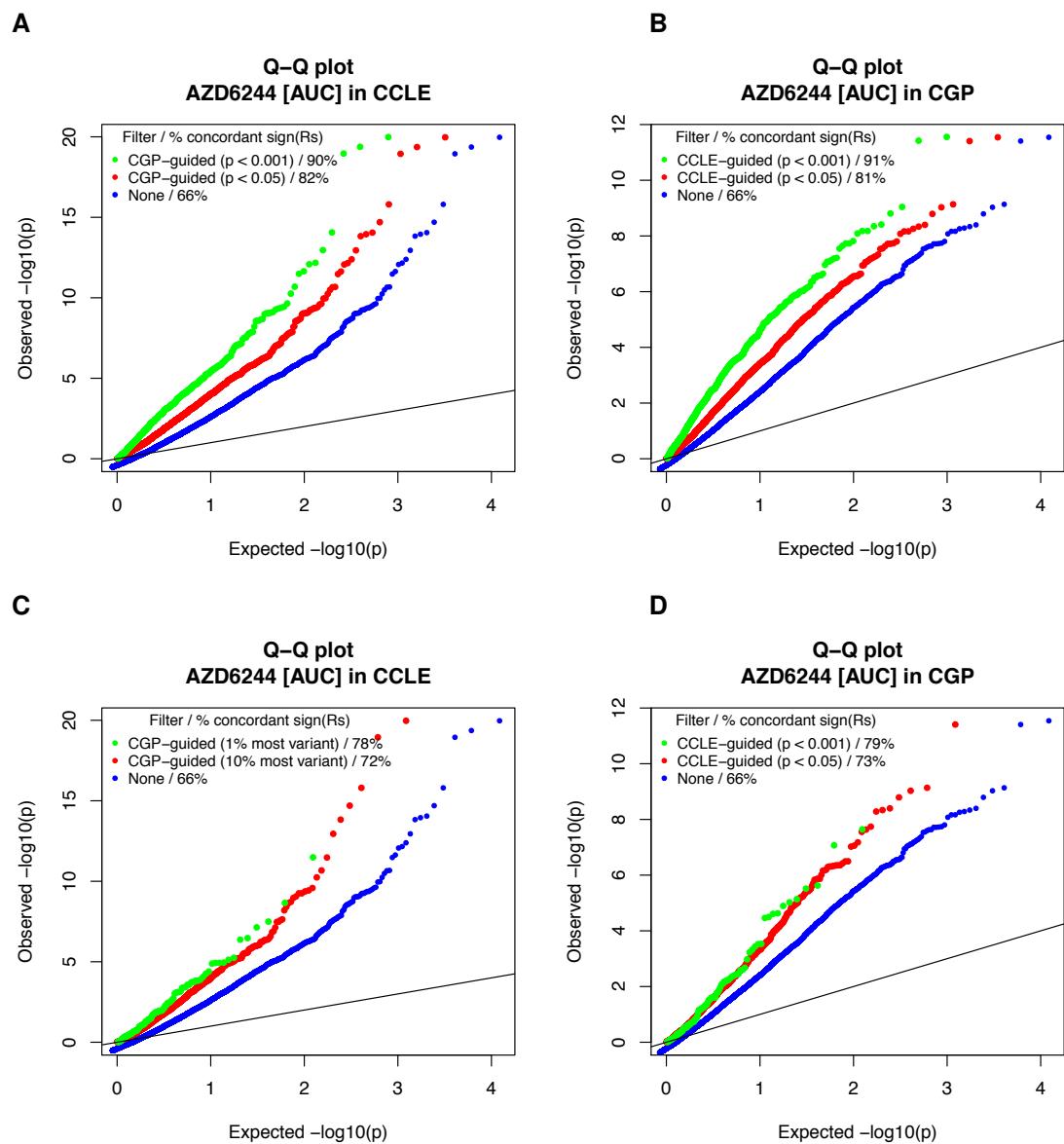
Supplementary Figure 22: Results from source study can improve discovery of novel gene-drug associations in target study for sorafenib. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



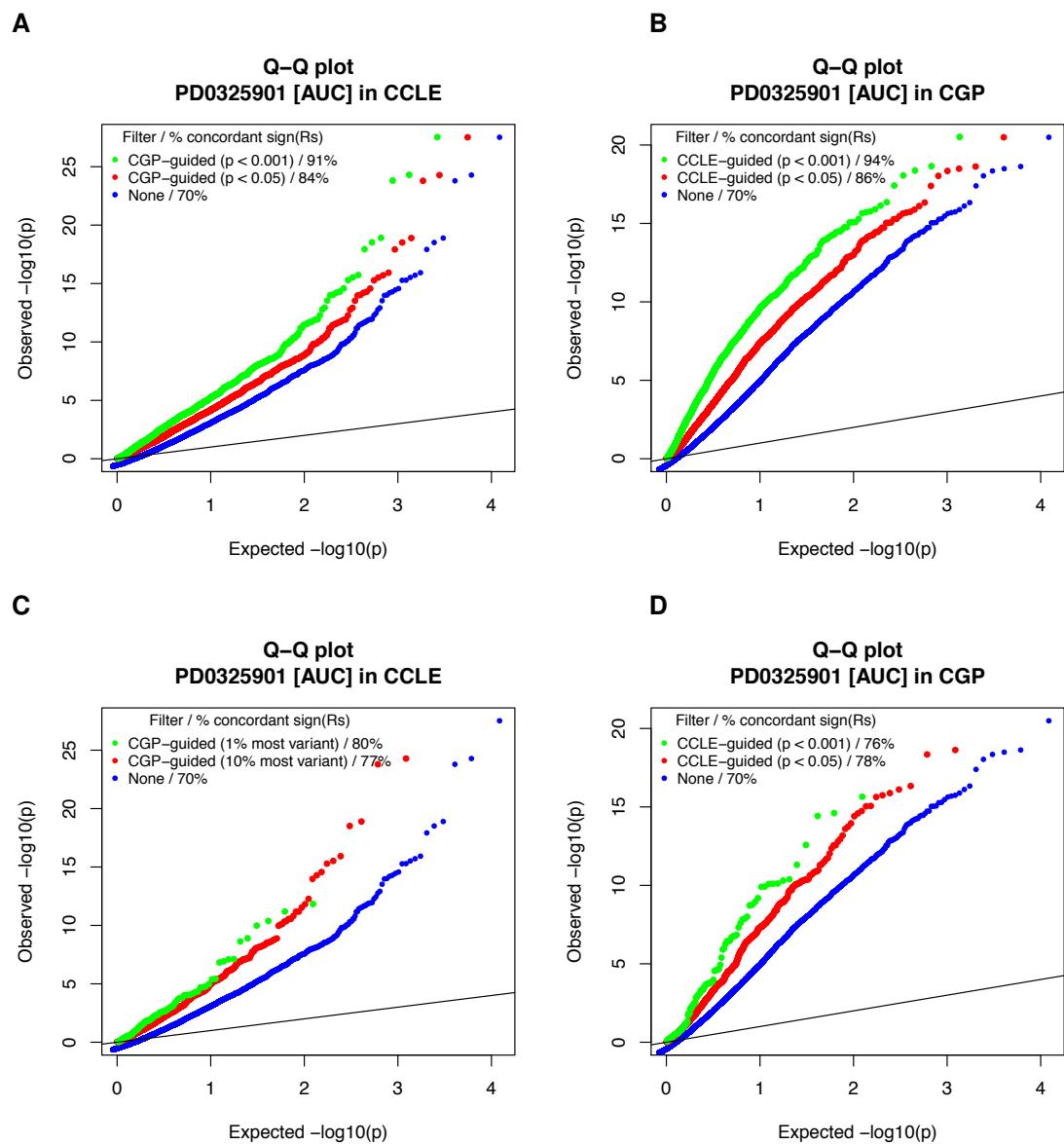
Supplementary Figure 23: Results from source study can improve discovery of novel gene-drug associations in target study for TAE684. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



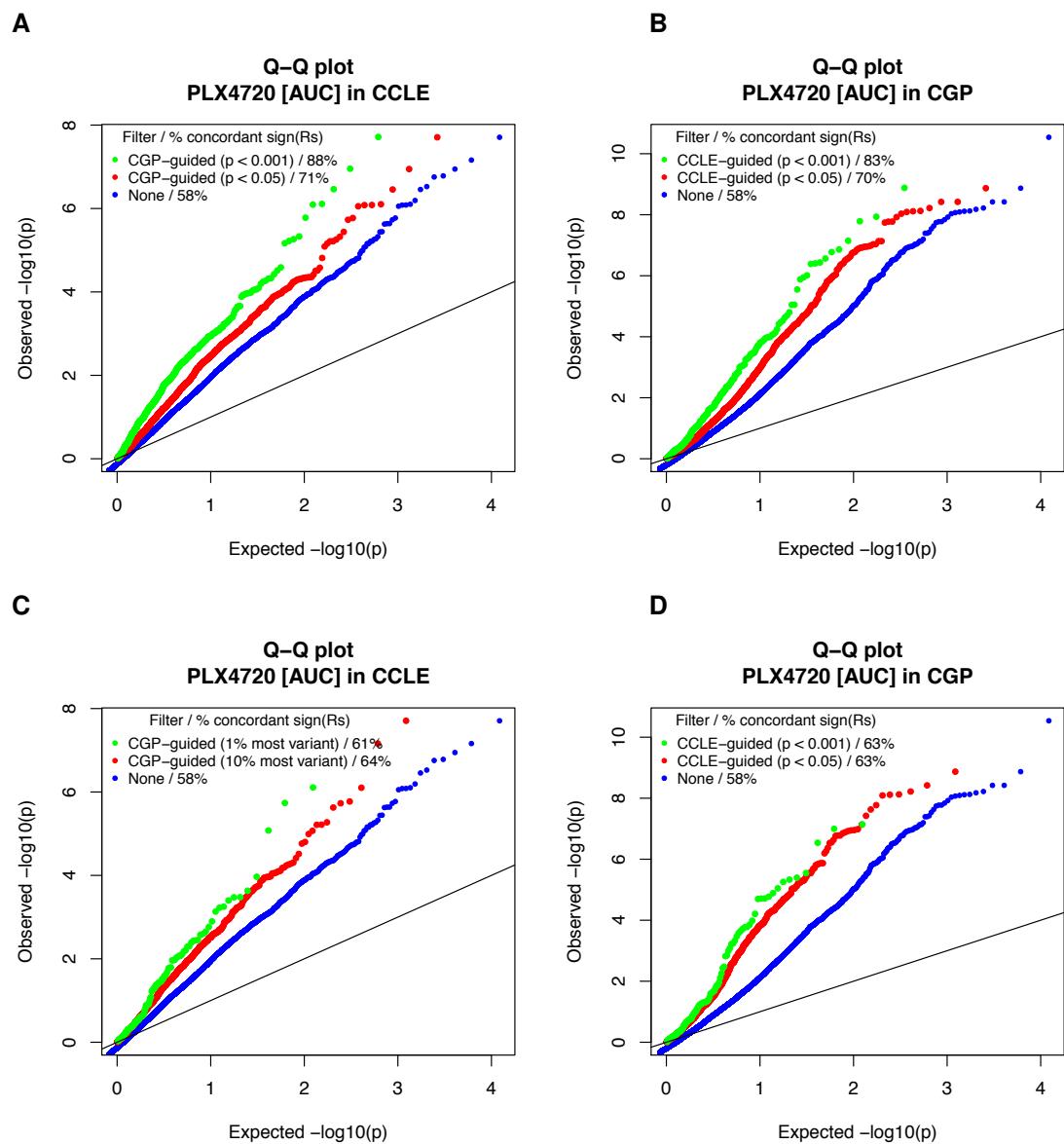
Supplementary Figure 24: Results from source study can improve discovery of novel gene-drug associations in target study for AZD6244. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



Supplementary Figure 25: Results from source study can improve discovery of novel gene-drug associations in target study for PD0325901. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



Supplementary Figure 26: Results from source study can improve discovery of novel gene-drug associations in target study for PLX4720. The Q-Q plots show the expected and observed p-values for the gene-drug Spearman correlations from the full target dataset (blue), and those filtered from the source dataset. **(A,B)** Gene drug-association filtered by significance in the source dataset ($p < 0.05$ and $p < 0.001$ in red and green, respectively); **(A)** Source = CGP, Target = CCLE; **(B)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/'). **(C,D)** Similar analysis but gene drug-association are filtered by variance in expression of the corresponding gene in the source dataset (top 10% and 1% most variant genes in red and green, respectively); **(C)** Source = CGP, Target = CCLE; **(D)** Source = CCLE, Target = CGP. The percentage (%) of gene-drug associations with concordant *direction* (sign of the Spearman's correlation coefficient R_s) is reported on the right side of the legend (separated by '/').



References

- [1] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F Berger, John E Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jané-Valbuena, Felipa A Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H Engels, Jill Cheng, Guoying K Yu, Jianjun Yu, Peter Aspesi, Melanie de Silva, Kalpana Jagtap, Michael D Jones, Li Wang, Charles Hatton, Emanuele Palescandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P. Mesirov, Stacey B Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E Myer, Barbara L Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L Harris, Matthew Meyerson, Todd R. Golub, Michael P Morrissey, William R Sellers, Robert Schlegel, and Levi A. Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, March 2012.
- [2] Mathew J Garnett, Elena J Edelman, Sonja J Heidorn, Chris D Greenman, Anahita Das-tur, King Wai Lau, Patricia Greninger, I Richard Thompson, Xi Luo, Jorge Soares, Qing-song Liu, Francesco Iorio, Didier Surdez, Li Chen, Randy J Milano, Graham R Bignell, Ah T Tam, Helen Davies, Jesse A Stevenson, Syd Barthorpe, Stephen R Lutz, Fiona Kogera, Karl Lawrence, Anne McLaren-Douglas, Xeni Mitropoulos, Tatiana Mironenko, Helen Thi, Laura Richardson, Wenjun Zhou, Frances Jewitt, Tinghu Zhang, Patrick O'Brien, Jessica L Boisvert, Stacey Price, Wooyoung Hur, Wanjuan Yang, Xianming Deng, Adam Butler, Hwan Geun Choi, Jae Won Chang, Jose Baselga, Ivan Stamenkovic, Jeffrey A Engelman, Sreenath V Sharma, Olivier Delattre, Julio Saez-Rodriguez, Nathanael S Gray, Jeffrey Settleman, P Andrew Futreal, Daniel A Haber, Michael R Stratton, Sridhar Ramaswamy, Ultan McDermott, and Cyril H Benes. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, March 2012.
- [3] Paul Gelehrter, Nancy J Cox, and R Stephanie Huang. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome biology*, 15(3):R47, March 2014.
- [4] M O Lorenz. Methods of measuring the concentration of wealth. *Journal of American Statistical Association*, 9(70), June 1905.

Consistency in large pharmacogenomic studies

Paul Geeleher¹, Eric R. Gamazon¹, Cathal Seoighe², Nancy J. Cox^{1*}, and R. Stephanie Huang^{1*}

¹Department of Medicine, The University of Chicago, Chicago, IL 60637, USA; and

²Department of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland

* Address for correspondence and reprints:

Nancy J. Cox, 900 E 57th street, KCBD room 3100, the University of Chicago, Chicago, IL 60637.

Telephone: (773) 702-; Fax: (773) 702-; E-mail: ncox@medicinebsd.uchicago.edu

R. Stephanie Huang, 900 E 57th street, KCBD room 7148, the University of Chicago, Chicago, IL

60637. Telephone: (773) 702-9363; Fax: (773) 702-0963; E-mail:

rhuang@medicinebsd.uchicago.edu

Haibe-Kains *et al.*¹ reported inconsistency between two large-scale pharmacogenomic studies (the Cancer Cell Line Encyclopedia² (CCLE) and the Cancer Genome Project³ (CGP)). Upon careful analysis of the same data we arrived at quite different and much more positive conclusions. Here, we highlight the most important reasons for this.

The authors reported correlation *between* cell lines for gene expression but, inconsistently, *across* cell lines for drug sensitivity (see Methods). On reanalysis, we found much higher correlations between cell lines than across cell lines for both gene expression and drug sensitivity measures (median $r_s = 0.88$ between cell lines, $r_s = 0.56$ across cell lines for expression; median $r_s = 0.62$ between cell lines and $r_s = 0.35$ across cell lines for AUC, a drug sensitivity measure). Thus, by correcting this inconsistency, the correlations for expression and drug sensitivity data were far more similar than was originally reported, severely undermining the authors' interpretation of the relative quality of expression and drug sensitivity datasets.

For gene-drug association analysis, the response data include severe outliers ($P < 3.9 \times 10^{-10}$ from Shapiro-Wilk test for all drug AUC in CGP) and the residuals in the linear regression models deviate dramatically from the normal distribution. Violation of the “normality assumption” may cause unstable models⁴ rendering the concordance analysis premised on the fit of these models statistically questionable.

Another important issue is that because of the highly targeted nature of many of the drugs assessed, the reported Spearman correlation coefficients do not reflect the true concordance of drug sensitivity between the two studies. To see why correlation is not an appropriate measure of biological concordance for these data, consider the extreme case of a drug that is not effective against any cell lines. In such a case the measurement error, inherent in biological assays, will dominate over the (non-existent) biological variability between the drug sensitivities of the assays and there could be no expectation of correlation between the drug sensitivity measures from the two studies. Many of the drugs studied were highly-targeted agents requiring, by design, specific molecular targets for response, resulting in low biological variability across most cell lines and consequently low Spearman correlation values. Consider nilotinib, which was reported as exhibiting “poor consistency” between CGP and CCLE ($r_s = 0.1$ for AUC); it targets the *BCR-ABL1* fusion-gene and in CGP, *BCR-ABL1* status was associated with drug sensitivity ($P = 2.54 \times 10^{-65}$), accurately reflecting the known biology. *BCR-ABL1* status was not reported by CCLE, however, three *BCR-ABL1*-positive cell lines were among the 189 nilotinib-treated cell lines that overlap CCLE, where we also found them to be the three most sensitive samples ($P = 9 \times 10^{-7}$). Despite the fact that these drug sensitivity data were accurately reflecting biological expectations in both studies, the authors’ criteria surprisingly describe nilotinib sensitivity as *discordant*. Of 577 cell lines screened in CGP only four were *BCR-ABL1* positive (across all cell lines median AUC = 0.99; $\sigma^2 = 4 \times 10^{-3}$; AUC of 1 represents no drug response; Fig. 1(a)). Clearly, given no expectation of response in almost all samples, there is little biological

variability in drug response across most cell lines, resulting in low correlation. Furthermore, given that most drugs compared in the initial publication were targeted agents, this lack of drug response variability is *common*; indeed, for 9 of the 15 drugs, median AUC is greater than 0.95 in CGP. Indeed, there is a *systematic* relationship between variability in drug response in either study and correlation between the two studies (Fig. 1(b)). A valid comparison of CGP and CCLE must consider the pharmacology of the drugs screened and in particular the differences in the variability induced by different drugs. Nilotinib is not an isolated case; in fact expected⁵⁻¹⁰ genomic associations were recapitulated in both CCLE and CGP for many drugs including *ERBB2* expression for lapatinib, *NQO1* expression for 17-AAG and *BRAF* mutation for PD-0325901, PLX4720 and AZ6244. Finally, recent findings that expression and drug sensitivity data in CGP could reliably predict drug response in multiple clinical trials further support the utility of these data¹¹.

Methods

In CGP and CCLE, using ordered data common to both studies, gene expression and drug sensitivity (AUC) values can be arranged in $n_1 \times m$ and $n_2 \times m$ matrices respectively where m is the number of cell lines, n_1 is the number of genes and n_2 is the number of drugs common to both studies. Correlations “between” cell lines are calculated by the correlation of matching columns of CGP and CCLE matrices (vectors of length n_1 for expression or n_2 for AUC). Correlations “across” cell lines are the correlations of matching rows (vectors of length m for both data).

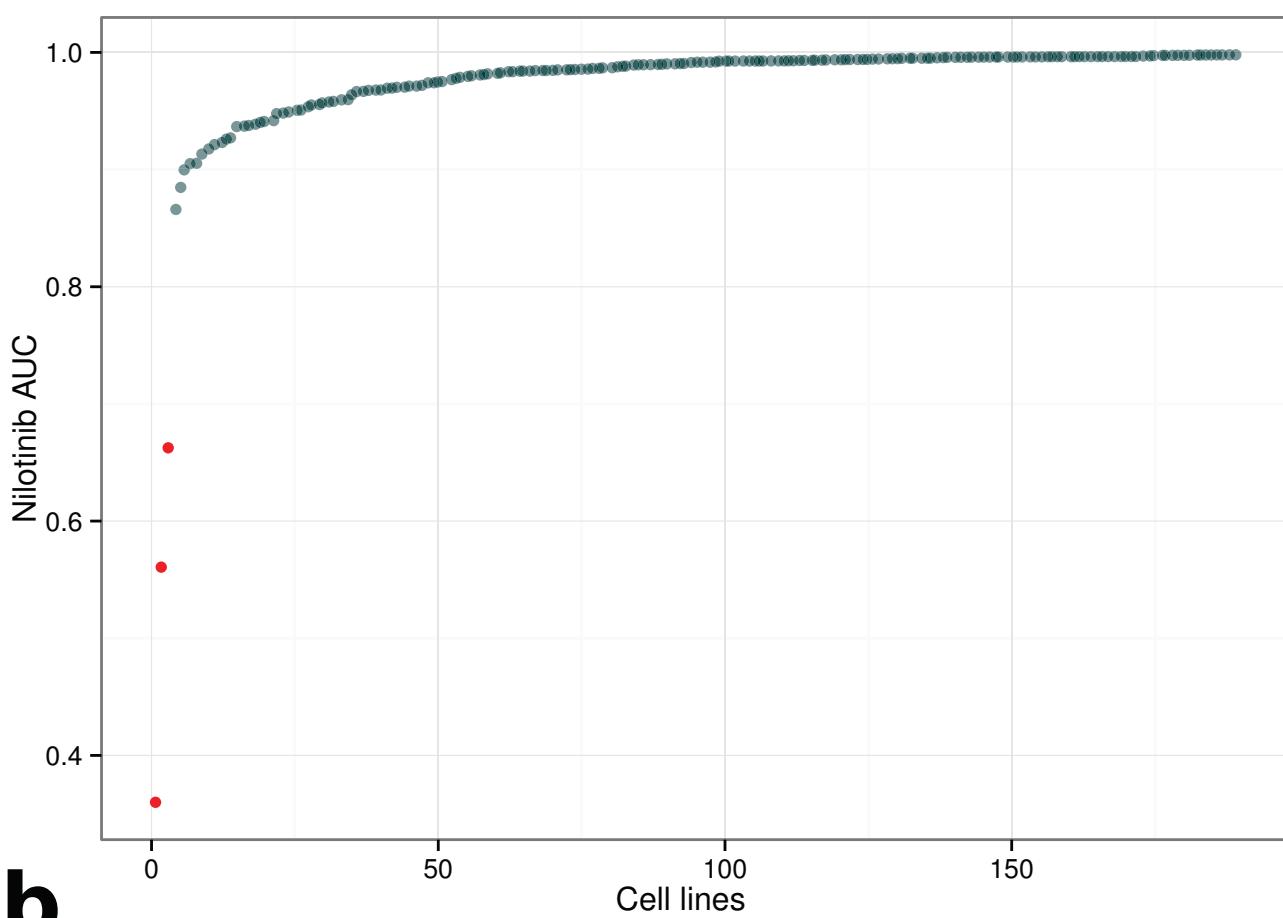
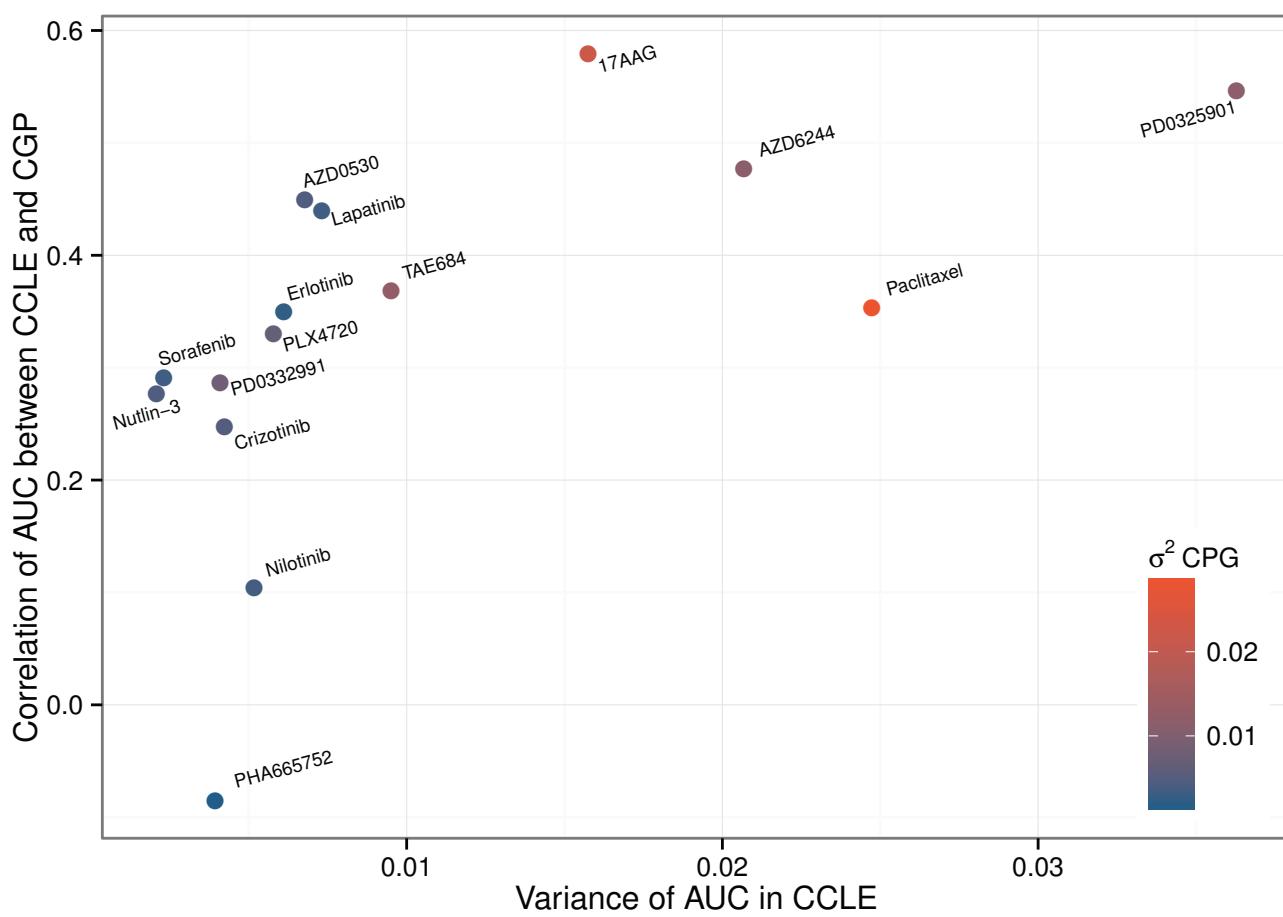
Figure Legends

(a) Highly targeted agents (nilotinib) highlight a major limitation of the authors' test for concordance. Scatterplot showing the nilotinib AUC values (in CGP) for the 189 cell lines that were screened by both CGP and CCLE. Only a very small proportion of cell lines achieve a response, e.g. the three BCR-ABL1 positive cell lines, which are highlighted in red. This almost complete lack of biological variability renders a Spearman correlation ineffective as a means to assess concordance.

(b) The authors' test for concordance is confounded by variability in drug response. Scatterplot showing the strong association between “Spearman’s correlation of AUC between CCLE and CGP” and “variance of AUC in CCLE”. Drugs whose AUC is more variable are more likely to be highly correlated between CCLE and CGP ($r_s = 0.83, P = 1.9 \times 10^{-4}$). The points have been color coded by their “variance of AUC in CGP”, which is also significantly associated with both “variance of AUC in CCLE” and “Spearman’s correlation of AUC between CCLE and CGP”.

References:

1. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–93 (2013).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–7 (2012).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–5 (2012).
4. Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to linear regression analysis*. **821**, (John Wiley & Sons, 2012).
5. Kelland, L. R., Sharp, S. Y., Rogers, P. M., Myers, T. G. & Workman, P. DT-Diaphorase expression and tumor cell sensitivity to 17-allylaminoo, 17-demethoxygeldanamycin, an inhibitor of heat shock protein 90. *J. Natl. Cancer Inst.* **91**, 1940–9 (1999).
6. McDermott, U. *et al.* Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19936–41 (2007).
7. Konecny, G. E. *et al.* Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res.* **66**, 1630–9 (2006).
8. Müller, C. R. *et al.* Potential for treatment of liposarcomas with the MDM2 antagonist Nutlin-3A. *Int. J. Cancer* **121**, 199–205 (2007).
9. Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3041–6 (2008).
10. Solit, D. B. *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **439**, 358–62 (2006).
11. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).

a**b**

Re: Nature 2014-03-03927

Dear Dr. Marte,

Thank you for overseeing the review of our Brief Communications Arising manuscript entitled “Consistency in large pharmacogenomic studies”. After consideration of the points made by Haibe-Kains and colleagues, we feel strongly that our concerns with the study still stand. We appreciate you granting us the opportunity to revise our article. Please see below our detailed responses to the points raised by Haibe-Kains and colleagues.

Point 1: While we agree that there are anecdotal cases for some targeted agents, such as crizotinib and nilotinib, that demonstrate consistent identification of a few highly sensitive cell lines shared by the CGP [1] and CCLE [2], our initial conclusions about the broader inconsistency in reported phenotypes between the CGP and CCLE still holds [3].

Response 1: We are glad that the authors now acknowledge that the cell lines that were expected to be targeted by nilotinib and crizotinib were consistently identified in CCLE and CGP, despite the fact that Spearman correlation for these drug sensitivity data were originally described in their study as showing “poor consistency” (based on the qualitative description of the data defined by the authors). Nilotinib represents an especially interesting example because the Spearman correlation for drug sensitivity data reported in the authors’ Nature publication is particularly low ($r_s = 0.1$ for AUC; the second worst correlation of all drugs and not statistically significant). However, for the cell lines that overlap both studies, concordance is clearly high and in these data both studies are reflecting exactly what would be expected; i.e. that a small number

of BCR-ABL1 positive cell lines are sensitive to nilotinib, but the majority of cell lines are resistant. The association with BCR-ABL1 is also highly statistically significant in both studies.

However, despite the acknowledgement, the authors dismiss the nilotinib observation as “anecdotal”; but this is clearly **not** the case (see also our “Response 4” below for expected associations identified by both studies for other drugs). In fact, the majority of the drugs are *targeted* agents as emphasized in our revised manuscript, and therefore by design, there is very little variance in the sensitivity values for these drugs across most of the cell lines. Of the 15 drugs for which the authors have assessed concordance, the median AUC value in CGP is greater than 0.95 in 9 of them (an AUC of 1 represents no drug response), indicating that (as expected) very high numbers of cell lines are unresponsive to these drugs. **In this context, the Spearman correlation is not a sensible metric for the concordance between the two studies;** for the resistant cell lines (the majority of the cell lines for most drugs), the only measured difference in AUC or IC₅₀ is technical variability, which exists in all biological assays and is not expected to be correlated between repeated measurements. Put equivalently, in these data involving highly targeted agents, the measurement error, inherent in most biological assays, will dominate over the (non-existent) biological variability between the drug sensitivities of the assays and there could be no expectation of correlation between the drug sensitivity measures from the two studies. In our manuscript, we present this crucial observation in Fig 1(b), which clearly demonstrates that there is a *systematic* relationship between the level of variability in the drug response phenotype (AUC) and the observed Spearman correlation between CCLE and CGP; with the higher the drug response variability, the more correlated these phenotypes are from CCLE and CGP. Therefore, **these low Spearman correlations reported in the authors’**

original publication are easily explained by low variability in the drug response phenotypes and are *not* a result of discordance of true drug sensitivity. This is an extremely important point and is the key oversight in the authors' interpretation of their results.

Nilotinib represents an example where the level of variability in drug response is extremely low, but is not anecdotal or an isolated case. The low variability stems from the fact that this drug (like many of the 15 drugs assessed) is designed to target only a very small proportion of the cell lines screened (i.e. those containing BCR-ABL1). Hence, the overwhelming majority of cell lines are unaffected by treatment, as would be expected given the drug's mechanism of action. Again, this is crucially important because any biological assay that measures a continuous phenotype (such as gene expression or drug sensitivity) will have some associated technical variability (i.e. measurement error). If there is no biological variability (for example, all cell lines are resistant to a drug) there is no expectation that repeated measurements will be correlated.

Point 2: In our initial publication, we computed the correlation between gene expression and mutation profiles of cell lines to assess whether large transcriptomic changes and/or genetic drift might be the cause of the observed inconsistency in drug sensitivity data, and subsequently the gene-drug associations [3]. We agree with the authors that correlations across and between cell lines should be thoroughly compared. Overall concordance across cell lines was lower than concordance between cell lines; however we still observed that gene expression and mutation data were significantly more concordant than IC₅₀ and AUC values in all comparisons (Wilcoxon rank sum test p<0.01), except for agreement of presence/absence of mutations and AUC sensitivity calls across cell lines (Cohen's kappa: 0.28 vs. 0.23 for mutations and AUC

sensitivity calls, p=0.16). In light of these results, our conclusion that genomic data are significantly more concordant than pharmacological data still holds, although the difference under these assumptions is less pronounced than reported in our initial publication.

Response 2: We thank the authors for acknowledging that the correlations for drug and expression data should be reported consistently and that for the drug sensitivity data “overall concordance across cell lines [as originally reported] was lower than concordance between cell lines [as needed to be reported for consistency]”. This simple oversight, however, has turned out to have substantial implications for the interpretation of their results and thus cannot be easily dismissed. When the results are calculated consistently, the correlations for drug and expression data are *far* more similar than was reported in the authors’ original paper. If expression data are reported in the same direction (across samples) as drug sensitivity data, the median Spearman correlation drops from 0.85 to 0.56, which instead of satisfying the authors’ ad-hoc definition of “almost perfect consistency”, now satisfies their definition of “fair consistency” and is only 0.06 short of their definition of “poor consistency”. In light of the authors acknowledging this key analytical error, the research community deserves to learn about this correction and the impact of these new results in the re-interpretation of their findings.

Furthermore, this analytical error becomes problematic for the authors’ overall interpretation because it serves to highlight the authors’ seemingly arbitrary qualitative descriptions of the correlations, which seems to have at least in part been defined based on the Spearman correlation of 0.85 that was originally reported for expression. Also, neither the concordance assessment for gene expression nor drug sensitivity account for the fact that the expression of a very large

number of genes would be expected to be highly variable across a diverse set of cancer cell lines that arise from a total of 51 different tissue types (defined by CGP). Unsurprisingly, for expression data (just like drug sensitivity, as in Response 1), correlation of any given gene between the two studies is strongly dependent on the “across sample” variability of gene expression ($r_s = 0.69$, $P < 2.2 \times 10^{-16}$ in CGP and $r_s = 0.67$, $P < 2.2 \times 10^{-16}$ in CCLE). The Wilcoxon rank sum test based statistical test presented by the authors above may demonstrate that *correlation* is slightly higher for expression data (in this particular dataset), however our key point (as explained in Response 1 and, indeed, also as relevant to the authors’ reply here) is that ***correlation cannot be used to fairly assess concordance*** between these data, due to the dependence of correlation on “across sample” variability. Thus, the new tests presented by the authors here to justify a higher concordance for gene expression than for drug sensitivity are also spurious and, indeed, cannot be used to show that *concordance* is higher for either dataset. Here again, in their reply, the authors repeat the same analytical error as in the original publication and fail to take into account “across sample” variability. Finally, we should note that this failure (as in the failure described in Response 1) highlights a key difference between “correlation” and “concordance”, and, in these cases, the correlation coefficient used is not sensible for the concordance test intended.

Point 3: We agree with the authors that the normality assumption of linear regression is violated for sensitivity data of highly targeted drugs. However our goal was to follow the approach used by the CGP and CCLE investigators to infer gene-drug associations (elasticnet linear regression). The authors of the current manuscript also used linear regression models (ridge) [4] to develop genomic predictors of response to cytotoxic drugs (such as paclitaxel in our

publication), and presented an ad hoc logistic ridge regression model, using the 15 most sensitive vs. the 55 most resistant CGP cell lines, for targeted drugs. This suggests that discretizing AUC measurements into sensitivity calls would yield more robust gene-drug associations. However, we reported in our comparative study that the concordance of AUC sensitivity calls (computed using the Waterfall method [2]) is poor for all drugs [3]. Therefore, selecting extreme cases for model fitting is unlikely to result in robust predictors of drug response for the vast majority of cases.

Response 3: We thank the authors for acknowledging that the gene-drug association analysis breaks the assumptions of linear regression. It is clearly important that this issue is highlighted if only to prevent future studies from incorrectly fitting linear models to drug sensitivity data (for highly targeted drugs) that are severely skewed. Importantly for the authors' paper, their failure to confirm that the assumptions of linear regression are met in their gene-drug association analyses has made an already weakened study (on the basis of the analytic errors and other considerations discussed above) even more suspect analytically. The authors correctly state that some in our group have previously proposed a logistic regression model for highly targeted drugs, although there are many other methods that could potentially be used for this type of data. Possible approaches include random forests and artificial neural networks, which do not impose assumptions on the distribution of the response variable and this will likely become a focus of future research. However, the discussion of which methods may be most suitable depends on the objective of a study (e.g., prediction, as in our published study). This is not relevant to this debate, because the authors have used linear models to assess concordance and **it remains clearly questionable, and undermines the validity of the conclusions drawn by the authors,**

to use the fit of a linear model to assess the concordance given the skewed distribution of these drug sensitivity data.

The authors have ignored the pharmacology of the drugs evaluated (i.e. their highly targeted nature and the resulting imbalanced data) and have consequently overlooked or failed to address some important statistical issues in their analyses. Although future research is likely to propose and test better analytical methods (such as, possibly, those that impose no strict requirement on the distribution of the response variable), it remains the case that the authors' study failed to consider the impact of the violations of the assumptions required for their gene-drug association analyses, and any far-reaching conclusions from the authors' study (such as how investment in this field of science should be allocated, as in the authors' conclusion concerning the development of better biological assays) must be re-cast in light of this failure.

Point 4: We agree that lack of variability in drug sensitivity measurements may prevent biologically meaningful assessment of concordance between pharmacogenomic datasets. The authors focused their discussion on nilotinib for which there are three sensitive cell lines in both datasets. Even among these cell lines the AUC values are not concordant; the least sensitive cell line in CGP is actually the most sensitive one in CCLE. Therefore the only way to consider these results concordant is to classify these three cell lines as sensitive and the rest as resistant, which cannot easily be done using an the Waterfall approach [3] and as noted previously, limits sensitivity and robustness. We agree with the authors that the strong association between BCR-ABL1 gene fusion and sensitivity to nilotinib is reproduced in both studies. However this observation is anecdotal and is not generalizable to the other targeted and cytotoxic drugs

investigated in our original report. For these drugs, there are multiple cell lines exhibiting high sensitivity in only one of the datasets. Moreover, for drugs with large variability in measured sensitivity, we observed only fair to poor correlation for these drugs ($r_s < 0.6$), which supports our report of inconsistencies between CGP and CCLE datasets [3].

Response 4: The authors state that “We agree with the authors that the strong association between BCR-ABL1 gene fusion and sensitivity to nilotinib is reproduced in both studies” but worryingly dismiss the observation as “anecdotal”. As stated above, Fig 1(b) from our manuscript shows that the nilotinib example is NOT anecdotal; the observed “high concordance but low correlation” is easily explained by this drug’s very low variability in drug response due to its highly-targeted mechanism of action. These data show that correlation of drug sensitivity data between CCLE and CGP is highly dependent on phenotypic variability in either study (Fig. 1(b)), thus illustrating a systematic relationship between higher level of variability in drug response and the tendency for that drug to be highly correlated between the two studies. Moreover, including nilotinib and crizotinib, many expected associations between genomic and drug sensitivity data were consistently recapitulated in both studies; this is even in spite of the fact that the genomic features interrogated by both studies were not identical, nor were they measured using identical technologies. Listed below are examples of some genomic associations with drug sensitivity that were reproducibly identified in both studies and were also consistent with biological expectation (references included):

17-AAG: *NQO1* expression was identified as most predictive by both studies [1].

AZ6244: *BRAF* mutation was identified by both studies [2].

Lapatinib: *ERBB2* expression was identified as most predictive by both studies [3].

Nutlin-3: The gene *MDM2* was identified by both studies [4].

PD-0325901: *BRAF* mutation was identified by both studies [5].

PLX4720: *BRAF* mutation was identified by both studies [6].

It should be noted that the authors appear to agree with our key point in their extended reply, stating that “We concur with the authors that highly targeted drugs tend to yield lower Spearman correlation coefficients, likely due to the lack of variability of the corresponding sensitivity data”. As we have been emphasizing throughout, the pharmacology of the drugs evaluated was overlooked in the authors’ study; this clearly affects the final conclusion of the initial publication and therefore needs to be brought to the attention of the readers of Nature.

The authors also claim that they observed only “fair to poor” correlation ($r_s < 0.6$) for drugs with higher variability; however, as we highlighted above (in Response 2), this correlation is approximately equal to the median correlation observed for expression data ($r_s = 0.56$) when the results are consistently reported. Worryingly, the qualitative descriptions defined by the authors are misaligned with correlations previously reported for similar types of data; for example the median Spearman correlation (across samples) was 0.2 of repeated measured gene expression from the HapMap YRI cell lines [7][8]. This comparatively low correlation was observed despite the fact that both studies used RNA-seq, a technology which is generally accepted to provide a more accurate estimate of true expression levels than microarrays (which were used by CCLE and CGP). The lower correlation likely owes to less variability in gene expression in

homogenous lymphoblastoid cell lines compared to a highly heterogeneous set of cancer cell lines.

However, the key point remains that even if the level of *correlation* was much lower for drug sensitivity (which it is not), it still could not be concluded that *concordance* for drug sensitivity data is lower, again due to the dependence of correlation of these measurements on variability.

Additional Note: The authors state above that “Even among these cell lines the AUC values are not concordant; the least sensitive cell line in CGP is actually the most sensitive one in CCLE”. To clarify, this statement is referring to the *three* BCR-ABL1 positive cell lines that were the most sensitive cell lines in both CCLE and CGP for the nilotinib-treated cell lines overlapping both studies. The statement is highlighting that the most sensitive of the 189 cell lines in CGP is third most sensitive in CCLE (not “least sensitive” of all 189) and the third most sensitive cell line in CGP is the most sensitive in CCLE. It seems difficult to accept that this could be thought of as representing discordance, particularly considering, for example, that the AUC values for these cell lines in CGP are 0.66, 0.36 and 0.56 compared to a median value of 0.99 for all BCR-ABL1 negative cell lines. The key point is that, as would have been expected, the BCR-ABL1 positive cell lines were the most sensitive and the fusion gene status was highly significantly associated with drug sensitivity in both studies. This aside, the authors do not seem to dispute that the results for nilotinib are highly concordant between the two studies, but do highlight some of the problems with the “waterfall method”, a novel method that was included as a minor part of the original CCLE Nature publication. However, the shortcomings of the “waterfall method” or indeed any existing or future methods for modeling these data are not relevant to the key points

that we have raised and, importantly, do not support the authors' claim that drug sensitivity data are discordant.

Point 5: Figure 1c. The authors attempted to leverage all data in both studies to test whether findings informed by one study can improve discovery of gene-drug associations in the other study. We are seriously concerned about multiple aspects of this analysis. First, the authors state that one should not use a correlation metric when there is lack of variability in the measurements, which is the case for nilotinib. However the authors appear to contradict themselves by using this metric to compute gene-drug associations. Second, the authors focused on nilotinib, for which the sensitivity data are somewhat reproducible between CGP and CCLE. We were not able to reproduce the authors' results; however we found that the Q-Q plot for nutlin3 is almost identical to Figure 1c, suggesting that drug names might have been switched, which cannot be confirmed as the author's software code was unavailable. Third, the Q-Q plot for nilotinib does not support the authors' conclusions. Indeed, we observed only a weak leftward shift compared to the unfiltered p-value distribution. Such a shift can easily be obtained by filtering genes based on variance to keep only the genes whose expression is part of the top 10% and top 1% most variant in the source dataset. Strikingly, we observed a stronger leftward shift for such a filtering process that is, by construction, fully independent from the drug sensitivity data. We made similar observations for 11 out of the common set of 15 drugs. These results demonstrate that the authors' analysis should be controlled for basic filtering procedures that are not based on drug sensitivity data. For the remaining 4 drugs, we observed a strong leftward shift for the filtered p-value distributions compared to the unfiltered distribution, suggesting that the source study is highly informative for the target one. We are not at all

surprised that there is value in integrating the CGP and CCLE datasets; it is well known that increasing sample size is important in filtering signal from noise (if everything were perfectly consistent across the datasets, integrating the two datasets would not be necessary since they would contain the same information). In our original report, we found statistically significant non-zero correlations between phenotype measurements for almost all drugs, supporting the fact that there is relevant signal in these datasets; the main point of our study was that more work is necessary to improve consistency of phenotypic measures to make these large-scale projects more useful for developing robust predictors of drug response.

Response 5: We agree with the authors' comments above that their original analysis reported "relevant signal in these datasets", and hence that improved power to discover associations with drug sensitivity by integrating both datasets could not be considered inconsistent with their findings. However, we would highlight that this statement is seemingly *contradictory* to one of the conclusions of their original paper which was that "predictive models of response developed using data from one study are almost guaranteed to fail when validated on data from another study". In our revised manuscript, as a more novel and interpretable way of highlighting the utility of the data, we have instead emphasized the genomic associations that were recapitulated in both studies.

We are extremely grateful for the opportunity to address the points raised by the original authors and to clarify and improve our manuscript. While we appreciate the additional analysis performed by Haibe-Kains and colleagues, none of their reanalysis change our initial stand, with

some of their findings further supporting our original points. We are concerned that the same statistical issues we raised concerning the original publication continue to mark the analyses presented in their reply (see, for example, Response 2). In light of the analytical errors in the authors' original publication that the authors have now acknowledged in their reply, we believe that the proper interpretation of these data has been severely compromised, and the far-reaching conclusions made by the authors on the basis of some key analytical errors necessitate that the issues be brought to the attention of the scientific community. We look forward to receiving reviews and comments from the outside referees. Thank you for your time and consideration in this matter.

R. Stephanie Huang, PhD
Assistant Professor of Medicine
Section of Hematology-Oncology
The University of Chicago
900 E. 57th St.,
KCBD, Room 7148
Chicago, IL 60637
phone: 773-702-9363
email: rhuang@medicinebsd.uchicago.edu

Nancy J. Cox, PhD
Professor of Medicine
Chair of Genetic Medicine
The University of Chicago
900 E. 57th St.,
KCBD, Room 3000
Chicago, IL 60637
Phone: 773-834-1001
email: ncox@medicinebsd.uchicago.edu

REFERENCES

- [1] Kelland, Lloyd R., et al. "DT-Diaphorase expression and tumor cell sensitivity to 17-allylaminoo, 17-demethoxygeldanamycin, an inhibitor of heat shock protein 90." *Journal of the National Cancer Institute* 91.22 (1999): 1940-1949.
- [2] McDermott, Ultan, et al. "Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling." *Proceedings of the National Academy of Sciences* 104.50 (2007): 19936-19941.
- [3] Konecny, Gottfried E., et al. "Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells." *Cancer research* 66.3 (2006): 1630-1639.
- [4] Müller, Christoph R., et al. "Potential for treatment of liposarcomas with the MDM2 antagonist Nutlin-3A." *International Journal of Cancer* 121.1 (2007): 199-205.
- [5] Solit, David B., et al. "BRAF mutation predicts sensitivity to MEK inhibition." *Nature* 439.7074 (2005): 358-362.
- [6] Tsai, James, et al. "Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity." *Proceedings of the National Academy of Sciences* 105.8 (2008): 3041-3046.
- [7] Pickrell, Joseph K., et al. "Understanding mechanisms underlying human gene expression variation with RNA sequencing." *Nature* 464.7289 (2010): 768-772.
- [8] Lappalainen, Tuuli, et al. "Transcriptome and genome sequencing uncovers functional variation in humans." *Nature* 501.7468 (2013): 506-511.

Response to “Consistency in large pharmacogenomic studies”

Benjamin Haibe-Kains, Zhaleh Safikhani, Nehme El-Hachem, Petr Smirnov, Nicolai Juul Birkbak,
Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush

We welcome the opportunity to respond to the report by Geeleher et al. While we agree that the anecdotal case of nilotinib demonstrates consistent identification of a few highly sensitive cell lines shared by the CGP¹ and CCLE², our initial conclusions about the broader inconsistency in reported phenotypes between the CGP and CCLE still holds³. Below we provide our response to the three criticisms raised by Geeleher et al.

1. In our initial publication, we computed the correlation between gene expression and mutation profiles of cell lines to assess whether large transcriptomic changes and/or genetic drift might be the cause of the observed inconsistency in drug sensitivity data, and subsequently the gene-drug associations³. We agree with the authors that correlations *across* and *between* cell lines should be thoroughly compared. Overall correlation across cell lines was lower than correlation between cell lines; however we still observed that gene expression data were significantly more concordant than IC₅₀ and AUC values in all comparisons (Wilcoxon rank sum test p<0.01; Supplementary Figure 1). In light of these results, our conclusion that gene expression data are significantly more correlated than pharmacological data still holds, although the difference between cell lines is smaller than reported in our initial publication.

2. CGP and CCLE investigators identified gene-drug associations using the elasticnet linear regression that assumes normality of the error distribution. In our comparative study we used the same approach to investigate the reproducibility of univariate gene-drug associations. The presence

of outliers in drug response violates the normality of the error distribution which may cause unstable models. This issue affects the CGP and CCLE analyses and produces inconsistent gene-drug associations, as we reported in our study. To address this issue, we fitted a logistic regression model using the 15 most sensitive and the 55 most resistant cell lines in CGP and CCLE separately, following the approach used Geeleher and co-workers⁴. We observed poor reproducibility of the most significant gene-drug associations (Supplementary Fig 2), consistent with our previous report. We agree with Geeleher et al. that the use of alternative statistical approaches should be investigated in hopes that they might yield more reproducible biomarkers of drug response extracted from large pharmacogenomic studies. However, more innovative biomarker discovery methods will not solve the problem of inconsistent measurement and reporting of phenotypic response to drug treatment. Simply put, if the phenotypes disagree, one would not expect any method to learn parameters relevant to the observed phenotype in one cell line and then predict a different phenotype for the same cell line in a different study with similar gene expression data than the first.

3. We agree that lack of variability in drug sensitivity measurements may prevent biologically meaningful assessment of concordance between pharmacogenomic datasets. Geeleher et al. focused their discussion on nilotinib for which there are three sensitive cell lines in both datasets. Even among these cell lines the AUC values are not concordant; the least sensitive of the three cell lines in CGP is actually the most sensitive one in CCLE (Supplementary Fig 3; Extended response in Supplementary Information). Therefore the only way to consider these results concordant is to classify these three cell lines as sensitive and the rest as resistant, which cannot easily be done using the Waterfall approach³ (Supplementary Figure 4). To check the authors' claim that nilotinib is not an isolated case, we used the adaptive Matthews correlation coefficient (AMCC; see Supplementary Methods) to estimate the concordance in measurements composed of a potentially

small set of informative outliers. AMCC can be used to estimate the concordance between drug phenotypes where only a few cell lines are sensitive, or between gene expressions where the gene of interest is rarely expressed. As expected, nilotinib yielded an AMCC of 1 which denotes perfect consistency between the two studies (Supplementary Fig 5); however, the rest of the drugs yielded much lower AMCC, with only PLX4720 yielding substantial AMCC (0.7), and 4 other drugs (lapatinib, crizotinib, AZD0530, PD0325901) yielding moderate AMCC. It should be noted that the inter-laboratory replicates of the measurements of camptothecin sensitivity performed using the same experimental protocols at two different locations within CGP yielded only an AMCC of 0.55 (Supplementary Fig 6). Again consistent with our previous report, gene expression data were significantly more concordant than drug sensitivity data across cell lines (median AMCC of 0.66 vs. 0.54 for gene expression and AUC, respectively; Supplementary Figure 7).

Our analysis confirms that nilotinib is indeed an anecdotal case and that drug sensitivity measurements for the rest of the drugs (cytotoxic or targeted) remain inconsistent. In our original report, we found statistically significant non-zero correlations between phenotype measurements for almost all drugs, supporting the fact that there is relevant signal in these datasets; the main point of our study was that more work is necessary to improve consistency of phenotypic measures to make data from these large-scale projects more useful for development of robust predictors of drug response.

Geeleher et al. report a few associations that were identified in both CGP and CCLE. However, all these associations were reported as consistent in our study (Supplementary Files 3, 4, 12, and 13 in our original report; detailed statistics provided in Supplementary Information); these findings, in fact, validate the methods we used. More importantly, the authors do not comment on the numerous associations that were not consistent.

Finally, Geeleher et al. state that their recently published study⁴ supports the utility of the data in these studies, implying that their results also support a consistency between CGP and CCLE. However, a true test of this would be to validate the models they trained on CGP to predict phenotypes reported by the CCLE on the shared cell lines. If they could predict the CCLE drug response phenotype with high accuracy, this would provide some quantitative evidence of a consistency between the two datasets in the context of their predictive models. However, we⁵ and others⁶ showed that such an analysis remains a major challenge.

In conclusion, Geeleher et al. do not provide any new way to assess the consistency of drug sensitivity data as they claim to have done in their report. Their only attempt, formerly Figure 1c where the authors showed that results from *source* study can improve discovery of novel gene-drug associations in *target* study, was flawed as we demonstrated in our previous review. We confirmed that gene expression data are significantly more concordant across studies than drug sensitivity data. Importantly, we demonstrated that the case of nilotinib is anecdotal as it is the only drug with consistent outliers in CGP and CCLE, therefore invalidating the authors' claim of a general consistency between these two large pharmacogenomic studies.

References

1. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
3. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).

4. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).
5. Papillon-Cavanagh, S. *et al.* Comparison and validation of genomic predictors for anticancer drug sensitivity. *J. Am. Med. Inform. Assoc.* **20**, 597–602 (2013).
6. Jang, I. S., Neto, E. C., Guinney, J., Friend, S. H. & Margolin, A. A. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac. Symp. Biocomput.* 63–74 (2014).

Supplementary Information

Response to

"Consistency in large pharmacogenomic studies"
from Geeleher et al.

Benjamin Haibe-Kains, Zhaleh Safikhani, Nehme El-Hachem, Petr Smirnov,
Nicolai Juul Birkbak, Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush

Contents

1 Extended Response	3
2 Supplementary Methods	5
2.1 Data retrieval and curation	5
2.1.1 CGP (release 4, March 2013)	5
2.1.2 CCLE (release March 2013)	5
2.2 Cell line annotations	5
2.3 Gene expression	6
2.4 Measures of concordance	7
2.4.1 Spearman rank-based correlation	7
2.4.2 Adaptive Matthews correlation	7
2.5 Gene-drug associations	8
3 Full Reproducibility of the Analysis Results	9
3.1 Set up the software environment	9
3.2 Run the R scripts	11
3.3 Additional parameters	11
3.4 Generate the Supplementary Information	11
4 List of Abbreviations	12
5 Supplementary Tables	13
6 Supplementary Figures	14

1 Extended Response

Targeted versus cytotoxic drugs

We agree that lack of variability in drug sensitivity measurements may prevent biologically meaningful assessment of concordance between pharmacogenomic datasets. Indeed, if the vast majority of cell lines are resistant to a given drug, the concordance, as estimated using Spearman's correlation, will be low even if the noise present in the two datasets is well controlled (but still present). This is the case for highly targeted drugs, such as nilotinib, for which only a few cell lines are expected to be sensitive, while cytotoxic drugs exhibit broader effects (Supplementary Figure 8). We compared the distribution of variability, estimated by the median absolute deviation (MAD), for targeted and cytotoxic drugs in CGP and CCLE (Supplementary Figure 9A,B). As expected the variability of drug sensitivity data for cytotoxic drugs is significantly greater than for targeted drugs; a MAD cutoff of 0.10 enables to roughly discriminate between drugs with highly targeted or broader effect (Supplementary Figures 9C). Based on this cutoff, paclitaxel, 17AAG and PD0325901 are classified as broad effect drugs in both datasets; similar observation for TAE684 and PD0332991 in CGP, and AZD6244 in CCLE. Highly targeted drugs include erlotinib, lapatinib, PHA665752, nilotinib, nutilin3, PLAX4720, and sorafenib (MAD of AUC < 0.10; Supplementary Figure 9C). We concur with the authors that highly targeted drugs tend to yield lower Spearman correlation coefficients, likely due to the lack of variability of the corresponding sensitivity data (Supplementary Figure 10A,B). However, high value of the AMCC statistic, which allows to detect consistent outliers across studies (see Section 2.4.2), is not associated with variability in drug response, as shown in Supplementary Figure 10C,D.

The authors focused their discussion on nilotinib. Given the prior knowledge that BCR-ABL1 fusion is the primary target of nilotinib and given that this gene fusion is rare in cancer cell lines, only few sensitive cell lines are expected. There are three such cell lines that have been screened with nilotinib both in CGP and CCLE:

Cell line	AUC in CGP	AUC in CCLE
EM-2	0.37	0.76
KU812	0.64	0.65
MEG-01	0.44	0.58

Among the three sensitive cell lines the AUC values are not quite concordant as the least sensitive cell line in CGP is actually the most sensitive one in CCLE. Therefore the only way to consider these results concordant is to classify these three cell lines as sensitive and the rest as resistant, yielding a perfect agreement between the two datasets. However, the use of an automated method for drug sensitivity calling, such as the waterfall approach introduced in the CCLE study, do not allow to reach high level of concordance (Supplementary Figure 4). However, the AMCC statistic, by optimizing the cutoff in CGP and CCLE, yield perfect consistency as shown in Supplementary Figure 3.

It is worth pointing out that nilotinib is not a representative example of all targeted drugs investigated in CCLE and CGP. As can be seen in Supplementary Figure 3, there are several cell lines for erlotinib, PHA665752, AZD0530, PLX4720, and nutilin3 that exhibit high sensitivity in only one of the datasets. The authors state that given that most drugs investigated in our comparative study are targeted (Supplementary Table 1), such a lack of variability in drug sensitivity data was widely observed. However, in addition to paclitaxel (the only cytotoxic drug in our study), we observed that several targeted drugs exhibited broad effect in the set of 504 cell lines screened both in CGP and CCLE (17AAG and PD0325901, as well as TAE684, PD0332991 and AZD6244, yielded MAD of AUC values > 0.10 in at least one dataset; Supplementary Figure 9A,B). For these drugs, one could argue that the variability of drug sensitivity measures warrants the use of a correlation metric for assessing the concordance between CGP and CCLE studies. However we observed only poor to fair correlation for these drugs ($r_s < 0.6$, Supplementary Figure 3). Alternatively, we adapted the Matthews correlation coefficient (AMCC) to assess the concordance

across studies by accounting for the lack of variability of sensitivity measurements for targeted drugs (see Section 2.4.2). While we agree with the authors that the sensitivity to nilotinib is reproduced in both studies (AMCC = 1; Supplementary Figure 5); this observation is anecdotal and is not generalizable to the other targeted and cytotoxic drugs investigated in our original report. Indeed, nilotinib was the only drug yielding an AMCC value > 0.8 (Supplementary Figure 5). Further, much of the motivation for the CGP and CCLE studies was the creation of large datasets through the screening of hundreds of cell lines that could be used to develop predictive models of drug response; if such studies are only useful in assessing targeted agents, it undermines the value of such large-scale pharmacogenomic studies.

Consistent known gene-drug associations

ERBB2 expression for lapatinib → 783th biomarker in CGP ($p=0.04$; Supplementary File 4); top biomarker for CCLE ($p=8.4E-15$; Supplementary File 5)

NQ01 expression for 17-AAG → top biomarker in both CGP ($p=2.4E-13$; Supplementary File 4) and CCLE ($p=6.2E-14$; Supplementary File 5)

BRAF mutation for PD-0325901 → top biomarker in both CGP ($p=0.009$; Supplementary File 12) and CCLE ($p=5.2E-5$; Supplementary File 5)

BRAF mutation for PLX4720 → top biomarker in both CGP ($p=1.7E-7$; Supplementary File 12) and CCLE ($p=2.6E-11$; Supplementary File 13)

BRAF mutation for AZ6244 → top biomarker in both CGP ($p=0.004$; Supplementary File 12) and CCLE ($p=6.3E-7$; Supplementary File 13)

2 Supplementary Methods

2.1 Data retrieval and curation

2.1.1 CGP (release 4, March 2013)

Gene expression, mutation data and cell line annotations were downloaded from ArrayExpress. Drug sensitivity measurements and drug information were downloaded from the CGP website ([link](#)) and the Nature website ([link](#)), respectively.

Minimum and maximum screening concentrations for each drug/cell line were extracted from `gdsc_compounds_conc_w2.csv` (μM).

The natural logarithm of IC_{50} measurements were retrieved from `gdsc_manova_input_w2.csv` in column " $*_{\text{IC_50}}$ " (referred to as x) and subsequently transformed using $-\log_{10}(\exp(x))$; high values are representative of cell line sensitivity to drugs.

The AUC measurements were retrieved from `gdsc_manova_input_w2.csv` in column " $*_{\text{AUC}}$ " (referred to as x); high values are representative of cell line sensitivity to drugs.

2.1.2 CCLE (release March 2013)

Gene expression, mutation data cell line annotations and drug information were downloaded from the CCLE website ([link](#)). Drug sensitivity data were downloaded from the Nature website ([link](#));

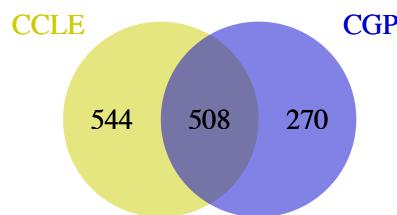
Screening concentrations for each drug/cell line were extracted from Supplementary Table 11 in column E (μM).

IC_{50} measurements were retrieved from Supplementary Table 11 in column J (" $\text{IC}_{50} \mu\text{M} (\text{norm})$ ") (referred to as x) and subsequently transformed into logarithmic scale, $-\log_{10}(x)$; high values are representative of cell line sensitivity to drugs.

AUC measurements were retrieved from Supplementary Table 11 in column L (" $\text{ActArea} (\text{norm})$ ") and subsequently divided by the number of drug concentrations tested (8); high values are representative of cell line sensitivity to drugs.

2.2 Cell line annotations

Cell line names were harmonized in both CGP and CCLE to match identical cell lines; this was done through manual search over alternative names of cell lines, as reported in CGP and CCLE cell line annotation files and online databases such as [hyperCLDB](#) and [BioInformationWeb](#). In our comparative analysis published in Nature [4], we focused on the set of 471 cell lines for which both gene expression and drug sensitivity were available. In the present work we extended our curation to all the cell lines for which at least one data type (gene expression, mutation or drug sensitivity) is available, increasing the shared set of cell lines to 508:



Tissue type nomenclature from CGP [2] was chosen throughout this study, CCLE tissue type information [1] was therefore updated to follow this nomenclature, which resulted in 24 tissue types:

Tissue type	Number of cell lines
lung	109
haematopoietic_and_lymphoid_tissue	77
breast	39
central_nervous_system	36
large_intestine	33
skin	30
oesophagus	21
urinary_tract	17
ovary	16
pancreas	16
stomach	16
autonomic_ganglia	12
soft_tissue	12
upper_aerodigestive_tract	12
liver	11
kidney	10
bone	8
endometrium	8
thyroid	8
pleura	6
prostate	4
biliary_tract	1
salivary_gland	1
small_intestine	1

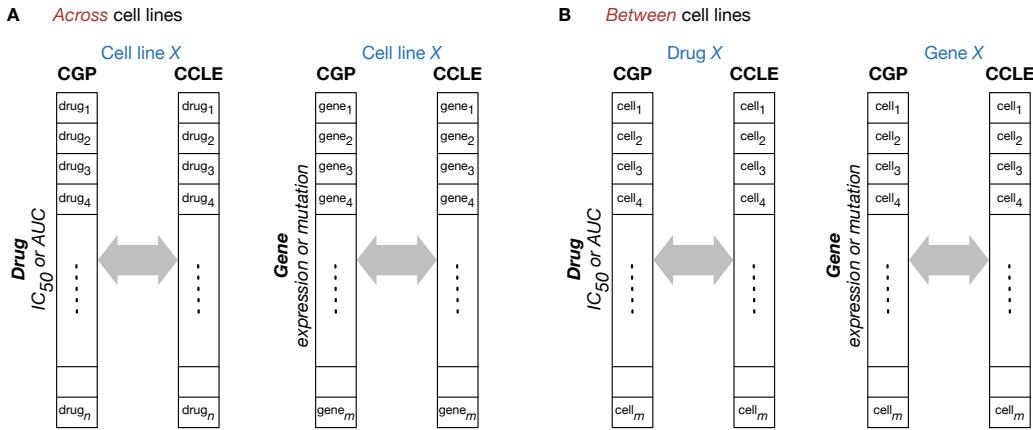
All the curation steps have been documented in the scripts `cdrug2_normalization_cgpr.R`, `cdrug2_normalization_ccle.R`, and `cdrug2_format.R`.

2.3 Gene expression

Raw gene expression profiles (Affymetrix CEL format) for 789 CGP and 1,067 CCLE cell lines were downloaded, respectively, from ArrayExpress and CCLE websites. Gene expression data were normalized with frozen RMA [7] using the Bioconductor Chip Description File (CDF) definitions (`hthgu133a` from CGP, and `hgu133plus2` for CCLE). We then used the R package `jetset` [5], to map Affymetrix probe sets to unique Entrez gene IDs by selecting the best probe set for each gene (overall jetset score > 0.20); subsequent analyses were restricted to the 10,312 probe sets common to the CGP and CCLE arrays. For replicates, the CEL files were ordered by hybridization date and the first experiment was selected.

2.4 Measures of concordance

We assessed the concordance of the gene expression, mutation and drug sensitivity of CGP and CCLE studies *across* and *between* cell lines, as illustrated in the figure below. When data are compared across cell lines, we assess whether, for a given gene expression or drug, the cell line data were concordant (a gene is expressed at a similar level or similar response to a drug is observed in the same set of cell lines for instance; panel A). When data are compared between cell lines, we assessed whether, for a given cell line, the genomic and pharmacological profiles were concordant in the two studies (a given cell line harbours similar gene expression patterns or pharmacological responses for instance; panel B).



2.4.1 Spearman rank-based correlation

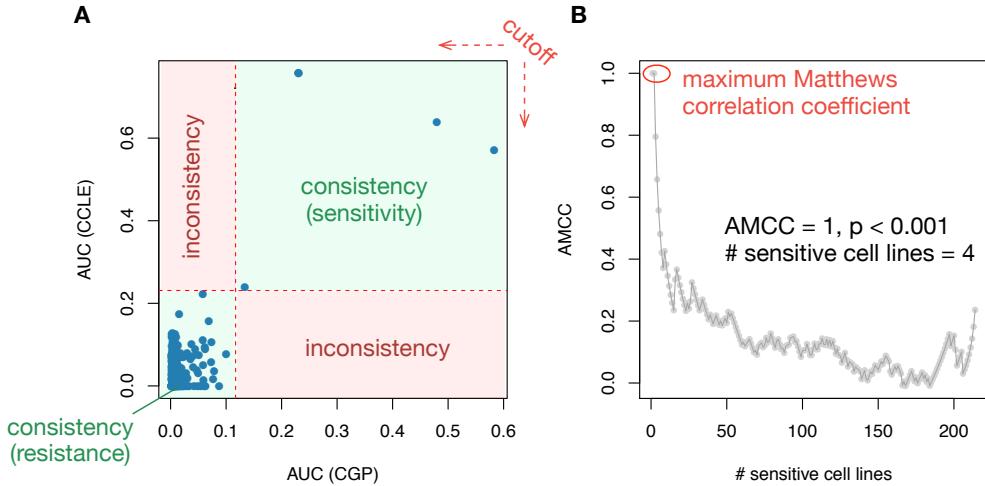
We assessed consistency between gene expressions or pharmacological responses in CGP and CCLE by computing Spearman rank-ordered correlations when ≥ 10 measures were available. The use of Spearman correlation is appropriate when high biological variability is observed, as opposed to pure technical noise. For instance, Spearman correlation can be used as a measure of concordance for drugs inducing growth inhibition in a substantial set of cell lines, with a sufficient dynamic range (median absolute deviation of AUC > 0.10 ; Supplementary Figure 9). This is the case for 17-AAG, paclitaxel, TAE684, and PD-0325901 in both CCLE and CGP, PD0332991 in CGP only, and AZD6244 in CCLE only (Supplementary Figure 9C).

Typically qualitative descriptions of correlation coefficients are associated with intervals: $r_s < 0.5$, poor correlation; $0.5 \leq r_s < 0.6$, fair correlation; $0.6 \leq r_s < 0.7$, moderate correlations; $0.7 \leq r_s < 0.8$, substantial correlation; and $r_s \geq 0.8$, almost perfect correlation.

2.4.2 Adaptive Matthews correlation

When the variability of the measurements is mostly due to technical noise, correlation is not an appropriate measure of concordance. This is the case for drugs where only a small subset of cell lines are expected to be sensitive. Similarly, genes that are expressed in a few cell lines are also problematic. We therefore developed the adaptive Matthews correlation coefficient (AMCC), a statistic specifically designed to detect consistent outliers across datasets. The rationale behind AMCC is illustrated in the figure below using sensitivity measurements for nilotinib as an example. The idea is to rank each measurement in the two dataset of interest (here CGP on the x-axis and CCLE on the y-axis; see panel A). The cutoffs are then varied in the two studies to include one more measurement at each step in order to define inconsistent and consistent observations and subsequently estimate the Matthews correlation coefficient [6]. Finally, the maximum coefficient

is selected and its significance is computed using a permutation test where the cell line labels are permuted 1000 times in each dataset (see panel **B**).



Note that AMCC optimizes the rank cutoff to yield the highest consistency between two studies. Ideally, drug sensitivity calling should be computed in each dataset separately to enable assessment of concordance of the resulting calls using well-established χ^2 tests. Therefore AMCC may yield overoptimistic estimates of concordance.

2.5 Gene-drug associations

We assessed the association between gene expression and AUC values (referred to as gene-drug association), using a similar approach than Geeleher and co-workers [3]:

1. For each drug, select the 15 most sensitive and 55 most resistant cell lines in the full CGP and CCLE datasets separately.
2. Fit a logistic regression model controlled for tissue type:

$$Y = \beta_0 + \beta_i G_i + \beta_t T$$

where Y denote the drug sensitivity call (sensitive vs. resistant), G_i is the expression of gene i . T is the tissue type, and β s are the regression coefficients. The strength of gene-drug association is quantified by β_i , above and beyond the relationship between drug sensitivity and tissue type. The gene expression measurements G are scaled (standard deviation equals to 1) to estimate standardized coefficients from the linear model.

3. Significance of the gene-drug association is estimated by the statistical significance of β_i (t statistic).

3 Full Reproducibility of the Analysis Results

We will describe how to fully reproduce the figures and tables reported in the main manuscript. We automated the analysis pipeline so that minimal manual interaction is required to reproduce our results. To do this, one must simply:

1. Set up the software environment
2. Run the R scripts
3. Generate the Supplementary Information

The code and associated files are publicly available on GitHub: <https://github.com/bhaibeka/cdrug2>.

3.1 Set up the software environment

We developed and tested our analysis pipeline using R running on linux and Mac OSX platforms.

To mimic our software environment the following R packages should be installed:

- R version 3.1.1 (2014-07-10), x86_64-unknown-linux-gnu
- Base packages: base, datasets, graphics, grDevices, grid, methods, parallel, splines, stats, utils
- Other packages: affxparser 1.36.0, affy 1.42.3, affyio 1.32.0, amap 0.8-12, AnnotationDbi 1.26.0, Biobase 2.24.0, BiocGenerics 0.10.0, BiocInstaller 1.14.3, biomaRt 2.20.0, bitops 1.0-6, corpcor 1.6.6, DBI 0.3.0, epibasix 1.3, fingerprint 3.5.2, Formula 1.1-2, frma 1.16.0, gdata 2.13.3, genefu 1.14.0, GenomeInfoDb 1.0.2, gplots 2.14.1, Hmisc 3.14-5, hthgu133acdf 2.14.0, hthgu133afrmavecs 1.1.0, igraph 0.7.1, inSilicoDb 2.1.1, jetset 1.6.0, lattice 0.20-29, lsa 0.73, MASS 7.3-34, mclust 4.3, MetaGx 0.0.2, mgcv 1.8-3, mRMRe 2.0.5, nlme 3.1-117, OptimalCutpoints 1.1-3, org.Hs.eg.db 2.14.0, Pharmacogenomics 0.0.3, prodlm 1.4.5, rcdk 3.3.0, RCurl 1.95-4.3, rjson 0.2.14, R.methodsS3 1.6.1, R.oo 1.18.0, RSQLite 0.11.4, R.utils 1.33.0, SnowballC 0.5.1, survcomp 1.15.1, survival 2.37-7, sva 3.10.0, vcd 1.3-2, VennDiagram 1.6.8, WriteXLS 3.5.0, XML 3.98-1.1, xtable 1.7-4
- Loaded via a namespace (and not attached): acepack 1.3-3.3, Biostrings 2.32.1, bit 1.1-12, bootstrap 2014.4, caTools 1.17.1, cluster 1.15.3, codetools 0.2-9, colorspace 1.2-4, ff 2.2-13, foreach 1.4.2, foreign 0.8-61, GenomicRanges 1.16.4, gtools 3.4.1, IRanges 1.22.10, iterators 1.0.7, KernSmooth 2.23-13, latticeExtra 0.6-26, lava 1.2.6, Matrix 1.1-4, nnet 7.3-8, oligo 1.28.2, oligoClasses 1.26.0, png 0.1-7, preprocessCore 1.26.1, rcdklibs 1.5.8.4, RColorBrewer 1.0-5, rJava 0.9-6, rmeta 2.16, rpart 4.1-8, stats4 3.1.1, SuppDists 1.1-9.1, survivalROC 1.0.3, tools 3.1.1, XVector 0.4.0, zlibbioc 1.10.0

All these packages are available on CRAN¹ or Bioconductor², except for jetset which is available on the CBS website³.

Run the following commands in a R session to install all the required packages:

¹<http://cran.r-project.org>

²<http://www.bioconductor.org>

³<http://www.cbs.dtu.dk/biotools/jetset/>

```

source("http://bioconductor.org/biocLite.R")
biocLite(c("AnnotationDbi", "affy", "affyio", "hthgu133acdf",
  "hthgu133afrmavecs", "hgu133plus2cdf", "hgu133plus2frmavecs",
  "org.Hs.eg.db", "genefu", "biomaRt", "frma", "Hmisc", "vcd",
  "epibasix", "amap", "gdata", "WriteXLS", "xtable", "plotrix",
  "R.utils", "DBI", "GSA", "gplots"))

```

Note that you may need to install Perl⁴ and its module Text::CSV_XS for the WriteXLS package to write xls file; once Perl is installed in your system, use the following command to install the Text::CSV_XS module through CPAN⁵:

```
cpan Text/CSV_XS.pm
```

Lastly, follow the instructions on the CBS website to properly install the jetset package or use the following commands in R:

```

download.file(url="http://www.cbs.dtu.dk/biotools/jetset/current/jetset_1.4.0.tar.gz",
  destfile="jetset_1.4.0.tar.gz")
install.packages("jetset_1.4.0.tar.gz", repos=NULL, type="source")

```

Once the packages are installed, clone the cdrug2 GitHub repository (<https://github.com/bhaibeka/cdrug2>) This should create a directory on the file system containing the following files:

`cdrug2_foo.R` Script containing the definitions of all functions required for the analysis pipeline.

`cdrug2_normalization_cgp.R` Script to curate, annotate and normalize of CGP data.

`cdrug2_normalization_ccle.R` Script to curate, annotate and normalize of CCLE data.

`cdrug2_format.R` Script to identify common cell lines, tissue types and drugs investigated both in CGP and CCLE.

`cdrug2_analysis_huang.R` Script generating all the figures and tables reported in the manuscript.

`cdrug2_pipeline.R` Master script running all the scripts listed above to generate the analysis results.

`matching_cell_line_CCLE_CGP.csv` Curation of cell line name to match CGP and CCLE nomenclatures.

`matching_tissue_type_CCLE_CGP.csv` Curation of tissue type name to match CGP and CCLE nomenclatures.

`HaibeKains_Nature_2013_common_cellines.csv` Set of 504 cell lines shared between CGP and CCLE as published in Haibe-Kains *et al.*, Nature, 2013.

`response_supplinfo_huang_v3.tex` The L^AT_EX file of the present supplementary information

All the files required to run the automated analysis pipeline are now in place. It is worth noting that raw gene expression and drug sensitivity data are voluminous, please ensure that at least 25GB of storage are available.

⁴<http://www.perl.org/get.html>

⁵<http://www.cpan.org/modules/INSTALL.html>

3.2 Run the R scripts

Open a terminal window and go to the `cdrug2` directory. You can easily run the analysis pipeline either in batch mode or in a R session. Before running the pipeline you can specify the number of CPU cores you want to allocate to the analysis (by default only 1 CPU core will be used). To do so, open the script `cdrug2_pipeline.R` and update line #41:

```
nbcore <- 4
```

to allocate four CPU cores for instance.

To run the full pipeline in batch mode, simply type the following command:

```
R CMD BATCH code/cdrug2_pipeline.R Rout &
```

The progress of the pipeline could be monitored using the following command:

```
tail -f Rout
```

To run the full analysis pipeline in an R session, simply type the following command:

```
source("code/cdrug2_pipeline.R")
```

Key messages will be displayed to monitor the progress of the analysis.

The analysis pipeline was developed so that all intermediate analysis results are saved in the directories `data` and `saveres`. Therefore, in case of interruption, the pipeline will restart where it stopped.

3.3 Additional parameters

download.method Method to download all the data from the CCLE and CGP websites; options are 'wget' (default), 'curl', 'lynx', or 'auto'.

cosmic.version Version of the COSMIC database ('v68' as for 2014-04-07; see Sanger's FTP)

saveres Path to the directory where all teh results should be stored. Default is "./saveres"

max.cellfiles Maximum number of CEL files taht can be processed at once by `frma`.

minsample Minimum number of samples to compute the Spearman correlation. Default is 10.

genedrugm Method to estimate the association gene-drug, controlled for tissue type. Method "lm" = linear regression or logistic regression, depending on the output variable

concordance.method Estimator for concordance between gene-drug associations; possible options are "spearman" (default), "cosine", and "pearson" for Spearman's rank-ordered correlation coefficient, cosine similarity and Pearson's correlation coefficient, respectively.

3.4 Generate the Supplementary Information

After completion of the analysis pipeline a directory `saveres` will be created to contain all the intermediate results, tables and figures reported in the main manuscript and this Supplementary Information.

4 List of Abbreviations

AMCC	Adaptive Matthews Correlation Coefficient.
AUC	Area under the drug sensitivity curve.
CGP	Cancer Genome Project initiated by the Wellcome Sanger Institute.
CCLE	The Cancer Cell Lines Encyclopedia initiated by Novartis and the Broad Institute.
IC ₅₀	Concentration in micro molar [μM] at which the drug inhibited 50% of the cellular growth.
FDR	False Discovery Rate
MAD	Median Absolute Deviation.
R _s	Spearman correlation coefficient

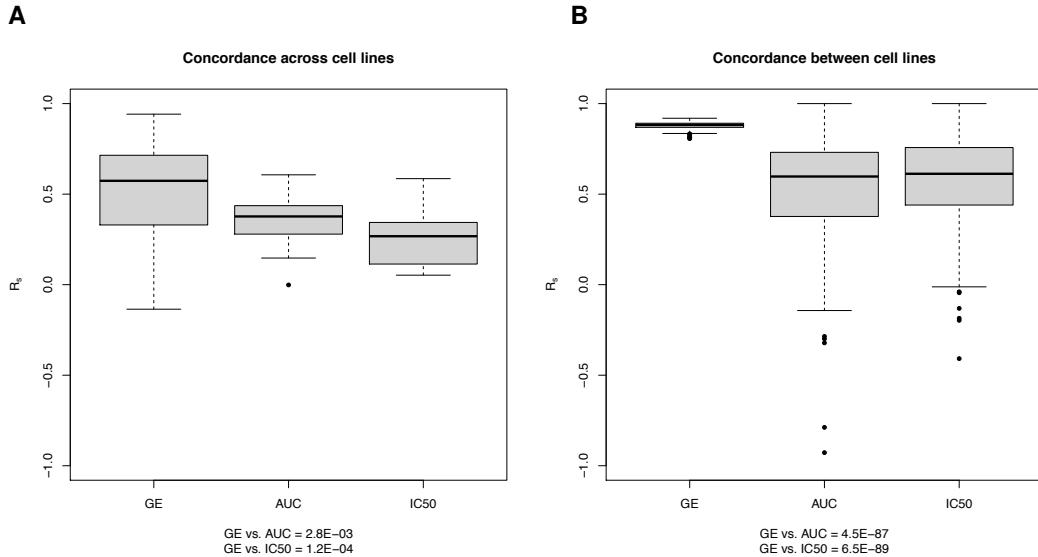
5 Supplementary Tables

Supplementary Table 1: Description of the 15 anticancer drugs screened both in CGP and CCLE studies.

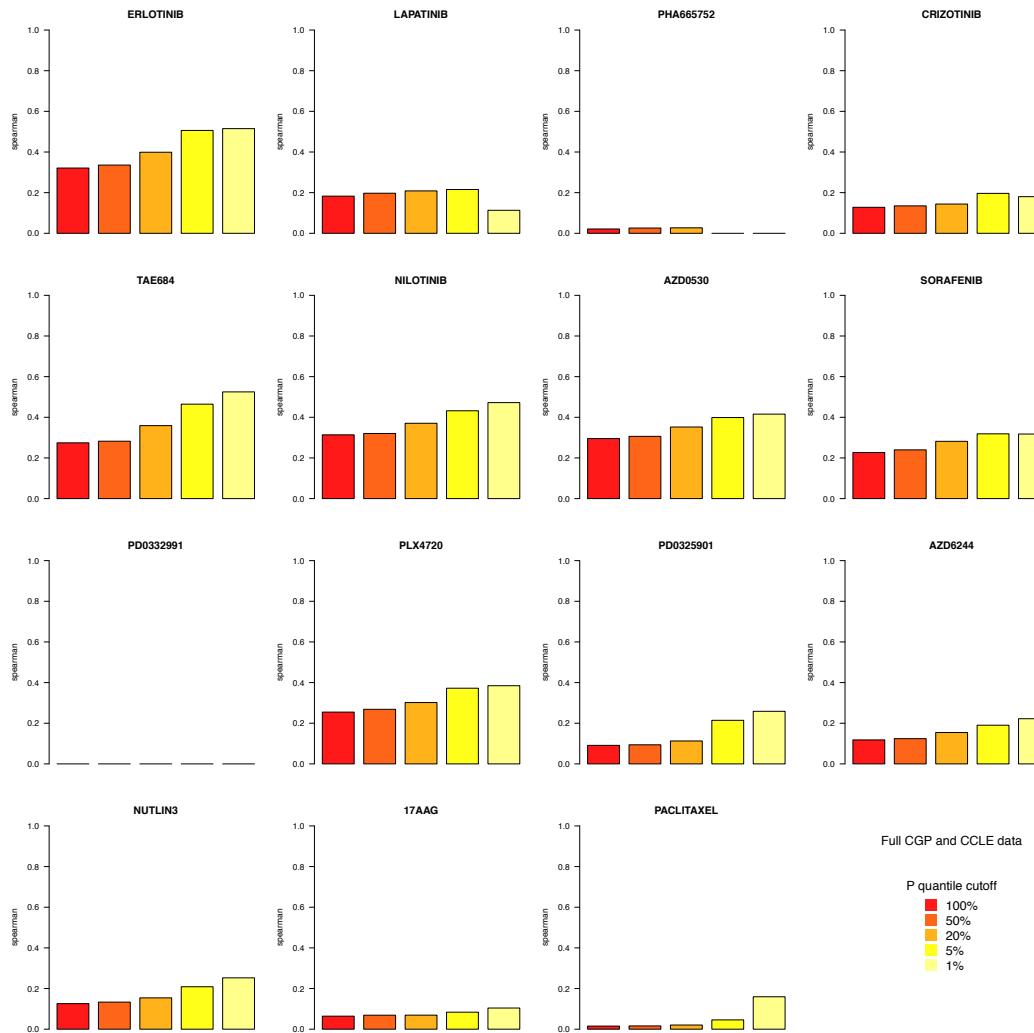
Compound	Class	Target(s)	Class	Organization
Erlotinib	Targeted	EGFR	Kinase inhibitor	Genentech
Lapatinib	Targeted	EGFR, HER2	Kinase inhibitor	GlaxoSmithKline
PHA-665752	Targeted	c-MET	Kinase inhibitor	Pfizer
Crizotinib	Targeted	c-MET, ALK	Kinase inhibitor	Pfizer
TAE684	Targeted	ALK	Kinase inhibitor	Novartis
Nilotinib	Targeted	Abl/Bcr-Abl	Kinase inhibitor	Novartis
AZD0530	Targeted	Src, Abl/Bcr-Abl, EGFR	Kinase inhibitor	AstraZeneca
Sorafenib	Targeted	Flt3, C-KIT, PDGFRbeta, RET, Raf kinase B, Raf kinase C, VEGFR-1, KDR, FLT4	Kinase inhibitor	Bayer
PD-0332991	Targeted	CDK4/6	Kinase inhibitor	Pfizer
PLX4720	Targeted	RAF	Kinase inhibitor	Plexxikon
PD-0325901	Targeted	MEK	Kinase inhibitor	Pfizer
AZD6244	Targeted	MEK	Kinase inhibitor	AstraZeneca
Nutlin-3	Targeted	MDM2	Other	Roche
17-AAG	Targeted	HSP90	Other	Bristol-Myers Squibb
Paclitaxel	Cytotoxic	beta-tubulin	Cytotoxic	Bristol-Myers Squibb

6 Supplementary Figures

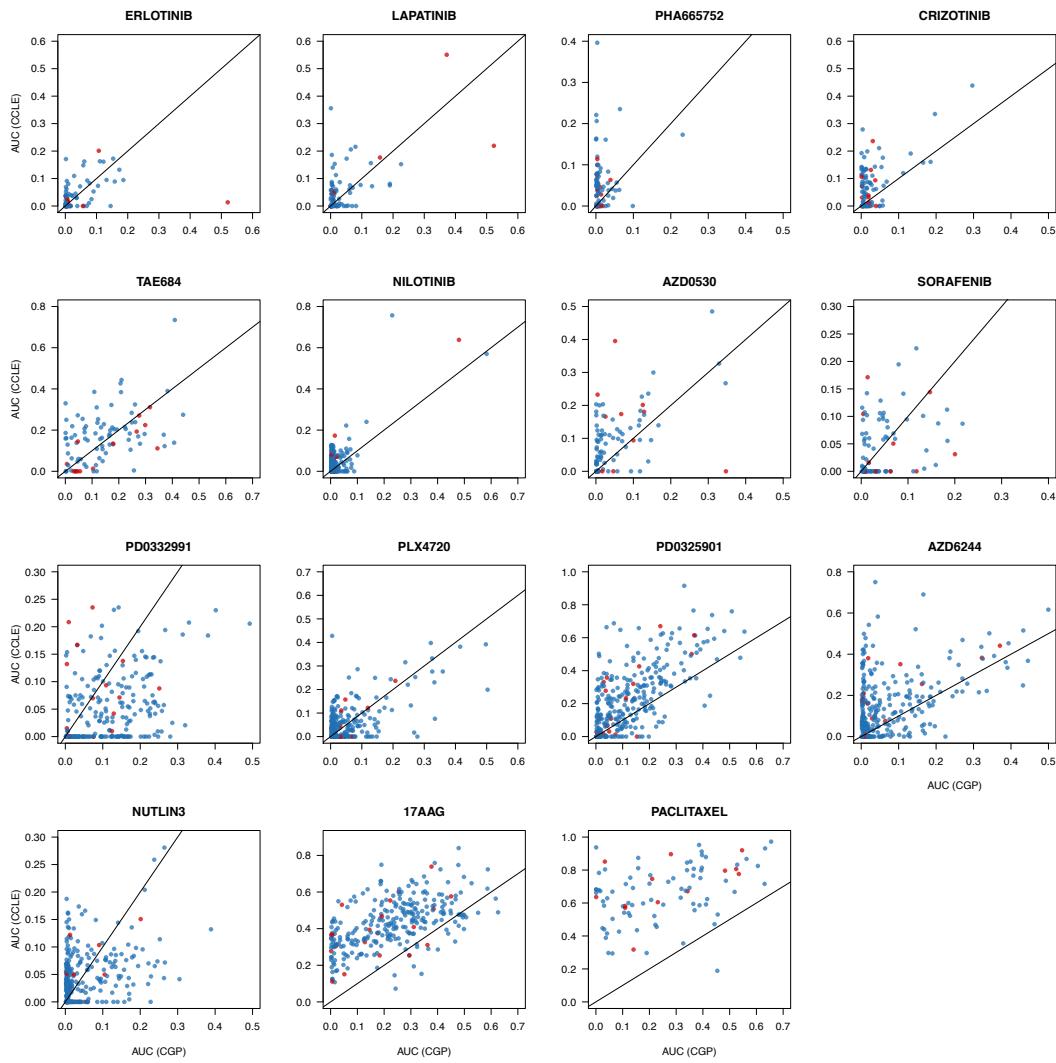
Supplementary Figure 1: Box plot comparing Spearman correlation coefficients estimating the concordance of gene expressions, IC_{50} and AUC measures **(A)** across and **(B)** between cell lines in CGP and CCLE. Significance of the difference between concordance observed for genomic and pharmacological data, as computed using the Wilcoxon rank sum test, is provided under each plot. GE: gene expression; R_s : Spearman's rank-ordered correlation.



Supplementary Figure 2: Concordance across cell lines of gene-drug associations computed with logistic regression analysis based on extreme AUC sensitivity calls (15 most sensitive and 55 most resistant) for all the cell lines investigated either in CGP or CCLE. The model is adjusted for cell lines' tissue of origin. The concordance of gene-drug association ranking is estimated by Spearman correlation for multiple p-value thresholds.



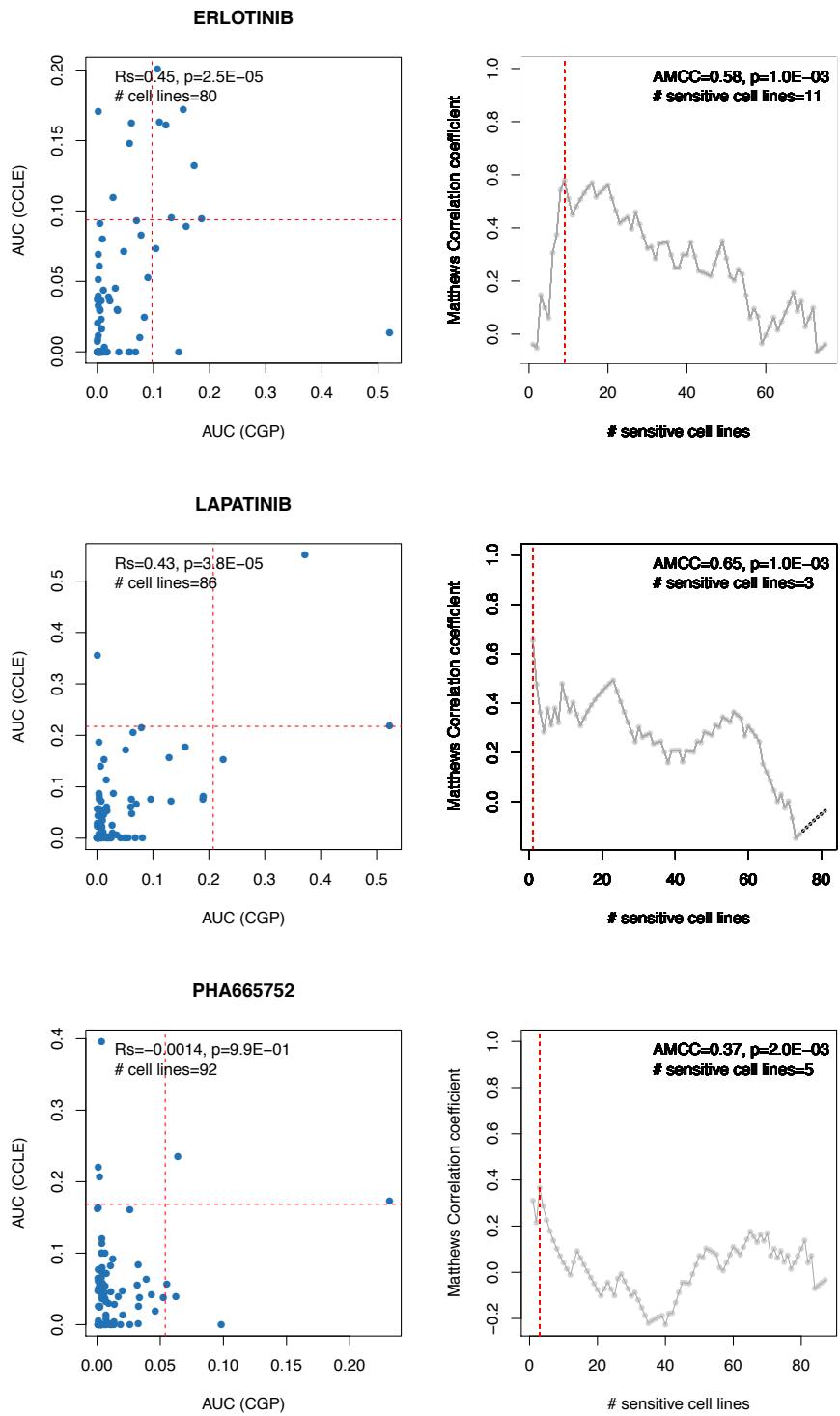
Supplementary Figure 3: Consistency across cell lines of AUC values between CGP and CCLE for the 504 cell lines and 15 drugs investigated both in CGP and CCLE. The scatter plot represents each data point, where the 471 cell lines analyzed in our original study are represented by blue points, while the new cell lines identified after additional curation of the data are presented in red. In case of perfect consistency, all the points should lie on the diagonal.



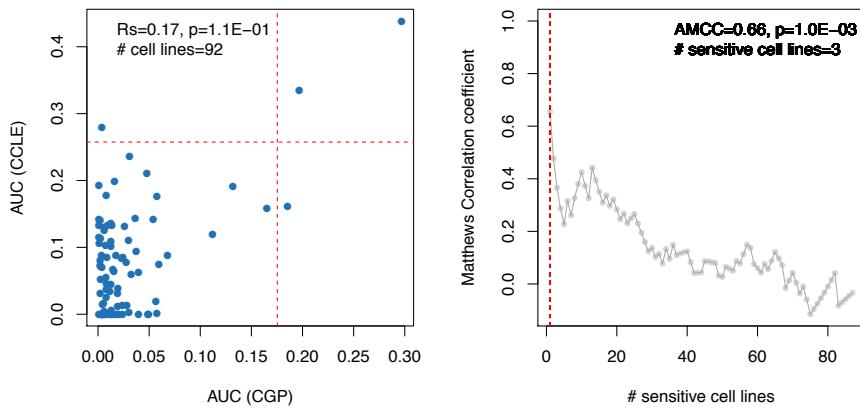
Supplementary Figure 4: Contingency tables comparing the (extreme) sensitivity calls (res, inter, and sens standing for resistant, intermediate and sensitive drug phenotypes, respectively) computed using the Waterfall approach based on AUC measures for each of the 15 drugs screened both in CGP and CCLE. The Kappa coefficient, its confidence interval and its significance are reported below each contingency table (Kappa and Kappa2 representing the Kappa coefficient of the contingency table with and without the intermediate cases, respectively). The intermediate drug sensitivity calls are made large to enable to identify the extreme sensitive and resistant cell lines. It is worth to note that the sensitivity calls have been computed using the Waterfall approach on the full CGP and CCLE datasets.

AUC sensitivity calling ERLOTINIB	AUC sensitivity calling LAPATINIB	AUC sensitivity calling PHA665752
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res [65 1 4] inter [1 0 2] sens [3 3 1]	res [63 6 4] inter [3 0 2] sens [3 2 3]	res [67 7 4] inter [6 0 0] sens [6 0 2]
Kappa=0.26, 95%CI [0.061,0.45], p=6.1E-04 Kappa2=0.17, 95%CI [-0.2,0.55], p=2.3E-01	Kappa=0.23, 95%CI [0.026,0.44], p=1.0E-02 Kappa2=0.41, 95%CI [0.052,0.77], p=6.1E-03	Kappa=0.043, 95%CI [-0.18,0.26], p=2.0E-01 Kappa2=0.22, 95%CI [-0.11,0.55], p=1.0E-01
AUC sensitivity calling CRIZOTINIB	AUC sensitivity calling TAE684	AUC sensitivity calling NILOTINIB
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res [64 4 1] inter [10 2 2] sens [6 0 3]	res [28 2 2] inter [15 3 1] sens [25 11 5]	res [197 6 0] inter [10 0 0] sens [3 0 4]
Kappa=0.25, 95%CI [0.031,0.46], p=1.0E-02 Kappa2=0.42, 95%CI [0.07,0.77], p=2.3E-03	Kappa=0.089, 95%CI [-0.056,0.23], p=1.0E-01 Kappa2=0.1, 95%CI [-0.11,0.31], p=2.2E-01	Kappa=0.26, 95%CI [0.029,0.5], p=2.8E-06 Kappa2=0.72, 95%CI [0.42,1], p=4.8E-08
AUC sensitivity calling AZD0530	AUC sensitivity calling SORAFENIB	AUC sensitivity calling PD0332991
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res [66 6 2] inter [7 6 0] sens [1 1 3]	res [67 2 10] inter [3 1 2] sens [1 1 1]	res [68 20 58] inter [13 4 11] sens [14 10 29]
Kappa=0.44, 95%CI [0.22,0.66], p=4.8E-05 Kappa2=0.64, 95%CI [0.27,1], p=2.1E-04	Kappa=0.19, 95%CI [-0.018,0.4], p=8.9E-02 Kappa2=0.12, 95%CI [-0.17,0.4], p=2.2E-01	Kappa=0.092, 95%CI [-0.017,0.2], p=1.3E-01 Kappa2=0.16, 95%CI [0.024,0.3], p=1.4E-02
AUC sensitivity calling PLX4720	AUC sensitivity calling PD0325901	AUC sensitivity calling AZD6244
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res [201 13 17] inter [11 2 6] sens [10 4 16]	res [119 6 0] inter [32 2 2] sens [60 33 28]	res [157 21 19] inter [18 6 6] sens [23 2 17]
Kappa=0.33, 95%CI [0.21,0.45], p=1.9E-09 Kappa2=0.48, 95%CI [0.31,0.65], p=3.6E-10	Kappa=0.22, 95%CI [0.14,0.3], p=0.0E+00 Kappa2=0.35, 95%CI [0.23,0.47], p=2.0E-13	Kappa=0.22, 95%CI [0.1,0.34], p=5.7E-05 Kappa2=0.33, 95%CI [0.17,0.49], p=9.4E-06
AUC sensitivity calling NUTLIN3	AUC sensitivity calling 17AAG	AUC sensitivity calling PACLITAXEL
CCLE vs CGP res inter sens	CCLE vs CGP res inter sens	CCLE vs CGP res inter sens
res [171 13 49] inter [15 3 12] sens [12 1 8]	res [63 12 17] inter [25 24 72] sens [7 10 54]	res [13 4 5] inter [21 6 24] sens [3 0 13]
Kappa=0.11, 95%CI [0.0058,0.21], p=7.1E-02 Kappa2=0.096, 95%CI [-0.038,0.23], p=9.1E-02	Kappa=0.28, 95%CI [0.19,0.36], p=0.0E+00 Kappa2=0.66, 95%CI [0.54,0.78], p=1.1E-16	Kappa=0.14, 95%CI [0.003,0.28], p=4.6E-03 Kappa2=0.53, 95%CI [0.25,0.81], p=1.3E-03

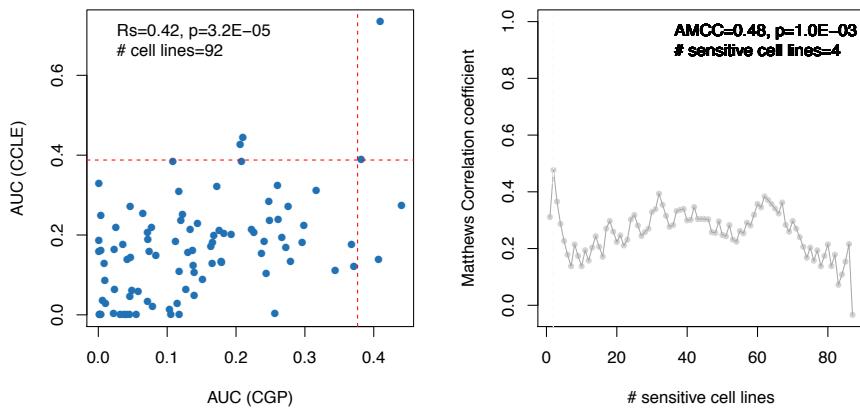
Supplementary Figure 5: Scatterplots of AUC measures in CGP and CCLE with cutoffs determined by the AMCC statistic and evolution of the Matthews correlation coefficient with respect to the cutoff values.



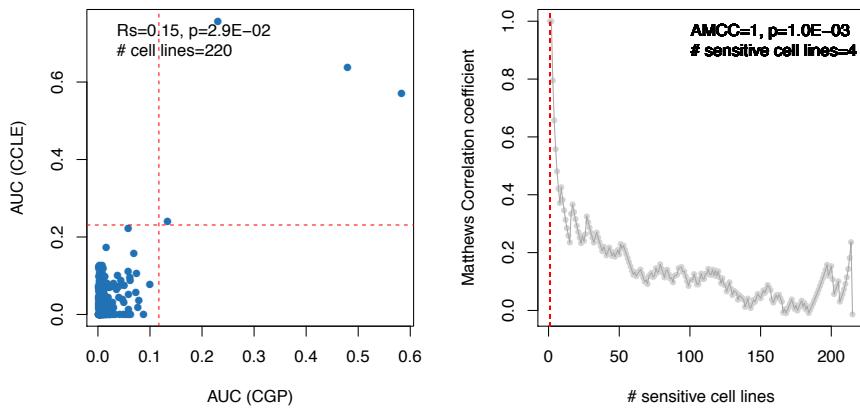
CRIZOTINIB



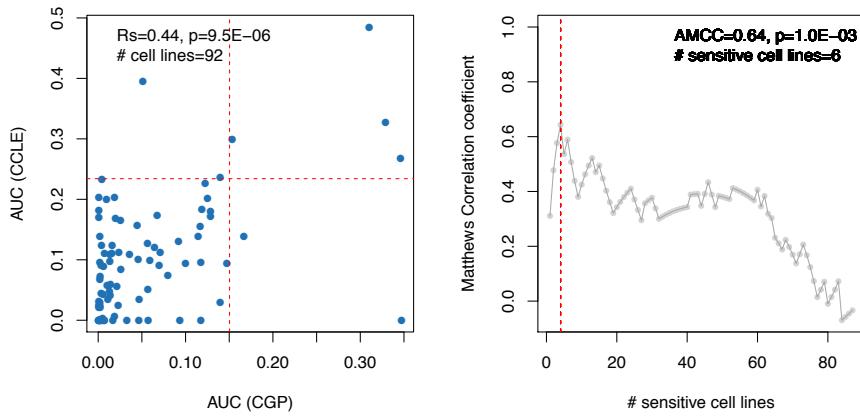
TAE684



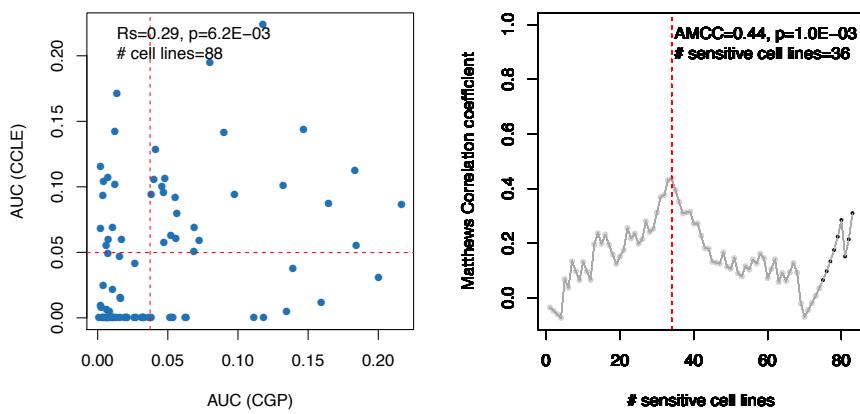
NILOTINIB



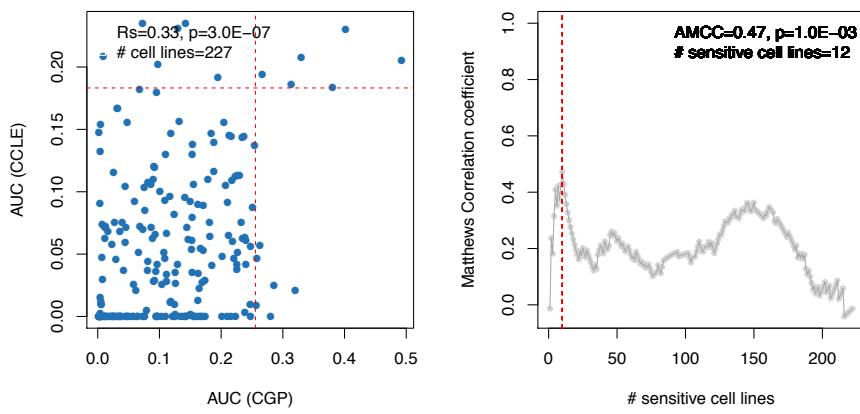
AZD0530



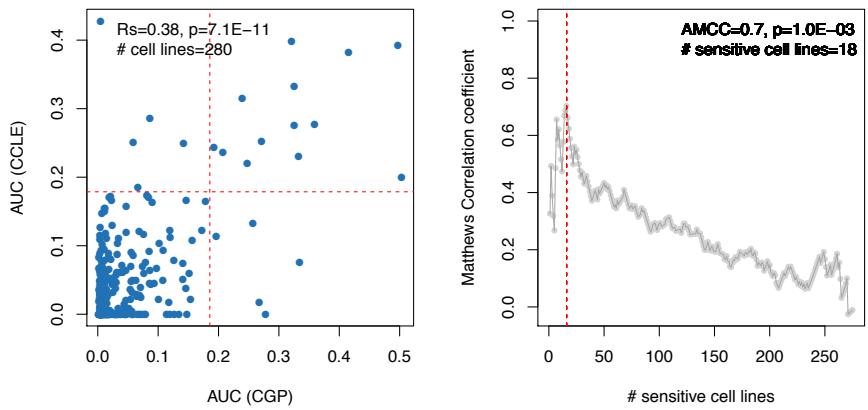
SORAFENIB



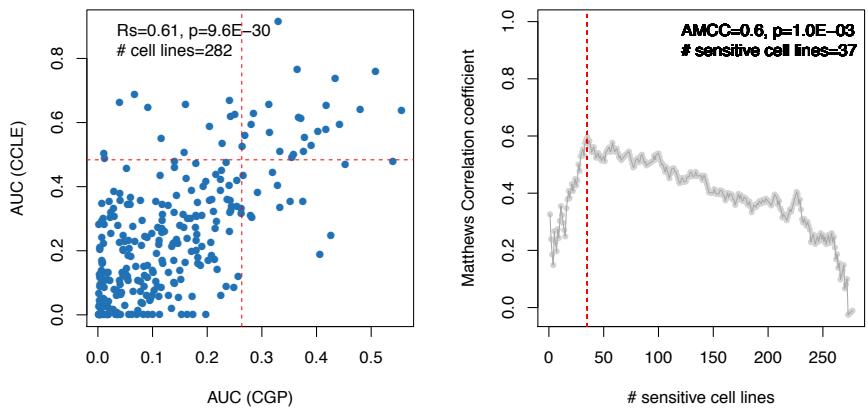
PD0332991



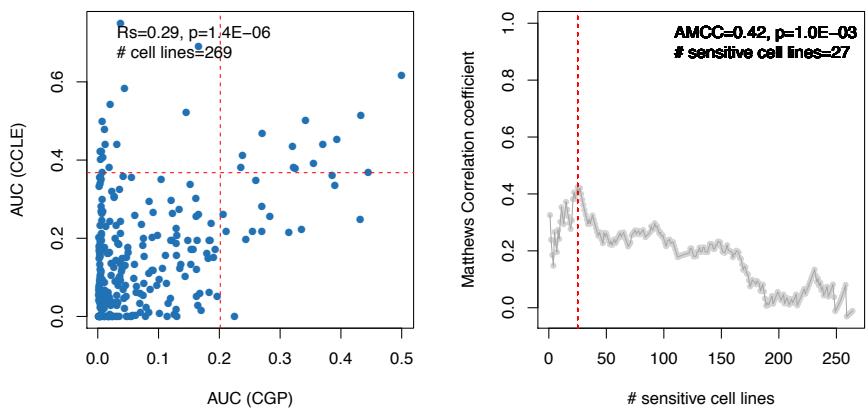
PLX4720



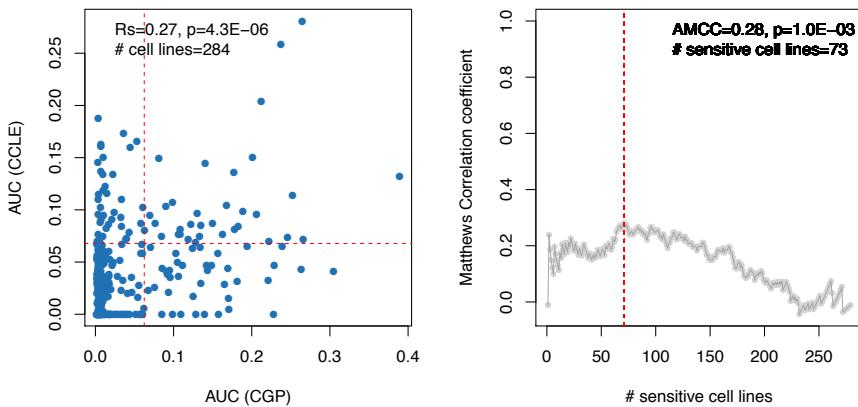
PD0325901



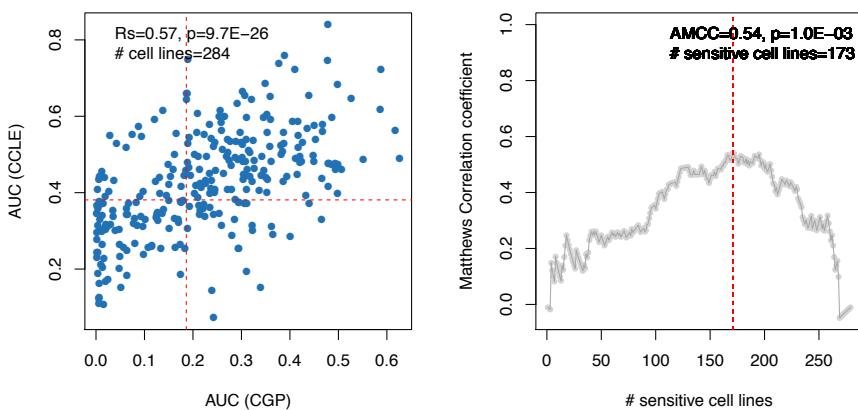
AZD6244



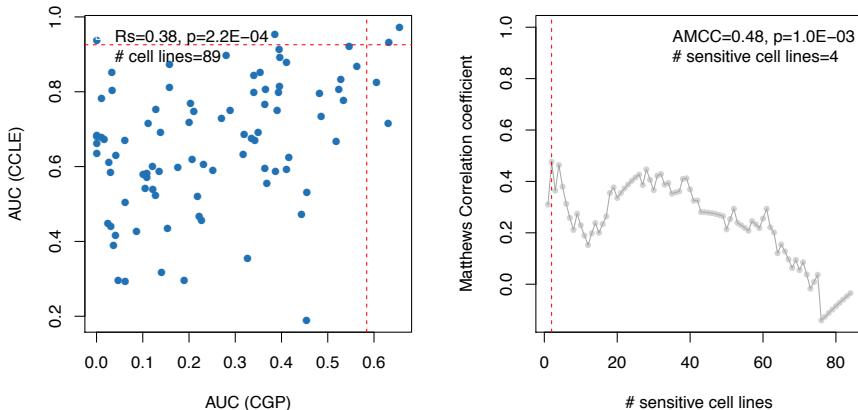
NUTLIN3



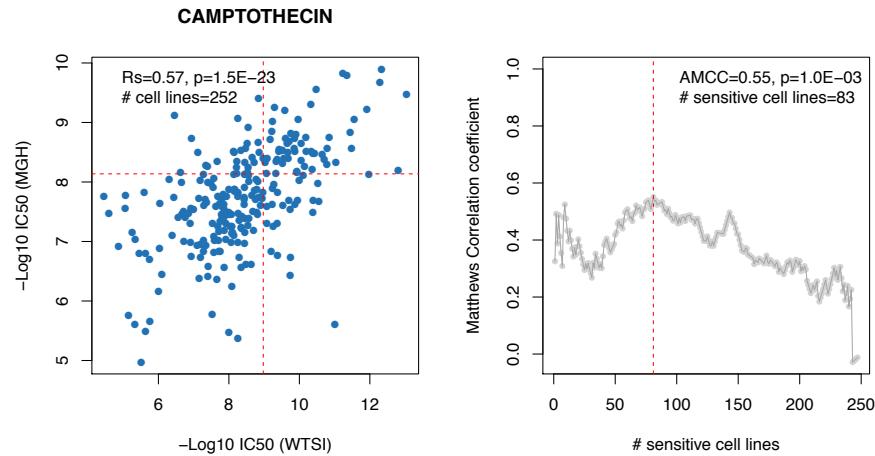
17AAG



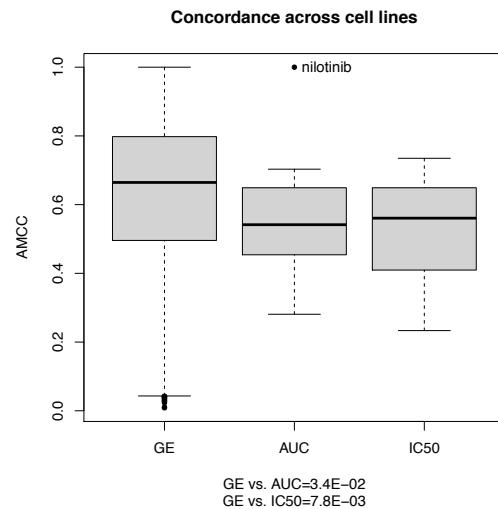
PACLITAXEL



Supplementary Figure 6: Scatterplot of IC_{50} measures for camptothecin replicated at the Wellcome Trust Sanger Institute (WTSI) and the Massachusetts General Hospital (MGH) with cutoffs determined by the AMCC statistic and evolution of the Matthews correlation coefficient with respect to the cutoff values.

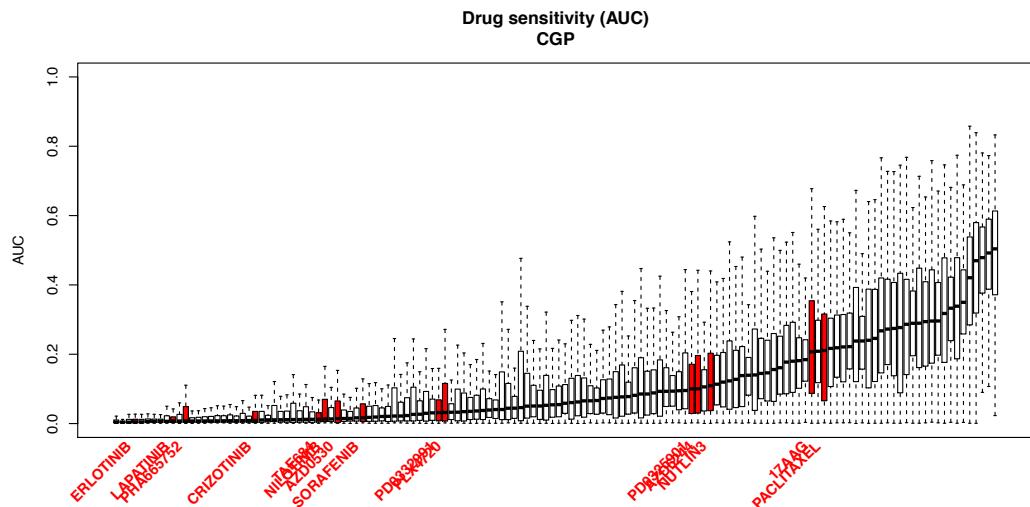


Supplementary Figure 7: Concordance across cell lines using AMCC for gene expression, AUC and IC₅₀ measurements.

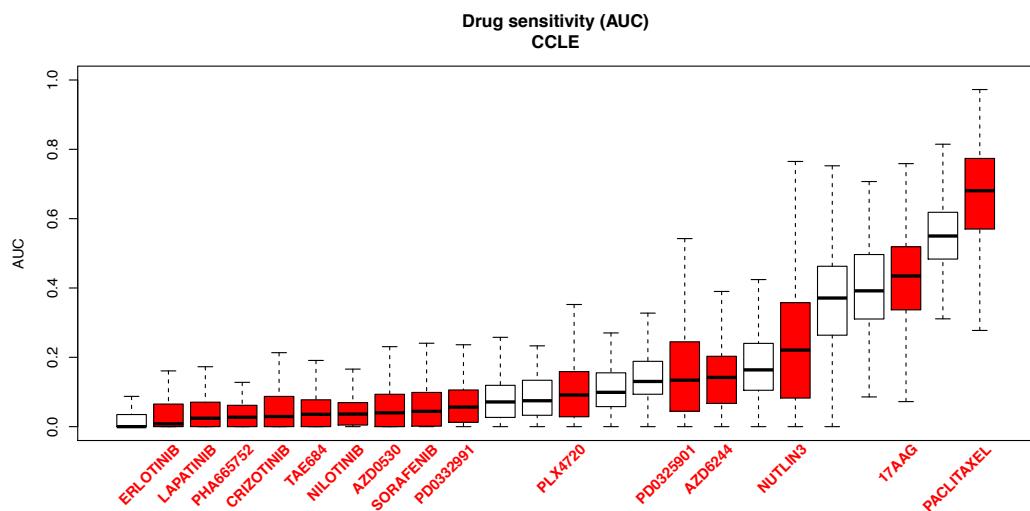


Supplementary Figure 8: Variability of drug sensitivity data. Box plot showing the distribution of AUC measures across cell lines for all the drugs screened in **(A)** CGP and **(B)** CCLE for the shared set of 504 cell lines. In red are highlighted the drugs screened both in CGP and CCLE.

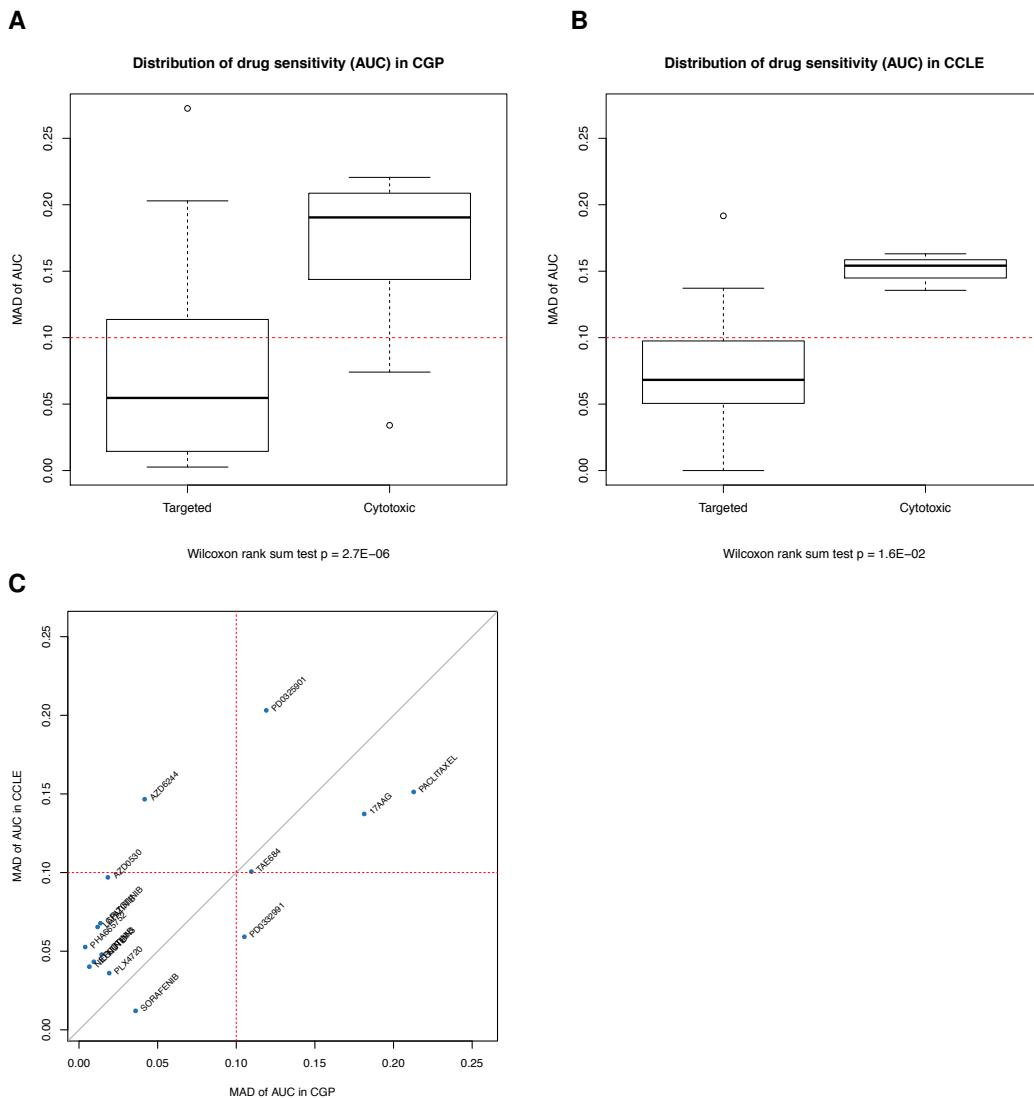
A



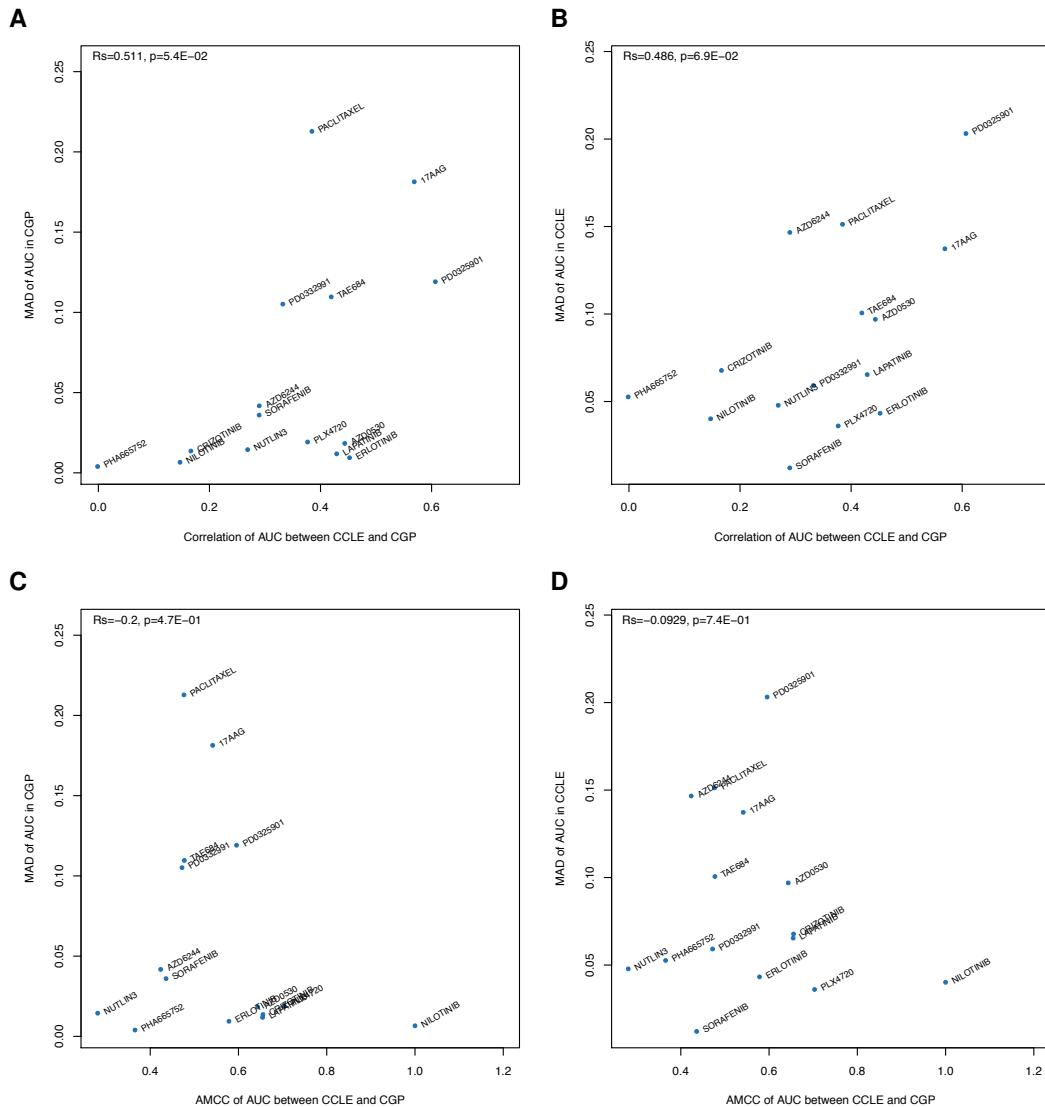
B



Supplementary Figure 9: Box plot comparing median absolute deviation (MAD) of AUC across cell lines for all drugs classified as targeted vs. cytotoxic in **(A)** CGP and **(B)** CCLE, and **(C)** comparison between MAD of AUC for common drugs in CGP and CCLE. A cutoff of 0.10 for MAD AUC is selected to discriminate between targeted and cytotoxic drugs



Supplementary Figure 10: Association between concordance of drug sensitivity, computed using **(A,B)** Spearman correlation coefficient and **(C,D)** AMCC, and variability of AUC across cell lines computed using median absolute deviation (MAD).



References

- [1] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F Berger, John E Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jané-Valbuena, Felipa A Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H Engels, Jill Cheng, Guoying K Yu, Jianjun Yu, Peter Aspesi, Melanie de Silva, Kalpana Jagtap, Michael D Jones, Li Wang, Charles Hatton, Emanuele Palescandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P. Mesirov, Stacey B Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E Myer, Barbara L Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L Harris, Matthew Meyerson, Todd R. Golub, Michael P Morrissey, William R Sellers, Robert Schlegel, and Levi A. Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, March 2012.
- [2] Mathew J Garnett, Elena J Edelman, Sonja J Heidorn, Chris D Greenman, Anahita Das-tur, King Wai Lau, Patricia Greninger, I Richard Thompson, Xi Luo, Jorge Soares, Qing-song Liu, Francesco Iorio, Didier Surdez, Li Chen, Randy J Milano, Graham R Bignell, Ah T Tam, Helen Davies, Jesse A Stevenson, Syd Barthorpe, Stephen R Lutz, Fiona Kogera, Karl Lawrence, Anne McLaren-Douglas, Xeni Mitropoulos, Tatiana Mironenko, Helen Thi, Laura Richardson, Wenjun Zhou, Frances Jewitt, Tinghu Zhang, Patrick O'Brien, Jessica L Boisvert, Stacey Price, Wooyoung Hur, Wanjuan Yang, Xianming Deng, Adam Butler, Hwan Geun Choi, Jae Won Chang, Jose Baselga, Ivan Stamenkovic, Jeffrey A Engelman, Sreenath V Sharma, Olivier Delattre, Julio Saez-Rodriguez, Nathanael S Gray, Jeffrey Settleman, P Andrew Futreal, Daniel A Haber, Michael R Stratton, Sridhar Ramaswamy, Ultan McDermott, and Cyril H Benes. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, March 2012.
- [3] Paul Geeleher, Nancy J Cox, and R Stephanie Huang. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome biology*, 15(3):R47, 2014.
- [4] Benjamin Haibe-Kains, Nehme El-Hachem, Nicolai Juul Birkbak, Andrew C Jin, Andrew H Beck, Hugo J W L Aerts, and John Quackenbush. Inconsistency in large pharmacogenomic studies. *Nature*, 504(7480):389–393, December 2013.
- [5] Qiyuan Li, Nicolai J Birkbak, Balázs Győrffy, Zoltan Szallasi, and Aron C Eklund. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*, 12(1):474, 2011.
- [6] B W Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et biophysica acta*, 405(2):442–451, October 1975.
- [7] Matthew N McCall, Benjamin M Bolstad, and Rafael A Irizarry. Frozen robust multiarray analysis (frMA). *Biostatistics (Oxford, England)*, 11(2):242–253, April 2010.

Consistency in large pharmacogenomic studies

Paul Geeleher¹, Eric R. Gamazon^{1,2}, Cathal Seoighe³, Nancy J. Cox^{1,2*}, and R. Stephanie

Huang^{1*}

¹Department of Medicine, The University of Chicago, Chicago, IL 60637, USA;

²Division of Genetic Medicine, Vanderbilt University, Nashville, TN 37232; and

³Department of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland

* Address for correspondence and reprints:

Nancy J. Cox, 900 E 57th street, KCBD room 3100, the University of Chicago, Chicago, IL 60637.

Telephone: (773) 702-; Fax: (773) 702-; E-mail: *nancy.cox@vanderbilt.edu*

R. Stephanie Huang, 900 E 57th street, KCBD room 7148, the University of Chicago, Chicago, IL

60637. Telephone: (773) 702-9363; Fax: (773) 702-0963; E-mail:

rhuang@medicine.bsd.uchicago.edu

Haibe-Kains *et al.*¹ reported inconsistency between two large-scale pharmacogenomic studies (the Cancer Cell Line Encyclopedia² (CCLE) and the Cancer Genome Project³ (CGP)). Upon careful analysis of the same data we arrived at much more positive conclusions. Here, we highlight the most important reasons for this.

To assess the concordance of two large studies of the efficacy of cancer drugs, Haibe-Kains *et al.* compared the correlation in drug sensitivities with correlation in gene expression values measured on the same human cancer cell lines. The authors reported correlation “*between*” cell lines for gene expression but, inconsistently, “*across*” cell lines for drug sensitivity (see Methods). On reanalysis, we found much higher correlations between cell lines than across cell lines for both gene expression and drug sensitivity measures (median $r_s = 0.88$ between cell lines, $r_s = 0.56$ across cell lines for expression; median $r_s = 0.62$ between cell lines and $r_s = 0.35$ across cell lines for AUC, a drug sensitivity measure). Thus, by correcting this inconsistency, the correlations for expression and drug sensitivity data were far more similar than was originally reported, severely undermining the authors' interpretation of the relative quality of expression and drug sensitivity datasets.

However, the fundamental issue is that the authors' reported Spearman correlation coefficients do not fairly reflect the concordance of drug sensitivity between the studies, because of the lack of variability in drug response, which arises due to the highly-targeted nature of many of the drugs assessed. To see why correlation is not an appropriate measure of biological concordance for these data, consider the hypothetical example of a drug that is not effective against any cell lines (a possibility for an experimental drug): In such a case randomly-fluctuating measurement error, inherent in biological assays, will dominate over the (non-existent) biological variability, meaning that there could be no expectation of correlation between repeated measures of drug sensitivity (assuming other experimental variables are held constant). In this study, many of the drugs were highly-targeted agents, which by design require specific (and often rare) molecular targets for response (see Supplementary Table). Consider nilotinib, which the authors reported as exhibiting “poor consistency” between CGP and CCLE ($r_s = 0.1$ for AUC); nilotinib targets *BCR-ABL1*, a fusion-gene. In CGP, *BCR-ABL1* status was reported strongly associated with drug sensitivity ($P = 2.54 \times 10^{-65}$), accurately reflecting the known biology. *BCR-ABL1* status was not reported by CCLE, however, upon re-analysis we identified three *BCR-ABL1*-positive cell lines among the 189 nilotinib-treated cell lines that overlapped CGP and these were also the three most sensitive samples ($P = 9 \times 10^{-7}$). Hence, despite the fact that these drug sensitivity data were accurately recapitulating biological expectations in both studies, the authors' criteria incorrectly classify nilotinib sensitivity as *discordant*. Of the 577 cell lines screened in CGP 573 do not harbour the nilotinib target, i.e. the *BCR-ABL1* fusion gene. Thus, given (as expected) no drug response in almost all cell lines screened (median AUC across all cell lines = 0.99; AUC of 1 represents no drug response; Fig. 1(a); Supplementary Table), there was little biological variability across the vast majority of cell lines, resulting in low correlation between the repeated measurements made by CCLE and CGP, despite clearly concordant results. Similarly, most other

drugs that the authors compared were also targeted agents, meaning this lack of drug response was *common*; indeed, for 10 of the 15 drugs, median AUC was greater than 0.90 in CGP and 8 of these 10 also have median AUC greater than 0.9 in CCLE, resulting in little variability across most cell lines when treated with these drugs. Indeed, we identified a *systematic* relationship between variability in drug response in either study and correlation between the two studies (Fig. 1(b)). A valid comparison of CGP and CCLE must consider the pharmacology of the drugs screened and in particular the differences in the variability induced by different drugs. Nilotinib was not an isolated case; in fact, despite the highly experimental nature of many of the compounds screened by CCLE/CGP, we still identified multiple expected associations that were consistently reported by both studies including *ERBB2* for lapatinib⁴, *NQO1* expression for 17-AAG⁵, *BRAF* mutation for PD-0325901⁶, AZD6244⁷, and PLX4720⁸, *MDM2* for Nutlin-3a⁹, and *MET* for Crizotinib¹⁰ (Supplementary Table). Finally, the utility of these pharmacogenomic datasets is now further supported by findings that models fit using data from CGP could reliably predict drug response in multiple clinical trials^{11,12}.

In summary, our analysis shows that Haibe-Kains *et al.*'s conclusions are unsubstantiated and we propose that a fair assessment of concordance between large pharmacogenomic datasets will require the development or adaptation of methods that account for the issues raised here, although great care will be required to ensure that such methods do not introduce their own unforeseen biases.

Methods

In CGP and CCLE, using ordered data common to both studies, gene expression and drug sensitivity (AUC) values can be arranged in $n_1 \times m$ and $n_2 \times m$ matrices respectively where m is the number of cell lines, n_1 is the number of genes and n_2 is the number of drugs common to both studies. Correlations “between” cell lines are calculated by the correlation of matching columns of CGP and CCLE matrices (vectors of length n_1 for expression or n_2 for AUC). Correlations “across” cell lines are the correlations of matching rows (vectors of length m for both data).

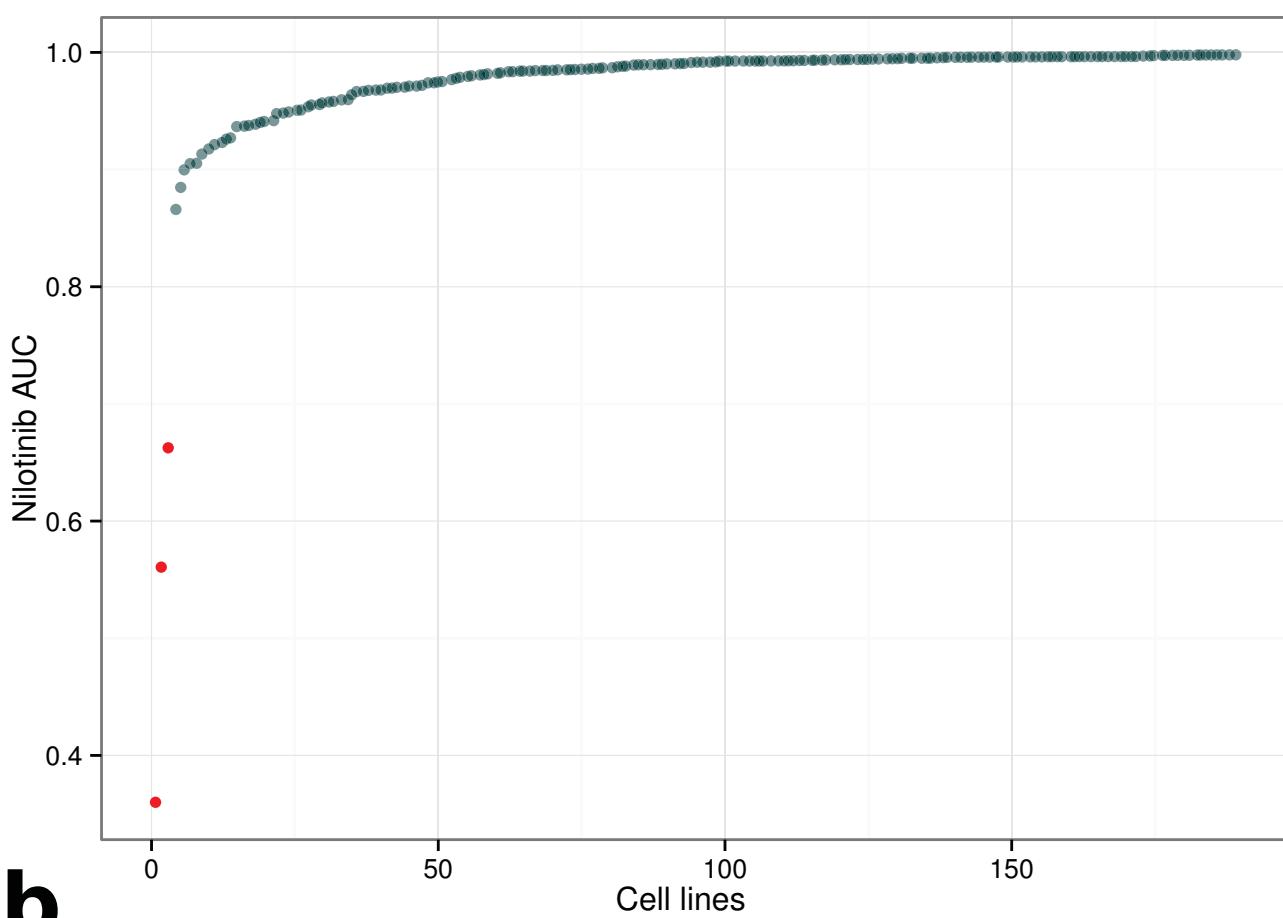
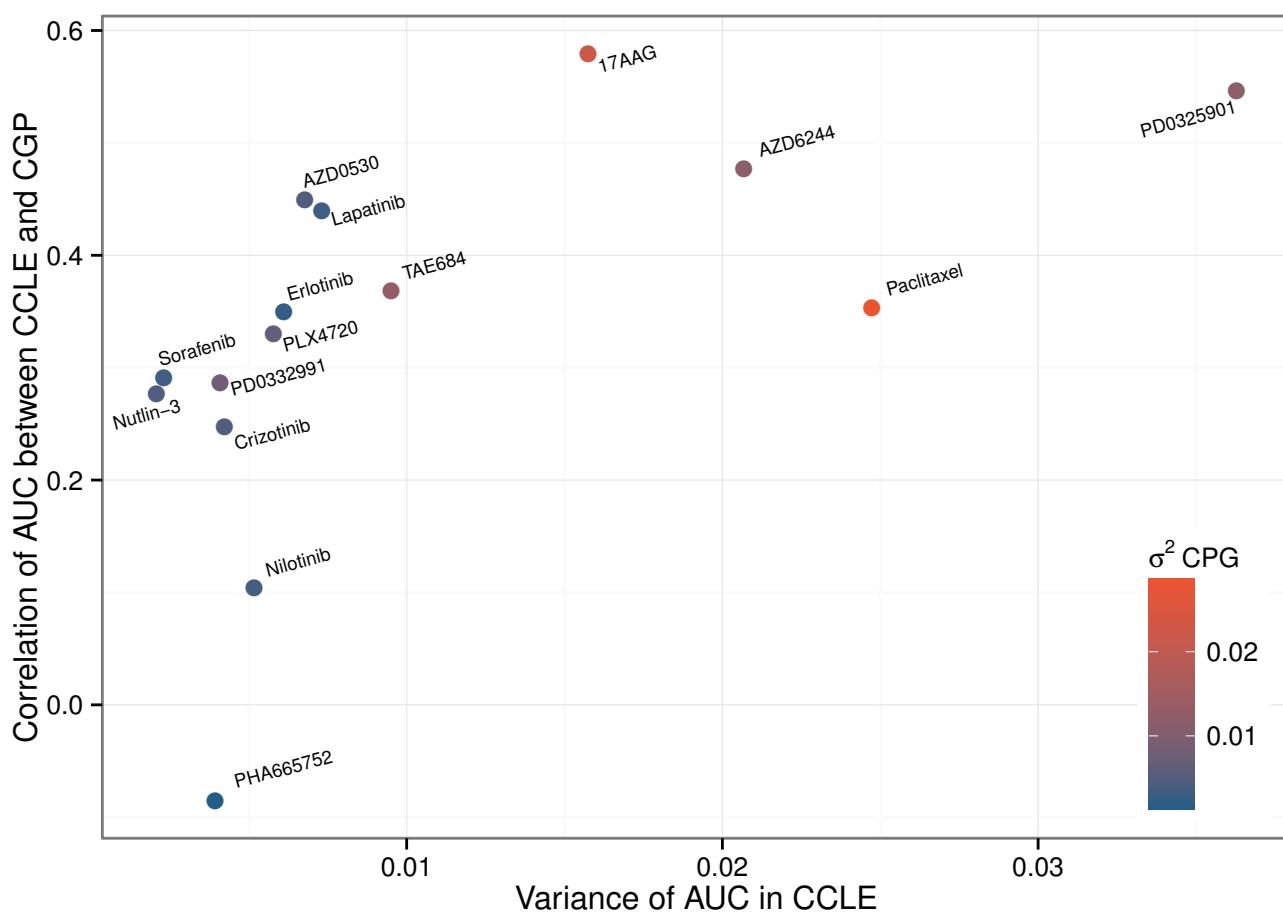
To achieve easy reproduction of our results, we have made the source code for our analysis available in a GitHub repository (https://github.com/paulgeeleher/nature_bca).

Figure Legends

- (a) Highly targeted agents (nilotinib) highlight a major limitation of the authors' test for concordance.** Scatterplot showing the nilotinib AUC values (in CGP) for the 189 cell lines that were screened by both CGP and CCLE. Only a very small proportion of cell lines achieve a response, e.g. the three BCR-ABL1 positive cell lines highlighted in red. This almost complete lack of biological variability renders a Spearman correlation ineffective as a means to assess concordance.
- (b) The authors' test for concordance is confounded by variability in drug response.** Scatterplot showing the strong association between “Spearman’s correlation of AUC between CCLE and CGP” and “variance of AUC in CCLE”. Drugs whose AUC is more variable are more likely to be highly correlated between CCLE and CGP ($r_s = 0.83, P = 1.9 \times 10^{-4}$). The points have been color coded by their “variance of AUC in CGP”, which is also significantly associated with both “variance of AUC in CCLE” and “Spearman’s correlation of AUC between CCLE and CGP”.

References:

1. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–93 (2013).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–7 (2012).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–5 (2012).
4. Konecny, G. E. *et al.* Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res.* **66**, 1630–9 (2006).
5. Kelland, L. R., Sharp, S. Y., Rogers, P. M., Myers, T. G. & Workman, P. DT-Diaphorase expression and tumor cell sensitivity to 17-allylaminoo, 17-demethoxygeldanamycin, an inhibitor of heat shock protein 90. *J. Natl. Cancer Inst.* **91**, 1940–9 (1999).
6. Solit, D. B. *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **439**, 358–62 (2006).
7. Dry, J. R. *et al.* Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). *Cancer Res.* **70**, 2264–73 (2010).
8. Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3041–6 (2008).
9. Müller, C. R. *et al.* Potential for treatment of liposarcomas with the MDM2 antagonist Nutlin-3A. *Int. J. Cancer* **121**, 199–205 (2007).
10. Timm, A. & Kolesar, J. M. Crizotinib for the treatment of non-small-cell lung cancer. *Am. J. Health. Syst. Pharm.* **70**, 943–7 (2013).
11. Geleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).
12. Falgreen, S. *et al.* Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models. *BMC Cancer* **15**, 235 (2015).

a**b**

Reviewer's general comments:

The authors have reasonably addressed the points raised by Haibe-Kains and colleagues (HK/Q). Specifically, they convincingly rebut the main point on the inadequacy of the Spearman correlation coefficient as a metric for the concordance of drug responses - if there is low variability among the cell lines and/or the number of sensitive cell lines is low (imbalanced data). They also highlight a critical flaw in the Haibe-Kains et al. manuscript where gene and drug correlations were computed inconsistently making their comparison meaningless. The strong dependence of correlation on variability across cell lines is clearly illustrated in the revised manuscript. Importantly, Haibe-Kains et al. seem to be still ignoring this point in the additional analyses they provided.

It is still unclear what level of correlation, accuracy and reproducibility, is possible for large pharmacogenomic screens, so it is difficult to judge how consistent or inconsistent CCLE vs CGP really is. Regardless, it is clear that there are analytical shortcomings of the Haibe-Kains manuscript.

Our response: We appreciate the referee's unbiased evaluation of our work and we are glad to see that the reviewer has agreed that the conclusions of Haibe-Kains *et al* do not hold up to scrutiny. Again, the primary goal of our manuscript was to highlight the shortcomings of the methods employed by Haibe-Kains *et al*. Given the reviewer's remaining concern is that "it is difficult to judge how consistent or inconsistent CCLE vs CGP really is", we have addressed this issue in this revision. The additional results included in our revised manuscript clearly demonstrate that the findings reported by CCLE and CGP were far more consistent than Haibe-Kains *et al* have led readers to believe.

Reviewer's comment 1. While the authors make some interesting points regarding a lack of variability causing a problem with correlation analysis, they have substantially undermined the conclusions made in the Haibe-Kains paper. For example, if correlation is not the best measure, then there should be a Fig. 1c that illustrates a superior metric that indicates a high degree of concordance for most samples; Or the authors should provide an explanation in a table with target names and citation for associated publication for the targeted nature of all drugs with low correlation values and proof that they do in fact have a strong and reproducible effect among the relevant cell lines. Given that there are only 15 drug treatments being studied across both studies, it is feasible to provide this information. This is done for nilotinib, lapatinib, and PD-0325901/PLX4720/AZ6244, but these limited examples apply to the majority of drugs tested.

Our response: We agree with the reviewer that it would be helpful to present the results of another more suitable concordance metric, to illustrate the concordance between drug sensitivity

data in CCLE and CGP. However, as we discussed in previous correspondence, our analysis of various other concordance metrics suggests that novel methods development will likely be necessary to address this problem in this way. However, even if it is possible, it may take some time before a consensus is reached on how to fairly assess concordance in these and similarly complex pharmacogenomics datasets and of course, any novel method could easily introduce additional unforeseen issues. Nevertheless, we have highlighted this opportunity in the revised manuscript. Indeed, we say:

“In summary, our analysis shows that Haibe-Kains et al.’s conclusions are unsubstantiated and we propose that a fair assessment of concordance between large pharmacogenomic datasets will require the development or adaptation of methods that account for the issues raised here, although great care will be required to ensure that such methods do not introduce their own unforeseen biases.”

Per the reviewer’s excellent suggestion and with the kind permission of the editor, we have compiled a new Supplementary Table. The table, summarizes the results reported by CCLE and CGP (information which is all publicly available) and compares them with expectation, given pre-existing literature. Of the 14 targeted agents studied (we exclude cytotoxic-agent paclitaxel), at least one *identical* gene target(s) and drug sensitivity relationships (consistent with previous literature) were recapitulated by both studies for 9 drugs, despite the “discordant” drug sensitivity data claimed by Haibe-Kains *et al.* Interestingly, for the remaining 5 drugs, at least one “expected” drug target relationship was recapitulated by one of the studies, suggesting these data are also valuable. Again, while labelling these results as “consistent”, “moderately consistent”, “highly consistent” etc. is a subjective matter, it seems impossible to reconcile this number of reproducible associations with the conclusions of Haibe-Kains *et al.*

In fact, the results in this Supplementary Table were achieved even despite many obvious factors that may limit reproducibility (“discordant” drug sensitivity data aside). For example, different sequencing and microarray platforms, different analytical methods, different drug screening concentration windows, the suitability of the phenotype screened (cellular viability) for some of the drugs screened (e.g. sorafenib – an angiogenesis inhibitor) and in particular, the highly experimental nature of some of the compounds (e.g. PHA-665752), which means that the evidence supporting their canonical targeting action is sometimes weak. Nevertheless, this table will allow a reader to make a far more unbiased subjective judgement as to the level of consistency evident in these data, than from the conclusions previously presented by Haibe-Kains *et al.*

If 17-AAG does in fact only affect a minority of cell lines, then it contradicts the authors' point: 17-AAG has the best correlation among any of the drugs.

Our response: Consistent with our conclusions, 17-AAG affects a very large number of cell lines. For example, in CGP over 300 cell lines achieved a measurable IC₅₀ value within the drug concentrations screened (0.00391 to 1.00 μM). As shown in the Supplementary Table, median AUC 0.77 and 0.56 in CGP and CCLE, respectively. As the reviewer pointed out, 17-AAG has the highest correlation between the two studies, and consistent with our findings (and despite this experimental compound's supposed "targeted" nature), this drug affected many cell lines and also induced the 4th highest level of variability across the CGP cell lines and the 2nd highest level of variability across the CCLE.

Reviewer's comment 2. In the rebuttal, the authors state "Of the 15 drugs for which the authors have assessed concordance, the median AUC value in CGP is greater than 0.95 in 9 of them (an AUC of 1 represents no drug response), indicating that, (as expected) very high numbers of cell lines are unresponsive to these drugs." However, this point is not made in the actual manuscript nor is it supported by the single main figure. Also, is there an equally convincing statistic for the CCLE samples? Why only mention the CGP samples? The statistic for low variability does not prove that there is in fact a high degree of concordance that is explained by known biological mechanisms.

Our response: We agree with the reviewer that high median AUCs are useful metrics for demonstrating the general lack of response of many drugs to the compounds assessed. Therefore, we have added the mean AUC values for all 15 drugs in both CGP and CCLE into the Supplementary Table and commented on these results in the revised manuscript.

In the manuscript, we now state this explicitly:

"Similarly, most other drugs that the authors compared were also targeted agents, meaning this lack of drug response was common; indeed, for 10 of the 15 drugs, median AUC was greater than 0.90 in CGP and 8 of these 10 also have median AUC greater than 0.9 in CCLE"

Reviewer's comment 3. In order to achieve "very easy" reproduction of results, access to the input data, such as the "CDRUG_cgpcCLE.RData" file, is needed. These data can be produced using the scripts from the original study (<https://github.com/bhklab/cdrug>); however, the scripts included with this rebuttal should be made publicly available and it would be nice if the authors could provide the relevant input files, since a few minor modifications had to be made to get the original code to work to produce the Rdata files for the author's code. At the very least, the authors should describe what steps need to be run before using their code. The "Part1" code was not able to be run to completion due to an error message "Error in as.data.frame.list(lapply(DF.LIST[[i]], as.character), stringsAsFactors = FALSE, : supplied 15

row names for 0 rows". The "minsample" and "myfdr" are not defined and these variables are not loaded as part of CDRUG cgp ccle.Rdata. When testing the code, the original CDRUG pipeline.R code to define values for these variables was used. Similarly, the code does not currently import the necessary 'myScatterPlot" function from the original CDRUG foo.R file, and the 'ggplot2' library has to be loaded.

Our response: As suggested, we have also made the revised code publicly available in a GitHub repository (https://github.com/paulgeeleher/nature_bca). To maximize transparency, we have maintained our original pipeline, which makes the fewest possible edits to Haibe-Kains' original code. However, we agree with the reviewer that the original Haibe-Kains *et al.* code is very difficult to run and may contain bugs; thus, as the reviewer has suggested, we have now provided an additional "very easy" pipeline to reproduce results. As per the reviewer's suggestion, this additional simplified pipeline provides the required ".RData" files, without need to run most of the Haibe-Kains' scripts. We also provide instructions on how to run both of these pipelines on the GitHub page. We are grateful to the reviewer for this very helpful suggestion.

Reviewer's comment 4. The authors state "measurement error, inherent in biological assays, will dominate over the (non-existent) biological variability". Great care should be taken before stating that there is no biological variability. Are the authors referring to measurements taken from the same cell lines? It is quite plausible that differences in growth protocols, drug treatment concentrations, etc. will affect gene expression values.

Our response: The text "*measurement error, inherent in biological assays, will dominate over the (non-existent) biological variability*" was referring to a *hypothetical* example, whereby the reader was asked to consider a drug with zero efficacy against all of the cell lines. The reason for this example is to help a reader understand the key relationship between lack of drug response, lack of biological variability and lack of expectation of a correlation of repeated measurements. We believe that such an example is important, thus, we apologize for the lack of clarity and we have thus taken steps to improve this example in the text. To address the reviewers concerns we now state:

"To see why correlation is not an appropriate measure of biological concordance for these data, consider the hypothetical example of a drug that is not effective against any cell lines (a possibility for an experimental drug): In such a case randomly-fluctuating measurement error, inherent in biological assays, will dominate over the (non-existent) biological variability, meaning that there could be no expectation of correlation between repeated measures of drug sensitivity (assuming other experimental variables are held constant)."

Reviewer's comment 5. Paragraph 2 of the rebuttal can be omitted. Although HK/Q should have tested the validity of the assumptions of their statistical models, their biggest error was a failure to recognize that most of the drugs are targeted and have only off-target effects on most cell lines.

Our response: We agree with the reviewer that this is the least important of the points that we have raised. Thus we have removed this paragraph and used the space freed up to address the reviewer's more important points.

Response to “Consistency in large pharmacogenomic studies”

Zhaleh Safikhani, Nehme El-Hachem, Petr Smirnov, Mark Freeman, Anna Goldenberg, Nicolai Juul Birkbak, Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush, Benjamin Haibe-Kains

Geeleher et al. claim to have discovered overall consistency between the CGP¹ and CCLE² by describing the reproducible response of three cell lines harboring BCR-ABL1 gene fusion and highly sensitive to a highly targeted drug, nilotinib (Supplementary Figure 1). They use this example to argue that, due to the targeted nature of many of the 15 drugs screened in both CGP and CCLE, the rest of the drugs should show consistent sensitivity measurements if one limits analysis to the highly sensitive cell lines. However, as described in more detail below, this isolated example does not invalidate our initial finding³ of a broader inconsistency in reported phenotypes between CGP and CCLE³.

Nonetheless, Geeleher and colleagues raise two potential issues with our published study and we welcome the opportunity to address them here.

1. In our initial publication, we computed the correlation of gene expression and mutation profiles between cell lines to assess whether large transcriptomic changes and/or genetic drift might be the cause of the observed inconsistency in drug sensitivity data (which was calculated across cell lines). We agree with Geeleher et al. that correlations *across* and *between* cell lines should be compared in a consistent manner.

Indeed, in reanalyzing our data, we do find differences in the correlation between cell lines and across cell lines. Overall, correlation across cell lines is lower than correlation between cell lines (Supplementary Figure 2). However, we still find that gene expression data are significantly more

concordant between studies than the drug response summary statistics (IC_{50} and AUC) values in all comparisons (Wilcoxon rank sum test $p<0.002$). Consequently, our conclusion that gene expression data are significantly more correlated than pharmacological response still holds although the even lower correlation of gene expression values across cell lines might lead one to argue that there is even less consistency between the CCLE and CGP studies than we initially reported.

2. We agree with the assertion by Geeleher et al. that lack of variability in drug sensitivity measurements may complicate biologically meaningful assessment of concordance between pharmacogenomic datasets. Geeleher et al. base their arguments on the sole example of nilotinib for which there are three sensitive cell lines out of the 200 cell lines screened in both datasets. However, even among these three cell lines, the AUC values are not concordant; the least sensitive of the three cell lines in CGP is actually the most sensitive one in CCLE (Supplementary Figure 1; Supplementary Information). Therefore the only way to consider these results concordant is to classify these three cell lines as sensitive and the remainder as resistant, which cannot easily be done using the Waterfall approach used in the CCLE study² (Supplementary Tables 1 and 2 in our initial study).

The authors claim that nilotinib is not an isolated example. To test this, we adapted the Matthews correlation coefficient (MCC)⁴ to automatically select the *optimal* cutoff for consistency. We used this measure of association, referred to as adaptive MCC (AMCC; see Supplementary Methods) to estimate concordance between drug sensitivity calls, where only a few cell lines may be sensitive, or between gene expressions where the gene of interest may be rarely expressed. As expected, nilotinib yielded an AMCC of 1 which denotes perfect consistency between the two studies (Supplementary Figure 3); however, the rest of the drugs yielded much lower AMCC, with only AZD0530, lapatinib and crizotinib yielding AMCC around 0.65, and five other drugs (17-AAG, AZD6244, erlotinib, PD0325901 and PLX4720) yielding moderate consistency ($0.5 \leq AMCC < 0.6$). It should be noted that the

inter-laboratory replicates of the measurements of camptothecin and AZD6482 sensitivity performed using the same experimental protocols at two different locations within CGP yielded an AMCC of only 0.55 and 0.41 for AUC and IC₅₀ respectively, indicating a lack of reproducibility of drug phenotype measures between biological replicates (Supplementary Figure 4). Consistent with our previous report, gene expression data were significantly more concordant than drug sensitivity data across cell lines (Wilcoxon rank sum test p-values < 0.006, Supplementary Figure 5). This re-analysis confirms that nilotinib is an anecdotal case and that drug sensitivity measurements for the rest of the drugs (cytotoxic or targeted) remain inconsistent.

Geeleher et al. also state that their recent paper in Genome Biology⁵ supports the utility of the data from the CCLE and CGP and they imply that their findings provide evidence for a consistency between CGP and CCLE. However, a true test of this assertion would be to train their models on the CGP and to use these to predict phenotypes reported by CCLE (and vice versa). If they could predict the drug response phenotype in the independent validation set with high accuracy, this would provide some quantitative evidence of a consistency between the two datasets in the context of Geeleher et al.'s predictive models. However, we⁶ and others⁷⁻⁹ have shown that such an analysis does not yield robust predictions for most drugs.

In our original report, we found statistically significant non-zero correlations between phenotype measurements for almost all drugs, supporting the fact that there is biologically relevant signal in these datasets, albeit confounded by significant noise. Concurring with Goosspeed et al.¹⁰, we identified known gene-drug associations that are reproducible between CGP and CCLE (Supplementary Files 3, 4, 12, and 13 in our original report; detailed statistics provided in Supplementary Information). In a recent re-analysis of the updated CGP and CCLE datasets, we found that the majority of these associations, but not all, are reproducible¹¹. We also reported that discovery of new biomarkers, which

was the main goal of the CGP and CCLE studies, was much more challenging due to inconsistency in pharmacological phenotypes¹¹. In the conclusion of our original paper, we argued that additional work is necessary to improve consistency of phenotypic measures with the ultimate goal of making data from these large-scale projects more useful for development of robust predictors of drug response, and we believe that this conclusion still holds upon our reanalysis. We and others are actively working on identifying stable measures that could lead to improved consistency across inter-lab experiments.

Like Geeleher et al., we originally hoped to use the CCLE and CGP to develop robust biomarkers that could predict response to treatment. While we could use methods to find some consistency in selected subsets of the data, we found no general methods that identified an overall consistency between these studies. Geeleher et al. have shown that if one appropriately discretized the drug sensitivity data, one can find consistency for nilotinib—but not for the other compounds tested by the CCLE and CGP. Despite the single example of nilotinib, we would conclude as we did originally, that drug response phenotypes measured by the CCLE and CGP are inconsistent.

References

1. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
3. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
4. Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451 (1975).
5. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).
6. Papillon-Cavanagh, S. *et al.* Comparison and validation of genomic predictors for anticancer drug sensitivity. *J. Am. Med. Inform. Assoc.* **20**, 597–602 (2013).

7. Dong, Z. *et al.* Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* **15**, 489 (2015).
8. Jang, I. S., Neto, E. C., Guinney, J., Friend, S. H. & Margolin, A. A. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac. Symp. Biocomput.* 63–74 (2014).
9. Cortés-Ciriano, I. *et al.* Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btv529
10. Goodspeed, A., Heiser, L. M., Gray, J. W. & Costello, J. C. Tumor-derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Mol. Cancer Res.* (2015). doi:10.1158/1541-7786.MCR-15-0189
11. Safikhani, Z. *et al.* Revisiting inconsistency in large pharmacogenomic studies. *bioRxiv* 026153 (2015).

Supplementary Information

Response to
"Consistency in large pharmacogenomic studies"
from Geeleher et al.

Zhaleh Safikhani, Nehme El-Hachem, Petr Smirnov, Mark Freeman, Anna Goldenberg, Nicolai Juul Birkbak, Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush, Benjamin Haibe-Kains

Contents

1 Known gene-drug associations	3
2 Targeted versus cytotoxic drugs	3
3 Supplementary Methods	5
3.1 PharmacoGx	5
3.2 Measures of concordance	7
3.2.1 Spearman rank-based correlation	7
3.2.2 Adaptive Matthews correlation	7
4 Full Reproducibility of the Analysis Results	9
4.1 Set up the software environment	9
4.2 Run the R scripts	9
4.3 Generate the Supplementary Information	10
5 List of Abbreviations	11
6 Supplementary Tables	12
7 Supplementary Figures	13

1 Known gene-drug associations

As recently reported by Goodspeed et al [3], several known biomarkers for targeted therapies have been shown to be predictive in both CGP and CCLE. In our initial comparative study we also found the following known gene-drug associations:

- BRAF mutations were significantly associated with sensitivity to MEK inhibitors (AZD6244 and PD-0325901) and BRAFV600E inhibitor (PLX4720) with nominal p-values < 0.01; see Supplementary Files 10-13 of our initial study.
- ERBB2 expression was significantly associated with sensitivity to Lapatinib with nominal p-value = 0.04 and 8.4×10^{-15} for CGP and CCLE, respectively; see Supplementary Files 4 and 5 of our initial study.
- NQO1 expression was significantly associated with sensitivity to 17-AAG with nominal p-value = 2.4×10^{-13} and 6.2×10^{-14} for CGP and CCLE, respectively; see Supplementary Files 4 and 5 of our initial study.
- MDM2 expression was significantly associated with sensitivity to Nutlin-3 with nominal p-value = 7.7×10^{-18} and 7×10^{-8} for CGP and CCLE, respectively; see Supplementary Files 4 and 5 of our initial study.
- ALK expression was significantly associated with sensitivity to TAE684 with nominal p-value = 1.6×10^{-9} and 1.7×10^{-9} for CGP and CCLE, respectively; see Supplementary Files 4 and 5 of our initial study.

In our recent re-analysis of the updated CGP and CCLE datasets, we found that not all known gene-drug associations were reproducible between the two studies [8].

2 Targeted versus cytotoxic drugs

The authors limited their discussion to nilotinib. Given the prior knowledge that BCR-ABL1 fusion is the primary target of nilotinib and given that this gene fusion is rare in cancer cell lines, only a few sensitive cell lines are expected. There are three such cell lines that were screened with nilotinib by both the CGP and CCLE:

Cell line	AUC in CGP	AUC in CCLE
EM-2	0.37	0.76
KU812	0.64	0.65
MEG-01	0.44	0.58

Among the three sensitive cell lines the AUC values are not quite concordant; the least sensitive cell line in CGP is actually the most sensitive one in CCLE. Therefore, the only way to consider these results concordant is to classify these three cell lines as sensitive and the remaining cell lines as resistant, yielding a perfect agreement between the two datasets. However, the use of an automated method for drug sensitivity calling, such as the waterfall approach introduced in the CCLE study, does not allow to reach high level of concordance. However, the AMCC statistic, by optimizing the cutoff in CGP and CCLE, yield perfect consistency as shown in Supplementary Figure 3.

It is worth pointing out that nilotinib is not a representative example of all targeted drugs investigated in CCLE and CGP. As can be seen in Supplementary Figure 1, there are several cell lines for erlotinib, PHA665752, AZD0530, PLX4720, and Nutlin3 that exhibit high sensitivity in only one of the datasets despite these being highly targeted drugs whose specificity should be predictable based on the genetic background of the cells.

Geeleher and colleagues claim that most drugs in our comparative study are targeted (Supplementary Table 1), and that absence of variability in drug sensitivity data was widely observed. However, in addition to paclitaxel (the only cytotoxic drug in our study), we observed that several supposedly targeted drugs exhibited broad effect in the set of 512 cell lines screened both in CGP and CCLE (17-AAG, PD0325901 and TAE684 yielded MAD of AUC values > 0.12 in at least one dataset; Supplementary Figure 6A,B). For these drugs, one could argue that the variability of drug sensitivity measures warrants the use of a correlation metric for assessing the concordance between CGP and CCLE studies. However we observed only poor to fair correlation for these drugs ($r_s < 0.6$, Supplementary Figure 1).

Alternatively, we adapted the Matthews correlation coefficient (AMCC) to assess the concordance across studies by accounting for the lack of variability of sensitivity measurements for targeted drugs (see Section 3.2.2). While we agree with the authors that the sensitivity to nilotinib is reproduced in both studies (AMCC = 1; Supplementary Figure 3); this observation is anecdotal and can not be generalized to the other targeted and cytotoxic drugs investigated in our original report. Indeed, nilotinib was the only drug yielding an AMCC value > 0.8 (Supplementary Figure 3). Further, much of the motivation for the CGP and CCLE studies was the creation of large datasets through the screening of hundreds of cell lines that could be used to develop predictive models of drug response; if such studies are only useful in assessing targeted agents, it undermines the value of such large-scale pharmacogenomic studies.

3 Supplementary Methods

3.1 PharmacoGx

The lack of standardization of cell line and drug identifiers hinders comparison of molecular and pharmacological data between large-scale pharmacogenomic studies, such as the CGP and CCLE datasets. To address this issue we developed *PharmacoGx*, a computational platform enabling users to download and interrogate large pharmacogenomic datasets that were extensively curated to ensure maximum overlap and consistency. *PharmacoGx* provides (*i*) a new object class, called *PharmacoSet*, that acts as a container for the high-throughput pharmacological and molecular data generated in large pharmacogenomics studies; and (*ii*) a set of parallelized functions to assess the reproducibility of pharmacological and molecular data and to identify molecular features associated with drug effects. The *PharmacoGx* package is open-source and publicly available on the Comprehensive R Archive Network (<https://cran.r-project.org/web/packages/PharmacoGx/>).

Structure of the *PharmacoSet* class

```
@ annotation:  
  $ name: Acronym of the pharmacogenomic dataset.  
  $ dateCreated: When the object was created.  
  $ sessionInfo: Software environment used to create the object.  
  $ call: Set of parameters used to create the object.  
  
@ datasetType: 'sensitivity'  
@ cell: data frame annotating all cell lines investigated in the study.  
@ drug: data frame annotating all the drugs investigated in the study.  
@ sensitivity:  
  $ n: Number of experiments for each cell line treated with a given drug  
  $ info: Metadata for each pharmacological experiment.  
  $ phenotype: Drug sensitivity values summarizing each dose-response curve (IC50, AUC, etc.)  
  
@ molecularProfiles: List of ExpressionSet objects containing the molecular profiles of the cell lines, such as mutations, gene expressions, or copy number variations.
```

Creation of the CGP_Nature2013 *PharmacoSet* Gene expression, mutation data and cell line annotations were downloaded from ArrayExpress. Drug sensitivity measurements and drug information were downloaded from the CGP website ([link](#)) and the Nature website ([link](#)), respectively.

Minimum and maximum screening concentrations for each drug/cell line were extracted from `gdsc_compounds_conc_w2.csv` (μM).

The natural logarithm of IC₅₀ measurements were retrieved from `gdsc_manova_input_w2.csv` in column "`*_IC_50`" (referred to as *x*) and subsequently transformed using $-\log_{10}(\exp(x))$ in the analysis (high values are representative of cell line sensitivity to drugs).

The AUC measurements were retrieved from `gdsc_manova_input_w2.csv` in column "`*_AUC`" (referred to as *x*); high values are representative of cell line sensitivity to drugs.

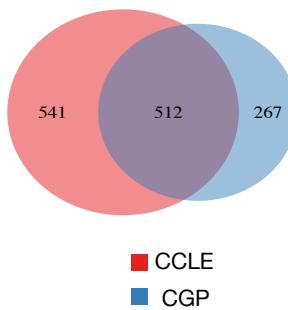
Creation of the CCLE_Nature2013 *PharmacoSet* Gene expression, mutation data cell line annotations and drug information were downloaded from the CCLE website ([link](#)) Drug sensitivity data were downloaded from the Nature website ([link](#));

Screening concentrations for each drug/cell line were extracted from Supplementary Table 11 in column E (μM).

IC_{50} measurements were retrieved from Supplementary Table 11 in column J (" IC_{50} μM (norm)") (referred to as x) and subsequently transformed into logarithmic scale, $-\log_{10}(x)$ in the analysis (high values are representative of cell line sensitivity to drugs).

AUC measurements were retrieved from Supplementary Table 11 in column L ("ActArea (norm)") and subsequently divided by the number of drug concentrations tested (8); high values are representative of cell line sensitivity to drugs.

Cell line annotations Cell line names were harmonized in both CGP and CCLE to match identical cell lines; this was done through manual search over alternative names of cell lines, as reported in CGP and CCLE cell line annotation files and online databases such as [hyperCLDB](#) and [BioInformationWeb](#). In our comparative analysis published in Nature [4], we focused on the set of 471 cell lines for which both gene expression and drug sensitivity were available. In the present work we extended our curation to all the cell lines for which at least one data type (gene expression, mutation or drug sensitivity) is available, increasing the shared set of cell lines to 512:



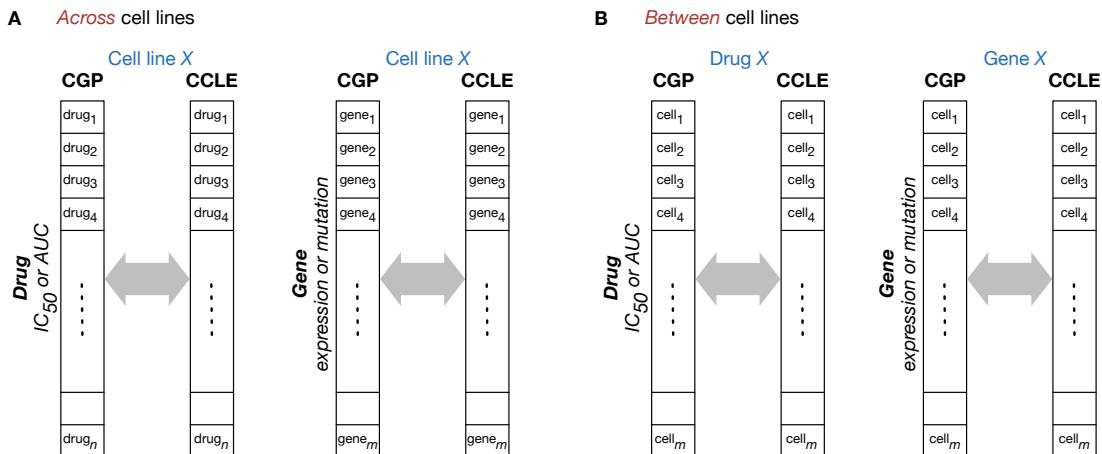
Tissue type nomenclature from CGP [2] was chosen throughout this study, CCLE tissue type information [1] was therefore updated to follow this nomenclature, which resulted in 23 tissue types:

Tissue type	Number of cell lines
Lung	109
Haematopoietic and lymphoid tissue	85
Breast	39
Central nervous system	35
Large intestine	34
Skin	27
Oesophagus	21
Urinary tract	17
Ovary	16
Pancreas	16
Stomach	16
Upper aerodigestive tract	13
Autonomic ganglia	12
Soft tissue	12
Endometrium	11
Kidney	10
Liver	10
Bone	8
Thyroid	8
Pleura	6
Prostate	4
Biliary tract	2
Small intestine	1

Gene expressions Raw gene expression profiles (Affymetrix CEL format) for 789 CGP and 1,067 CCLE cell lines were downloaded, respectively, from ArrayExpress and CCLE websites. Gene expression data were normalized with frozen RMA [7] using the Bioconductor Chip Description File (CDF) definitions (hthgu133a from CGP, and hgu133plus2 for CCLE). We then used the R package jetset [5], to map Affymetrix probe sets to unique Entrez gene IDs by selecting the best probe set for each gene. For replicates, the median of each gene expression was computed.

3.2 Measures of concordance

We assessed the concordance of the gene expression, mutation and drug sensitivity of CGP and CCLE studies *across* and *between* cell lines, as illustrated in the figure below. When data are compared across cell lines, we assess whether, for a given gene expression or drug, the cell line data were concordant (a gene is expressed at a similar level or similar response to a drug is observed in the same set of cell lines for instance; panel **A**). When data are compared between cell lines, we assessed whether, for a given cell line, the genomic and pharmacological profiles were concordant in the two studies (a given cell line harbours similar gene expression patterns or pharmacological responses for instance; panel **B**).



3.2.1 Spearman rank-based correlation

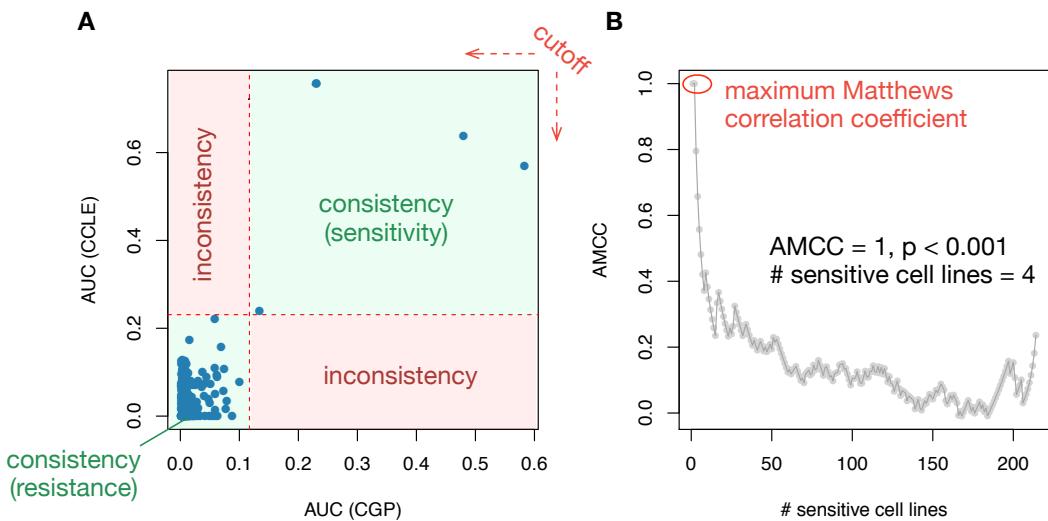
We assessed consistency between gene expressions or pharmacological responses in CGP and CCLE by computing Spearman rank-ordered correlations when ≥ 10 measures were available. The use of Spearman correlation is appropriate when high biological variability is observed, as opposed to pure technical noise. For instance, Spearman correlation can be used as a measure of concordance for drugs inducing growth inhibition in a substantial set of cell lines, with a sufficient dynamic range (median absolute deviation of AUC > 0.10 ; Supplementary Figure 6). This is the case for 17-AAG, Paclitaxel, TAE684, and PD-0325901 in both CCLE and CGP, PD0332991 in CGP only, and AZD6244 in CCLE only (Supplementary Figure 6C).

Typically qualitative descriptions of correlation coefficients are associated with intervals: $r_s < 0.5$, poor correlation; $0.5 \leq r_s < 0.6$, fair correlation; $0.6 \leq r_s < 0.7$, moderate correlations; $0.7 \leq r_s < 0.8$, substantial correlation; and $r_s \geq 0.8$, almost perfect correlation.

3.2.2 Adaptive Matthews correlation

When the variability of the measurements is mostly due to technical noise, correlation is not an appropriate measure of concordance. This is the case for drugs where only a small subset of cell lines are expected to be sensitive. Similarly, genes that are expressed in a few cell lines are also

problematic. We therefore developed the adaptive Matthews correlation coefficient (AMCC), a statistic specifically designed to detect consistent outliers across datasets. The rationale behind AMCC is illustrated in the figure below using sensitivity measurements for nilotinib as an example. The idea is to rank each measurement in the two dataset of interest (here CGP on the x-axis and CCLE on the y-axis; see panel **A**). The cutoffs are then varied in the two studies to include one more measurement at each step in order to define inconsistent and consistent observations and subsequently estimate the Matthews correlation coefficient [6]. Finally, the maximum coefficient is selected and its significance is computed using a permutation test where the cell line labels are permuted 1000 times in each dataset (see panel **B**).



Note that AMCC optimizes the rank cutoff to yield the highest consistency between two studies. Ideally, drug sensitivity calling should be computed in each dataset separately to enable assessment of concordance of the resulting calls using well-established χ^2 tests. Therefore AMCC may yield overoptimistic estimates of concordance.

4 Full Reproducibility of the Analysis Results

We will describe how to fully reproduce the figures and tables reported in the main manuscript. We automated the analysis pipeline so that minimal manual interaction is required to reproduce our results. To do this, one must simply:

1. Set up the software environment
2. Run the R scripts
3. Generate the Supplementary Information

The code and associated files are publicly available on GitHub: <https://github.com/bhklab/cdrug-rebuttals>.

4.1 Set up the software environment

We developed and tested our analysis pipeline using R running on linux and Mac OSX platforms.

To mimic our software environment the following R packages should be installed. All these packages are available on CRAN¹ or Bioconductor². Run the following commands in a R session to

install all the required packages:

```
source("http://bioconductor.org/biocLite.R")
biocLite(c("PharmacoGx", "VennDiagram", "xtable", "Hmisc"))
```

Note that PharmacoGx requires that several packages are installed. However, all dependencies are available from CRAN or Bioconductor.

Once the packages are installed, clone the cdrug-rebuttals GitHub repository (<https://github.com/bhklab/cdrug-rebuttals>) This should create a directory on the file system containing the following file:

`cdrug_analysis_huang.R` Script generating all the figures and tables reported in the manuscript.

`HaibeKains_Nature_2013_common_cellines.csv` Set of 504 cell lines shared between CGP and CCLE as published in Haibe-Kains *et al.*, Nature, 2013.

All the files required to run the automated analysis pipeline are now in place. It is worth noting that 500 MB storage for downloading CGP and CCLE Psets which is done Automatically when user runs `cdrug_analysis_huang.R` script.

4.2 Run the R scripts

Open a terminal window and go to the `cdrug` directory. You can easily run the analysis pipeline either in batch mode or in a R session. You should pass the number of CPU cores you want to allocate to the analysis (by default only 1 CPU core will be used).

To run the full pipeline in batch mode, simply type the following command:

```
Rscript code/cdrug_analysis_huang.R 4 > Rout
#to allocate four CPU cores for instance.
```

¹<http://cran.r-project.org>

²<http://www.bioconductor.org>

The progress of the pipeline could be monitored using the following command:

```
tail -f Rout
```

To run the full analysis pipeline in an R session, simply type the following command:

```
nbcore <- 4  
#to allocate four CPU cores for instance.  
source("code/cdrug_analysis_huang.R")
```

Key messages will be displayed to monitor the progress of the analysis.

The analysis pipeline was developed so that all intermediate analysis results are saved in the directories `data` and `saveres`. Therefore, in case of interruption, the pipeline will restart where it stopped.

Based on our new curation the number of shared cell lines between studies have been increased from 471 to 512. In case user wants to limit the analyses to previous 471 cell lines in our original study they should set the flag variable `confine.analyses.to.nature.common.cell.lines` to TRUE in the script.

4.3 Generate the Supplementary Information

After completion of the analysis pipeline a directory `saveres` will be created to contain all the intermediate results, tables and figures reported in the main manuscript and this Supplementary Information.

5 List of Abbreviations

AMCC	Adaptive Matthews Correlation Coefficient.
AUC	Area under the drug sensitivity curve.
CGP	Cancer Genome Project initiated by the Wellcome Sanger Institute.
CCLE	The Cancer Cell Lines Encyclopedia initiated by Novartis and the Broad Institute.
IC_{50}	Concentration in micro molar [μM] at which the drug inhibited 50% of the cellular growth.
FDR	False Discovery Rate
MAD	Median Absolute Deviation.
R_s	Spearman correlation coefficient

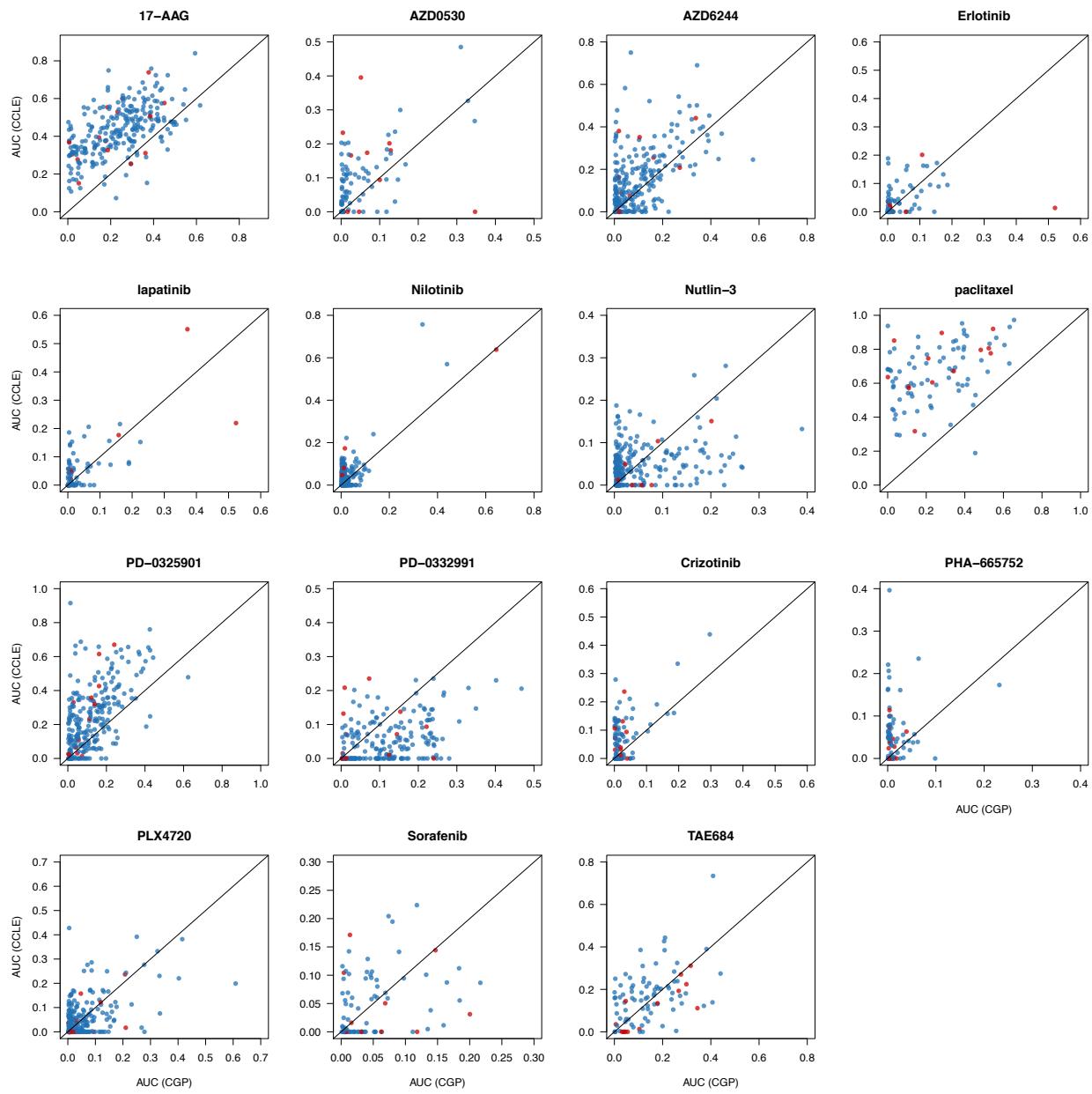
6 Supplementary Tables

Supplementary Table 1: Description of the 15 anticancer drugs screened both in CGP and CCLE studies.

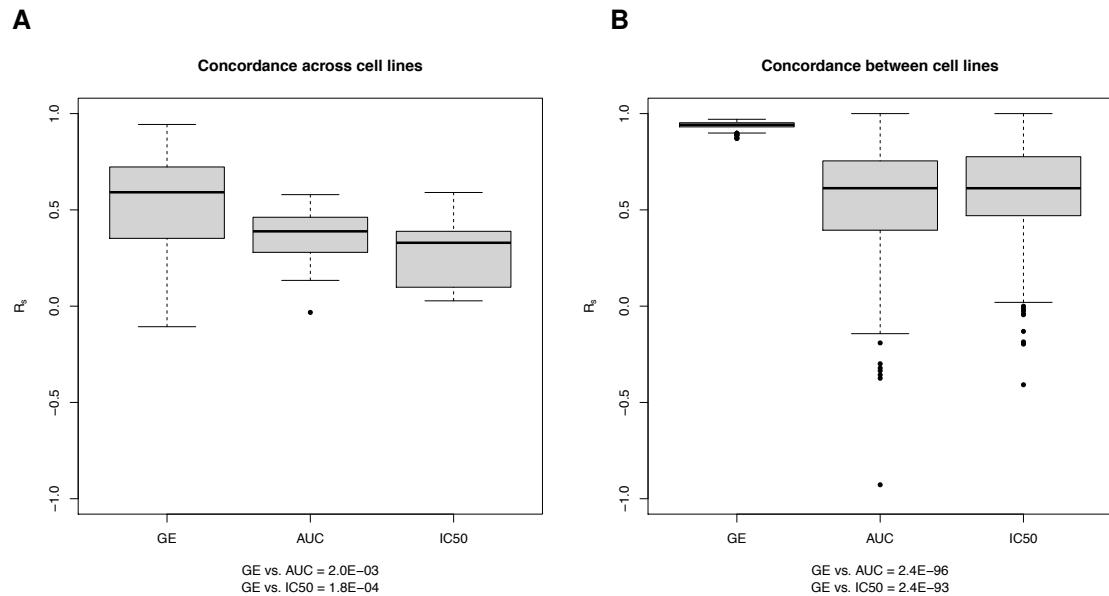
Compound	Class	Target(s)	Class	Organization
Erlotinib	Targeted	EGFR	Kinase inhibitor	Genentech
Lapatinib	Targeted	EGFR, HER2	Kinase inhibitor	GlaxoSmithKline
PHA-665752	Targeted	c-MET	Kinase inhibitor	Pfizer
Crizotinib	Targeted	c-MET, ALK	Kinase inhibitor	Pfizer
TAE684	Targeted	ALK	Kinase inhibitor	Novartis
nilotinib	Targeted	Abl/Bcr-Abl	Kinase inhibitor	Novartis
AZD0530	Targeted	Src, Abl/Bcr-Abl, EGFR	Kinase inhibitor	AstraZeneca
Sorafenib	Targeted	Flt3, C-KIT, PDGFRbeta, RET, Raf kinase B, Raf kinase C, VEGFR-1, KDR, FLT4	Kinase inhibitor	Bayer
PD-0332991	Targeted	CDK4/6	Kinase inhibitor	Pfizer
PLX4720	Targeted	RAF	Kinase inhibitor	Plexxikon
PD-0325901	Targeted	MEK	Kinase inhibitor	Pfizer
AZD6244	Targeted	MEK	Kinase inhibitor	AstraZeneca
Nutlin-3	Targeted	MDM2	Other	Roche
17-AAG	Targeted	HSP90	Other	Bristol-Myers Squibb
Paclitaxel	Cytotoxic	beta-tubulin	Cytotoxic	Bristol-Myers Squibb

7 Supplementary Figures

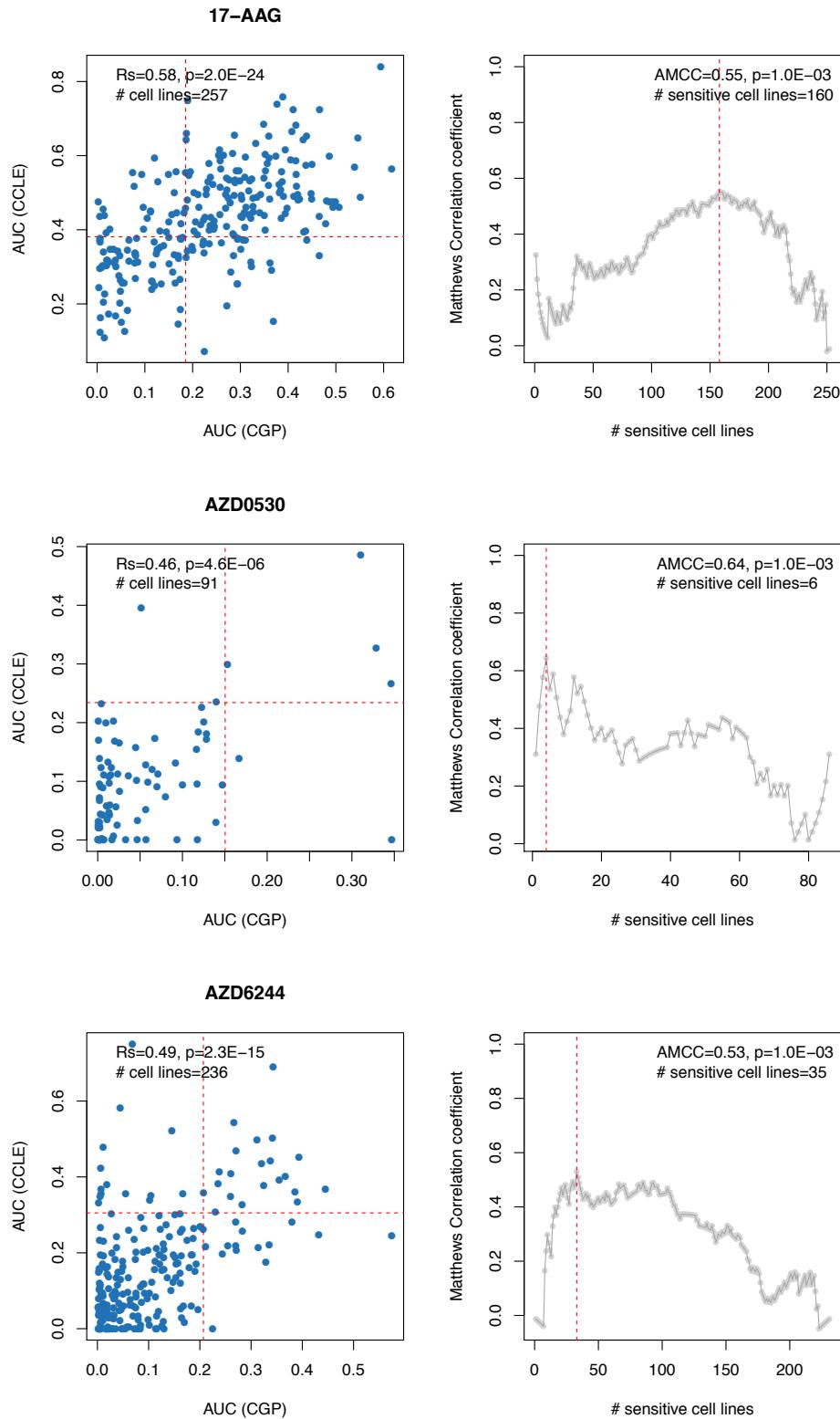
Supplementary Figure 1: Consistency across cell lines of AUC values between CGP and CCLE for the 512 cell lines and 15 drugs investigated both in CGP and CCLE. The scatter plot represents each data point, where the 471 cell lines analyzed in our original study are represented by blue points, while the new cell lines identified after additional curation of the data are presented in red. In case of perfect consistency, all the points should lie on the diagonal.



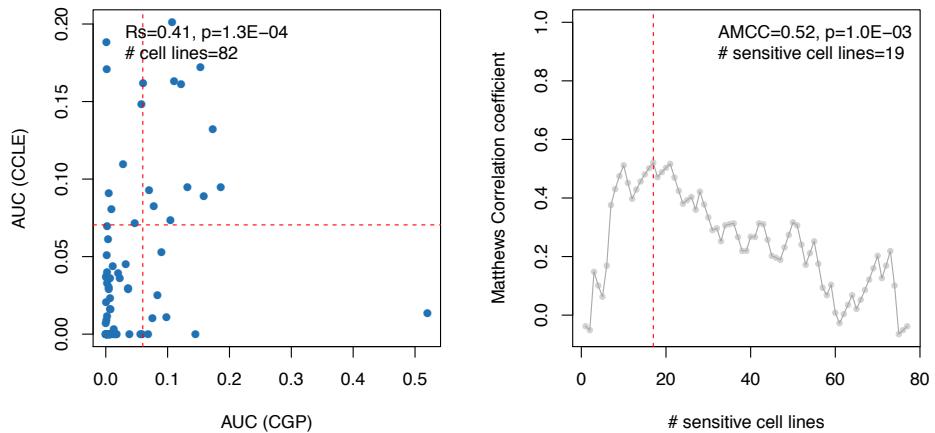
Supplementary Figure 2: Box plot comparing Spearman correlation coefficients estimating the concordance of gene expressions, IC_{50} and AUC measures **(A)** across and **(B)** between cell lines in CGP and CCLE. Significance of the difference between concordance observed for genomic and pharmacological data, as computed using the Wilcoxon rank sum test, is provided under each plot. GE: gene expression; R_s : Spearman's rank-ordered correlation.



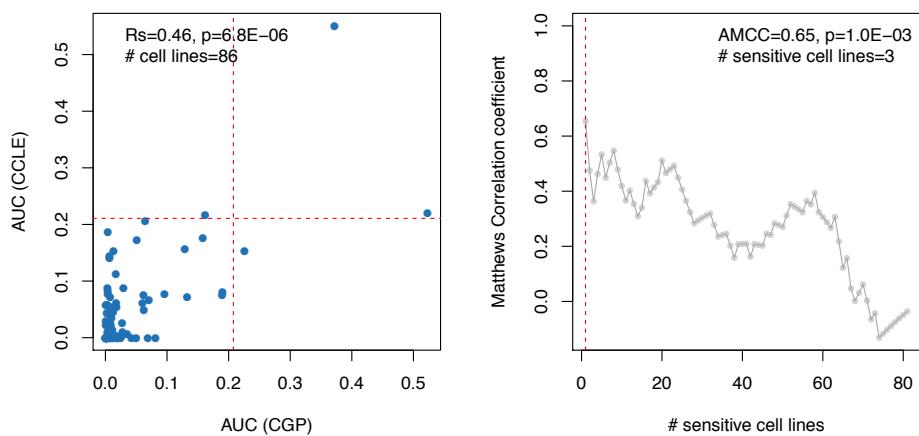
Supplementary Figure 3: Scatterplots of AUC measures in CGP and CCLE with cutoffs determined by the AMCC statistic and evolution of the Matthews correlation coefficient with respect to the cutoff values.



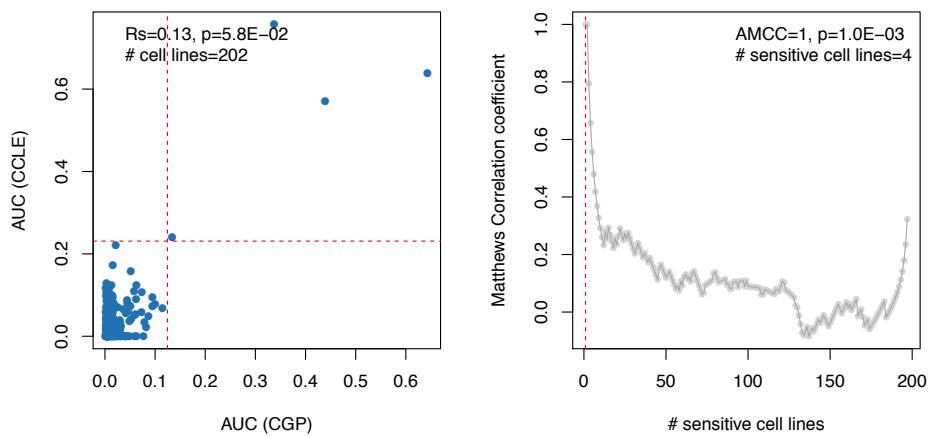
Erlotinib



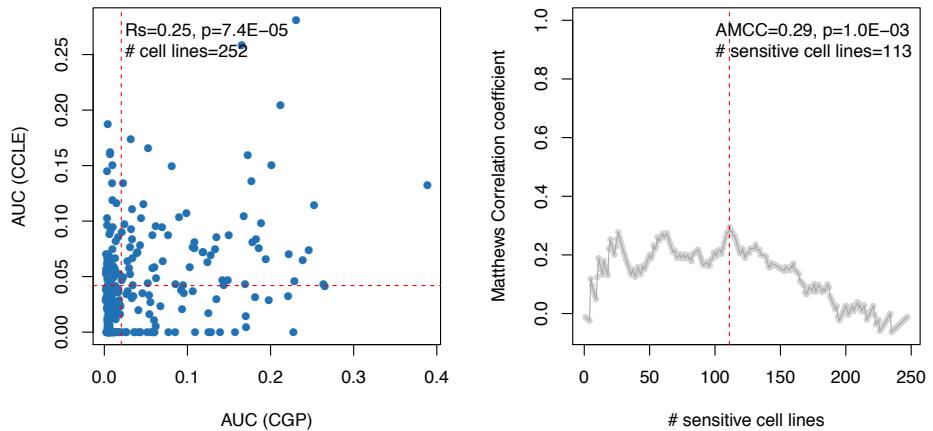
Ipatinib



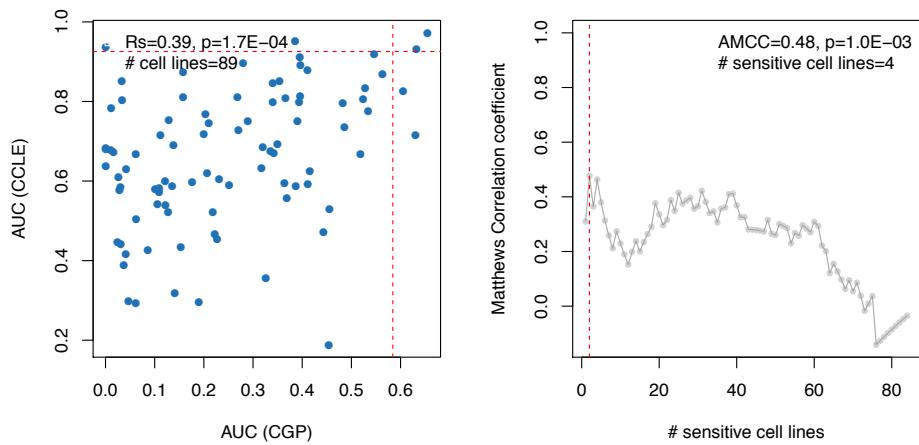
Nilotinib



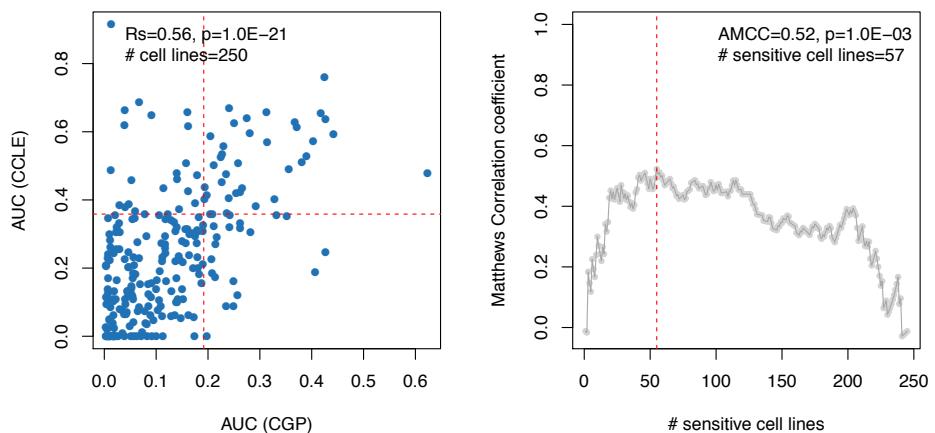
Nutlin-3



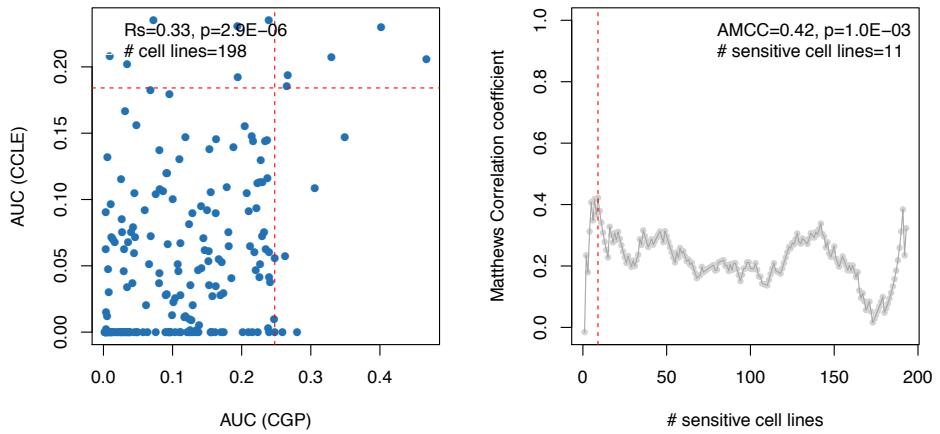
paclitaxel



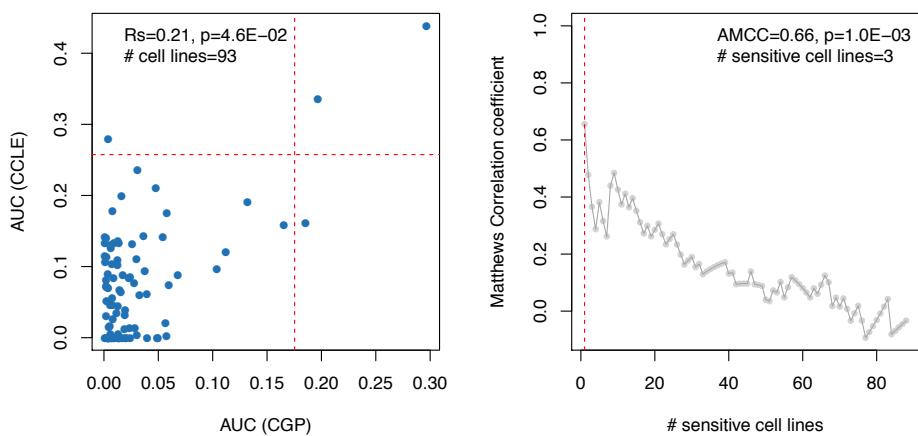
PD-0325901



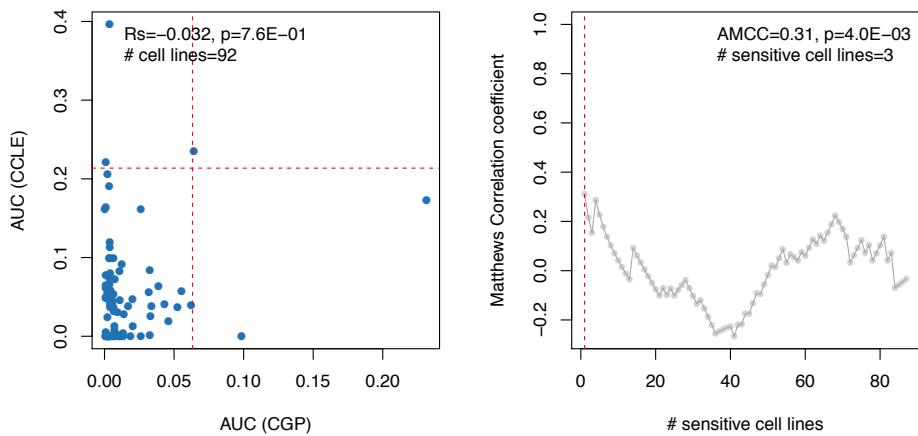
PD-0332991



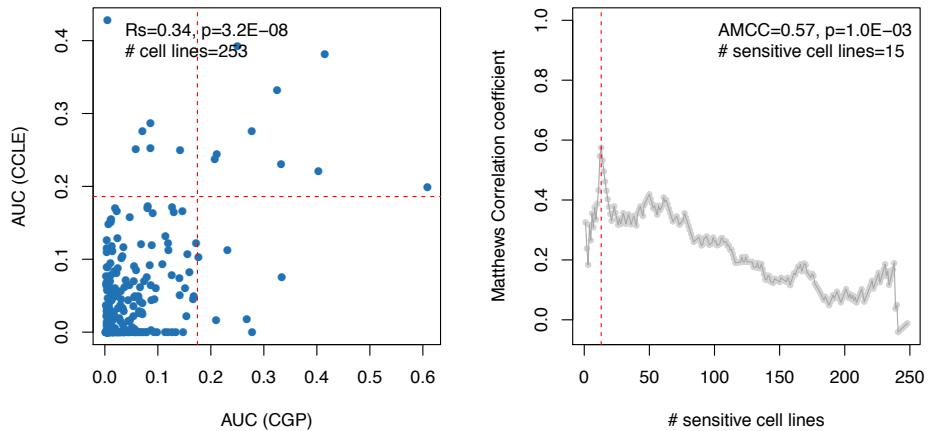
Crizotinib



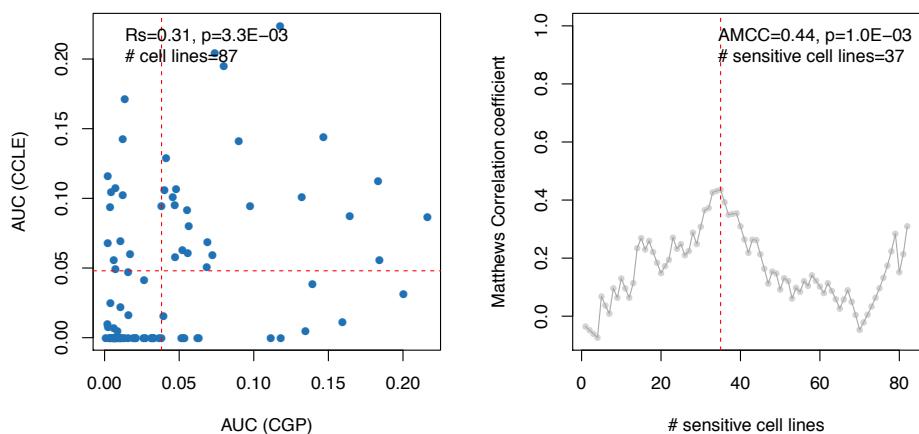
PHA-665752



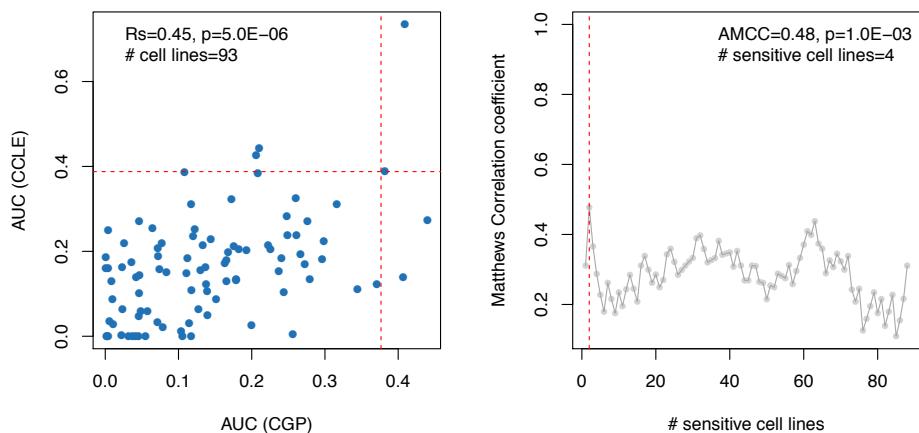
PLX4720



Sorafenib

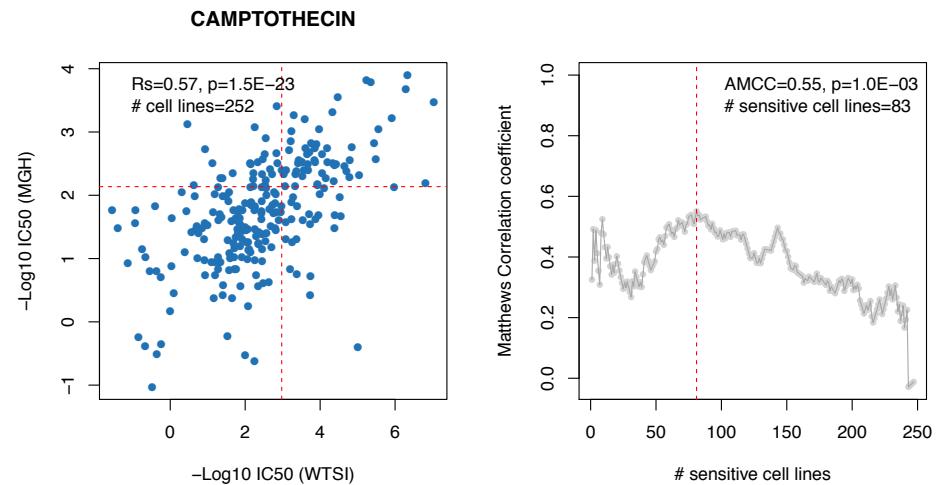


TAE684

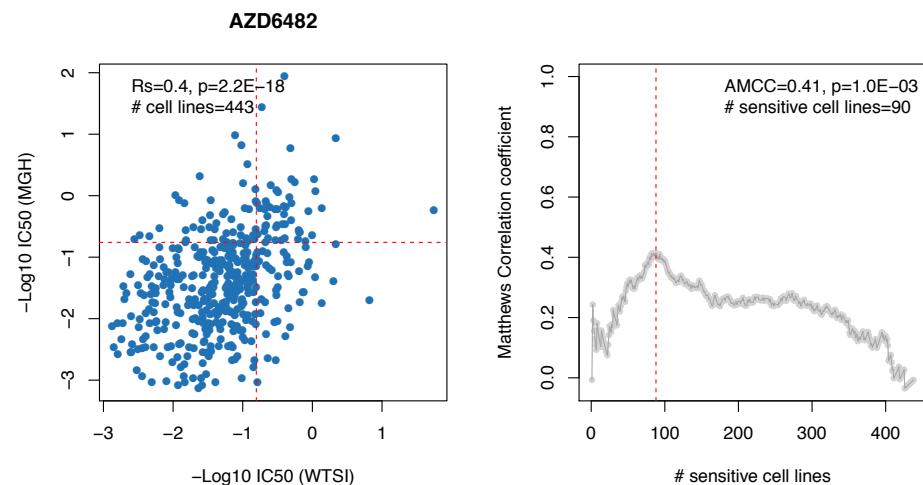


Supplementary Figure 4: Scatterplot of IC_{50} measures for **(A)** camptothecin and **(B)** AZD6482 replicated at the Wellcome Trust Sanger Institute (WTSI) and the Massachusetts General Hospital (MGH) with cutoffs determined by the AMCC statistic and evolution of the Matthews correlation coefficient with respect to the cutoff values.

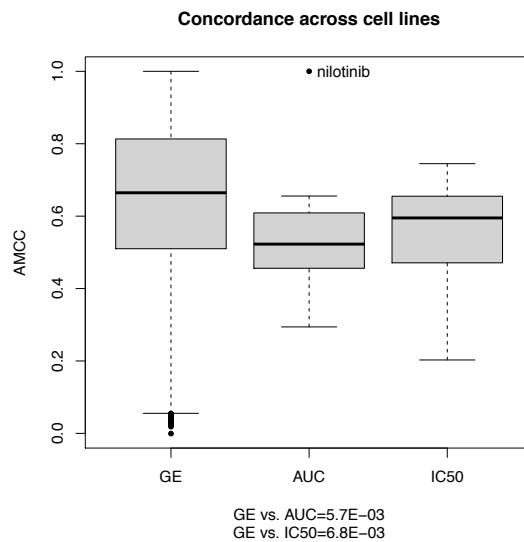
A



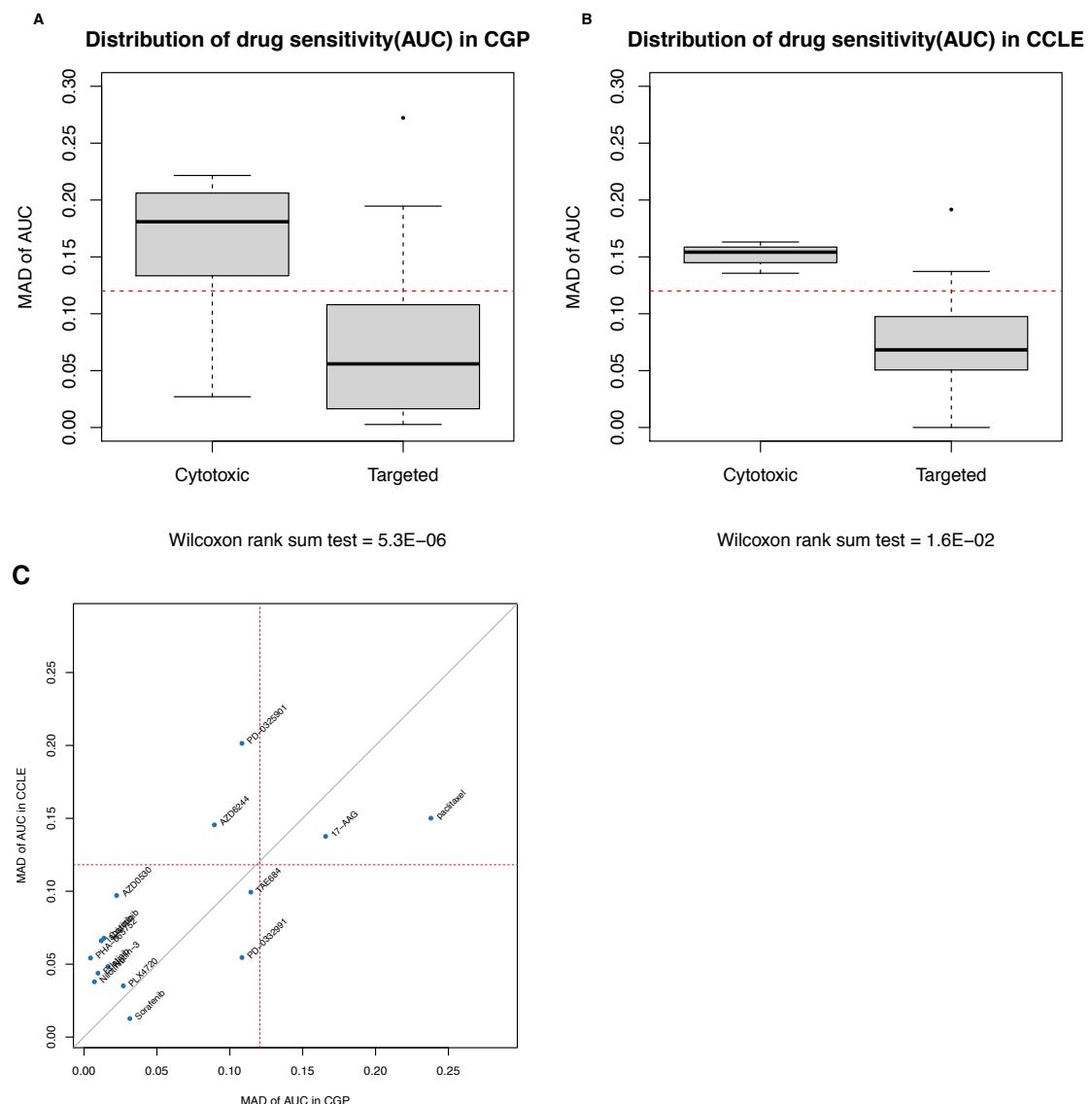
B



Supplementary Figure 5: Concordance across cell lines using AMCC for gene expression, AUC and IC₅₀ measurements.

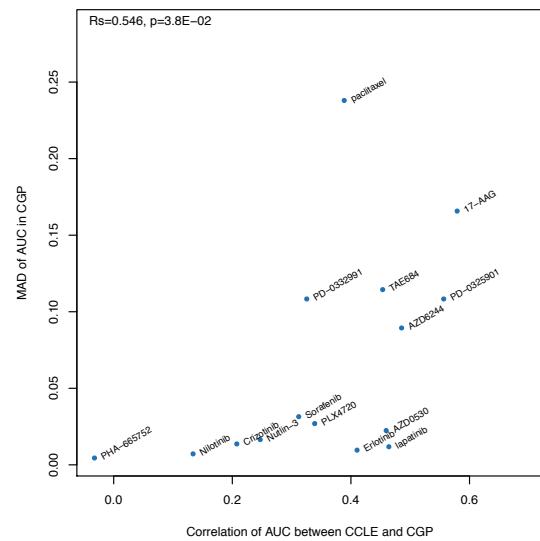


Supplementary Figure 6: Box plot comparing median absolute deviation (MAD) of AUC across cell lines for all drugs classified as targeted vs. cytotoxic in **(A)** CGP and **(B)** CCLE, and **(C)** comparison between MAD of AUC for common drugs in CGP and CCLE. An optimized cutoff of 0.12 for MAD AUC is computed using Youden method to discriminate targeted and cytotoxic drugs

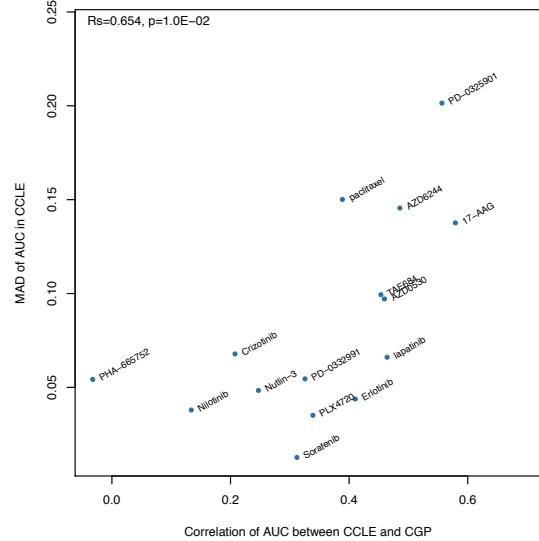


Supplementary Figure 7: Association between concordance of drug sensitivity, computed using (A,B) Spearman correlation coefficient and (C,D) AMCC, and variability of AUC across cell lines computed using median absolute deviation (MAD).

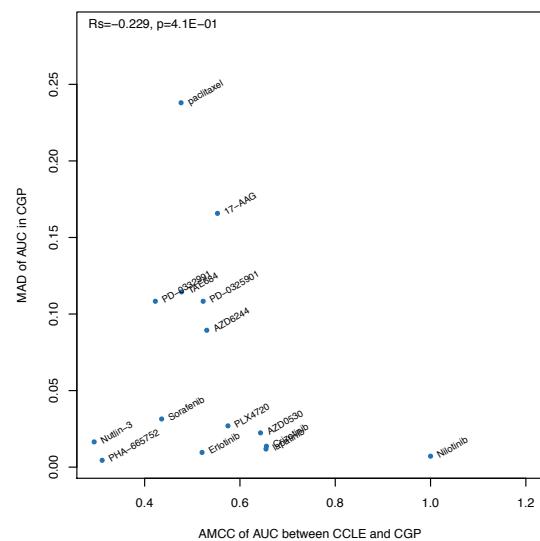
A



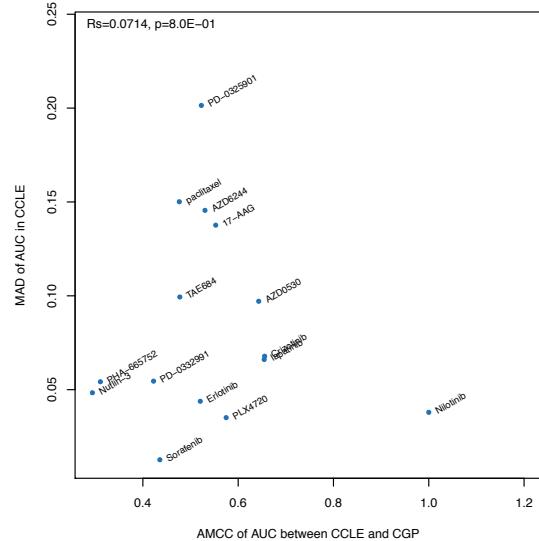
B



C



D



References

- [1] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F Berger, John E Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jané-Valbuena, Felipa A Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H Engels, Jill Cheng, Guoying K Yu, Jianjun Yu, Peter Aspasia, Melanie de Silva, Kalpana Jagtap, Michael D Jones, Li Wang, Charles Hatton, Emanuele Palascandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P. Mesirov, Stacey B Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E Myer, Barbara L Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L Harris, Matthew Meyerson, Todd R. Golub, Michael P Morrissey, William R Sellers, Robert Schlegel, and Levi A. Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, March 2012.
- [2] Mathew J Garnett, Elena J Edelman, Sonja J Heidorn, Chris D Greenman, Anahita Das-tur, King Wai Lau, Patricia Greninger, I Richard Thompson, Xi Luo, Jorge Soares, Qing-song Liu, Francesco Iorio, Didier Surdez, Li Chen, Randy J Milano, Graham R Bignell, Ah T Tam, Helen Davies, Jesse A Stevenson, Syd Barthorpe, Stephen R Lutz, Fiona Kogera, Karl Lawrence, Anne McLaren-Douglas, Xeni Mitropoulos, Tatiana Mironenko, Helen Thi, Laura Richardson, Wenjun Zhou, Frances Jewitt, Tinghu Zhang, Patrick O'Brien, Jessica L Boisvert, Stacey Price, Wooyoung Hur, Wanjuan Yang, Xianming Deng, Adam Butler, Hwan Geun Choi, Jae Won Chang, Jose Baselga, Ivan Stamenkovic, Jeffrey A Engelman, Sreenath V Sharma, Olivier Delattre, Julio Saez-Rodriguez, Nathanael S Gray, Jeffrey Settleman, P Andrew Futreal, Daniel A Haber, Michael R Stratton, Sridhar Ramaswamy, Ultan McDermott, and Cyril H Benes. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, March 2012.
- [3] Andrew Goodspeed, Laura M Heiser, Joe W Gray, and James C Costello. Tumor-derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Molecular Cancer Research*, page molcanres.0189.2015, August 2015.
- [4] Benjamin Haibe-Kains, Nehme El-Hachem, Nicolai Juul Birkbak, Andrew C Jin, Andrew H Beck, Hugo J W L Aerts, and John Quackenbush. Inconsistency in large pharmacogenomic studies. *Nature*, 504(7480):389–393, December 2013.
- [5] Qiyuan Li, Nicolai J Birkbak, Balázs Győrffy, Zoltan Szallasi, and Aron C Eklund. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*, 12(1):474, 2011.
- [6] B W Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et biophysica acta*, 405(2):442–451, October 1975.
- [7] Matthew N McCall, Benjamin M Bolstad, and Rafael A Irizarry. Frozen robust multiarray analysis (fRMA). *Biostatistics (Oxford, England)*, 11(2):242–253, April 2010.
- [8] Zhaleh Safikhani, Mark Freeman, Petr Smirnov, Nehme El-Hachem, Adrian She, Rene Quevedo, Anna Goldenberg, Nicolai Juul Birkbak, Christos Hatzis, Leming Shi, Andrew H Beck, Hugo JW Aerts, John Quackenbush, and Benjamin Haibe-Kains. Revisiting inconsistency in large pharmacogenomic studies. *bioRxiv*, page 026153, September 2015.