Dr.Paso:

Drug response prediction and analysis system for oncology research

Francisco Azuaje [1,+,*], Tony Kaoma[2], Céline Jeanty, Petr V. Nazarov, Arnaud Muller, Sang-Yoon Kim, Anna Golebiewska[1], Gunnar Dittmar[2], Simone P. Niclou[1].

[1]NorLux Neuro-Oncology Laboratory, Department of Oncology, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg.

[2] Proteome and Genome Research Unit, Department of Oncology, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg.

+ Current affiliation: [2]

* Corresponding Author, email: Francisco.Azuaje@lih.lu

Summary

The prediction of anticancer drug response is crucial for achieving a more effective and precise treatment of patients. Models based on the analysis of large cell line collections have shown potential for investigating drug efficacy in a clinically-meaningful, cost-effective manner. Using data from thousands of cancer cell lines and drug response experiments, we propose a drug sensitivity prediction system based on a 47-gene expression profile, which was derived from an unbiased transcriptomic network analysis approach. The profile reflects the molecular activity of a diverse range of cancer-relevant processes and pathways. We validated our model using independent datasets and comparisons with published models. A high concordance between predicted and observed drug sensitivities was obtained, including additional validated predictions for four glioblastoma cell lines and four drugs. Our approach can accurately predict anti-cancer drug sensitivity and will enable further pre-clinical research. In the longer-term, it may benefit patient-oriented investigations and interventions.

Introduction

The unbiased, large-scale prediction of anticancer drug activity using tumor-derived molecular data is crucial to deliver on the promise of a more personalized, precise treatment of cancer patients (Caponigro and Sellers, 2011; Ross and Wilson, 2014). The prediction of drug sensitivity based on the analysis of large collections of cell lines offers significant opportunities for investigating clinical efficacy in a biologically-meaningful, cost-effective manner (Geeleher et al., 2014; Goodspeed et al., 2016; Wilding and Bodmer, 2014). Computational models for predicting anticancer drug sensitivity can aid in the selection and prioritization of candidate compounds for pre-clinical research (Costello et al., 2014; Rees et al., 2016; Reinhold et al., 2012; Stetson et al., 2014).

Although cell line-based models may not fully recapitulate tumor biology, appropriately validated models may accelerate patient-oriented research, and have already shown potential to generate clinically-relevant predictions in different oncology domains. Such models may complement and in some cases offer an early substitute for *in vivo* models that tend to be expensive, time consuming and less scalable. In the short-term, this could enable the generation of novel biological hypotheses in the lab and, in the longer term, guide therapeutic decision-making in the clinic.

Over the past few years, the investigation of cell line-based computational models for anti-cancer drug sensitivity prediction has been accelerated by publicly-funded efforts of large research consortia (Barretina et al., 2012; Iorio et al., 2016b; Reinhold et al., 2012; Yang et al., 2013). In particular, the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Garnett et al., 2012; Yang et al.,

2013) projects represented significant steps forward for the oncology and pharmacogenomics research communities. These projects have generated and shared (untreated) molecular data from thousands of cancer cell lines and their accompanying treatment sensitivity measurements for hundreds of experimental and clinically-approved drugs. To date, computational models have mainly emphasized the application of different widely-investigated multivariable statistical and machine learning models, such as linear models and support vector machines, with various versions of feature selection methodologies (Dong et al., 2015; Haverty et al., 2016; Jang et al., 2014). Despite their potential for accurately predicting drug sensitivity across multiple types of cancer cell lines, less attention has been given to the investigation of biological importance of the proposed drug sensitivity markers, which have ranged from one to hundreds of gene-based features. Moreover, the majority of reported models have not been evaluated on independently generated datasets (Azuaje, 2017). Although different studies have tested the resulting prediction models on independent cell line datasets, e.g., models trained and tested on the GDSC and CCLE dataset respectively, there is a lack of studies that experimentally validate predicted anticancer sensitivity on independent biological samples, including cell lines that were not included in the training and initial testing datasets (Cortes-Ciriano et al., 2016; Gupta et al., 2016; Jang et al., 2014).

Here, we present Dr.Paso: Drug response prediction and analysis system for oncology research (Figure 1A). Dr.Paso predicts drug sensitivity responses based on the (baseline) expression patterns of 47 genes, which represented “hubs” in a pan-cancer transcriptomic network extracted from more than 1K cell lines and are substantially implicated in a diversity of cancer-relevant biological processes. A computational prediction model based on the multiple-linear regression of the 47-gene expression values measured in hundreds of cell

lines provided both accurate and robust prediction performance. First, the model was trained and cross-validated on a (discovery) dataset consisting of more than 10K cell line-drug experiments for 24 (targeted and cytotoxic) drugs. Next, the resulting model was tested on a second, more recently-released, (validation) dataset comprising almost 10K cell line-drug experiments that included 16 drugs also found in the discovery dataset. Dr.Paso's prediction performance is comparable to, and in some cases outperforms, previously published computational models. Motivated by these findings, Dr.Paso next predicted sensitivity scores for 4 glioblastoma (GBM) cell lines, including three (stem-like) cell lines that were not included in the discovery and validation datasets, against 24 drugs. We selected the top three drugs predicted as highly effective together with a drug predicted as lowly effective (negative control), and performed *in vitro* tests on the 4 cell lines. As in the case of the public datasets, the sensitivity scores estimated by Dr.Paso were highly concordant with the observed *in vitro* responses. To further facilitate research, we offer Dr.Paso through a Web-based interface that allows users to predict drug sensitivity scores for their own samples and expression data. The following sections will describe in detail these research phases, which are outlined in Figure 1B.

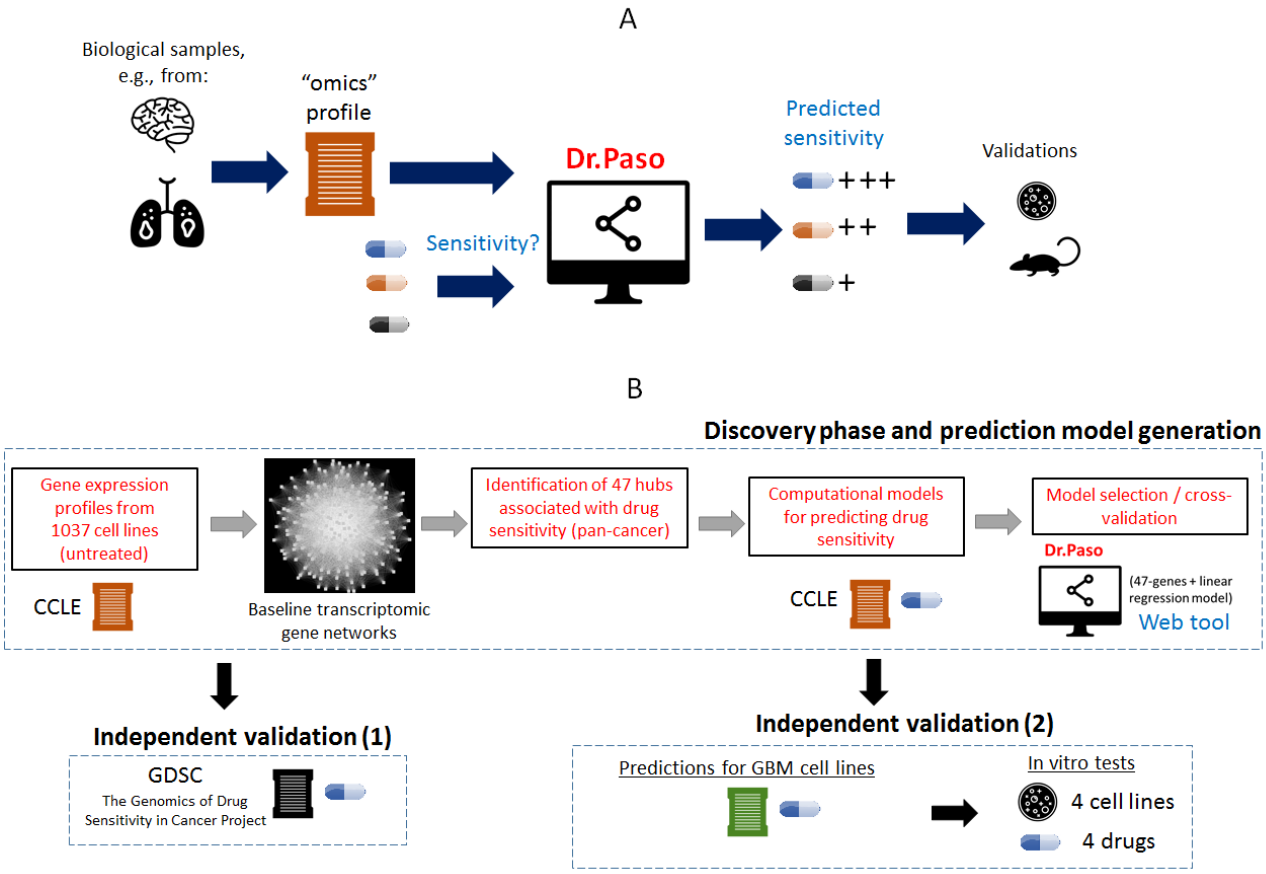

Figure 1. Dr.Paso: Overview of problem, application scenario and methodologies investigated. A. Outline of the general problem and application scenarios envisioned for the application of Dr.Paso. B. Workflow of the discovery phase, model generation and validation steps reported in this article.

Results

Hubs in a pan-cancer transcriptomic network display drug sensitivity predictive potential

Motivated by evidence indicating the drug sensitivity prediction power of gene expression profiles (Barretina et al., 2012; Iorio et al., 2016a), we investigated the predictive potential

of such data in the context of a pan-cancer transcriptomic correlation network. Our hypothesis was that genes highly connected within such networks, i.e., hubs, may be reflective of molecular activity across biological processes and tissue sites. To test this hypothesis, we analyzed the CCLE gene expression dataset, which was derived from 1037 (untreated) cell lines representing different cancer types in 18 tissue sites. To reduce network complexity while aiming at preserving potentially relevant information across all samples, we selected genes with highly variable expression pattern across cell lines (i.e., 177 genes with standard deviation of expression values across cell lines located above the 99th percentile). Using the pan-cancer expression profiles from these genes, we calculated all the between-gene (Pearson) correlation values and merged them into a fully-connected weighted network (Figure 2A), which included 177 nodes and more than 15K edges (correlations).

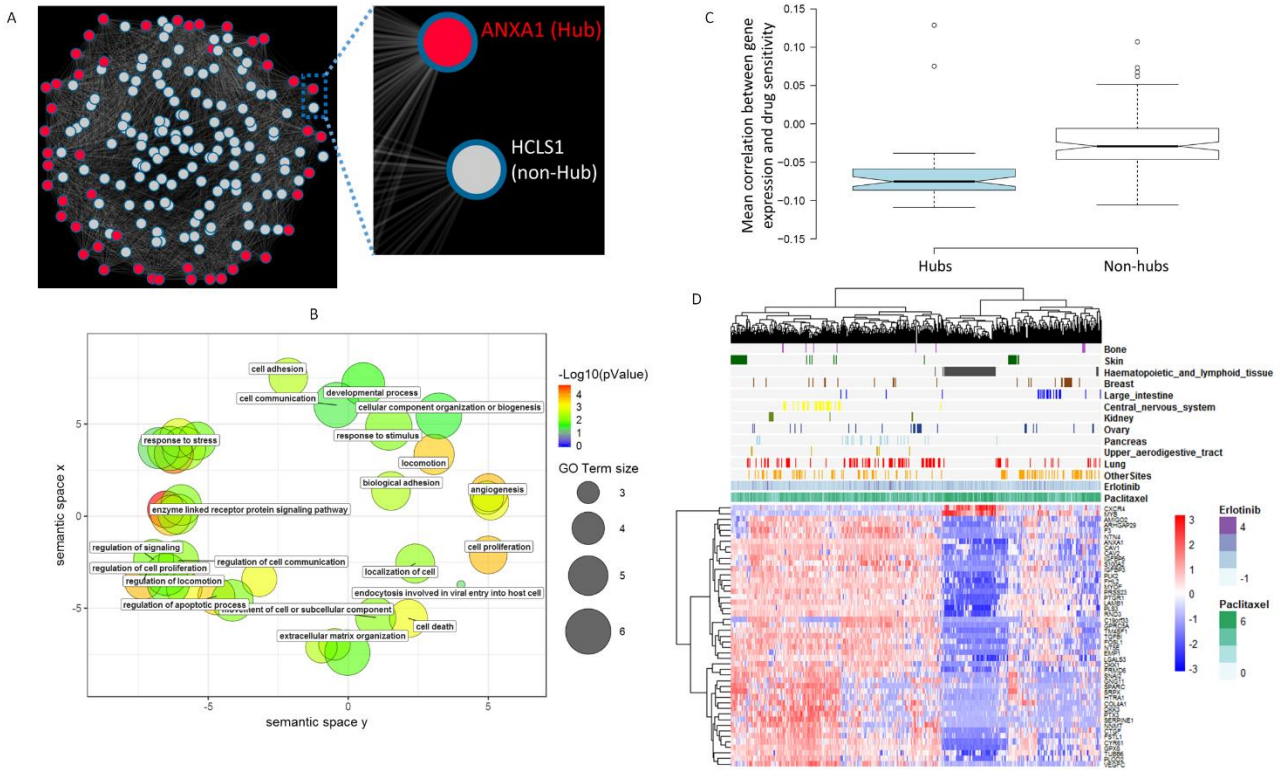

Figure 2. Hubs in a pan-cancer transcriptomic network display drug sensitivity predictive potential. A. Snapshot of a (fully connected) weighted gene correlation network. Nodes and

edges representing genes and their correlations respectively. Network hubs and non-hubs are colored in red and white respectively. A zoom-in view of examples of hub and non-hub nodes. The color intensity of edges reflect the expression correlations between such nodes and others in the network. B. Graphical summary of (non-redundant) Gene Ontology terms statistically over-represented in the list of 47 hub genes. Terms are projected onto a semantic similarity space with REViGO (Supek et al., 2011), in which similar terms are positioned closer to each other. Each term is represented by a bubble with color and size indicating the term's level of statistical enrichment in our list and frequency in the GO database respectively. C. Comparison of hubs vs. non-hubs on the basis of their individual associations with drug sensitivity. The boxplot depicts the mean correlation between the gene expression and the AA values across CCLE cell lines. Box notches indicate 95% confidence interval for each median value. D. Cell line-drug experiments are visualized in terms of the 47-gene expression data. The panel above the gene expression heatmap illustrates the AA values observed for selected sets of cancer cell lines (grouped by tissue site) and two compound examples (Erlotinib and Paclitaxel) for illustration purposes.

We identified network hubs by extracting those genes with statistically detectable connectivity scores (i.e., weighted degree values) using WiPer (Azuaje, 2014). This resulted in 47 hubs (WiPer adjusted-P < 0.05, online resource and Figure S1), one of which (*ANAX1*) is illustrated in Figure 2A together with an example of a non-hub node (*HCLS1*). The 47 hub genes are significantly implicated in a wide diversity of biological processes and pathways of relevance to cancer progression and therapeutic response. They include cell proliferation, death, migration, adhesion, angiogenesis, kinase signaling and the extracellular matrix (Figures 2B and S1).

Next, we analyzed drug sensitivity data (activity areas, AA) associated with these cell lines (11670 cell line-drug experiments) available in the CCLE. The AA, which is inversely correlated with the IC50 (the drug concentration at which an inhibition of 50% of cell viability is achieved), was defined by the CCLE to approximate the efficacy and potency of a drug simultaneously (Barretina et al., 2012). We stress that such data were not considered during the network generation and analysis steps. For each gene in the network, we calculated the correlation between gene expression and AA across all available (cell line-drug) data, and observed that: a. hubs tend to be anti-correlated with drug sensitivity, and b. such an anti-correlation is significantly stronger than in the case of non-hub genes. Moreover, such an association is considerably different to that displayed by non-hubs (Figure 2C). The 47 hub genes did not include previously reported markers of drug sensitivity: *ALK, BRAF, ERBB2, EGFR, HGF, NQ01, MDM2, MET and VEGFR*s (Barretina et al., 2012; Safikhani, 2017). To further demonstrate the potential relevance of these genes, we clustered the samples (available cell line-drug experiment data) based on their 47-gene (baseline) expression profiles and verified that these genes could, in principle, segregate samples according to cancer types (tissue sites) and highlight differential drug responses across samples (Figure 2D). Using an alternative visualization and (unsupervised) clustering technique (Figure S2), we verified the potential of these 47 genes' expression data to segregate samples in terms of their drug sensitivity. Overall, these results suggest that our 47 hubs represented a potentially novel, biologically-meaningful gene set with drug sensitivity prediction potential.

Predicting drug sensitivity based on the network-derived 47-gene expression profiles

We used the expression values from the 47 network hubs and drug sensitivity data (n = 10981, cell line-drug experiments, i.e., samples, with full expression and AA data available) to generate a drug sensitivity prediction model based on multiple linear regression (Methods). For a given sample (47-gene expression profile) and drug (identity of one of the 24 CCLE drugs), the model estimates a sensitivity score that approximates the AA values observed in the CCLE. For model training and testing, we used separate datasets respectively through a 10-fold cross-validation sampling procedure. Prediction capability was evaluated with multiple performance indicators that compare the predicted and observed sensitivity values: Pearson, Spearman and Kendall correlations, root-mean-squared errors (RMSE) and a concordance index (Figure 3).

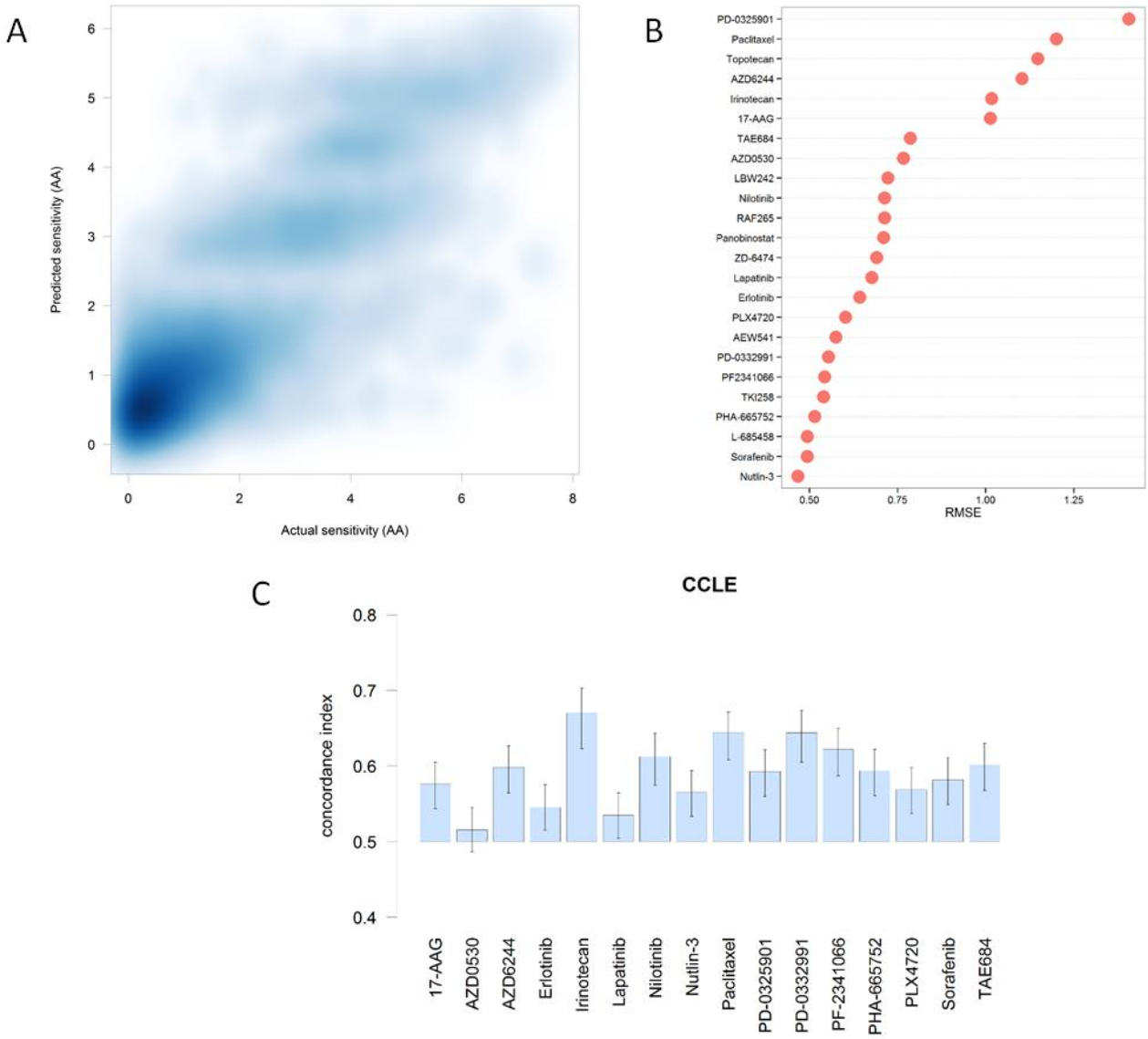

Figure 3. Alternative views of our model's predictive capacity on the CCLE dataset using alternative performance indicators. A. Density plot of predicted vs. actual sensitivity values (n=10981). Pearson, Spearman and Kendall, correlations coefficients: 0.86, 0.73 and 0.54 respectively. B. Plot of root-mean-square errors (RMSE) observed for each drug. C. Concordance indices between the predicted and the observed AA values for a selected set of drugs. An index value = 0.5 is the expected value from random prediction. Error bars: 95% confidence interval (CI) of the estimated concordance index.

Figures 3A and S3 show that the predicted and actual AA values are positively correlated (Pearson, Spearman and Kendall, correlations coefficients: 0.86, 0.73 and 0.54 respectively), which corroborates the predictive potential of our model. Such performance measures are comparable to, and in many cases outperform, those obtained from other published models trained and cross-validated on the same dataset. For example, a comprehensive analysis of different machine learning techniques (Jang et al., 2014), e.g., multiple-linear regression techniques, support vector machines and random forests, displayed (median) Pearson correlation coefficients falling into the range from 0.4 to 0.6, including top-performing models generated with gene expression data or their combination with other data types (Jang et al., 2014).

Figure 3B offers an alternative assessment of our model's prediction capability based on the RMSE obtained for each CCLE drug. This plot offers two key insights: 1. There are drugs for which our model can make relatively very accurate sensitivity predictions (e.g., Nutlin-3, an inhibitor of p53-Mdm2 complexes, and Sorafenib, a muti-kinase inhibitor) in comparison to other drugs (e.g., PD-0325901, a MEK inhibitor, and Paclitaxel, a mitotic inhibitor). 2. Our model's (drug-specific) prediction performance is competitive in relation to other published approaches trained and tested on the same dataset. For example, our model made predictions with a median RMSE = 0.70 (range: [0.47, 1.40]), which compares well with top-performing machine learning models that have reported median and minimum RMSE values above 0.80 and 0.65 respectively (Neto et al., 2014). For drugs such as Sorafenib, Nutlin-3 and PHA-665752, Dr.Paso tends to outperform models based on elastic-net and other variations of multiple-linear regression (Neto et al., 2014). Conversely, such models tend to offer relatively more accurate predictions for drugs such as Irinotecan and PD-0325901. These results corroborate previous findings about the lack of generalized solutions for highly

accurately predicting sensitivity across all types of drugs (Fersini et al., 2014; Haverty et al., 2016; Jang et al., 2014).

To provide further insights into our model's prediction capacity, Figure 3C displays the concordance index for a selected set of drugs. For a random pair of samples, the concordance index estimates the probability of correctly predicting the relative sensitivities of such samples (e.g., sample X is more sensitive than sample Y) in relation to the actual observed relative sensitivities. Perfect and random prediction performances are indicated by concordance indices equal to 1 and 0.5 respectively. Our model reported concordance indices with median values above 0.5, which compares favorably with the results obtained by (Papillon-Cavanagh et al., 2013) with different alternative models, including multiple linear regression with elastic net, and applied to the same dataset. For instance, Papillon-Cavanagh et al. obtained concordance indices lower than 0.7, including predictions with concordance indices falling below 0.5 for different drugs (e.g., Nutlin-3 and TAE684). These results suggest that our model can accurately predict drug sensitivity and provide, in relation to previously published models, promising predictive capability that we further investigated as follows.

Model evaluation on an independent dataset

We tested our 47-gene sensitivity prediction model on the 2016 release of the GDSC dataset (Iorio et al., 2016a). To allow our CCLE-derived model to make predictions on this dataset, we focused on the 16 drugs that are found in both datasets. First, as in the case of the CCLE data, we show that the (baseline) expression profiles of the 47 genes can, in principle, cluster the GDSC samples according to cancer types (tissue sites) and highlight differential

drug responses across samples (Figure 4A) in an unsupervised manner. Note that in the GDSC dataset drug sensitivity is represented as the logarithm of IC50 (LNIC50) values (AA values were not provided in this dataset). Using an alternative visualization and (unsupervised) clustering technique (Figure S2), we verified the potential of these 47 genes' expression data to segregate GDSC cell line-drug samples in terms of drug sensitivity.

Next, we applied our (CCLE-derived) prediction model to the GDSC data and made sensitivity predictions (AA values) for all the samples (cell line-drug experiments) available (Methods). The resulting predictions were then compared with the actual sensitivity values in the GDSC dataset (Figures 4B and S3). The predicted (AA) and actual sensitivity (LNIC50) values for these samples (n = 9984) are anti-correlated (Pearson, Spearman and Kendall, correlations coefficients: -0.72, -0.71 and -0.50 respectively). This indicates that our model is, in general, estimating sensitivity values that are in agreement with those observed in the test dataset, i.e., higher predictive agreement is reached when high AA (prediction) relates to a low LNIC50 (actual) values, and vice versa.

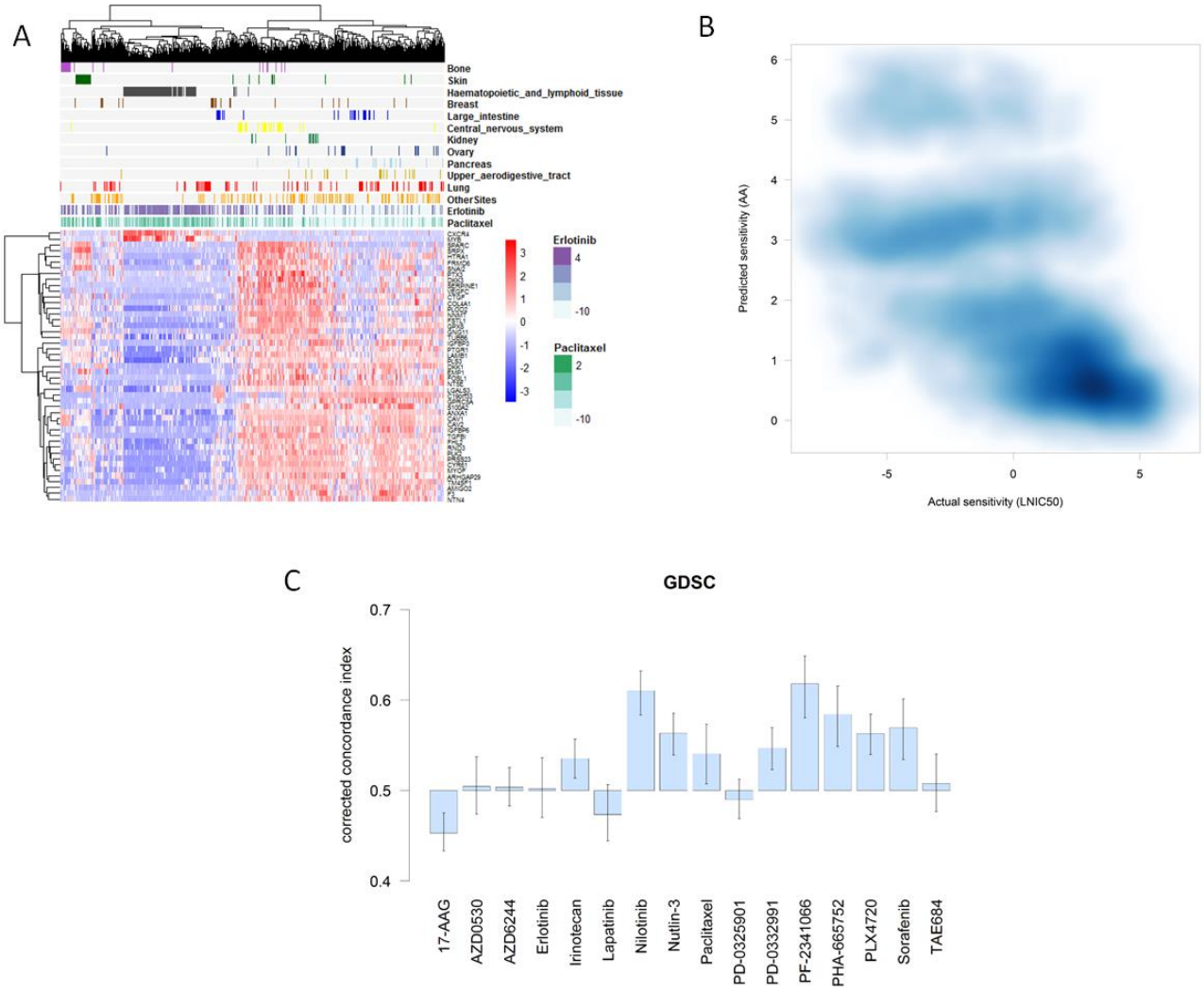

Figure 4. Alternative views of our model's prediction capacity on the GDSC dataset. A. Cell line-drug experiments are visualized in terms of the 47-gene expression data. The panel above the gene expression heatmap illustrates the LNIC50 (μM) values observed for selected sets of cancer cell lines (grouped by tissue site) and two compounds (Erlotinib and Paclitaxel). B. Application of CCLE-derived model to the GDSC data. Density plot of predicted (AA) vs. actual sensitivity (LNIC50) values for drugs that are common between the CCLE and GDSC (n = 9984). Pearson, Spearman and Kendall, correlations coefficients: -0.72, -0.71 and -0.50 respectively. C. Concordance indices between the predicted and the observed sensitivity values. An index value = 0.5 is the expected value from random prediction. Indices are corrected to account for the notion that higher concordance is

reached when high AA (prediction) corresponds to a low LNIC50 (observed) values, and vice versa. Error bars: 95% confidence interval (CI) of the estimated concordance index.

Figure 4C summarizes the assessment of our model's predictive performance on the GDSC dataset based on drug-specific concordance indices, as done for the CCLE dataset (Figure 3). Concordance indices > 0.5 were obtained for twelve out of the 16 drugs, and (among those 12 drugs) concordance estimates for 9 drugs can be reliably interpreted as larger than 0.5 (95% confidence intervals of the estimated indices). The predictive performances for several drugs (e.g., Nilotinib, Nutlin-3 and Sorafenib) are very similar to those estimated in the CCLE dataset. As in the CCLE dataset, the sensitivity observed in samples treated with AZD0530 and Lapatinib proved to be more difficult to accurately to predict. Although concordance indices > 0.5 were obtained for Irinotecan and Paclitaxel predictions, this represented a reduction of prediction performance in comparison to the predictions made for CCLE samples. The prediction performance of 17-AAG, PD-0325901 and TAE684 were also diminished. A previous study, which also used the GDSC dataset, consistently reported concordance indices < 0.5 for Sorafenib (Papillon-Cavanagh et al., 2013). Moreover, in comparison to that study's models, our model reported comparable or higher concordance indices for other drugs, such as Nilotinib and PF-2341066 (Crizotinib). Conversely, such a previous study reported better prediction performances for 17-AAG, Lapatinib and PD-0325901. Such comparisons should, nevertheless, be interpreted with caution as Papillon-Cavanagh et al.'s concordance indices were obtained with an older version of the GDSC dataset, which was used for both model training and testing. Overall, our findings further corroborate the predictive potential of our model, and highlight strengths and challenges in a drug-specific context.

Independent *in vitro* validation on several cell lines and compounds

To further validate our model's predictive capability on independently-generated data, we generated predictions and performed *in vitro* tests for several GBM cell lines and compounds. First, we measured the (baseline) expression profiles of 4 GBM cell lines that have been well-characterized in our lab: U87, NCH644, NCH601 and NCH421k (Methods). While the CCLE and GDSC datasets included U87, the latter three are stem-like GBM cell lines that were not included in the previous model training and test phases.

Although genome-wide expression (microarray) data can appropriately cluster multiple samples (biological replicates) from such cell lines, we found that the expression profiles of our model's 47 genes are sufficient to achieve the same biologically-meaningful segregation while offering a clearer, fine-grained view of their differences (Figure S4). We also verified the platform-independent replicability of these results with another 47-gene expression dataset derived from 3 of these cell lines measured with qPCR (Figure S4). These results corroborate the biologically-relevant discriminatory capacity and reproducibility of our model's 47-gene expression patterns.

Next, our model predicted the sensitivity of our 4 GBM cell lines (18 samples in total, Methods) against the 24 drugs included in our model. The 47-gene (microarray) expression profiles of these cells were input to the prediction model (6 U87, 3 NCH644, 3 NCH601 and 6 NCH421k gene expression profiles). Figure 5A summarizes the 432 predicted sensitivity (AA) values according to drug (18 predictions per drug). To investigate such predictions *in vitro*, we focused on the top-3 drugs associated with the highest predicted sensitivities (Paclitaxel, Panobinostat and 17-AAG), as well as on Erlotinib, which was predicted as a

relatively ineffective compound. These drugs correspond to four different drug classes: cytotoxic, histone deacetylase inhibitor, antibiotic derivative and an EGFR inhibitor respectively. In the case of Erlotinib, the predictions are consistent with the fact that the tested cells do not (NCH644, NCH421k) or very lowly (U87, NCH601) express EGFR. The Figures 5B and S5 show a more focused view of the predicted sensitivity values for our samples against these 4 drugs.

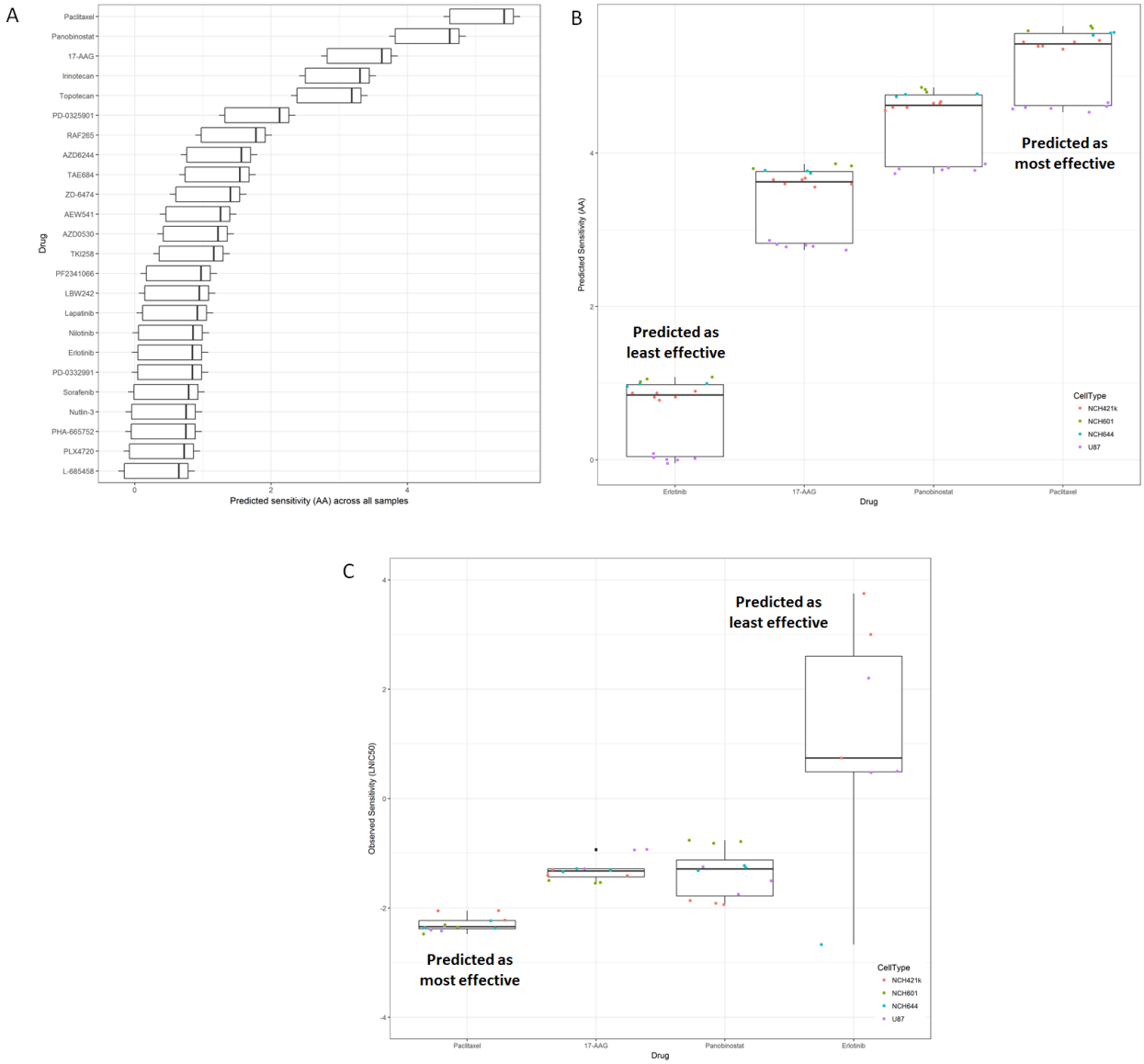

Figure 5. Drug sensitivity predictions and *in vitro* validation for different glioblastoma cell lines and compounds. A. Sensitivity predictions (horizontal axis) for 24 drugs (vertical axis). Box plot summarizes the (432) predicted sensitivity (AA, as defined in the prediction model)

values for 4 glioblastoma cell lines: U87, NCH644, NCH601 and NCH421k. The 47-gene expression profiles of multiple biological replicates (18 samples in total) were input to the prediction model (6 U87, 3 NCH644, 3 NCH601 and 6 NCH421k samples). B. Alternative boxplot summary of the prediction results for 4 drugs (Erlotinib, 17-AAG, Panobinostat and Paclitaxel) and the different cell lines. These drugs, which were selected for subsequent *in vitro* tests, were predicted to be relatively highly (17-AAG, Panobinostat and Paclitaxel) and lowly (Erlotinib) effective against the 4 cell lines. C. Summary of *in vitro* test results. The selected drugs were tested on each cell line in triplicates, relative viability (vs. vehicle-treated samples) was measured for 8 drug concentration values (µM) and IC50 values were estimated for each drug-sample experiment. The boxplot shows the resulting LNIC50 values obtained. Drug response data for NCH601 samples and Erlotinib are not available, and for NCH644 samples and Erlotinib not shown because of lack of effect. Boxes show the median, the 25th and 75th percentiles (lower and upper hinges), and (1.5 x) inter-quartile ranges.

We tested the selected drugs on each cell line, in triplicates, and measured their response based on their relative viability (i.e., normalized to vehicle-treated samples) for 8 drug concentration values (µM). For each treated cell line, we estimated the IC50 values and compared them on the basis of cell line and drug groups. Figure 5C summarizes the results with boxplots showing the LNIC50 values. Drug response data for NCH601 samples and Erlotinib were not available (not tested), and data for NCH644 samples and Erlotinib are not shown due to lack of effect. Figure S6 includes all the drug response curves and additional details.

As predicted by our model, all our cell lines exhibited the lowest sensitivity, i.e., the highest IC50 values, when treated with Erlotinib (median LNIC50 = 0.74 µM). U87 was the least

sensitive cell line in relation to all 4 drugs (median LNIC50 = -1.27 µM across all sample-drug experiments), in full agreement with the predictions. Our model consistently predicted NCH601 as the most sensitive cell line against all drugs (Figures S6). Our *in vitro* tests showed that NCH421k tended to be more sensitive than NCH601 (median logIC50: -1.64 vs. -1.54 µM). Despite this particular discrepancy, we found global agreement between predicted and observed sensitivities on the basis of cell type (Spearman correlation between the median sensitivity values, predicted (AA) vs. observed (LNIC50) in the 4 cell line groups: -0.40).

In accordance with the predictions, Paclitaxel was the most effective drug across all treated samples (median LNIC50 = -2.35 µM). Lesser agreement between predicted and observed sensitivities were obtained in the case of the remaining two drugs. For all samples, our model predicted overall higher sensitivity for Panobinostat than for 17-AAG (Figure 5B). *In vitro*, relatively similar responses were obtained for Panobinostat (median LNIC50 = -1.29) and 17-AAG (median LNIC50 = -1.33 µM), though a larger variability of sensitivity was observed in the former case. Nevertheless, predictions and *in vitro* tests concordantly showed that NCH421k and U87 samples treated with Panobinostat were consistently more sensitive than all samples treated with 17-AAG (Figures 5C and S6). Taken together, these results provide further evidence of the predictive capacity of our model. The resulting system, Dr.Paso, will enable the community to conduct further investigations.

Dr.Paso online

To share our model and enable further research, we developed a web-accessible tool that allows researchers to upload their own gene expression data, make sensitivity predictions

and visualize results in a few steps (Figure 6). The Help section of the website offers a guided application example using CCLE data. Users provide their input data as a text file containing the (baseline) 47-gene expression for different samples, and then can select all or specific drugs for making predictions (Figure 6A). Dataset re-scaling (feature standardization with means and standard deviations equal to 0 and 1 respectively) can be applied to harmonize the input dataset with the feature representation used in our model. Prediction results are presented with graphical displays and tables in different panels. Moreover, users can control the amount and focus of information at the drug and sample levels (Figure 6B to 6D). Results can be saved in different graphical and tabular file formats. The tool is freely available at www.drpaso.lu.

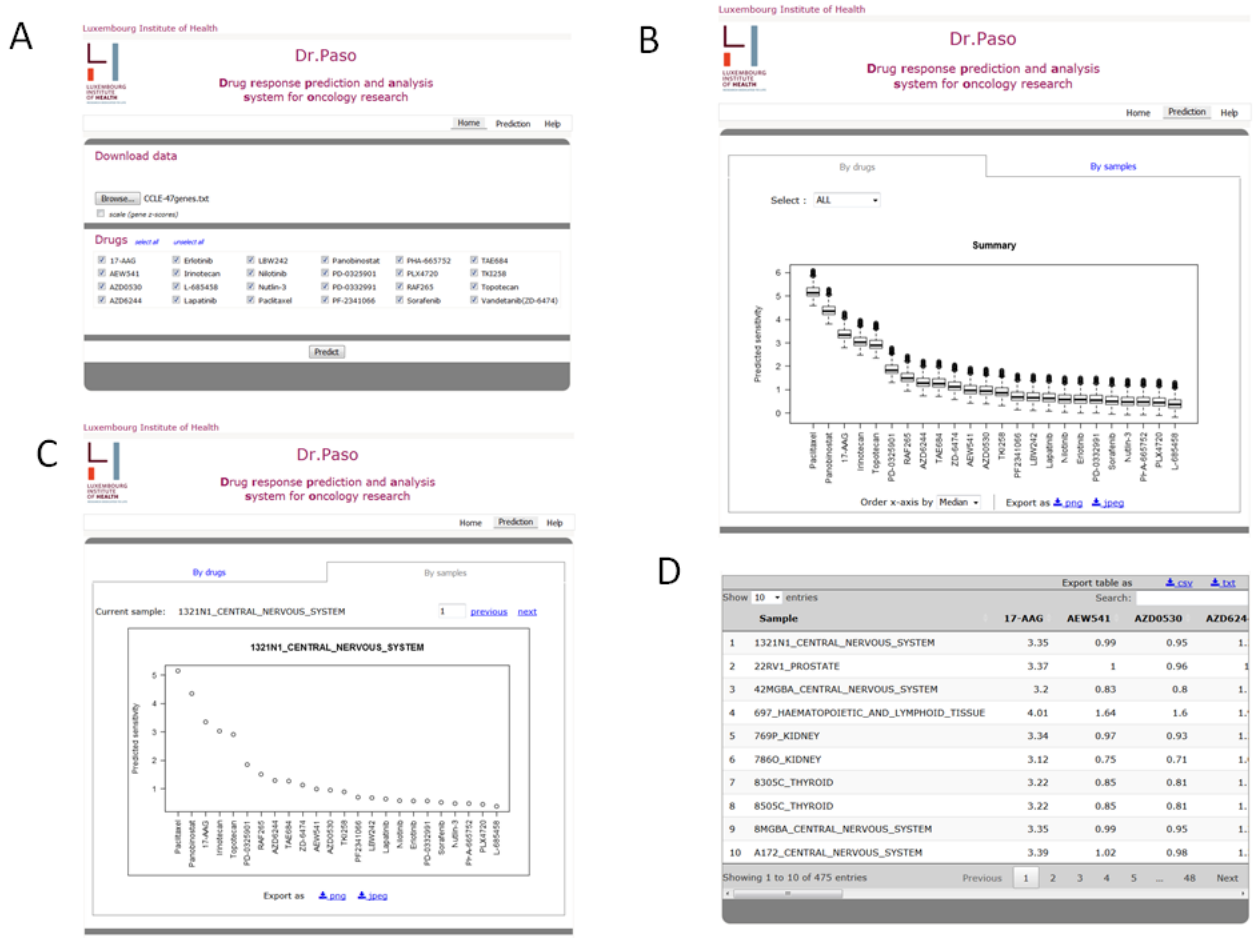

Figure 6. Dr.Paso online: a Web-based tool for predicting drug sensitivity and enabling further research. Screenshots of: A. Main page with user input and analysis options; B: Global view of predicted sensitivity values for a given input gene expression dataset and all drugs available in the CCLE; C: Alternative view of predictions focused on a specific input sample and all drugs; D. Tabular-based view of results. All views can be selected and downloaded according to user requirements.

Discussion

The development of computational models for estimating drug sensitivity based on the analysis of large and diverse collections of cancer cell lines is important to support pre-clinical research, and provides a basis for future clinically-oriented applications. Access to such models and their user-friendly application will enable new research across oncology domains and additional computational investigations. Our Drug Response Prediction and Analysis System for Oncology research, Dr.Paso, addresses such needs through the integration of network-based and statistical modeling approaches. For a given drug, our system predicts an anti-cancer sensitivity score based on the gene expression profile of 47 genes, which were shown to represent hubs in a pan-cancer transcriptomic network and to be prominently implicated in a variety of cancer-relevant biological processes. Dr.Paso was generated and evaluated on independent datasets, including our own *in vitro* validations of several cell line-compound combinations, and showed promising results in terms of predictive accuracy and concordance. Future research can apply Dr.Paso to 47-gene expression signatures from patient samples to investigate its potential relevance in the clinical setting.

Our study and other previous research highlight the challenges faced and complementary predictive capacity exhibited by different modeling approaches (Costello et al., 2014; Papillon-Cavanagh et al., 2013). No single model can consistently make accurate predictions for all drugs and cell lines available in the CCLE and GDSC datasets, including models that include genomic data (Gupta et al., 2016; Jang et al., 2014; Menden et al., 2013). Different models can offer more, or less, accurate predictions for certain drugs, and there is no conclusive evidence about the dominance of a particular modeling technique (Azuaje, 2017). Such limitations may be partially explained by a lack of sufficient molecular information to account for the complexity of cell lines and their drug responses, choice of surrogate measures of drug sensitivity and inconsistencies of sensitivity data between the CCLE and GDSC (Haverty et al., 2016; Investigators, 2015; Safikhani, 2017). The latter may also partly explain the overall degradation of predictive performance when training models on the CCLE and testing them on the GDSC.

Dr.Paso generates sensitivity scores based on a multiple linear regression model. We, as others elsewhere, have shown that relatively less complex regression models can offer comparable, and in some cases better, prediction performance than those models based on larger sets of learning parameters. Dr.Paso's predictive capacity is grounded in an unbiased network-guided selection of model inputs (47 genes) prior to the fitting of the regression model. Such a discovery process was shown to be both statistically- and biologically-meaningful. Apart from our multiple linear regression, we applied other regression techniques, e.g., support vector machines and neural networks, but decided to focus our investigation on a model with relatively lower complexity. Collectively, Dr.Paso is based on a biomarker discovery and prediction-making methodology that is both biologically-driven and statistically-powerful. New investigations, motivated by new datasets and clinically-

oriented questions, are certainly envisaged and are expected to include new biomarker discovery and prediction modeling strategies.

Notwithstanding recent advances in the field, there is a need to make executable models accessible to the research community to enable new investigations, including new applications and comparative analyses among different techniques. Here we offer Dr.Paso as a publicly-accessible online tool. While further investigations are needed, our study offers further evidence of the potential of computational models for predicting anti-cancer sensitivity. In the short-term, our findings will enable new pre-clinical research applications and may provide a new perspective for bringing such models closer to the clinic.

Methods

Identification of 47 genes with drug sensitivity predictive potential

The published pre-processed CCLE (microarray) gene expression and drug sensitivity datasets were obtained from the CCLE website. In the gene expression dataset, we focused on genes with symbols, calculated their standard deviation (SD) across all samples (1037) and ranked them based on their SD. For further analyses, we selected the most variable genes: 177 genes with SD values above the 99th percentile of the SD value distribution. We computed the gene-gene (Pearson) correlation coefficients between all the 177 genes and merged them into a single gene expression correlation network. We applied WiPer (Azuaje, 2014) to this fully-connected weighted network to detect highly connected nodes (hub genes). For each network node, WiPer computes the weighted degree and a corresponding P-value to assess the significance of the observed values, and adjusts it for multiple testing.

Genes exhibiting (Bonferroni adjusted) P< 0.05 were considered hubs (47 genes). Drug sensitivity information was not used to select hubs. The resulting 47 genes were examined with different Gene Ontology (GO) and biological pathway analysis tools (below). For each hub gene, we estimated the correlation of its expression profile (across all samples) with the activity area (AA) values available from all sample-drug combinations. The AA was formulated by the CCLE to approximate the efficacy and potency of a drug simultaneously and is inversely correlated with the IC50 (Barretina et al., 2012). We compared hubs and non-hubs on the basis of such individual expression-sensitivity correlations. Visualizations and unsupervised clustering of hubs and cell lines described by hub expression values were implemented with different open-source tools (below).

Training and testing of prediction model

We represented each CCLE sample (cell line-drug combination) with the expression values of the 47 hub genes and their corresponding AA values. We focused on samples with complete expression and AA data. The resulting set of 10981 samples was used for training and testing regression models. The dataset was standardized by re-scaling each gene so that each gene has mean and standard deviation of 0 and 1 respectively. For each model, we implemented 10-fold cross-validation (CV) for separating training from testing and for assessing prediction performance. We also used leave-one-out CV (LOOCV) and similar prediction performance results were obtained. Diverse regression techniques with different levels of complexity were investigated. We focused on a multiple linear regression model with Ridge regularization (Ridge parameter = 1E-08) because its performance (regression errors) was better than or comparable to those obtained with other techniques, such as support vector machines and k-nearest neighbors, and because of its interpretability in

comparison to relatively more complex models. The accuracy of model predictions was assessed by measuring their (Pearson, Spearman and Kendall) correlations with the observed values in the CCLE, the root-mean-squared error (RMSE) and a concordance index. The latter approximates, for a random pair of samples, the probability of correctly predicting which sample is more (or less) sensitivity than the other (Harrell et al., 1996). A concordance index equal to 0.5 indicates that the model's performance is comparable to that from a random predictor, while an index equal to 1 represents the perfect predictor.

Independent evaluation on the GDSC dataset

Raw expression data were obtained from the ArrayExpress database (accession number E-MTAB-3610) and drug sensitivity (natural logarithm of the IC50 in μM, LNIC50) were downloaded from GDSC database (http://www.cancerrxgene.org, release-5.0). We normalized raw expression data with the RMA function of R/oligo package (Carvalho and Irizarry, 2010). Then we averaged the resulting log2 probe-set intensities to estimate the expression of each gene. Associations between probe-sets and gene symbols were obtained through the hgu219.db annotation package (Carlson, 2016). For each cell line-drug experiment available (sample), we retrieved the expression data for the 47 genes used as inputs to our prediction model and retrieved the corresponding drug sensitivity. We focused on the 16 drugs found in both this and the CCLE dataset. This resulted in a dataset consisting of 9984 samples, each one represented by 47 gene expression values and one LNIC50 value. We standardized expression data as in the case of the CCLE dataset, reformatted the file and input it to the CCLE-derived prediction model (further information below). For each sample in the dataset, the model predicted a drug sensitivity score (approximation of AA). We compared predicted vs. observed values using the indicators

applied to the CCLE dataset analysis. We adapted the concordance index to account for the fact that AA and LNIC50 are expected to be anti-correlated, i.e., for a given sample, concordance is achieved when a high (predicted) AA value corresponds to a low (observed) LNIC50 value, and vice versa.

GBM cell lines and expression data for *in vitro* validations

U87 cells were obtained from the ATCC (Rockville, USA) and were cultured as monolayers in DMEM containing 10% FBS, 2 mM L-Glutamine and 100 U/ml Pen-Strep (Lonza). GBM stem-like cultures (NCH421k, NCH601 and NCH644) were kindly provided by Christel Herold-Mende (University of Heidelberg, Germany) and were cultured as 3D non-adherent spheres as previously described (Abdul Rahim et al., 2017; Sanzey et al., 2015).

We measured the (baseline) gene expression of 4 GBM cell lines using microarrays (6 U87, 6 NCH421k, 3 NCH644 and 3 NCH601 biological replicates), as reported in (Sanzey et al., 2015). For our model's 47 genes, we also replicated gene expression measurements using qPCR for U87, NCH421k and NCH644 cell lines (each one in triplicate). RNA was extracted from 10^6 cells using TRI Reagent® (Sigma-Aldrich). RNA isolated in the aqueous phase with a Phase lock gel-Heavy (5 Prime) was precipitated with 100% isopropanol and purified using RNeasy® Mini kit combined with an on-column DNase treatment (Qiagen). For the qPCR, RNA was reverse-transcribed into cDNA using Superscript III™ (Invitrogen) following manufacturer's instructions. qPCR was performed in 96-well plates using SYBR® Green Master Mix (Bio-Rad) and CFX-96 thermal cycler (Bio-Rad). Normalized gene expression levels were calculated using the CFX manager 3.1 software (Bio-Rad) via the delta-delta Cq method with "Hspcb, Rps13, 18sRNA" as reference genes and taking into account the

calculated amplification efficiency for each primers pair. We provide a MIQE-compliance checklist table as a supplemental item.

Drug sensitivity predictions and *in vitro* validation on GBM cell lines

The gene (microarray) expression dataset was standardized as above. Each sample, represented by a 47-gene (microarray) expression profile, was input to the prediction model and a drug sensitivity value was predicted for each one of them (18 samples in total), for each of the 24 drugs included in the model. Predicted values were compared between them to determine their relative differences in terms of cell lines and drugs. Next, these predictions were compared to the *in vitro* sensitivity values that were obtained as follows. We tested 4 drugs: Paclitaxel (Sigma-Aldrich), Panobinostat, 17-AAG and Erlotinib (Selleck Chemicals) independently on the selected 4 GBM cell lines with 8 drug concentrations. For each cell line and dose, we performed treatment experiments in triplicate (i.e., 3 treated biological replicates / dose). As a measurement of drug sensitivity, WST-1 (Sigma-Aldrich) cell viability assays were implemented. U87, NCH421k, NCH644 and NCH601 cell lines were seeded into 96-well plates at densities of 1,500, 5,000, 4,000, 6,000 cells per well, in appropriate culture medium (Sanzey et al., 2015). Cells were incubated, 24h hours after seeding, with the 8 different drug concentrations ranging from 10µM to 6.1 x 10-4 µM, with a final volume of DMSO not exceeding 0.1% and each condition was tested with 6 technical replicates. After 72h incubation, WST-1 reagent was added in medium to a final concentration of 10%. Adherent cell line (U87) was incubated at 37°C for 2 hours and 3D sphere stem-like cell lines (NCH421k, NCH644 and NCH601) were incubated at 37°C for 6-8 hours. Absorbance was measured against a background control at 450nm on a FLUOstar OPTIMA Microplate Reader (BMG LABTECH). Using the normalized viability measurements, we generated drug

dose-response curves and estimated IC50 values (μM) for each sample-drug combination. The dose-response curves were fitted with a four-parameter logistic regression model, whose parameters were calculated using GraphPad Prism 7 (GraphPad).

Software and Dr.Paso's Web-based tool

We used the R statistical environment for data analysis and visualization (www.r-project.org), packages: ggplot2, pheatmap, MASS and SNFtool (Wang et al., 2014). Concordance indexes (Harrell et al., 1996) were calculated based on rescaled Kendall rank correlation coefficients, which were also used to estimate confidence intervals (by Fisher's transformation). For network analyses, we applied Cytoscape for visualization (Shannon et al., 2003), MINE for similarity exploration (Reshef et al., 2011) and WiPer for network hub identification (Azuaje, 2014). REViGO (Supek et al., 2011) and g:Profiler (Reimand et al., 2007) were applied for biological process and pathway enrichment analyses. The Weka workbench was used for building and testing regression models (Frank, 2016; Hall, 2009), and GraphPad Prism (www.graphpad.com) for analyzing drug response curves. We provide researchers with a Web-based application to enable them to predict anticancer drug sensitivity using their own (47-gene) transcriptomic data. The tool is based on the R/Shiny package (https://shiny.rstudio.com/). Although this package offers useful functionality for generating an interactive user interface, we customized available code using the R/Shinyjs package (http://deanattali.com/shinyjs/). Users can input pre-processed expression datasets. Alternatively, our application can also implement z-score rescaling of the input data. Figures containing the prediction results can be downloaded and stored as either png or jpeg files. Results are also shown as tables with sample-specific predictions (in rows) with their corresponding drugs (in columns), and may be stored as either csv or tsv files.

Author Contributions

Conceptualization, F.A.; Methodology, F.A., P.V.N, C.J., T.K., A.G, S.P.N; Software, F.A., T.K., P.V.N, C.J., A.M., S.K.; Validation, T.K, C.J.; Resources, A.G, G.D, S.P.N, Supervision, F.A., A.G., S.P.N, G.D.; Writing - Original Draft, F.A.; Writing - Review & Editing, all authors.

Acknowledgments

This research was funded by (LIH-MESR) project Connect2Predict to F.A. For technical guidance to C.J, we thank H. Erasimus and S. Fritah (drug experiments) and V. Barthelemy, A. Bernard, J. Bohler and A. Dirkse (cell line manipulation) at the LIH NorLux Neuro-Oncology Laboratory. For helpful feedback on the manuscript, we thank L.C. Tranchevent at the LIH Proteome and Genome Research Unit.

References

Abdul Rahim, S.A., Dirkse, A., Oudin, A., Schuster, A., Bohler, J., Barthelemy, V., Muller, A., Vallar, L., Janji, B., Golebiewska, A*., et al.* (2017). Regulation of hypoxia-induced autophagy in glioblastoma involves ATG9A. British journal of cancer *117*, 813-825.

Azuaje, F. (2017). Computational models for predicting drug responses in cancer research. Briefings in bioinformatics *18*, 820-829.

Azuaje, F.J. (2014). Selecting biologically informative genes in co-expression networks with a centrality score. Biology direct *9*, 12.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D*., et al.* (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607.

Caponigro, G., and Sellers, W.R. (2011). Advances in the preclinical testing of cancer therapeutic hypotheses. Nature reviews Drug discovery *10*, 179-187.

Carlson, M. (2016). hgu219.db: Affymetrix Human Genome 219 Plate annotation data (chip hgu219). R package version 3.2.3.

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics (Oxford, England) *26*, 2363-2367.

Cortes-Ciriano, I., van Westen, G.J., Bouvier, G., Nilges, M., Overington, J.P., Bender, A., and Malliavin, T.E. (2016). Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. Bioinformatics (Oxford, England) *32*, 85-95.

Costello, J.C., Heiser, L.M., Georgii, E., Gonen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A*., et al.* (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nature biotechnology *32*, 1202-1212.

Dong, Z., Zhang, N., Li, C., Wang, H., Fang, Y., Wang, J., and Zheng, X. (2015). Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. BMC cancer *15*, 489.

Fersini, E., Messina, E., and Archetti, F. (2014). A p-Median approach for predicting drug response in tumour cells. BMC bioinformatics *15*, 353.

Frank, E.H., M.A; Witten, I.H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Fourth Edition edn (Morgan Kaufmann).

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J*., et al.* (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570-575.

Geeleher, P., Cox, N.J., and Huang, R.S. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome biology *15*, R47.

Goodspeed, A., Heiser, L.M., Gray, J.W., and Costello, J.C. (2016). Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. Molecular cancer research : MCR *14*, 3-13.

Gupta, S., Chaudhary, K., Kumar, R., Gautam, A., Nanda, J.S., Dhanda, S.K., Brahmachari, S.K., and Raghava, G.P. (2016). Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: A step towards personalized medicine. Scientific reports *6*, 23857.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P.; Witten, I.H. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations *11*.

Harrell, F.E., Jr., Lee, K.L., and Mark, D.B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine *15*, 361-387.

Haverty, P.M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., Neve, R.M., Martin, S., Settleman, J., Yauch, R.L*., et al.* (2016). Reproducible pharmacogenomic profiling of cancer cell line panels. Nature *533*, 333-337.

Investigators, T.C.C.L.E.a.G.o.D.S.i.C. (2015). Pharmacogenomic agreement between two cancer cell line data sets. Nature *528*, 84-87.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Goncalves, E., Barthorpe, S., Lightfoot, H*., et al.* (2016a). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740-754.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Goncalves, E., Barthorpe, S., Lightfoot, H*., et al.* (2016b). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740-754.

Jang, I.S., Neto, E.C., Guinney, J., Friend, S.H., and Margolin, A.A. (2014). Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 63-74.

Menden, M.P., Iorio, F., Garnett, M., McDermott, U., Benes, C.H., Ballester, P.J., and Saez-Rodriguez, J. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PloS one *8*, e61318.

Neto, E.C., Jang, I.S., Friend, S.H., and Margolin, A.A. (2014). The Stream algorithm: computationally efficient ridge-regression via Bayesian model averaging, and applications to pharmacogenomic prediction of cancer cell line sensitivity. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 27-38.

Papillon-Cavanagh, S., De Jay, N., Hachem, N., Olsen, C., Bontempi, G., Aerts, H.J., Quackenbush, J., and Haibe-Kains, B. (2013). Comparison and validation of genomic predictors for anticancer drug sensitivity. Journal of the American Medical Informatics Association : JAMIA *20*, 597-602.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E*., et al.* (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nature chemical biology *12*, 109-116.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic acids research *35*, W193-200.

Reinhold, W.C., Sunshine, M., Liu, H., Varma, S., Kohn, K.W., Morris, J., Doroshow, J., and Pommier, Y. (2012). CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer research *72*, 3499-3511.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., and Sabeti, P.C. (2011). Detecting novel associations in large data sets. Science (New York, NY) *334*, 1518-1524.

Ross, N.T., and Wilson, C.J. (2014). In vitro clinical trials: the future of cell-based profiling. Frontiers in pharmacology *5*, 121.

Safikhani, Z.S., P; Freeman, M; El-Hachem, N; She, A; et al. Revisiting inconsistency in large pharmacogenomic studies (2017). Revisiting inconsistency in large pharmacogenomic studies. F1000Research *5*

Sanzey, M., Abdul Rahim, S.A., Oudin, A., Dirkse, A., Kaoma, T., Vallar, L., Herold-Mende, C., Bjerkvig, R., Golebiewska, A., and Niclou, S.P. (2015). Comprehensive analysis of glycolytic enzymes as therapeutic targets in the treatment of glioblastoma. PloS one *10*, e0123544.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research *13*, 2498-2504.

Stetson, L.C., Pearl, T., Chen, Y., and Barnholtz-Sloan, J.S. (2014). Computational identification of multi-omic correlates of anticancer therapeutic response. BMC genomics *15 Suppl 7*, S2.

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one *6*, e21800.

Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., and Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. Nature methods *11*, 333-337.

Wilding, J.L., and Bodmer, W.F. (2014). Cancer cell lines for drug discovery and development. Cancer research *74*, 2377-2384.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R*., et al.* (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic acids research *41*, D955-961.

Supplemental Information

Dr.Paso:

Drug response prediction and analysis system for oncology research

Francisco Azuaje [1,+,*], Tony Kaoma[2], Peter V. Nazarov, Céline Jeanty, Arnaud Muller, Sang-Yoon Kim, Anna Golebiewska[1], Gunnar Dittmar[2], Simone P. Niclou[1].

[1]NorLux Neuro-Oncology Laboratory, Department of Oncology, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg.

[2] Proteome and Genome Research Unit, Department of Oncology, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg.

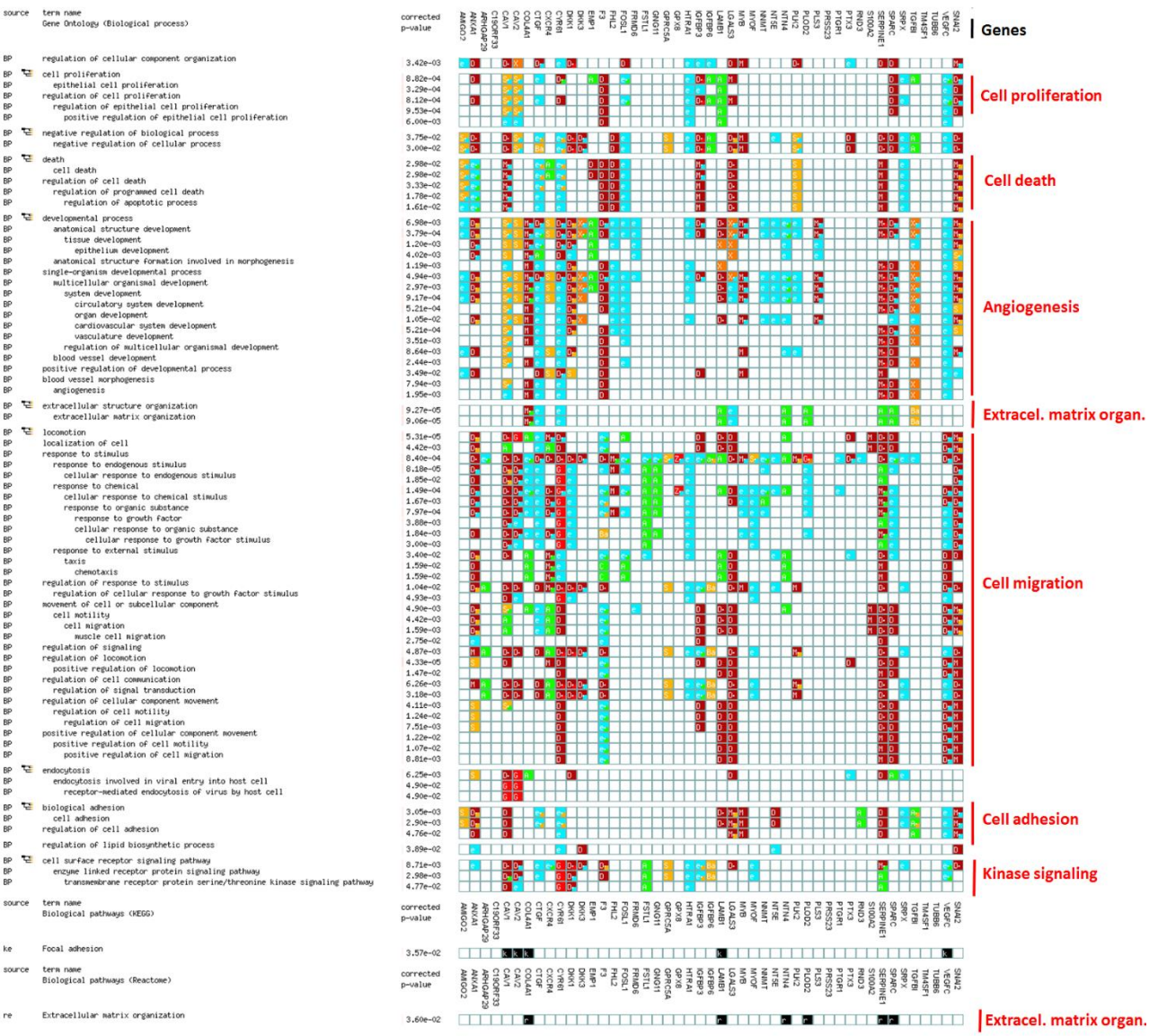

Figure S1. Statistical enrichment analysis of biological processes and pathways in the set of 47 network hubs. Related to Figure 2B. An alternative visualization of functional enrichments using G:Profiler (Reimand et al., 2007). As shown in Figure 2B, our set of 47 hub genes (columns) is significantly associated (at corrected P-value = 0.05) with a diversity of biological processes and pathways (rows). Each colored cell represents the association between an individual gene and a functional annotation. Colors are used to specify the evidence type of the observed association: Z: Inferred from experiment, D M: Direct assay / Mutant phenotype, A Traceable author, e: Electronic annotation, additional information at http://biit.cs.ut.ee/gprofiler.

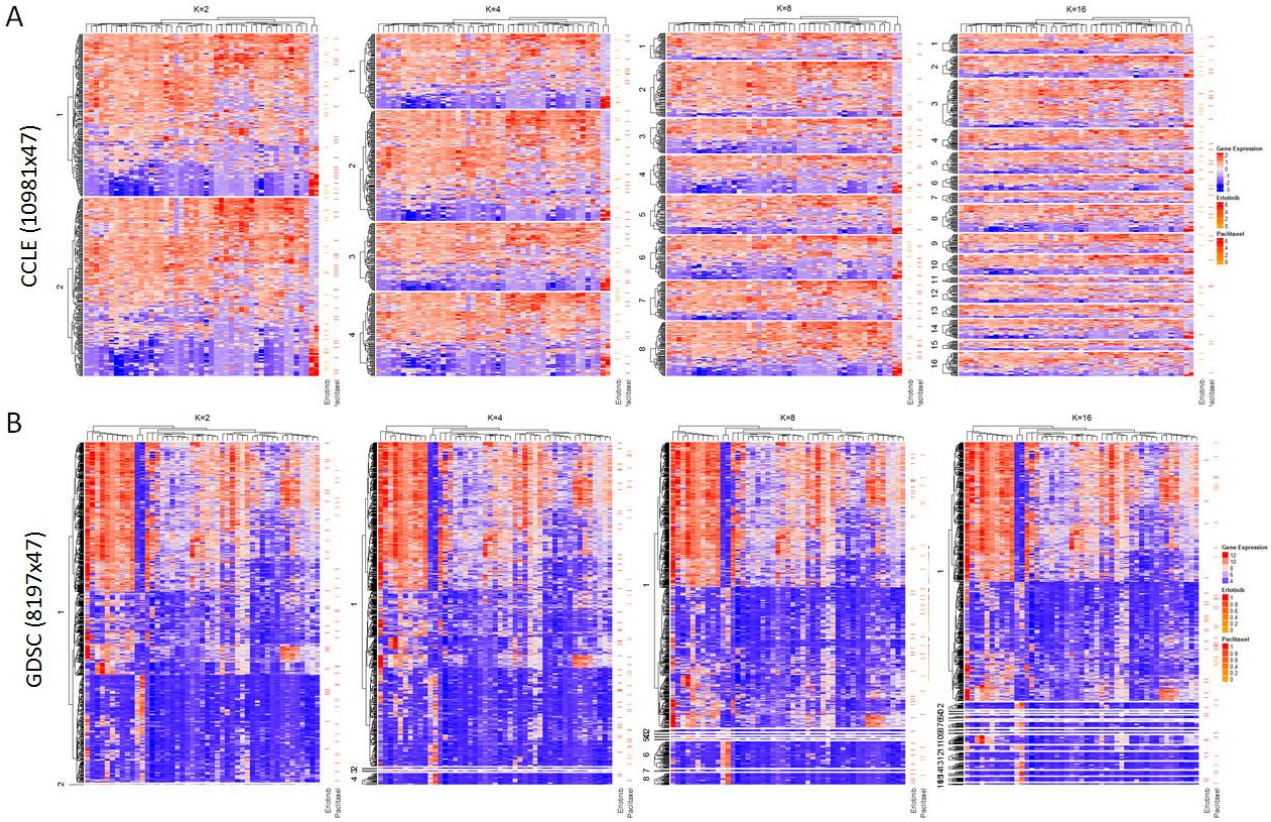

Figure S2. Alternative visualizations and unsupervised clustering of CCLE and GDSC cell lines on the basis of their 47-gene profiles. Related to Figures 2D and 4A. Spectral clustering analysis was applied using the SNFtool (Wang et al., 2014) to independently explore the potential of the 47 genes' expression data to segregate (cell line-drug experiment) samples. A. CCLE and B. GDSC results. In A. and B., rows and columns in each heatmap represent samples and genes respectively, and color represents gene expression intensity. To facilitate visualization, clustering results for different numbers of clusters (K) are provided as independent plots. Note that the order of the rows in each clustering (plot) is not preserved. In each plot, additional columns (right side) representing the drug sensitivity of the samples against Erlotinib ad Paclitaxel are illustrated.

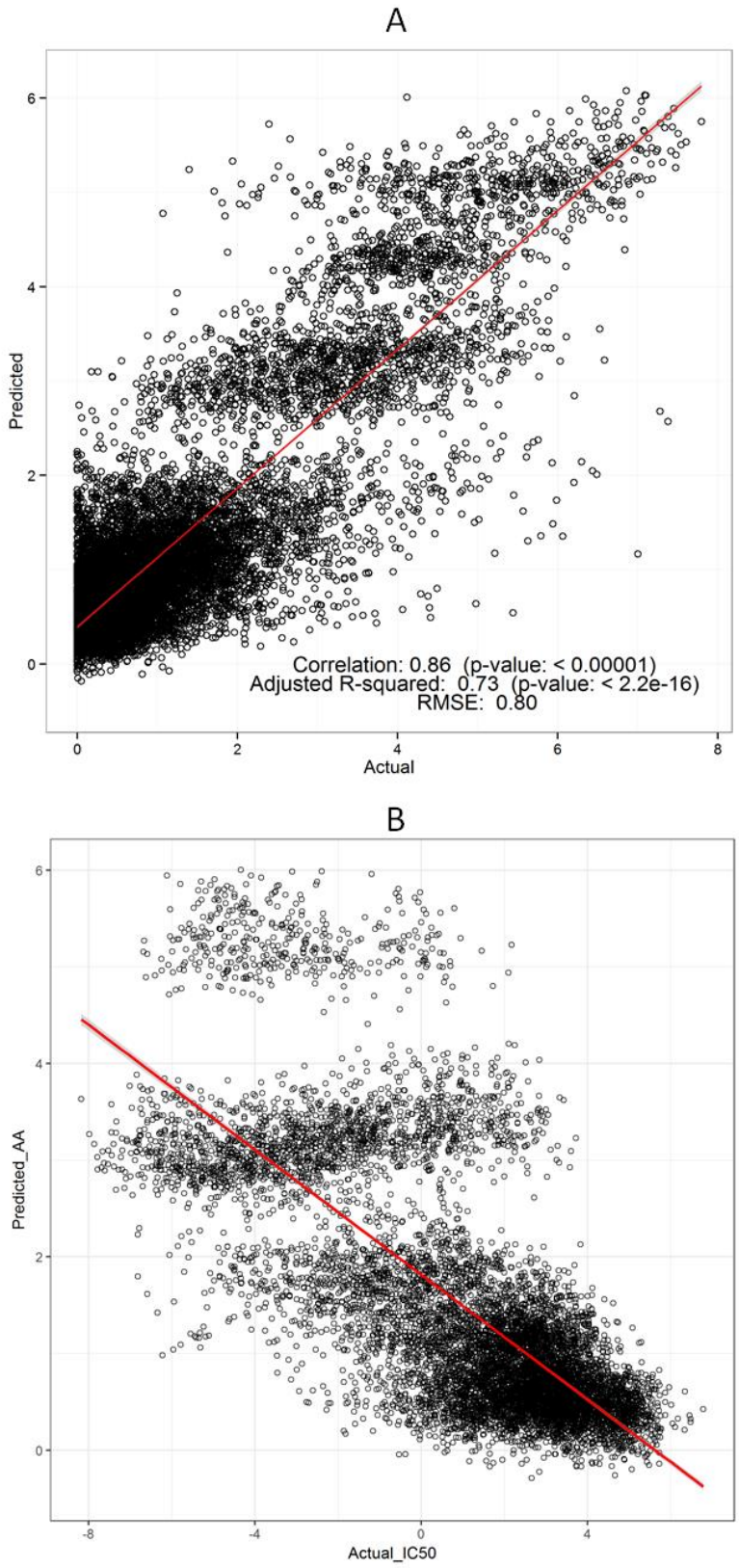

Figure S3. Predicted vs. actual sensitivity values in the CCLE and GDSC datasets Related to Figures 3A and 4B. Alternative visualization to those shown in Figure 3A. A. CCLE plot (n=10981). B. GDSC plot (n=9984).

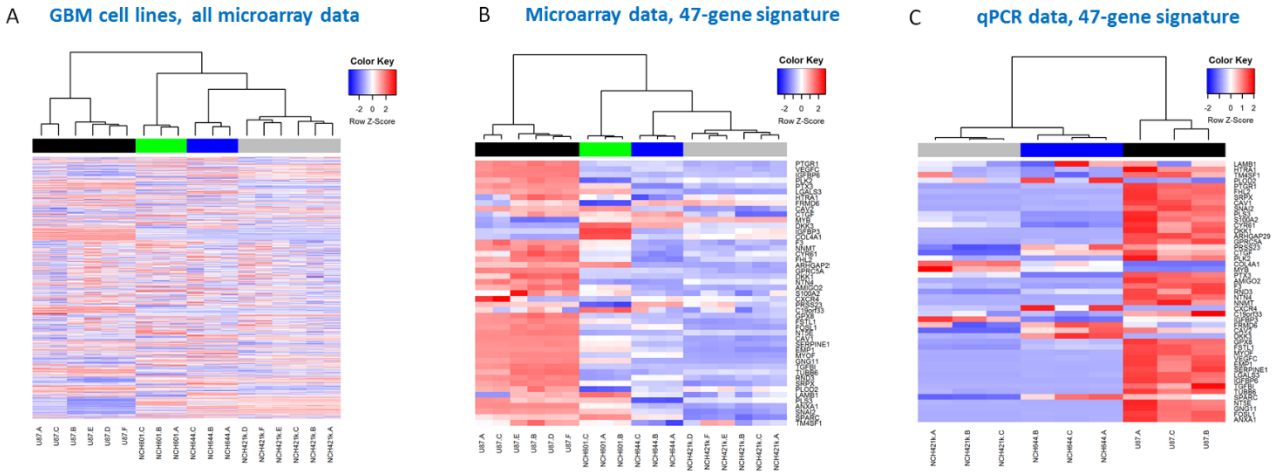

Figure S4. The 47-gene signature distinguishes cell types and is reproducible. Related to section: "Independent *in vitro* validation on several cell lines and compounds". Gene expression of 47 genes in 3 GBM cell lines using microarrays and qPCR. Analysis peformed to verify the robustness and platform-independent replicability of the 47-gene expression data and its capacitiy to distinguish between cell lines.

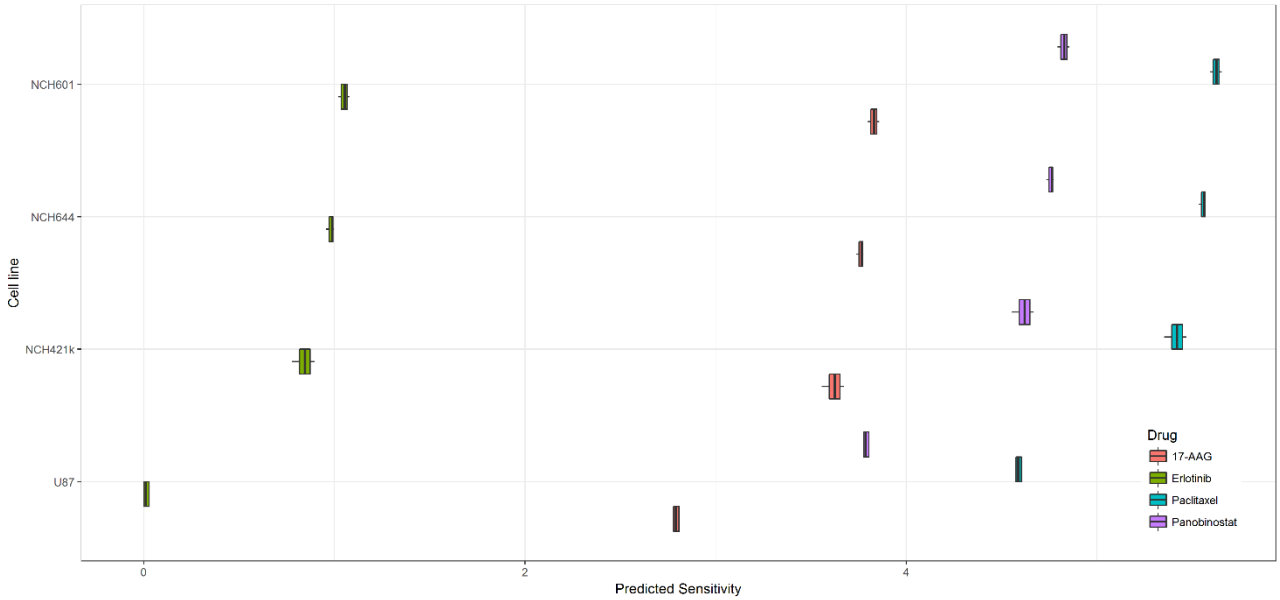

Figure S5. Boxplot summary of prediction results for 4 drugs (Erlotinib, 17-AAG, Panobinostat and Paclitaxel) and 4 GBM cell lines. Related to Figure 5B. Ech cell line type comprises multiple biological replicates (18 samples in total): 6 U87, 3 NCH644, 3 NCH601 and 6 NCH421k samples.

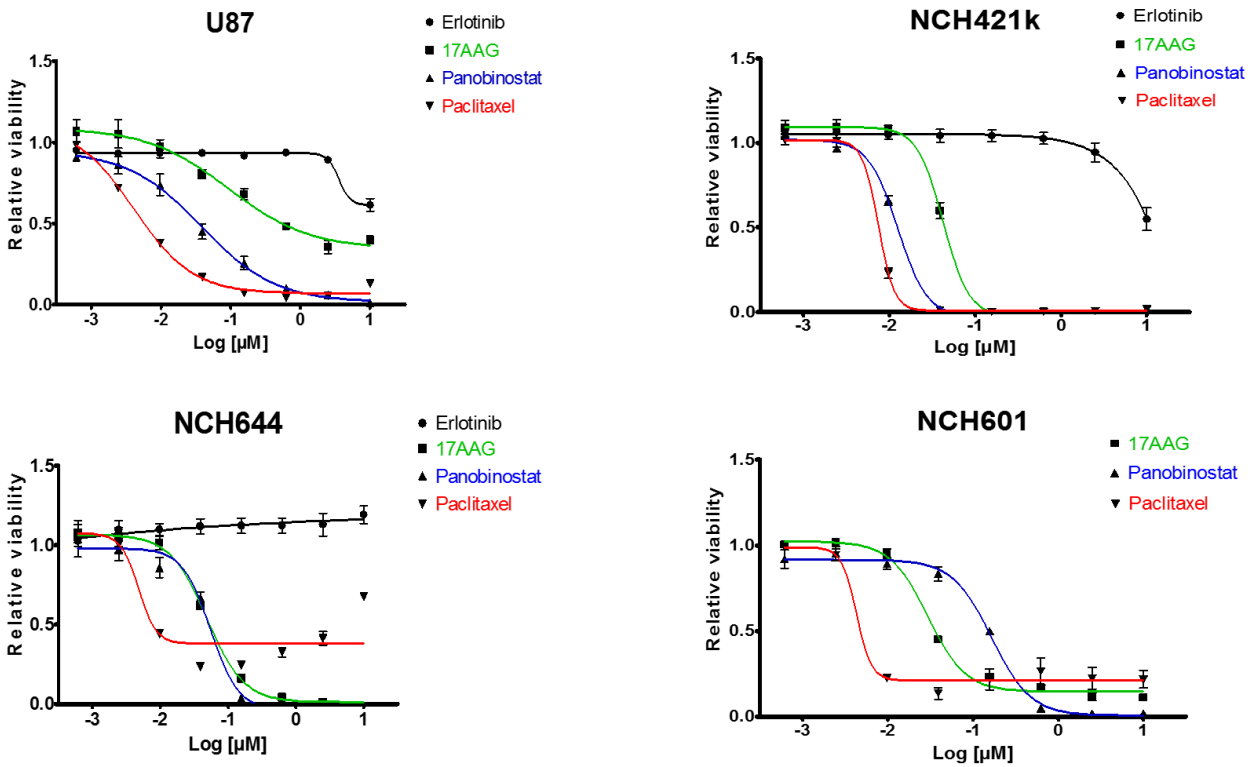

Figure S6. Drug response curves for the 4 drugs tested on the 4 GBM cell lines. Related to Figure 5C. Drugs were tested on each cell line in triplicates, and relative viability (vs. vehicle-treated samples) was measured for 8 drug concentration values (shown here as Log[μM]).

ITEM TO CHECK	IMPORTANCE	CHECKLIST
EXPERIMENTAL DESIGN		
Definition of experimental and control groups	**E**	U87, NCH421k, NCH644 cell lines
Number within each group	**E**	n=3
Assay carried out by core lab or investigator's lab?	D	
Acknowledgement of authors' contributions	D	
SAMPLE		
Description	**E**	U87, NCH421k, NCH644 cell lines cf. Sanzey M et al. PLoS One 2015 article.
Volume/mass of sample processed	D	1 x 10^6 cells per sample
Microdissection or macrodissection	**E**	N/A
Processing procedure	**E**	Cells were washed and counted in PBS (without Ca++ and Mg++). 1 x 106 cells were resuspended in TRI Reagent®, snap-frozen and then stored at -80°C
If frozen - how and how quickly?	**E**	Samples were snap-frozen in TRI Reagent® and stored at -80°C
If fixed - with what, how quickly?	**E**	N/A
Sample storage conditions and duration (especially for FFPE samples)	**E**	Samples were stored in TRI Reagent® at -80°C until RNA extraction
NUCLEIC ACID EXTRACTION		
Procedure and/or instrumentation	**E**	Total RNA was extracted from 1 x 10^6 cells with a TRI Reagent® (Sigma-Aldrich) isolation protocol. Aqueous phase was isolated with Phase lock gel-Heavy (5 Prime, Gaithersburg, MD). Total RNA was precipitated with 100% isopropanol and purified with a RNeasy® Mini kit combined with an on-column DNase treatment following the manufacturer's instructions(Qiagen, Valencia, CA).
Name of kit and details of any modifications	**E**	TRI Reagent® - RNeasy® Mini kit combined with an on-column DNase treatment following the manufacturer's instructions.
Source of additional reagents used	D	Chloroform (Merck); Isopropanol (Merk); Ethanol (Merck); Nuclease free water (Life Technologies)
Details of DNase or RNAse treatment	**E**	RNeasy® Mini kit combined with an on-column DNase treatment following the manufacturer's instructions.
Contamination assessment (DNA or RNA)	**E**	DNase treatement + Bioanalyzer + primers flanking intron + Negative controle (RT & qPCR)
Nucleic acid quantification	**E**	Nanodrop
Instrument and method	**E**	Nanodrop
Purity (A260/A280)	D	All RNA sample : Purity (A260/A280) = 2
Yield	D	N/A
RNA integrity method/instrument	**E**	Bioanalyzer
RIN/RQI or Cq of 3' and 5' transcripts	**E**	All RNA sample : RIN ≥ 9
Electrophoresis traces	D	N/A
Inhibition testing (Cq dilutions, spike or other)	**E**	The standard curve perfomed to check primers efficiency has been considerred sufficient to rule out the presence of inhibitors of reverse transcrition activity or PCR.
REVERSE TRANSCRIPTION		
Complete reaction conditions	**E**	1µg of RNA were reverse transcribed into cDNA using the SuperScript III (Invitrogen, Carlsbad, CA) reverse transcriptase with the following protocol: RNAs were mixed with random primers, oligo (dT)12-18 and dNTPs in a total volume of 13µl. Samples were heated to 65°C for 5 min and incubated on ice for at least 1min. Then the 5X RT buffer, DTT, RNaseOUT and SuperScript III was added to a total volume of 20µl. RT was allowed at 50°C for 60 min. and was followed by enzyme inactivation at 70°C for 15 min. Final concentrations were: 100ng of oligo(dT)12-18, 50ng of random primers, 0.5mM dNTPs, 50mM Tris-HCl, 75mM KCl, 3mM MgCl2, 5mM DTT, 40U of RNaseOUT and 200U of SuperScript III. To remove RNA complementary to the cDNA, 2U of *E.coli* RNaseH was added and incubated at 37°C for 20 minutes. In each RT-PCR a no template control (no RNA in RT) were performed.
Amount of RNA and reaction volume	**E**	1µg RNA / 20µl reaction volume
Priming oligonucleotide (if using GSP) and concentration	**E**	Random primers : 2.5ng/µl Oligo$(dT)_{12-18}$: 5ng/µl (final concentration)
Reverse transcriptase and concentration	**E**	SuperScript III (Invitrogen) : 10U/µl (final concentration)
Temperature and time	**E**	RNAs were mixed with random primers, oligo (dT)12-18 and dNTPs in a total volume of 13µl. Samples were heated to 65°C for 5 min and incubated on ice for at least 1min. Then the 5X RT buffer, DTT, RNaseOUT and SuperScript III was added to a total volume of 20µl. RT was allowed at 50°C for 60 min. and was followed by enzyme inactivation at 70°C for 15 min. To remove RNA complementary to the cDNA, 2U of *E.coli* RNaseH was added and incubated at 37°C for 20 minutes.
Manufacturer of reagents and catalogue numbers	D	Life Technologies : SuperScript III (Cat.18080-085), Oligo$(dT)_{12-18}$ primer (18418-012), Random primers (Cat. 48190-011), 10mM dNTP Mix (18427-013), RNaseOUT 40U/µl (10777-019), *E.coli* RNaseH (AM2293)
Cqs with and without RT	D*	N/A - DNase treatement + primers flanking intron + Melt Curve
Storage conditions of cDNA	D	-20°C
qPCR TARGET INFORMATION		
If multiplex, efficiency and LOD of each assay.	**E**	N/A
Sequence accession number	**E**	see additional file X-A
Location of amplicon	D	see additional file X-A
Amplicon length	**E**	see additional file X-A
In silico specificity screen (BLAST, etc)	**E**	Beacon Designer Pro 8.10 software (Premier Biosoft) + NCBI BLAST tool
Pseudogenes, retropseudogenes or other homologs?	D	
Sequence alignment	D	
Secondary structure analysis of amplicon	D	
Location of each primer by exon or intron (if applicable)	**E**	see additional file X-A
What splice variants are targeted?	E	see additional file X-A cf. Accession number
qPCR OLIGONUCLEOTIDES		
Primer sequences	**E**	see additional file X-A
RTPrimerDB Identification Number	D	N/A
Probe sequences	D**	N/A
Location and identity of any modifications	**E**	N/A
Manufacturer of oligonucleotides	D	EUROGENTEC (Seraing, Belgium)
Purification method	D	RP-Cartridge - Gold
qPCR PROTOCOL		
Complete reaction conditions	**E**	cDNAs obtained from RT of RNA were diluted 10-fold and 4µL were mixed with SYBR®Green Master Mix (Bio-Rad, Nazareth, Belgium) to a final volume of 20µL containing 300nM of each primer. Amplification was carried out in the CFX96 thermal cycler (Bio-Rad) under the following conditions: heating for 3 minutes at 95°C, 40 cycles of denaturation for 30 seconds at 95°C, followed by an annealing/extension for 1 min. After each run a Melting curve analysis was performed, ramping from 55°C to 95°C in 20min. A negative control without cDNA template was run in every assay and measures were performed in duplicates.
Reaction volume and amount of cDNA/DNA	**E**	4µl cDNa diluted 10fold / 20µl reaction volume
Primer, (probe), Mg++ and dNTP concentrations	**E**	300nM of each primer + SYBR®Green Master Mix
Polymerase identity and concentration	**E**	iTaq DNA polymerase in final concentration : 25 U/ml
Buffer/kit identity and manufacturer	**E**	SYBR®Green Master Mix (Biorad, Nazareth, Belgium) (Cat. 1708885)
Exact chemical constitution of the buffer	D	2x qPCR mix contains : dNTPs, 50 U/ml iTaq DNA polymerase, 6 mM MgCl2, SYBR Green I, enhancers, stabilizers, 20 nM fluorescein
Additives (SYBR Green I, DMSO, etc.)	**E**	N/A
Manufacturer of plates/tubes and catalog number	D	BioRad - HSP9655
Complete thermocycling parameters	**E**	heating for 3 minutes at 95°C, 40 cycles : denaturation for 30 seconds at 95°C, followed by an annealing/extension for 1 min. After each run a Melting curve analysis was performed, ramping from 55°C to 95°C in 20min.
Reaction setup (manual/robotic)	D	
Manufacturer of qPCR instrument	**E**	CFX thermal cycler (BioRad)
qPCR VALIDATION		
Evidence of optimisation (from gradients)	D	Annealing temperature gradients
Specificity (gel, sequence, melt, or digest)	**E**	Gene-specific amplification was confirmed by a single band in 4% E-Gel® (Life technologies). Melt Curve analysis were performed in each assay. No template controls(no cDNA in qPCR) were run for each gene to detect unspecific amplification and primer dimerization.
For SYBR Green I, Cq of the NTC	**E**	No amplifiction signal detected
Standard curves with slope and y-intercept	**E**	see additional file X-A
PCR efficiency calculated from slope	**E**	see additional file X-A
Confidence interval for PCR efficiency or standard error	D	
r2 of standard curve	**E**	see additional file X-A
Linear dynamic range	**E**	see additional file X-A
Cq variation at lower limit	**E**	see additional file X-A
Confidence intervals throughout range	D	
Evidence for limit of detection	**E**	see additional file X-A
If multiplex, efficiency and LOD of each assay.	**E**	N/A
DATA ANALYSIS		
qPCR analysis program (source, version)	**E**	CFX manager 3.1 software (Bio-Rad)
Cq method determination	**E**	The threshold is determined using the regression method. This mode applies a multi-variable, non-linear regression model to individual well traces and then uses this model to compute an optimal Cq value.
Outlier identification and disposition	**E**	Bad replicates were retested and measurements below LOD were discard
Results of NTCs	**E**	No amplifiction signal detected
Justification of number and choice of reference genes	**E**	8 reference genes were tested. Data normalization was carried out against three reference genes: Hspcb, Rps13, 18sRNA, see additional file X-B
Description of normalisation method	**E**	Normalized expression was calculated using the CFX manager 3.1 software (Biorad) via the ΔΔ-Cq method, taking into account the calculated amplification efficiency for each primers pair.
Number and concordance of biological replicates	D	
Number and stage (RT or qPCR) of technical replicates	**E**	qPCR reactions were performed in duplicates
Repeatability (intra-assay variation)	**E**	For each sample, standard deviation (SD) for the Cq variation between replicates has been used to express intra-assay variation. Instrument and liquid handeling variations were shown to be minimal.
Reproducibility (inter-assay variation, %CV)	D	
Power analysis	D	
Statistical methods for result significance	**E**	Refer to Methods section of article
Software (source, version)	E	Refer to Methods section of article
Cq or raw data submission using RDML	**D**	

Table 1. MIQE checklist for authors, reviewers and editors. All essential information (E) must be submitted with the manuscript. Desirable information (D) should be submitted if available. If using primers obtained from RTPrimerDB, information on qPCR target, oligonucleotides, protocols and validation is available from that source.

*: Assessing the absence of DNA using a no RT assay is essential when first extracting RNA. Once the sample has been validated as RDNA-free, inclusion of a no-RT control is desirable, but no longer essential.

**: Disclosure of the probe sequence is highly desirable and strongly encouraged. However, since not all commercial pre-designed assay vendors provide this information, it cannot be an essential requirement. Use of such assays is advised against.

A- MIQE qPCR primers informations

Gene Name	Accession number	Foward primer sequence (5'>3')	Reverse primer sequence (5'>3')	Primers location	Amplicon (pb)	Standard curves	PCR efficiency (%)	r2
18sRNA	NR_003286	CAGGATTGACAGATTGAT	TTATCGGAATTAACCAGAC	only one exon	97	y= 2.977 x -3.175	106.50%	0.999
AMIGO2	NM_001143668; NM_181847	TTCTGGATTCTGAGTGGATTC	TGCTGGTGATGTTGTTATGA	F : E2(ou3) - R: E2(ou3) (same exon)	78	y= 19.694 x -3.325	99.90%	0.996
ANXA1	NM_000700	TCGCAGAGTGTTTCAGAA	TCTCAATGTCACCTTTCAAC	F:E8/9 (Intron 887pb) - R:E9	86	y= 11.534 x -3.189	105.80%	0.999
ARHGAP29	NM_004815	AAGAACACTGACTCTATCG	CTCCAATTCCAAGTTAAGC	F:E7 - R:E7/8 (Intron 1066pb)	108	y= 18.879 x -3.448	95.00%	0.998
C19orf33	NM_033520	TCCAAAGCAAGGACACCA	TGGGACTTCACATCCGTG	F:E 2/3 (Intron 133pb) - R:E3/4 (Intron 158pb)	75	y= 18.655 x -3.268	102.30%	0.996
CAV1	NM_001172895; NM_001172896; NM_001172897; NM_001753	AGATCGACCTGGTCAACC	GCAATCACATCTTCAAAGTCAATC	F:E 2(ou 1) - R:E 2/3 (ou1/2) (Intron 32256pb)	76	y= 14.398 x -3.308	100.60%	0.999
CAV2	NM_001206747; NM_001206748; NM_001233; NM_198212	CAAGTCTATAATGTGAGTAGT	TTATTCCAGTTCAATCATCA	F : E3 - R: E3 (3'UTR)	190	y= 18.621 x -3.566	90.70%	0.997
COL4A1	NM_001845	AGGGACAAATGGGCTTAA	TTCTTGAACTTGAGCTTGT	F:E11/12 (Intron 501pb) - R:E13 (Intron 1359pb)	101	y= 18.405 x -3.331	99.60%	0.998
CTGF	NM_001901	GCTGACCTGGAAGAGAAC	AAACTTGATAGGCTTGGAGAT	F:E4 - R:E5 (Intron 388pb)	75	y= 17.593 x -3.42	96.10%	0.996
CXCR4	NM_001008540; NM_003467	GAGGCAGATGACAGATAT	AATACCAGGCAGGATAAG	F : E1(ou2) - R: E1(ou2) (same exon)	105	y= 17.432 x -3.254	102.90%	0.996
CYR61	NM_001554	AATGAATTGATTGCAGTTG	TGTAAAGGGTTGTATAGGA	F:E3 - R:E4 (Intron 131pb)	89	y= 17.355 x -3.158	107.30%	0.998
DKK1	NM_012242	TATCACACCAAAGGACAA	GTCTAGCACAACACAATC	F:E3-4 (Intron 118pb) - R:E4	76	y= 19.432 x -3.335	99.50%	0.999
DKK3	NM_015881; NM_013253; NM_001018057	AAAGCATCATCAGAAGTG	TGTTGGTTATCTTGTGAAT	F:E3 - R:E3/4 (Intron 3520pb)	124	y= 17.042 x -3.292	101.30%	0.999
EMP1	NM_001423	AATGTCTGGTTGGTTTCC	GCATCTTCACTGGCATAT	F:E2/3(Intron 1890pb) - R:E3/4 (Intron 105pb)	104	y= 13.964 x -3.231	104.00%	0.999
F3	NM_001178096; NM_001993	CGTCAATCAAGTCTACAC	TTCACATCCTTCACAATC	F:E2 - R:E3 (Intron 4092pb)	117	y= 19.464 x -3.087	110.80%	0.998
FHL2	NM_001039492; NM_001450; NM_201555; NM_201557	AGACTGCTATTCCAACGA	CCTTGTACTCCATCTTGC	F:E3/4/5 - R:E 4/5/6 (Intron 5819pb)	86	y= 15.448 x -3.295	101.20%	0.999
FOSL1	NM_005438	TCCCTAACTCCTTTCACC	CTGCTACTCTTGCGATGA	F : E4 - R: E4	86	y= 13.047x -3.245	103.30%	0.999
FRMD6	NM_001042481; NM_001267046; NM_001267047; NM_152330	CTACATCACAGAGGACAT	GACCCAATTTCTTTCACA	F:E3/12/13 - R:E 4/13/14 (Intron 3698pb)	75	y= 18.423 x -3.454	94.80%	0.999
FSTL1	NM_007085	CTGCCATCAATATTACAAC	TTCATCAGACAGTTCAAT	F:E7 - R:E8 (Intron 1498pb)	92	y= 15.296 x -3.351	98.80%	0.992
GNG11	NM_004126	GAAGATTTGCCAGAGAAG	ACATTTAGACACTTGTTGT	F:E1 - R:E1/2 (Intron 3857pb)	90	y= 12.903 x -3.283	101.60%	1.000
GPRC5A	NM_003979	CTCAACTCGTGAAGAAGA	GAAAATGTGTGGAATAGGG	F:E2 - R:E3 (Intron 2912pb)	117	y= 18.350 x -3.274	102.00%	0.999
GPX8	NM_001008397	CTTCCACAAGATTAAGATTC	GTTGACAAGATACTTCCA	F:E2 - R:E3 (Intron 2799pb)	109	y= 15.841 x -3.334	99.50%	0.999
HSPCB	NM_001271969; NM_001271970; NM_001271971; NM_001271972; NM_007355	AAGCATTCTCAGTTCATA	TCTTCTTCTTATCCTTACC	F:E5 - R:E 6 (Intron 136)	193	y= 11.415 x -3.323	99.90%	0.999
HTRA1	NM_002775	ATCATCAACTATGGAAAC	GATACCAATATACTTCTTCT	F:E4 - R:E6 (Intron 484-1253pb)	192	y= 15.354x -3.368	98.10%	1.000
IGFBP3	NM_000598; NM_001013398	CGGGAGACAGAATATGGT	CAGCACATTGAGGAACTT	F:E2-3 (Intron 545pb) - R:E3	75	y= 19.317 x -3.354	98.70%	0.999
IGFBP6	NM_002178	CCCTCCCAGCCCAATTCT	CAGCACTGAGTCCAGATGTCT	F:E2 - R:E3 (Intron 183pb)	75	y= 14.322 x -3.307	100.60%	1.000
LAMB1	NM_002291	ATTATCTGACACAACTTC	AATACTTGGTAATGCTATC	F:E26 - R:E27 (Intron 1436pb)	164	y= 15.632 x -3.315	100.30%	0.999
LGALS3	NM_001177388; NM_002306	ATGCTGATAACAATTCTGG	CAAACAATGACTCTCCTG	F:E4 - R:E5 (Intron 2220pb)	131	y= 14.376 x -3.436	95.50%	0.999
MYB	NM_001130172; NM_001130173; NM_001161656; NM_001161657; NM_001161658; NM_001161659; NM_001161660; NM_005375	CAACGACTATTCCTATTACC	CTGAGGGACATTGACTAT	F:E6 - R:E7 (Intron 1279pb)	100	y= 16.785 x -3.308	100.60%	0.998
MYOF	NM_013451; NM_133337	ATAGAAGACACGAGATACAC	GCTTTCGGATCTGAGTAT	F:E21/22 (Intron 2367pb) - R:E 22	79	y= 15.903 x -3.228	104.10%	1.000
NNMT	NM_006169	CAGTGGTGACCTATGTGT	CCTGTCTCAACTTCTCCT	F:E4 - R:E5 (Intron 13886pb)	75	y= 17.363 x -3.312	100.40%	1.000
NT5E	NM_001204813; NM_002526;	GGAATCGTTGGATACACT	ACTTATCTACTTCAGGTTGT	F:E2 - R:E3 (Intron 3954pb)	106	y= 13.536 x -3.344	95.10%	1.000
NTN4	NM_021229	CTGGAAGATGATGTTGTC	GGTTCTCTGTATCGTATG	F:E2 - R:E3 (Intron 48794pb)	121	y= 20.842 x -3.313	100.40%	0.994
PLK2	NM_006622; NM_001252226	GGATGCTATTCGGATGAT	ATGGTACTGTCTTCAAGG	F:E10 - R:E11 (Intron 246pb)	78	y= 17.666 x -3.423	95.90%	0.997
PLOD2	NM_182943; NM_000935	CTAGCAGACAAGTATCCT	GAACTATACGGTTGACATAT	F:E4 - R:E5 (Intron 3640pb)	94	y= 16.960 x -3.311	100.40%	0.999
PLS3	NM_001136025; NM_001172335; NM_001282337; NM_001282338; NM_005032	GCTGATGAGAAGATATACC	CTCTGAATGGAAGTTGAT	F:E13/14 (Intron 1014pb) - R:E14	129	y= 15.001 x -3.349	98.90%	0.999
PRSS23	NM_007173	CTCGGCGCGGAACAG	CCAACAGCACAGAGCAGAA	F:E1/2 (Intron 6979pb) - R:E2	79	y= 18.777 x -3.346	99.00%	0.997
PTGR1	NM_001146108; NM_001146109;NM_012212	GTTGGCTATCCTACTAAT	CATCATTGTATCACCTTC	F:E2 - R:E4 (Intron 3049/1239pb)	153	y= 15.472 x -3.356	98.60%	0.999
PTX3	NM_002852	TGAATTTGGACAACGAAATAGAC	ATTCCGAGTGCTCCTGAC	F:E1 - R:E2 (Intron 449pb)	84	y= 20.26 x -3.573	90.50%	0.997
RND3	NM_001254738; NM_005168	TGTTAGTACATTAGTAGAG	AGCATTCGATATAAGTAG	F:E5 - R:E6 (Intron 1388pb)	107	y= 18.532 x -3.148	107.80%	0.996
RPS13	NM_001017	GTCCCCACTTGGTTGAAG	CCATGTGAATCTCTCAGGAT	F:E2-3 (Intron182) - R:E 3	113	y= 12.185 x -3.344	99.10%	1.000
S100A2	NM_005978	CAAGTTCAAGCTGAGTAAG	CTCCTCATCCACTTTCTC	F:E2 - R:E3 (Intron 2142pb)	85	y= 18.328 x -3.244	103.40%	0.999
SERPINE1	NM_000602	TAGAGAACCTGGGAATGAC	GAGGCTCTTGGTCTGAAA	F:E6 - R:E 6/7 (Intron 120pb)	75	y= 14.046 x -3.274	102.00%	1.000
SNAI2	NM_003068	ACACATACAGTGATTATTTCC	GTAGTCCACACAGTGATG	F:E1-2/2-3 (Intron 745pb) - R:E2/3	113	y= 15.391 x -3.262	102.60%	1.000
SPARC	NM_003118	AGGTGACTGAGGTATCTGT	TGGTTCTGGCAGGGATTT	F:E 3/4 (Intron 1395pb) - R:E4/5 (Intron 1434pb)	115	y= 13.713 x -3.308	100.60%	0.996
SRPX	NM_001170750; NM_001170751; NM_001170752; NM_006307	ATTCTTACTGATGTCATTCT	TCTGTCATAGACTGTGTA	F:E 4/5 (Intron 3714pb) - R:E 5	90	y= 14.738 x -3.142	108.10%	0.999
TGFBI	NM_000358	AGAAGGTTATTGGCACTAAT	GCTGATGACTGTTGATTTG	F:E2 - R:E 2/3 (Intron 10196pb)	89	y= 14.290 x -3.345	99.00%	0.999
TM4SF1	NM_014220	ATTGTGGAATGGAATGTATC	ATATTGCTGTTGGTGAGA	F:E4 - R:E4/5 (Intron 1806pb)	138	y= 15.462 x -3.278	101.90%	0.999
TUBB6	NM_032525	AGAGAATCAACGTCTACTACAATG	GGCTCTAAGTCCACCAGG	F:E2 - R:E3 (Intron 2147pb)	76	y= 14.063 x -3.346	99.00%	1.000
VEGFC	NM_005429	AAGGACAGAAGAGACTAT	CACATCTATACACACCTC	F:E2 - R:E3 (Intron 1564pb)	118	y= 16.637 x -3.374	97.90%	0.999

B- MIQE data analysis informations

Target Stability

Target	Coefficient Variance	M-Value
Hspcb	0.136	0.333
Rps13	0.119	0.319
18sRNA	0.162	0.371

Average Coefficient Variance: : 0.139
Average M-Value: : 0.341

Coefficient of Variation (CV) of normalized reference gene relative quantities. A lower CV value denotes higher stability
M-value. A measure of the reference gene expression stability

Linear dynamic range	Cq variation at lower limit
from 1.00E-01 cDNA dilution to 1.00E-08	28.24 ± 0.389
from 1.00E-01 cDNA dilution to 1.56E-03	28.97 ± 0.242
from 1.00E-01 cDNA dilution to 2.96E-05	25.89 ± 0.061
from 1.00E-01 cDNA dilution to 1.56E-03	28.52 ± 0.097
from 1.00E-01 cDNA dilution to 3.91E-04	29.64 ± 0.258
from 1.00E-01 cDNA dilution to 1.00E-05	30.91 ± 0.323
from 1.00E-01 cDNA dilution to 3.91E-04	30.57 ± 0.095
from 1.00E-01 cDNA dilution to 1.56E-03	27.67 ± 0.007
from 1.00E-01 cDNA dilution to 3.91E-04	29.05 ± 0.198
from 1.00E-01 cDNA dilution to 3.91E-04	28.35 ± 0.255
from 1.00E-01 cDNA dilution to 7.72E-05	30.17 ± 0.209
from 1.00E-01 cDNA dilution to 1.56E-03	28.75 ± 0.093
from 1.00E-01 cDNA dilution to 1.60E-04	29.61 ± 0.057
from 1.00E-01 cDNA dilution to 1.00E-04	26.82± 0.109
from 1.00E-01 cDNA dilution to 3.91E-04	29.89 ± 0.118
from 1.00E-01 cDNA dilution to 7.72E-05	29.07 ± 0.121
from 1.00E-01 cDNA dilution to 1.00E-05	29.27 ± 0.35
from 1.00E-01 cDNA dilution to 3.91E-04	30.2 ± 0.140
from 1.00E-01 cDNA dilution to 1.00E-05	32.48 ± 0.861
from 1.00E-01 cDNA dilution to 1.00E-05	29.32 ± 0.249
from 1.00E-01 cDNA dilution to 8.00E-04	28.51 ± 0.063
from 1.00E-01 cDNA dilution to 1.95E-04	28.18 ± 0.26
from 1.00E-01 cDNA dilution to 1.98E-06	30.55 ± 0.236
from 1.00E-01 cDNA dilution to 1.95E-04	27.81 ± 0.046
from 1.00E-01 cDNA dilution to 3.91E-04	30.69 ± 0.038
from 1.00E-01 cDNA dilution to 1.00E-04	27.54 ± 0.008
from 1.00E-01 cDNA dilution to 1.95E-04	27.89 ± 0.209
from 1.00E-01 cDNA dilution to 7.72E-05	28.58 ± 0.21
from 1.00E-01 cDNA dilution to 1.60E-04	29.31 ± 0.285
from 1.00E-01 cDNA dilution to 2.44E-05	30.81 ± 0.018
from 1.00E-01 cDNA dilution to 8.00E-04	27.59 ± 0.083
from 1.00E-01 cDNA dilution to 1.00E-05	30.7 ± 0.112
from 1.00E-01 cDNA dilution to 3.70E-03	28.81 ± 0.324
from 1.00E-01 cDNA dilution to 3.91E-04	29.34 ± 0.337
from 1.00E-01 cDNA dilution to 7.72E-05	30.7 ± 0.098
from 1.00E-01 cDNA dilution to 7.72E-05	28.83 ± 0.049
from 1.00E-01 cDNA dilution to 3.91E-04	30.12 ± 0.089
from 1.00E-01 cDNA dilution to 7.72E-05	29.21 ± 0.25
from 1.00E-01 cDNA dilution to 6.25E-03	28.81 ± 0.23
from 1.00E-01 cDNA dilution to 8.00E-04	28.43 ± 0.042
from 1.00E-01 cDNA dilution to 1.00E-04	25.61 ± 0.076
from 1.00E-01 cDNA dilution to 1.60E-04	30.51 ± 0.112
from 1.00E-01 cDNA dilution to 7.72E-05	27.47 ± 0.162
from 1.00E-01 cDNA dilution to 1.95E-04	27.51 ± 0.144
from 1.00E-01 cDNA dilution to 1.00E-05	30.33 ± 0.855
from 1.00E-01 cDNA dilution to 7.72E-05	27.65 ± 0.171
from 1.00E-01 cDNA dilution to 1.00E-05	31 ± 0.249
from 1.00E-01 cDNA dilution to 1.95E-04	27.54 ± 0.116
from 1.00E-01 cDNA dilution to 7.72E-05	27.76 ± 0.036
from 1.00E-01 cDNA dilution to 4.63E-04	27.78 ± 0.026