

# Shifting from Population-wide to Personalized Cancer Prognosis with Microarrays

Li Shao<sup>1</sup>\*, Xiaohui Fan<sup>1</sup>\*, Ningtao Cheng<sup>2</sup>, Leihong Wu<sup>1</sup>, Haoshu Xiong<sup>1</sup>, Hong Fang<sup>3</sup>, Don Ding<sup>3</sup>, Leming Shi<sup>4</sup>, Yiyu Cheng<sup>1\*</sup>, Weida Tong<sup>4\*</sup>

**1** Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, **2** The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, United States of America, **3** ICF International at NCTR/FDA, Jefferson, Arkansas, United States of America, **4** National Center for Toxicological Research (NCTR), U. S. Food and Drug Administration, Jefferson, Arkansas, United States of America

## Abstract

The era of personalized medicine for cancer therapeutics has taken an important step forward in making accurate prognoses for individual patients with the adoption of high-throughput microarray technology. However, microarray technology in cancer diagnosis or prognosis has been primarily used for the statistical evaluation of patient populations, and thus excludes inter-individual variability and patient-specific predictions. Here we propose a metric called clinical confidence that serves as a measure of prognostic reliability to facilitate the shift from population-wide to personalized cancer prognosis using microarray-based predictive models. The performance of sample-based models predicted with different clinical confidences was evaluated and compared systematically using three large clinical datasets studying the following cancers: breast cancer, multiple myeloma, and neuroblastoma. Survival curves for patients, with different confidences, were also delineated. The results show that the clinical confidence metric separates patients with different prediction accuracies and survival times. Samples with high clinical confidence were likely to have accurate prognoses from predictive models. Moreover, patients with high clinical confidence would be expected to live for a notably longer or shorter time if their prognosis was good or grim based on the models, respectively. We conclude that clinical confidence could serve as a beneficial metric for personalized cancer prognosis prediction utilizing microarrays. Ascribing a confidence level to prognosis with the clinical confidence metric provides the clinician an objective, personalized basis for decisions, such as choosing the severity of the treatment.

**Citation:** Shao L, Fan X, Cheng N, Wu L, Xiong H, et al. (2012) Shifting from Population-wide to Personalized Cancer Prognosis with Microarrays. PLoS ONE 7(1): e29534. doi:10.1371/journal.pone.0029534

**Editor:** Christina Chan, Michigan State University, United States of America

**Received:** June 20, 2011; **Accepted:** November 30, 2011; **Published:** January 25, 2012

**Copyright:** © 2012 Shao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National Science Foundation of China (No. 30801556 and 30830121), National S&T Major Project (No. 2008ZX09312-001), Science Foundation of Chinese University (No. 2009QNA7031), and the Zhejiang Provincial Natural Science Foundation of China (No. R2080693). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: chengyy@zju.edu.cn (YC); weida.tong@fda.hhs.gov (WT)

† These authors contributed equally to this work.

## Introduction

Not all individuals respond to drug treatment in the same way. Accordingly, the development of personalized therapeutic regimens optimized for individual patients represents a major goal of 21<sup>st</sup>-century medicine [1]. Modern tools are being utilized to assist physicians in effectively treating patients as individuals and providing personalized drug intervention.

Inter-individual variation in response to drug treatment is strongly influenced by a patient's physiological state at the time of treatment. This state can be characterized by gene expression profiles [2]. Therefore, microarray technology can guide the selection of drugs or therapeutic regimens and be employed to assess the susceptibility of a patient to certain diseases, enabling a personalized plan for prevention monitoring and treatment [3]. The prospective benefits of microarray technology in clinical applications have been demonstrated by several landmark studies [4–7]. Microarray-based predictive models (or genomic signatures) have shown utility in associating different subgroups of breast cancer with distinct clinical outcomes [8–13], such as Mamma-

Print<sup>TM</sup> [4,5], a milestone in microarray-based prognosis for breast cancer [14].

The development of a microarray-based predictive model for tumor classification typically involves two sequential steps [4,15–17]. First, the model is developed based on a training set of patients with known class labels (e.g., tumor status) and gene expression data. Next, the training model is validated using a validation set that contains patients with known class labels. The validity of the training model in performance on the validation set has been the focus of 'class prediction' research. To ensure the training model can be used in real-world clinical applications, it was suggested that the model must be assessed on a large number of independent samples in this external validation process [18].

It is important to note that the aforementioned external validation strategy assesses the performance of a training model using a population defined by the validation set. The average performance (e.g., specificity, sensitivity) over the population is used to assess whether the model can be a reliable diagnostic or prognostic test. This strategy is performed under the assumption that the model performs equally for everyone without considering

the inter-individual variability. Thus, the average performance based on a population of patients cannot ensure its predictive ability for individual patients, which might result in potentially unreliable diagnoses or prognoses in the real-world application. This one-size-fits-all strategy needs to be modified from population-wide to personalized medicine in microarray data-based applications.

We propose a metric called clinical confidence that measures the model's reliability in prediction performance on an individual basis. Clinical confidence can be useful in determining appropriate treatments; for example, patients with high confidence and poor prognosis may be given more rigorous treatments. Additionally, patients with lower clinical confidences may be prime candidates for further evaluation of their conditions with alternative methods. The accuracy of the clinical confidence metric was investigated on three large clinical datasets with total of six clinical endpoints [19].

Specifically, we first divided each dataset into two, i.e., the training and validation set. To mimic real-world clinical scenarios, we made the validation set that contains only the patients whose microarray data were generated at a later date than those in the training set. We derived the clinical confidence from the training model, followed by the assessment of its correlation with prediction accuracy for prognosis and the survival time of the patients in the validation set. To the best of our knowledge, this is the first attempt to provide a measure of confidence for individual patients in microarray-based "class prediction" research, which is an important step forward in personalized medicine.

## Materials and Methods

### Datasets

Three large-scale, clinical cancer datasets were used in this study: breast cancer (BR) [20], multiple myeloma (MM) [21], and neuroblastoma (NB) [22]. A concise summary of the datasets is given in **Table 1**. More detailed information of these datasets can be found in the main paper of the second phase of MicroArray Quality Control project (MAQC-II) [19].

Each dataset has two clinical endpoints related to cancer prognosis (including survival data) or treatment: BR-pCR and BR-erpos in the treatment response dataset, NB-EFS and MM-EFS in the event-free survival dataset, and NB-OS and MM-OS in the overall survival dataset (**Table 1**). These three clinical datasets were studied in the MAQC-II project led by the FDA [19]. To emulate a real-world clinical scenario in applying genomic signatures, two independent populations of patients for each of the three clinical datasets were defined by the MAQC Consortium as the training and validation sets using a chronological approach where the samples in the validation sets were generated at a later date than those in the training sets. The sample sizes for the training sets varied between 130 and 340, and the ratio of positive events to negative events ranged from 0.18 to 1.60; meanwhile, the sample sizes in the validation sets ranged from 100 to 214, and the ratio of positive events to negative events varied between 0.14 and 1.56.

Two positive and two negative control endpoints were also used in this study. The positive control endpoints, i.e., NB-PC and MM-PC, were derived from the NB and MM datasets separately, with the samples denoted by the gender. For the two negative control endpoints, i.e., NB-NC and MM-NC (which correspond to the NB and MM datasets, respectively), the sample labels (i.e., positive or negative events) were randomly generated. Using these two controls allow us to assess the performance of the clinically relevant endpoints against the expected maximum and minimum performance provided by the controls.

### Clinical confidence

The clinical confidence measures the confidence of a sample being assigned to a specific class by a predictive model. For sample  $i$ , its clinical confidence metric ( $C_i$ ) is the confidence level of a sample in being correctly assigned by a predictive model and is defined as:

$$C_i = \begin{cases} |x_1|/(|x_1|+|x_2|) & |x_1| > |x_2| \\ |x_2|/(|x_1|+|x_2|) & |x_1| \leq |x_2| \end{cases} \quad (1)$$

where  $x_1$  and  $x_2$  are the similarity measures between sample  $i$  and samples in class 1 and class 2, respectively. The similarity measure varies according to classifiers used. Two well-studied classifiers for gene expression data were employed in this study, i.e., Nearest-Centroid classification rule (NC) [4] and  $k$ -nearest neighbors ( $kNN$ ,  $k=5$ ) [23]. For the NC classifier,  $x_1$  and  $x_2$  were defined as the correlation coefficients of the unknown sample to the centroids of class 1 and class 2, respectively. The centroid is defined as vectors of the average expression values. For the  $kNN$  classifier,  $x_1$  and  $x_2$  are defined to be the number of nearest neighbors to the unknown sample belonging to class 1 and class 2, respectively.

$C_i$  values range from 0.5 and 1 in which a value of 0.5 indicates that the prediction is due to chance. Increasingly larger  $C_i$  values correspond to increasingly higher prediction confidence. For the sake of simplicity, all of the analysis was based on three confidence levels: low confidence (LC;  $0.5 \leq C_i \leq 0.6$ ), medium confidence (MC;  $0.6 < C_i \leq 0.8$ ) and high confidence (HC;  $0.8 < C_i \leq 1.0$ ).

### Statistical analysis

The general analysis workflow is depicted in **Figure 1**; additional details are provided in **Methods S1**. The analysis protocol starts by developing a best classifier based on the training set, and ends by predicting the validation set. The predicted class and corresponding clinical confidences are recorded in matrices  $L$  and  $C$ , respectively. To ensure the statistical validity, the procedure is repeated 500 times, resulting in 500 different classifiers from the training sets and 500 predictions for the validation set. The performance of both training models and predictions is assessed using Matthews correlation coefficient (MCC) [24,25].

A permutation test was also employed to compare classifier prediction accuracy versus chance [26,27]. In each permutation, the analysis protocol shown in **Figure 1** was repeated with the exception that the class labels in the training set were randomized. In other words, models constructed with randomized training sets were utilized to predict the validation sets. After 500 repetitions, the degree of chance correlation and predictability of endpoints was computed with Cohen's  $d$  [28], which measures the standardized difference between two means.

## Results

The cross-validation performance measured in MCC values for all the training models along with the average prediction performance on the validation sets are summarized in **Table S1**. The model performance follows the order of NB-PC, MM-PC, BR-erpos, NB-EFS, NB-OS, BR-pCR, MM-EFS, MM-OS, MM-NC, and NB-NC. The two positive controls performed best while the two negative controls perform worst, which is consistent with expectations from the experiment design.

### Clinical confidence positively correlates with the model prediction performance

We first investigated the model performance on the validation set for patients falling into different categories of clinical

**Table 1.** A concise summary of datasets.

Data Set code	Endpoint Description	Endpoint Code*	Sample Size		Ratio of events		Microarray Platform (number of channel)
			Training	Validation	Training	Validation	
BR	Treatment Response	BR-pCR	130	100	0.34 (33/97) <sup>†</sup>	0.18 (15/85)	Affymetrix U133A (1)
		BR-erpos	130	100	1.60 (80/50)	1.56 (61/39)	
MM	Overall Survival Milestone Outcome	MM-OS	340	214	0.18 (51/289)	0.14 (27/187)	Affymetrix U133Plus2.0 (1)
	Event-free Survival Milestone Outcome	MM-EFS	340	214	0.33 (84/256)	0.19 (34/180)	
NB	Overall Survival Milestone Outcome	NB-OS	246	177	0.32 (59/187)	0.28 (39/138)	Agilent NB Customized Array (2)
	Event-free Survival Milestone Outcome	NB-EFS	246	193	0.65 (97/149)	0.75 (83/110)	
Control	Positive control	NB-PC	246	231	1.44 (145/101)	1.36 (133/98)	Agilent NB Customized Array (2)
		MM-PC	340	214	1.33 (194/146)	1.89 (140/74)	Affymetrix U133Plus2.0 (1)
	Negative control	NB-NC	246	253	1.44 (145/101)	1.30 (143/110)	Agilent NB Customized Array (2)
		MM-NC	340	214	1.43 (200/140)	1.33 (122/92)	Affymetrix U133Plus2.0 (1)

\*BR - Breast Cancer; MM - Multiple Myeloma; NB - Neuroblastoma; pCR - Pathologic Complete Response; erpos - ER Positive; OS - Overall Survival; EFS - Event-free Survival; PC - Positive Control; NC - Negative Control.

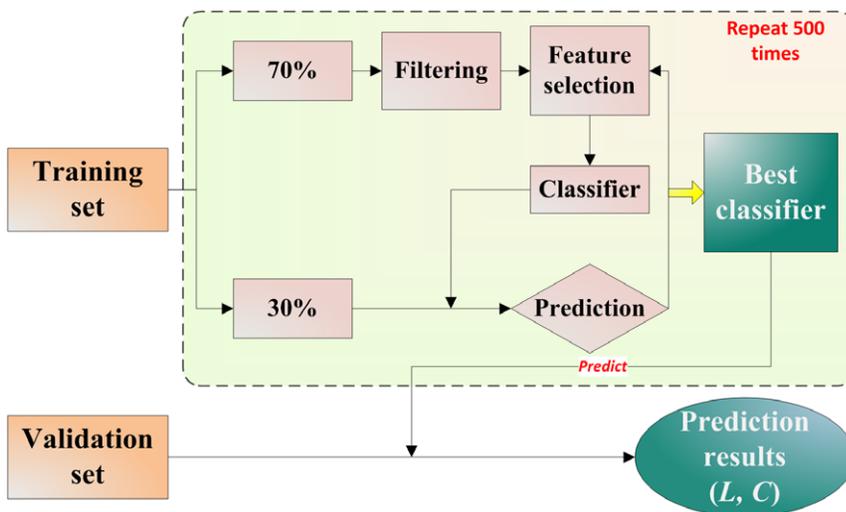
<sup>†</sup>Ratio of good to poor prognoses (i.e., good/poor prognoses).

doi:10.1371/journal.pone.0029534.t001

confidence. As depicted in **Figure 2**, a positive correlation is shown between the prediction accuracy and the confidence level for the six clinical and four control endpoints using the  $kNN$  classifier. Among the six clinical endpoints, the BR-erpos dataset showed the strongest correlation. For the BR-erpos, the average MCC value predictions with low confidence (LC) was only 0.19, while the average MCC value markedly increased to approximately 0.78 as the confidence level approached 1. Thus, compared to the overall MCC value (0.71) (**Table S1**), clinical confidence could successfully account for inter-individual variability in

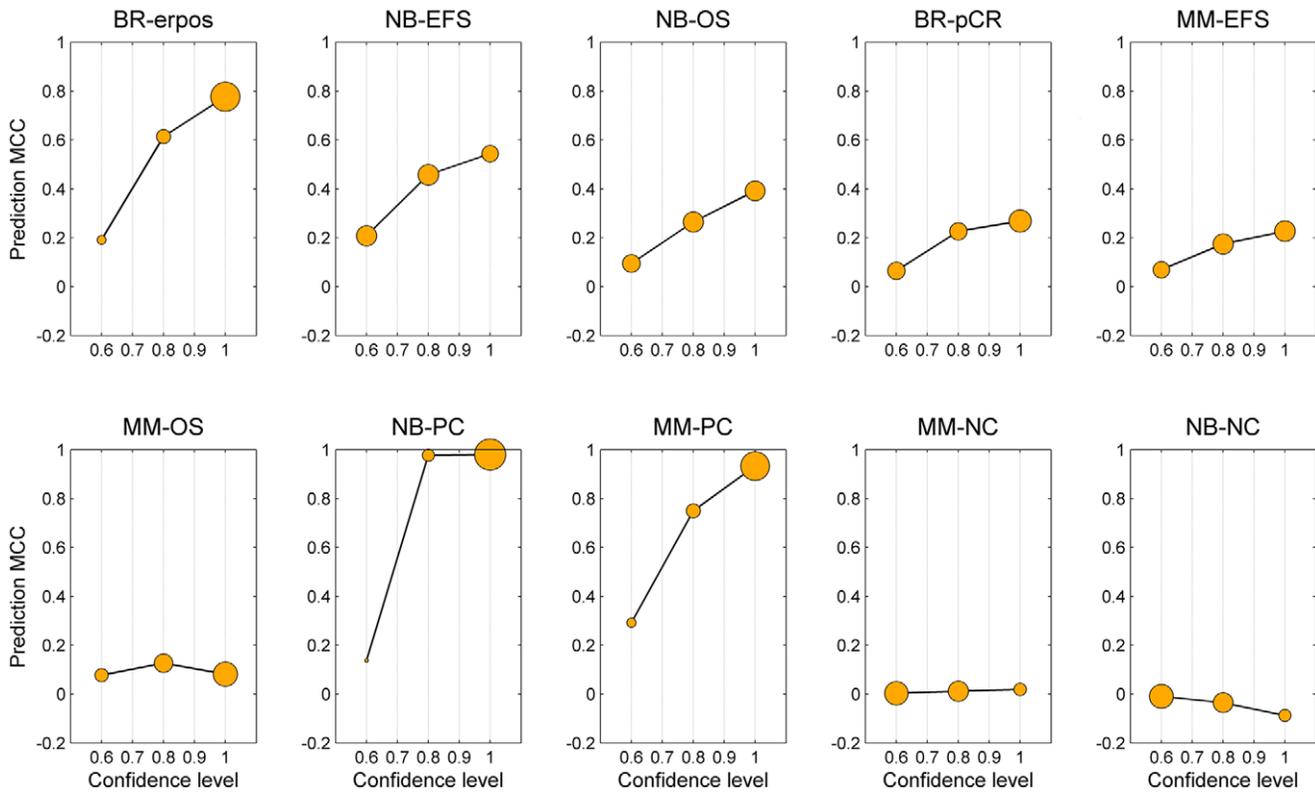
discriminating patients with lower or higher than average prediction accuracy.

It is clear that the intrinsic predictability by gene expression profiles varies for different endpoints, as evidenced by the gradual decrease in the steepness of model performance for six clinical endpoints over different confidence intervals (i.e., the slopes in **Figure 2**, data was shown in **Table S2**) and the number of samples distributed across different confidence regions (i.e., the marker size in **Figure 2**). As shown in **Figure 3**, a positive linear correlation was observed between the slope obtained from



**Figure 1.** Detailed workflow for correlation analysis of clinical confidence and model performance. Additional details are provided in **Methods S1**.

doi:10.1371/journal.pone.0029534.g001



**Figure 2. Prediction MCC as a function of clinical confidence for ten datasets using *kNN*.** The Circle radii are scaled to the percentage of total samples in the clinical confidence level. The confidence levels are '0.6', '0.8', and '1'. doi:10.1371/journal.pone.0029534.g002

**Figure 2** and the inherent predictability (quantified by Cohen’s *d* [28]) of the six clinical and four control endpoints. The predictable endpoints (e.g., BR-erpos, NB-EFS) tended to have a larger percentage of patients (represented as the marker size in **Figure 2**) in the high confidence regions with high prediction accuracy than the less predictable endpoints (e.g., MM-EFS, MM-OS). Detailed information about sample distribution in each confidence region was given in **Table S3**. These observations were further verified using a different pattern recognition method (i.e., *NC*) (**Figures S1 and S2**), and also a different sample splitting strategy (80/20 splitting, **Figures S4 and S5**).

The results demonstrate that a higher inherent predictability of an endpoint is related to a higher percentage of patients that fall into higher confidence levels when using microarray-based predictive models. As the correlation of a genomic signature with a clinical outcome is rarely perfect, the clinical confidence could be useful to separate the patients into different groups for whom specific treatment procedures can be developed.

### The relationship of clinical confidence with patient’s survival time

We also evaluated whether clinical confidence is predictive of the survival rate for the patients in the validation set. The patients were divided into two prognosis groups (i.e., good and poor prognosis) for both NB and MM datasets with endpoints OS (overall survival) and EFS (event-free survival), respectively (**Methods S1**). **Figure 4** presents the OS curves for patients with different clinical confidences for both prognosis groups. Patients with high clinical confidence exhibited an increased survival rate in the good prognosis group and a decreased survival

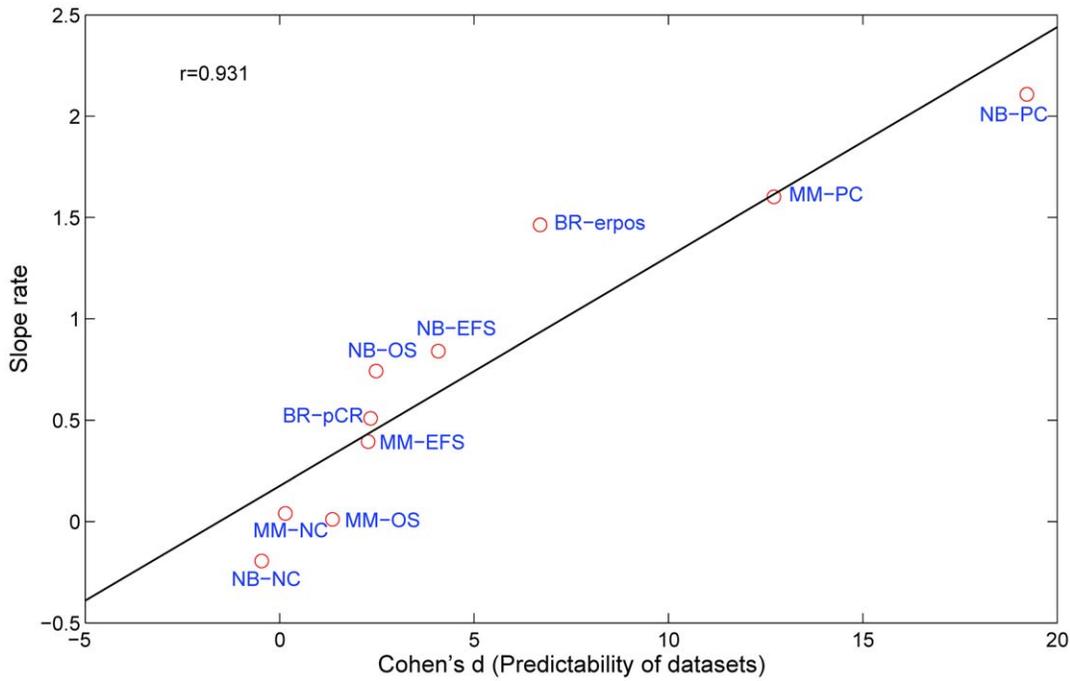
rate in the poor prognosis group, indicating that the clinical confidence enhanced the accuracy of prognosis derived from the predictive models. Taking MM-OS as an example, the survival rate is apparently higher for patients in the good prognosis group with high confidence (HC) compared to those with low (LC) (log-rank test *p* value < 0.01) and medium (MC) ones (log-rank test *p* value 0.13), especially for each day mark more than 1000 days (**Figure 4**). For patients with poor prognosis, more than 80% of those with low clinical confidence lived as long as 300 days, while approximately 30% of patients for those with high confidences survived at that time, respectively. Similar trends were also observed in the NB-OS dataset.

**Figure S3**, depicting the EFS curves for patients with different clinical confidences, demonstrates a similar trend as the OS curves presented in **Figure 4**. The positive correlation of clinical confidence with EFS rate is clearly shown in **Figures S3c and S3d** for the patients with good prognosis. However, the correlation is less significant for patients with poor prognosis (**Figures S3a and S3b**). Corresponding results for 80/20 splitting was shown in **Figures S6 and S7**, and conformed to those above-mentioned observations.

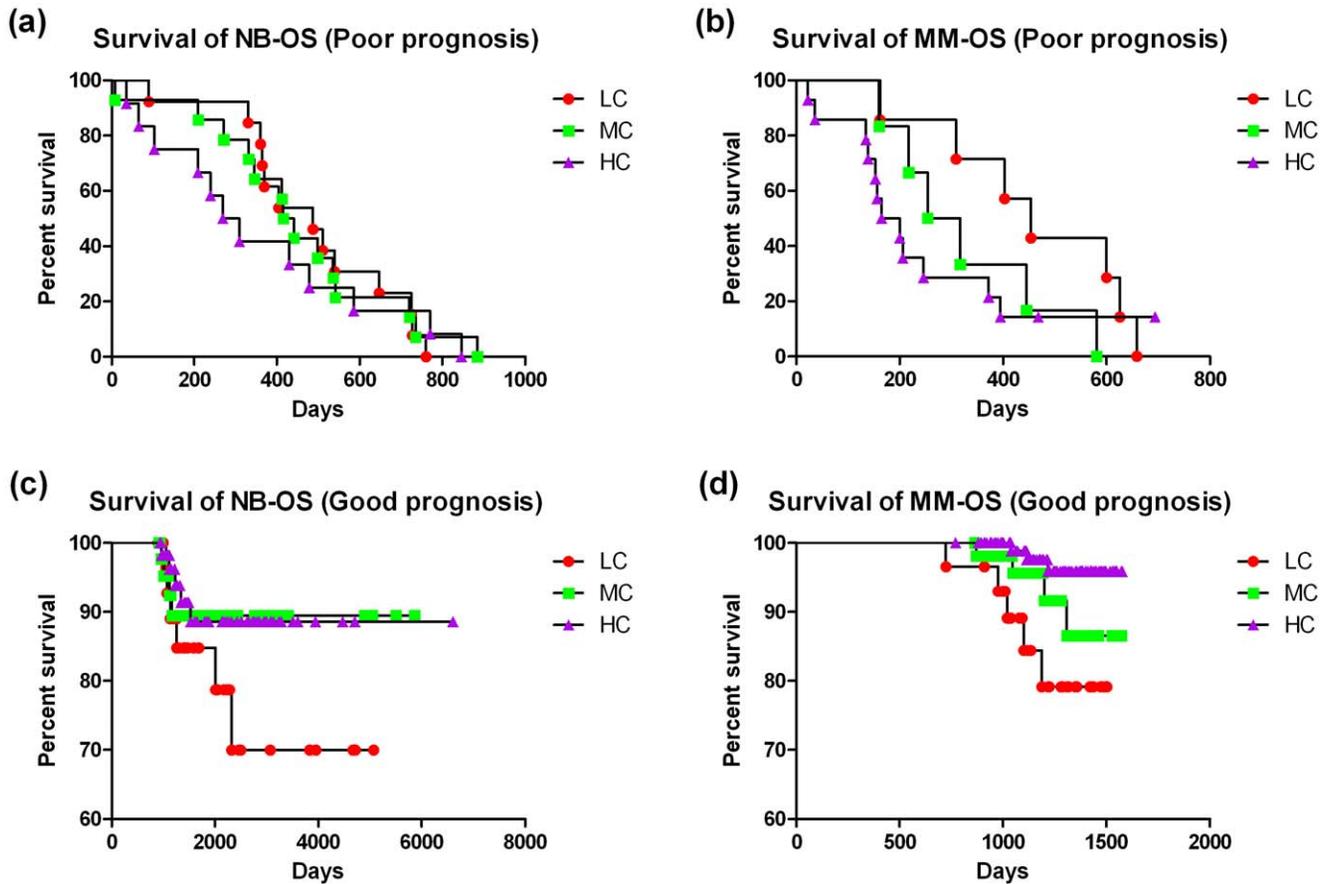
The results demonstrate that once the patients were grouped into either good or poor prognosis groups by the predictive models, the clinical confidence can further characterize the survival rate of individual patients in each prognosis group.

### Discussion

Several population-wide diagnostic/prognostic tests based on gene expression have been reported [4,6,7]. The population-based models provide only an average indication for the population with



**Figure 3. Correlation between slope rate and Cohen's *d* for the *kNN* classifier.** The slopes are obtained from regression analysis based on the linear portion of the confidence-MCC curve, while Cohen's *d* represents the inherent predictability of the dataset.  
doi:10.1371/journal.pone.0029534.g003



**Figure 4. Overall survival (OS) curves for patients with different clinical confidences using *kNN*,** where 'LC', 'MC', and 'HC' denote 'low confidence (0.6)', 'medium confidence (0.8)', and 'high confidence (1)', respectively.  
doi:10.1371/journal.pone.0029534.g004

corresponding average population accuracy. In this study, we demonstrated that clinical confidence is both capable of separating patients that can be more reliably predicted from those that are less accurately predicted, and predictive of the survival rate for the patients after they are grouped into different prognostic groups. Thus, ascribing a confidence level to prognosis with the clinical confidence metric will provide the clinician a more personalized, objective basis for decisions when using biomarkers derived from microarray data.

Specifically, we found that the clinical confidence provided a better estimation for the survival time when patients were classified into different prognosis categories based on both 70/30 and 80/20 sample assignments. For patients with good prognosis, higher clinical confidence was strongly correlated with longer survival time. Similarly, for patients with poor prognosis, the survival rate was significantly lower for those with high confidences than for the others. Taking endpoints MM-EFS and MM-OS as examples, despite the fact that they are rather difficult to be predicted, patients with high confidence display a significantly higher or lower survival rate when they are grouped in accordance with good or poor prognosis, respectively. Importantly, all patients in the high confidence group survived to 5000 days (**Figure S3c**), demonstrating that clinical confidence is an informative survival time prognosis tool.

An important aspect of this study is that two positive (NB-PC, MM-PC) and two negative control (NB-NC, MM-NC) datasets were involved, which is essential to assess the performance of the clinically relevant endpoints against the theoretical maximum and minimum performance provided by the controls. Specifically, the positive correlation between model performance and clinical confidence for the two positive control datasets shown in **Figure 2** confirmed the potential of clinical confidence to provide a measure of reliability for personalized medicine, while the negligible impact of clinical confidence in the two negative control datasets further limited the possibility of obtaining false positives. Thus, the inclusion of positive and negative control datasets in such an analysis would be of great help to ensure the reliability of the results.

It remains enigmatic why some of the endpoints were more difficult than others to predict. **Figure 2** and **Figure 3** compare predictability across the three datasets and corresponding six endpoints. Readily predictable endpoints have a high percentage of patients who fall into the high confidence region. For example, the percentage of patients that showed high clinical confidence (74.70%) for the BR-erpos endpoint is much higher than that of the MM-EFS endpoint (37.51%) (**Figure 2**), which may indicate that the BR-erpos endpoint contains a stronger gene expression signal than MM-EFS does. Additionally, the predictability of the dataset (Cohen's  $d$ ) is directly related with the correlation coefficient between the confidence level and MCC prediction performance (**Figure 3**).

The ability to quantify clinical confidences may greatly enhance clinical decision-making processes based on microarray-based prediction models, especially for personalized treatment options. For example, the models presented here could test for potential treatment response with the high confidence and low confidence predictions being used in different ways. Patients with good prognosis and high confidences are candidates for applying routine protocols to avoid over-treatment, while rigorous strategies should be selected for those with poor prognosis and high confidences to prolong survival time as long as possible. However, for patients in the low confidence regions, additional evaluation using alternative methods should be considered.

It is important to note that the strategy proposed in this study emphasizing the shift from population-based to personalized cancer prognosis does not negate the importance of population-based prediction, but rather builds upon its success. If the performance of a predictive model is not informative, such as seen in the two negative controls (i.e., MM-NC and NB-NC), the clinical confidence will not be predictive. Thus, model validation methods that include cross-validation and independent external validation are still essential to ensure the validity of microarray-based predictive models. However, since the population-based prediction does not provide an accurate assessment for each patient within the population, clinical confidence offers a means to measure reliability for individual predictions based on the population-based prediction.

The benefits of personalized medicine in health care are well recognized [1]. It allows both the patient and the physician to be more aware of the benefits and risks of possible treatments and potential outcomes affected by genetic make-up or other environmental influences. Thus, informed, tailored, and health-related decisions can be made for each person [29]. Combining microarray technology capable of profiling the expression levels of hundreds of thousands of genes with pattern recognition techniques has been an important step toward individualized decision-making [30]. We presented examples applying confidence assessment to cancer prognosis and survival time prediction for models developed from microarray data. However, the approach can be generalized to biomarkers and models built based on data from other high throughput platforms. Moreover, the concept is generally applicable for all supervised learning classification methodologies that can define a clinical confidence.

## Supporting Information

**Figure S1 Prediction MCC as a function of clinical confidence for ten datasets using NC.** Circle radii are scaled to the percentage of total samples in the clinical confidence level. The confidence levels are '0.5–0.6', '0.6–0.8' and '0.8–1', respectively.  
(TIF)

**Figure S2 Correlation between slope rate and Cohen's  $d$  for the NC classifier.** The slopes are obtained from regression analysis based on the linear portion of the confidence-MCC curve, while Cohen's  $d$  represents the inherent predictability of the dataset.  
(TIF)

**Figure S3 Event-free survival (EFS) curves for patients with different clinical confidences using  $kNN$  where 'LC', 'MC', and 'HC' denote 'low confidence (0.6)', 'medium confidence (0.8)', and 'high confidence (1)', respectively.**  
(TIF)

**Figure S4 Prediction MCC as a function of clinical confidence for ten datasets using 80/20 splitting and  $kNN$ .** The Circle radii are scaled to the percentage of total samples in the clinical confidence level. The confidence levels are '0.6', '0.8', and '1'.  
(TIF)

**Figure S5 Correlation between slope rate and Cohen's  $d$  for the  $kNN$  classifier based on 80/20 sample assignment.** The slopes are obtained from regression analysis based on the linear portion of the confidence-MCC curve, while Cohen's  $d$  represents the inherent predictability of the dataset.  
(TIF)

**Figure S6 Overall survival (OS) curves for patients with different clinical confidences using 80/20 splitting and  $kNN$ , where ‘LC’, ‘MC’, and ‘HC’ denote ‘low confidence (0.6)’, ‘medium confidence (0.8)’, and ‘high confidence (1)’, respectively.**  
(TIF)

**Figure S7 Event-free survival (EFS) curves for patients with different clinical confidences using 80/20 splitting and  $kNN$ , where ‘LC’, ‘MC’, and ‘HC’ denote ‘low confidence (0.6)’, ‘medium confidence (0.8)’, and ‘high confidence (1)’, respectively.**  
(TIF)

**Table S1** MCC performance for training and validation sets.  
(DOCX)

**Table S2** Slope and Cohen’s  $d$  for each dataset.  
(DOCX)

## Reference

- Loscalzo J, Kohane I, Barabasi AL (2007) Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol Syst Biol* 3: 124.
- Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. *Cell* 134: 714–717.
- Abrahams E (2009) Personalized Medicine Realizing Its Promise. *Genet Eng Biotechnol News* 29(15). Available: <http://www.genengnews.com/gen-articles/personalized-medicine-realizing-its-promise/3025/>. Accessed: 2012 Jan 2.
- van't Veer LJ, Dai HY, van de Vijver MJ, He YDD, Hart AAM, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530–536.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999–2009.
- Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, et al. (2004) Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 22: 2284–2293.
- Iwao-Koizumi K, Matoba R, Ueno N, Kim SJ, Ando A, et al. (2005) Prediction of docetaxel response in human breast cancer by gene expression profiling. *J Clin Oncol* 23: 422–431.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98: 10869–10874.
- Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, et al. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 100: 10393–10398.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100: 8418–8423.
- Bertucci F, Finetti P, Cervera N, Maraninchi D, Viens P, et al. (2006) Gene expression profiling and clinical outcome in breast cancer. *OMICS* 10: 429–443.
- Balleine RL, Webster LR, Davis S, Salisbury EL, Palazzo JP, et al. (2008) Molecular Grading of Ductal Carcinoma In situ of the Breast. *Clin Cancer Res* 14: 8244–8252.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. *Nature* 406: 747–752.
- Slodkowska EA, Ross JS (2009) MammaPrint (TM) 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* 9: 417–422.
- Roepman P, Wessels LFA, Kettelarij N, Kemmeren P, Miles AJ, et al. (2005) An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet* 37: 182–186.
- Williams PD, Cheon S, Havaleshko DM, Jeong H, Cheng F, et al. (2009) Concordant Gene Expression Signatures Predict Clinical Outcomes of Cancer Patients Undergoing Systemic Therapy. *Cancer Res* 69: 8302–8309.
- Shao L, Wu LH, Fang H, Tong WD, Fan XH (2010) Does applicability domain exist in microarray-based genomic research? *PLoS ONE* 5: e11055.
- Simon R, Radmacher MD, Dobbin K, McShane LM (2003) Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 95: 14–18.
- The MicroArray Quality Control Consortium (2010) The MAQC-II Project: A comprehensive study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* 28: 827–838.
- Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, et al. (2006) Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 24: 4236–4244.
- Shaughnessy JD, Zhan FH, Burington BE, Huang YS, Colla S, et al. (2007) A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* 109: 2276–2284.
- Oberthuer A, Berthold F, Warnat P, Hero B, Kahlert Y, et al. (2006) Customized oligonucleotide microarray gene expression based classification of neuroblastoma patients outperforms current clinical risk stratification. *J Clin Oncol* 24: 5070–5078.
- Theodoridis S, Koutroumbas K (2006) *Pattern Recognition*. San DiegoCA: Elsevier.
- Matthews BW (1975) Comparison of predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 405: 442–451.
- Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16: 412–424.
- Radmacher MD, McShane LM, Simon R (2002) A paradigm for class prediction using gene expression profiles. *J Comput Biol* 9: 505–511.
- Fan XH, Shi LM, Fang H, Cheng YY, Perkins R, et al. (2010) DNA microarrays are predictive of cancer prognosis: A re-evaluation. *Clin Cancer Res* 16: 629–636.
- Cohen J (1992) A power primer. *Psychol Bull* 112: 155–159.
- Ely S (2009) Personalized medicine: individualized care of cancer patients. *Trans Res* 154: 303–308.
- Cantor MN (2009) Enabling personalized medicine through the use of healthcare information technology. *Per Med* 6: 589–594.

## Acknowledgments

The authors would like to thank the MAQC Data providers for sharing their data and information to the MAQC Consortium. The views presented in this article do not necessarily reflect those of the U. S. Food and Drug Administration.

## Author Contributions

Conceived and designed the experiments: XF HF WT YC. Analyzed the data: LS (Li Shao) XF LW HX. Wrote the paper: LS (Li Shao) XF NC HF DD LS (Leming Shi) WT YC.