A Robust Predictor of Response to Preoperative Chemotherapy for Breast Cancer

Lin Zhang, Chunxiang Hao, and Zheng Guo

Abstract—Breast cancer patients with pathological complete response (pCR) to taxane and anthracycline containing preoperative chemotherapy tend to have excellent distant-free and overall survival. However, most of current pCR predictors for guiding chemotherapy tend to perform poor in inter-laboratory validation, most likely due to microarray experimental batch effects and insufficient training samples for feature selection. To tackle this difficulty, by merging three separate datasets of patients treated with paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide, we extracted gene pairs with consistent relative expression ordering in patients not achieving pCR and with relative expression reversal in patients achieving pCR. Then, based on combination of these pairs, a pCR predictor was built. The performance of this predictor achieved a sensitivity of 93% and specificity of 62% in an independent dataset, much better than the performances of three previously proposed predictors. In conclusion, the rank-based pCR predictor derived from a large cohort of samples can accurately and robustly predict patients with high probability of achieving pCR after chemotherapy.

Index Terms—Preoperative chemotherapy, pathological complete response, gene expression profile, breast cancer.

I. INTRODUCTION

Preoperative chemotherapy is widely used in the treatment of patients with breast cancer and patients who achieve a PCR after chemotherapy have excellent disease-free and overall survival [1]. For pCR in response to chemotherapy, especially for the most effective taxane and anthracycline containing regimens, although many drug targets-based predictors [2], [3] and discovery-based predictors using microarray gene expression profiles have been developed [4], [5], their performances often degrade greatly or fail to be validated in inter-laboratory datasets [6]-[8]. This irreproducibility could be due to that microarray measurement values of gene expression tend to be affected by experimental batch effects including differences in laboratory conditions, reagent lots and personnels [9].

More importantly, this irreproducibility could result from the conflict between enormous numbers of genes in microarray data and the relatively small numbers of samples used in the training set of each study [10]. Therefore, it is preferable to use a large number of samples to develop a robust predictor for pCR. Rank-based classifiers, that are built based on relative expression ordering of gene pairs, are rather robust to experimental batch effects on samples [11] and thus can be trained by using large samples from different data sources [10]. However, without refined feature combination, the current rank-based methods including the most widely used Top Scoring Pair (TSP) [11] and K disjoint Top Scoring Pair (KTSP) [12] method, may be unable to efficiently treat the molecular heterogeneity of breast cancers in responsiveness to treatments [13].

In this paper, we analyzed four microarray datasets of breast cancer patients treated with a standard taxane and anthracycline containing (paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide) chemotherapy regimen. Firstly, we used three of the four dataset to extract a combination of gene pairs each showing a reserve relative expression ordering between patients with pCR and residual disease (RD). Then, we generated a pCR predictor based on this combination of gene pairs. The robustness and accuracy of the pCR predictor was validated by an independent data from a different clinical trail and by comparison with three other previously proposed predictors.

II. METERIALS AND METHODS

A. Microarray Datases

The four gene expression microarray datasets using the Affymetrix GeneChip arrays, referred to as MDA133, MAQC100, MDA91 and ISPY79, were collected from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). As described in Table I, the datasets of MDA133, MAQC100 and MDA91 were used for generating the pCR predictor, and the independent dataset ISPY79 was used to validate this predictor.

B. The Rank-Based Classifier for pCR

For a single pair of genes (i, j), we indicated their relative expression as $R_i > R_j$ (or $R_i < R_j$) in an RD class if the relative expression was consistent in the majority of RD samples. Accordingly, the inverse relative expression was used to characterize the pCR class. For a combination of pairs, we classified a patient as pCR if more than one pair had a 'relative expression reversal', otherwise that patient was classified as RD. We used these rules to calculate the sensitivity and specificity scores for the pCR and RD classes, respectively.

TABLE I: DATASETS ANALYZED FOR PCR PREDICTION

Title	Usage	GEO acc.	Sample size
MDA133	Training	GSE20194	34:99
MAQC100	Training	GSE20194	15:85
MDA91	Training	GSE20271	19:72
ISPY79	validation	GSE25055	14:65

The harmonic score calculated as the harmonic mean of the sensitivity and specificity was used to evaluate a single pair or a combination of pairs. The optimal combination of gene pairs was selected using a greedy algorithm that maximize the harmonic score in the training dataset.

C. Evaluation of Response Predictors

Three-fold cross validation was performed by repeated iteration (n=1000) of stratified random sampling to estimate the prediction performance of the different classifiers in the training data and to facilitate selection of a single classifier for independent validation. Stratification was performed to insure that the relative proportion of outcomes sampled in both cross-validation training and test sets was similar to the original proportions for the full training data. We performed complete cross validation including gene pairs selection and combination in each iteration to avoid selection bias [14].

Finally, the performances of the predictors in the cross validation test sets or independent validation dataset were presented as sensitivity and specificity scores or the area under the curve (AUC) score. The corresponding 95% confidence intervals (95% CIs) for the sensitivity and specificity scores were calculated based on the exact binomial confidence intervals estimate.

III. RESULTS

A. Development of a pCR Predictor

Firstly, we estimate the prediction performances of different rank-based classifiers for pCR across three separate datasets by performing 1000 iterations of three-fold cross validations (see Methods). As shown in Fig. 1, the median value of sensitivity and specificity scores were 75% and 72%, the inter-quartile range (IQR) of sensitivity ranged from 70% to 83% and the IQR of specificity ranged from 68% to 76%, respectively. These results showed that our rank-based classifier for pCR can performance well on a dataset from different sources, thus it's reasonable for us to merge these three separate datasets together to generate a robust pCR predictor.

TABLE II: THE PERFORMANCES OF FOUR PCR PREDICTORS

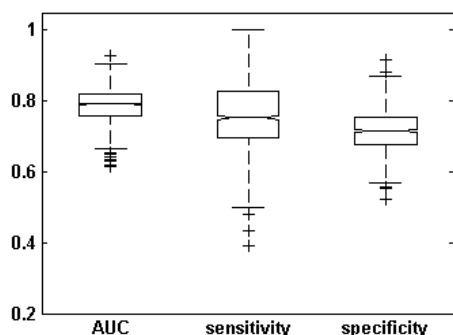| Predictor	Sensitivity		Specificity	
	Estimate	95% CI	Estimate	95% CI
CTSP-6	0.93	0.66-1	0.62	0.49-0.73
DLDA-30	0.71	0.42-0.92	0.63	0.50-0.75
RNAi-6	0.93	0.66-1	0.40	0.28-0.53
KTSP-4	0.50	0.23-0.77	0.66	0.53-0.77

Fig. 1. The prediction performance scores for different classifiers generated across 1000 repeats of three-fold cross validation.

Then, from the merged dataset containing 324 samples, we extracted 225 pairs with stable relative expression orderings among above 80% of patients in the RD class and a 'relative expression reversal' among over half of patients in the pCR class. Finally, from top 100 of these pairs ranked by the harmonic scores, we generated a CTSP-6 predictor using a combination of six top-scoring gene pairs (CCND1/LMO4), (PADI2/MYO5C), (BRPF1/KIAA0831), (CENPA/ZNF263), (GALNT1/DYNLT3) and (CENPA/ZNF552) by the method described in the Methods. The sensitivity was as high as 96% for the patients with pCR, and the specificity was 75% for the patients with RD in the training dataset, respectively.

B. Evaluation of pCR Predictors in Independent Dataset

Using an independent ISPY79 datasets from the I-SPY trial for validation, we compared our predictor with three other predictors, including a thirty-gene predictor generated by dimensional linear discriminant analysis (DLDA-30) [4], a six-gene relative predictor generated by RNAi screening (RNAi-6) [15] and a four-gene pairs predictor obtained by the KTSP algorithm (KTSP-4). Our predictor had a sensitivity of 93% and a specificity of 62% in the ISPY79 dataset (Table II).

In contrast, the DLDA-30 predictor generated from the first group of 82 samples of MDA133 showed a high sensitivity of 92% and a specificity of 71% in the second group of 51 samples [4]. However, its sensitivity and specificity scores decreased to 71% and 63% in the ISPY79 dataset (Table II). Similarly, when using fixed sensitivity score the same as our predictor, the specificity score of the RNAi-6 predictor were as low as 40% in the ISPY79 dataset (Table II). Finally, the most applied ranked-based KTSP method obtained an optimal K at four gene pairs from the 100 iterations of three-fold cross validation and achieved a sensitivity of 50% and a specificity of 66% in the training set. Low sensitivity and specificity score of the KTSP-4 predictor could be caused by the inefficiency of the feature selection given the molecular heterogeneity of breast tumors.

IV. DISCUSSION

The predictor developed in this work performed more accurately and robustly than three other predictors previously proposed. A large cohort of samples merged from different datasets was used to extract the gene pairs with robust gene expression ordering in the majority of RD samples. As shown in the large dataset, each pair showed 'relative expression reversal' in only a certain proportion of pCR patients, which reflected that the molecular heterogeneity in breast tumors and in their responsiveness to treatments is substantial. Thus, as indicated in the results, simply using K disjoint top scoring pairs without refined feature combination failed to predict pCR accurately in the independent dataset. Moreover, the six-gene marker proposed in [15], using the relative expression value of a mitosis metagene (i.e., a combination of functionally related genes) comparing to a ceramide metagene predefined by an RNA inference (RNAi) screen, was also lack of sufficient predictivity power in the independent dataset, largely due to that the magnitude of the relative expression measurements

of two metagenes could also be affected by experimental batch effects in inter-laboratory validation. The rank-based response predictor of pCR was also more accurate than the predictors generated by the DLDA method, due to that the rank-based method was rather robust to inter-array batch effects and that our predictor was trained from a large cohort of samples.

Although patients who achieve a pCR after neoadjuvant chemotherapy tend to have excellent distant-free and overall survival, the pCR rate with current breast neoadjuvant chemotherapy treatment is still quite low whereas a proportion of patients with minimal residual disease (RD) could also have good prognosis. Moreover, patients intrinsically at different risk levels have different response to chemotherapy and prognosis after chemotherapy. We would focus on more detailed prognostic value of preoperative chemotherapy in the future work.

REFERENCES

[1] M. Kaufmann *et al.*, "International expert panel on the use of primary (preoperative) systemic treatment of operable breast cancer: review and recommendations," *J Clin Oncol,* vol. 21, no. 13, pp. 2600-8, July 1, 2003.

[2] K. I. Pritchard *et al.*, "HER-2 and topoisomerase II as predictors of response to chemotherapy," *J Clin Oncol,* vol. 26, no. 5, pp. 736-44, Feb. 10, 2008.

[3] R. Rouzier *et al.*, "Microtubule-associated protein tau: a marker of paclitaxel sensitivity in breast cancer," in *Proc. Natl Acad Sci U S A,* vol. 102, no. 23, pp. 8315-20, Jun 7, 2005.

[4] K. R. Hess *et al.*, "Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer," *J Clin Oncol,* vol. 24, no. 26, pp. 4236-44, Sep. 10, 2006.

[5] C. Liedtke *et al.*, "Genomic grade index is associated with response to chemotherapy in patients with breast cancer," *J Clin Oncol,* vol. 27, no. 19, pp. 3185-91, July 1, 2009.

[6] F. J. Esteva and G. N. Hortobagyi, "Topoisomerase II{alpha} amplification and anthracycline-based chemotherapy: the jury is still out," *J Clin Oncol,* vol. 27, no. 21, pp. 3416-7, July 20, 2009.

[7] L. Pusztai *et al.*, "Evaluation of microtubule-associated protein-Tau expression as a prognostic and predictive marker in the NSABP-B 28 randomized clinical trial," *J Clin Oncol,* vol. 27, no. 26, pp. 4287-92, Sep. 10, 2009.

[8] A. Tabchy *et al.*, "Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer," *Clin Cancer Res,* vol. 16, no. 21, pp. 5351-61, Nov. 1, 2010.

[9] J. T. Leek *et al.*, "Tackling the widespread and critical impact of batch effects in high-throughput data," *Nat Rev Genet,* vol. 11, no. 10, pp. 733-9, Oct. 2010.

[10] L. Xu, A. C. Tan, R. L. Winslow, and D. Geman, "Merging microarray data from separate breast cancer studies provides a robust prognostic test," *BMC Bioinformatics,* vol. 9, pp. 125, 2008.

[11] D. Geman, C. d'Avignon, D. Q. Naiman, and R. L. Winslow, "Classifying gene expression profiles from pairwise mRNA comparisons," *Stat Appl Genet Mol Biol,* vol. 3, pp. Article19, 2004.

[12] A. C. Tan, D. Q. Naiman, L. Xu, R. L. Winslow, and D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics,* vol. 21, no. 20, pp. 3896-904, Oct 15, 2005.

[13] C. M. Perou *et al.*, "Molecular portraits of human breast tumours," *Nature,* vol. 406, no. 6797, pp. 747-52, Aug. 17, 2000.

[14] R. Simon, "Roadmap for developing and validating therapeutically relevant genomic classifiers," *J Clin Oncol,* vol. 23, no. 29, pp. 7332-41, Oct. 10, 2005.

[15] N. Juul *et al.*, "Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: a retrospective analysis of five clinical trials," *Lancet Oncol,* vol. 11, no. 4, pp. 358-65, Apr. 2010.

Lin Zhang was born on August 17, 1984, in Zhengjiang, China. She received the B.S. degree in Computer Science and Technology from Zhejiang Normal University and M.S. degree in Biophysics from University of Electronic Science and Technology of China (UESTC). She is currently a Ph.D. candidate in studying in Bioinformatics at UESTC. She is interested in individual therapy analysis in cancer genome.

Chunxiang Hao was born on January 18, 1989, in HengYang Province. She graduated from Central China Normal University and gained a bachelor's degree in Mathematics and Applied Mathematics. Now she is a postgraduate student in Biophysics of University of Electronic Science and Technology of China (UESTC). She is focused on expression and methylation analysis in cancer genome.

Zheng Guo was born on October 9, 1963. He obtained the PHD degree in computer science from Harbin Institute of Technology and is currently a Professor of Bioinformatics at Fujian Medical University. He has undertaken multiple National Natural Science Foundation projects. His research is focused on analysing microarray data and complex diseases at the functional module level.