pISSN 1598-2998, eISSN 2005-9256

Cancer Res Treat. 2017;49(1):116-128

https://doi.org/10.4143/crt.2016.085

Original Article

Open Access

Use of a Combined Gene Expression Profile in Implementing a Drug Sensitivity Predictive Model for Breast Cancer

Xianglan Zhang, MD, PhD[1]
In-Ho Cha, DDS, PhD[2]
Ki-Yeol Kim, PhD[3]

[1]*Department of Pathology, Yanbian University Medical College, Yanji City, China,* [2]*Department of Oral and Maxillofacial Surgery, College of Dentistry, Yonsei University, Seoul,*
[3]*BK21 PLUS Project, Yonsei University College of Dentistry, Yonsei University, Seoul, Korea*

Correspondence: Ki-Yeol Kim, PhD
BK21 PLUS Project, Yonsei University College of Dentistry, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: 82-2-2228-3039
Fax: 82-2-392-2959
E-mail: kky1004@yuhs.ac

Received February 24, 2016
Accepted May 2, 2016
Published Online May 18, 2016

Purpose
Chemotherapy targets all rapidly growing cells, not only cancer cells, and thus is often associated with unpleasant side effects. Therefore, examination of the chemosensitivity based on genotypes is needed in order to reduce the side effects.

Materials and Methods
Various computational approaches have been proposed for predicting chemosensitivity based on gene expression profiles. A linear regression model can be used to predict the response of cancer cells to chemotherapeutic drugs, based on genomic features of the cells, and appropriate sample size for this method depends on the number of predictors. We used principal component analysis and identified a combined gene expression profile to reduce the number of predictors

Results
The coefficients of determinanation (R^2) of prediction models with combined gene expression and several independent gene expressions were similar. Corresponding F values, which represent model significances were improved by use of a combined gene expression profile, indicating that the use of a combined gene expression profile is helpful in predicting drug sensitivity. Even better, a prediction model can be used even with small samples because of the reduced number of predictors.

Conclusion
Combined gene expression analysis is expected to contribute to more personalized management of breast cancer cases by enabling more effective targeting of existing therapies. This procedure for identifying a cell-type-specific gene expression profile can be extended to other chemotherapeutic treatments and many other heterogeneous cancer types.

Key words
Gene expression, Drug sensitivity, Combined predictor

Introduction

Breast cancer, the most common cancer in women, is a major cause of female mortality. Approximately 232,340 new cases of invasive breast cancer and 39,620 breast cancer deaths were reported in the United States in 2013 [1-3]. In Korea, breast cancer is the second most common cancer in women, accounting for 19.8% of female cancer cases. Approximately 13,400 breast cancer cases were reported in 2009 and the rate of new diagnosis has been rising steadily [4,5].

Treatment of breast cancer usually involves one or more drugs, surgery, and sometimes radiation. Positive clinical responses to anticancer therapies are often restricted to a subset of patients. Chemotherapy, a systematic drug regimen intended to stop cancer cells from dividing and growing, targets all rapidly growing cells, not only cancer cells, and is therefore associated with unpleasant side effects [6]. Mutated cancer genes contain biomarkers, making them good candi-

Table 1. Summary of the datasets used in this work

Data set	No. of samples	No. of probes
GSE51086 (Dataset was published on Nov 27, 2013)	14 Lu, 4 BaA, 8 BaB, 3 unknown (29 untreated breast cancer cell lines vs. pool of all 29 cell lines)	45,220 probes (Agilent-014850 Whole Human Genome Microarray 4x44K), 19,722 gene symbol
Garnett et al. [7]	43 Breast cancer cell lines (total 608 cell lines, 111 drugs)	

dates for drug targeting [7]. Therefore, study of the relationship between gene expression and drug sensitivity is required in order to administer the right drug to the right patient and reduce the side effects of cancer treatment.

Scherf et al. [8] examined gene-gene, gene-drug, and drug-drug interactions by combining gene expression data and drug sensitivity data. The first limitation of their research was that laboratory cell lines differ from tumor cells. Later, Kao et al. [9] reported that gene expression patterns were similar in their cell lines and primary breast cancer tumors. Therefore, many studies for predicting chemosensitivity with gene expression have used cell lines.

Several studies have demonstrated that genomic biomarkers can be used in prediction of chemotherapeutic responses in human cancer patients [7,10-15]. Many such studies used statistical methodologies or machine learning methods [10-12].

The aim of this study was to predict drug sensitivity using gene expression analysis and a linear regression model. The necessary sample size for linear regression depends on the number of predictive variables [16]. Sample size must increase as the number of predictors increases.

In this study, we suggest a method for implementing a chemosensitivity prediction model with a reduced number of predictors, which will be useful in cases with small sample sizes. We identify a combined gene expression pattern that requires fewer predictors than analysis involving several individual genes. Doxorubicin, an anthracycline widely used in the treatment of breast cancer, was the targeted chemotherapeutic agent [17].

Materials and Methods

1. Data preparation

Two publicly available datasets of gene expressions and drug sensitivities were used in this study. The first dataset was compiled by Felding-Habermann et al. [18] for examination of gene expression in cell lines; it is accessible from a public microarray database (gene expression omnibus [GEO], GSE51086). This dataset consists of 29 cell lines, including 14 luminal, four basal A, and eight basal B cell lines (the other three cell lines are of unknown type), and includes 45,220 probes, and it was summarized by 19,722 gene symbols for this study.

The other dataset, published by Garnett et al. [7], consists of 608 cell lines and reports the IC_{50} scores of 111 drugs [10] (inhibitory concentration, IC_{50}, represents the concentration of a drug required for 50% growth inhibition *in vitro*). Of those, we looked at 43 breast cancer cell lines and 98 drugs, because those missing more than 50% of data were excluded. The datasets are summarized in Table 1.

2. Statistical methods

IC_{50} scores were standardized before performing statistical analysis. Associations between gene expression and cell line, drug sensitivity and cell line, and drug sensitivity and gene expression in the combined dataset were examined. For this, analysis of variance (ANOVA) was performed for identification of significantly expressed genes among different types of breast cancer cells. Correlation analysis was performed to examine the relation of drug sensitivity to gene expression. Linear regression was used for identification of a drug sensitivity prediction model with gene expression as the predictor.

3. DG-matrix analysis (the association of drug sensitivity and gene expression)

The degree of similarity between the G-matrix (association of gene expression and cell line) and D-matrix (association of drug sensitivity and cell line) was calculated using the Pearson correlation coefficient, r, calculated as follows:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

, where x_i denotes the gene expression of i^{th} cell, y_i denotes

the IC_{50} of i^{ih} cell to drug y, $\bar{x}$ represents the average expression of gene x, and $\bar{y}$ represents the mean chemosensitivity of drug y.

Principal component analysis (PCA) was used to reduce the number of independent variables (the number of genes) in the prediction model for drug sensitivity [19]. PCA is a simple nonparametric statistical method that reduces data dimensionality for conversion of correlated variables into uncorrelated variables, which are termed *principal components*. Each principal component is represented in the form of linear combinations of original variables. Therefore, a value for each cell line can be calculated according to the following formula, which we call combined gene expression:

$$\text{Combined gene expression}=C_1g_1+C_2g_2+C_3g_3+\ldots+C_ng_n$$

, where g_1, g_2, g_3 … g_n are gene expressions, and C_1, C_2, C_3 … C_n are weights of each gene expression.

In examining the association of gene expression and drug sensitivity, cell lines which did not include gene expressions or IC_{50} scores were excluded. The usage of datasets and a schematic diagram of the study design are shown in Fig. 1.

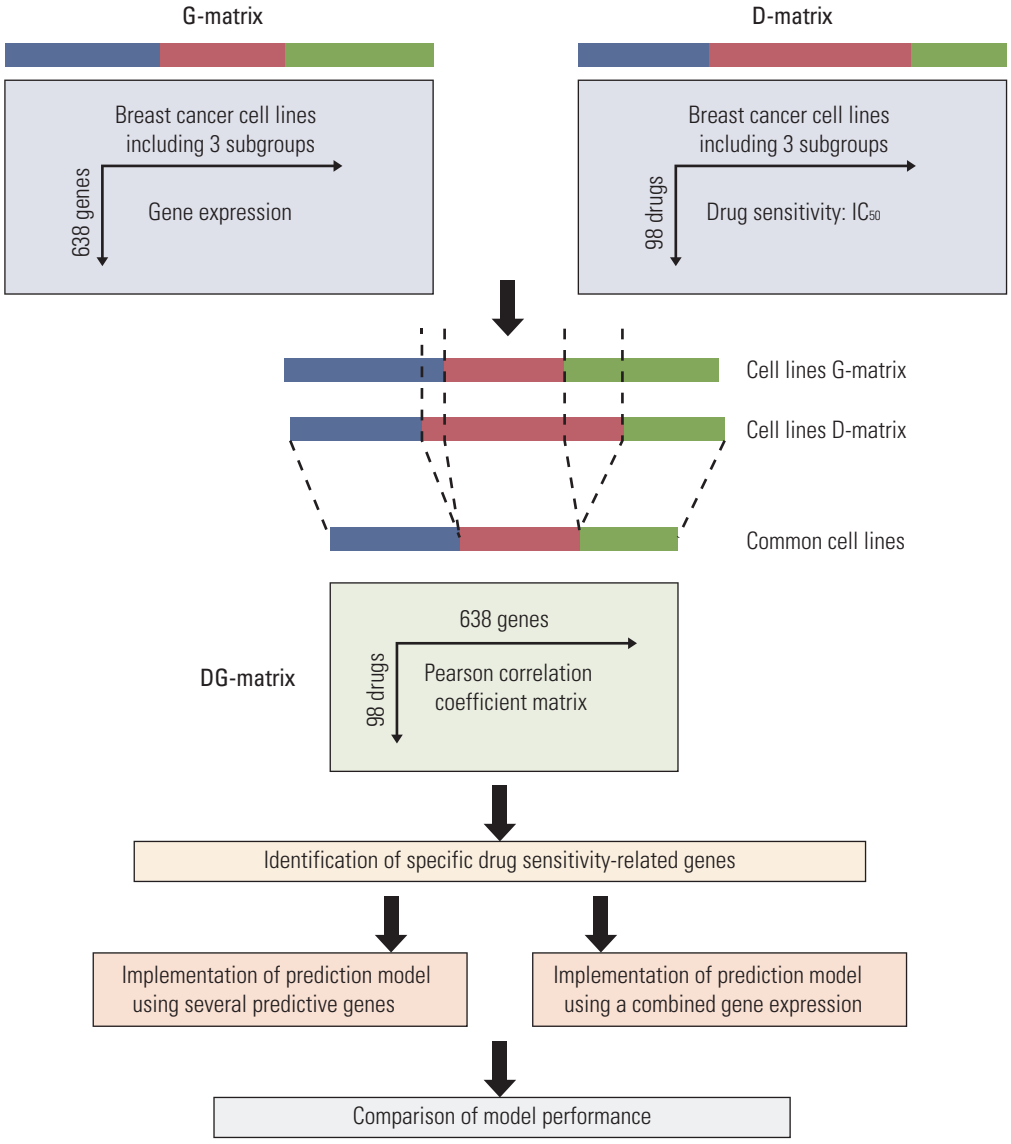

Fig. 1. The summary of the study plan using two published datasets. G-matrix includes 638 genes, which showed significantly different expression among subtypes of breast cancer cell lines.

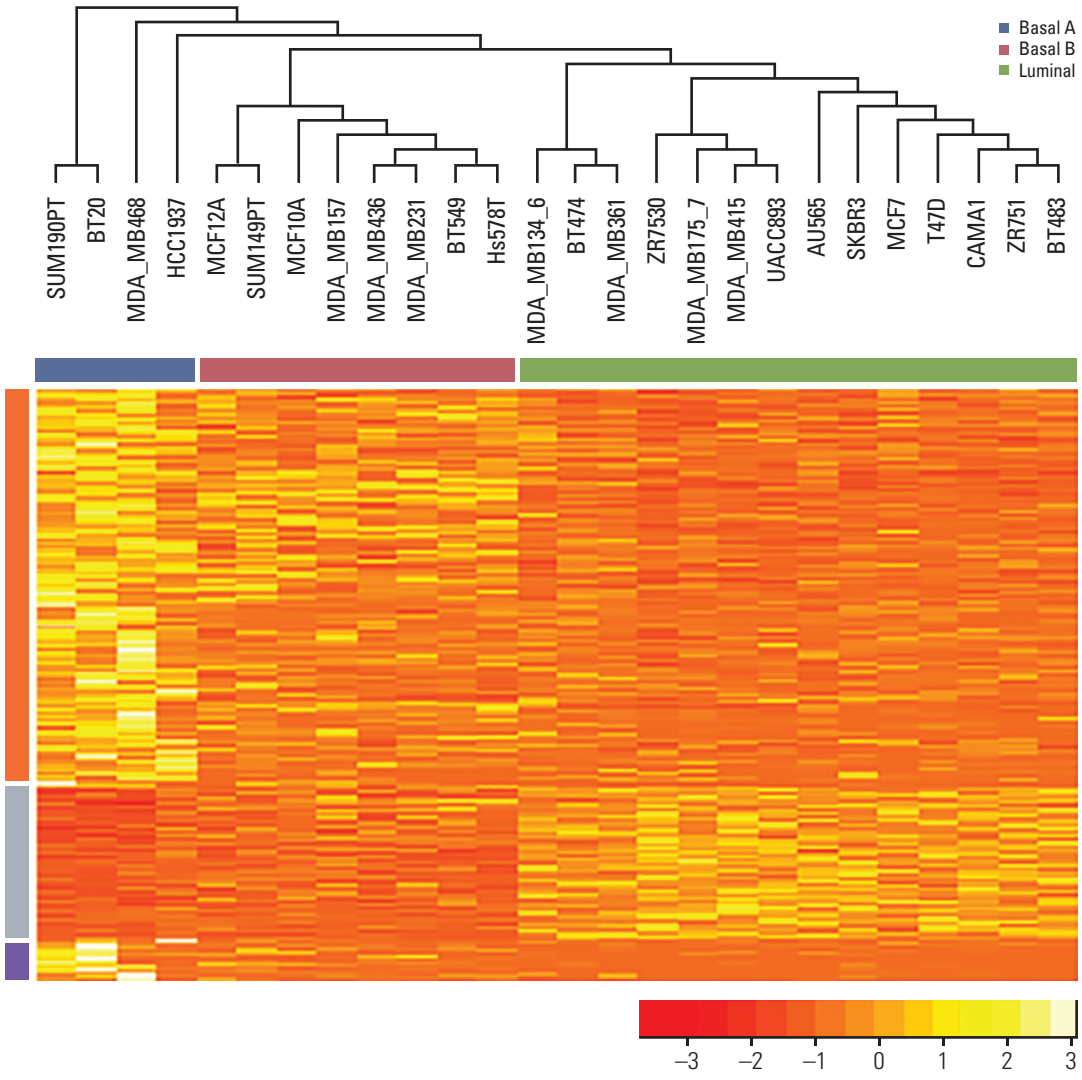

Fig. 2. Expression patterns of 638 genes in breast cancer cell lines, identified by ANOVA. The vertical and horizontal axes represent gene expressions and breast cancer cell lines, respectively.

Statistical analyses were performed using R ver. 3.1.1 (R Foundation for Statistical Computing, Vienna, Austria), $p < 0.05$ was considered statistically significant. The PCA algorithm was also performed using R [20].

Results

1. Gene expression profiling of breast cancer cell lines (G-matrix)

The GSE51086 microarray dataset was used to examine the relationship between gene expression and subtypes of breast cancer cell lines. The dataset included 45,220 probes, and it was summarized by 19,722 gene symbols for this study. The three cell lines in the dataset that did not include subtype information were excluded. A total of 638 genes were identified by significance among breast cancer subtypes. The sub-

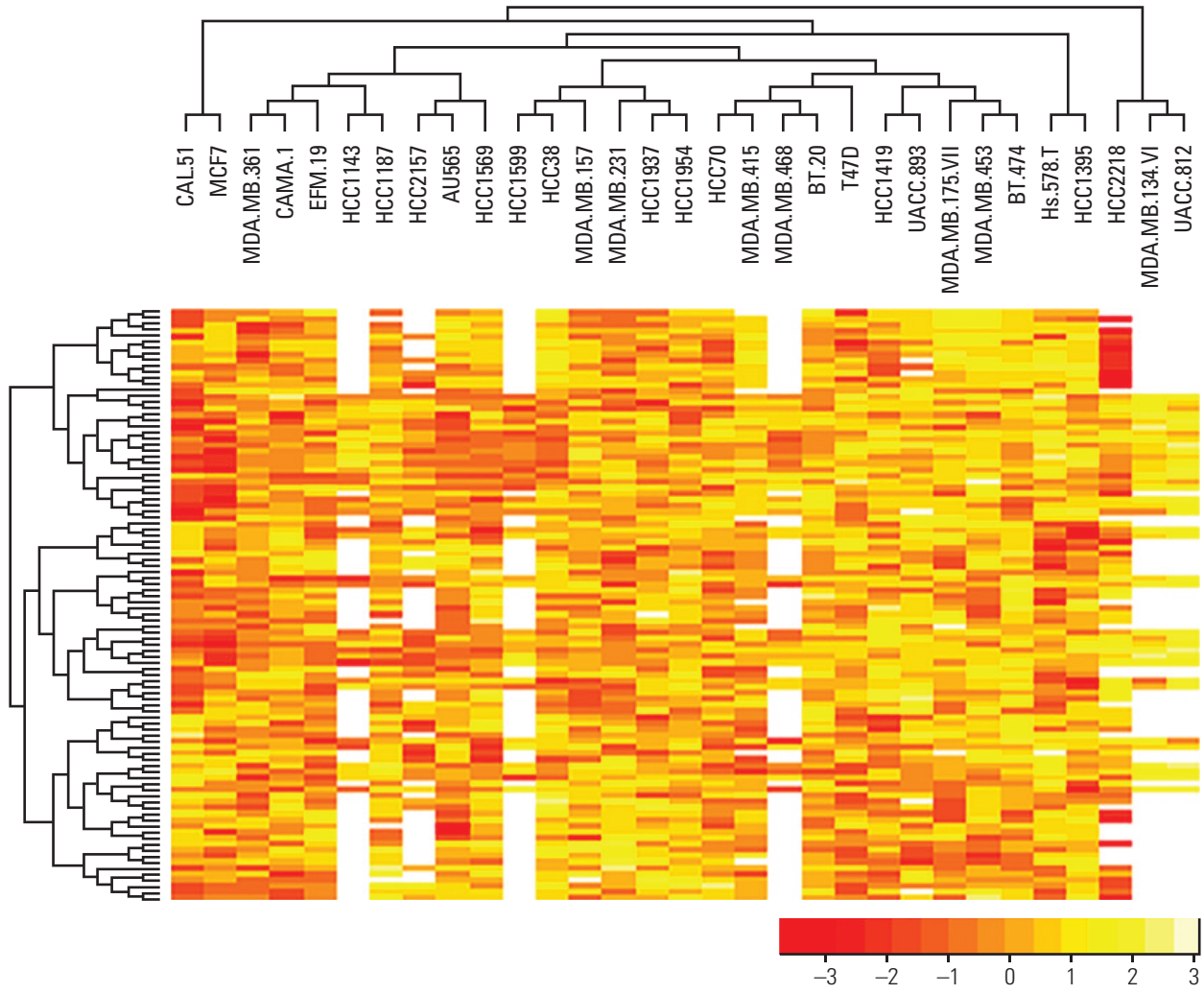

Fig. 3. Patterns of standardized IC_{50} scores in 31 breast cancer cell lines. The vertical and horizontal axes represent 98 drugs and 31 breast cancer cell lines, respectively.

types of breast cancer cell lines were 14 luminal, four basal A, and eight basal B.

Unsupervised hierarchical clustering performed to examine the genetic characteristics of each cell line revealed significant grouping of cell lines based on subtypes of breast cancer (Fig. 2).

Orange color denotes low expression and yellow denotes high expression. The gene expression pattern was clearly divided into several sections, which were strongly associated with subtypes of breast cancer cell lines. Expression of the first gene group was decreased in the order of basal A > basal B > luminal, and expression of the second group was increased according to the same order (i.e., basal A < basal B < luminal).

2. Chemosensitivity profiling of breast cancer cell lines (D-matrix)

The dataset from Garnett et al. [7] was used to examine the chemosensitivity of breast cancer cell lines. The dataset originally included 111 drugs and 43 breast cancer cell lines; however, cell lines and drugs with more than 50% missing data were excluded. Thus, 98 drugs and 31 breast cancer cell lines from this dataset were used in this study. The IC_{50} scores were standardized due to the large variation in the scale of IC_{50} scores for different drugs.

When unsupervised hierarchical clustering was applied to the dataset, there was no association between drug sensitivity and cell line (Fig. 3).

Each column represents a cell line, and each row repre-

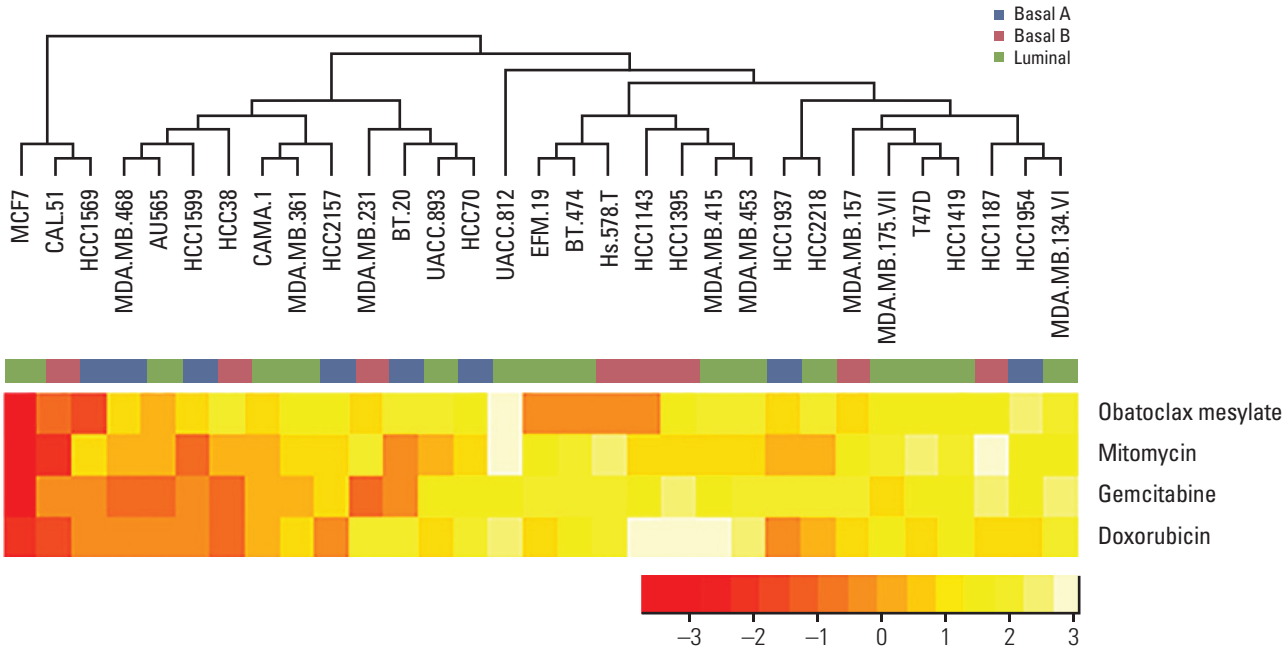

Fig. 4. Patterns of standardized IC_{50} of four drugs in 31 breast cancer cell lines. The vertical and horizontal axes represent four drugs and 31 breast cancer cell lines, respectively.

sents the IC_{50} score for the particular compound. Both rows and columns are clustered hierarchically. Yellow color represents high IC_{50} score (resistant), orange color represents low IC_{50} score (sensitive), and white indicates absence of data.

As shown in Fig. 3, no clear relationship was observed between drug sensitivity and breast cancer cell line. The cell lines CAL51 and MCF7 were sensitive to most drugs (orange color represents low IC_{50} scores), compared to other cell lines. By contrast, UACC812 was resistant to most drugs, though it should be noted that there were several drugs which did not include IC_{50} scores.

To examine the relationship between gene expression and drug sensitivity in detail, we selected four drugs, mitomycin, doxorubicin, gemstabine and obatoclax mesylate, all of which have been used for the treatment of breast cancer.

Unsupervised hierarchical clustering showed an association between drug sensitivity and cell line (Fig. 4).

The cell lines were divided into three groups by unsupervised hierarchical clustering, which were independent of breast cancer subtype, indicating that drug sensitivity is associated with specific gene expression rather than subtype of breast cancer.

Given this association pattern, doxorubicin was selected as a specific drug for implementing the drug sensitivity prediction model. Doxorubicin, which was discovered in 1969 by Farmitalia Research Laboratories in Italy, is used in treatment of early-stage or node-positive breast cancer, HER2-positive breast cancer, and metastatic disease. The first two cell line groups were sensitive to doxorubicin while the third was resistant, as shown in the tree diagram (Fig. 4).

3. Correlation between drug sensitivity and gene expression (DG-matrix)

For screening of the genes associated with chemosensitivity, the databases for gene expression and for chemosensitivity were integrated into one database matrix (DG-matrix) using the Pearson correlation coefficient. Using the DG-matrix, hierarchical clustering was performed based on the correlation between drug sensitivity and gene expression (Fig. 5).

Significantly expressed genes among subgroups of breast cancer cell lines were used to examine the association of gene expression and drug sensitivity. Several sectional blocks (indicated by orange and yellow) are shown in Fig. 5, suggesting a relationship between gene expression and drug sensitivity. The orange and yellow colors represent the negative and positive relationship between gene expression and drug sensitivity, respectively. Therefore, it can be interpreted that the drugs shown in the dashed box could be sensitive when the genes shown in the box are overexpressed (negative relationship).

It also shows that the genes related to subtypes of breast

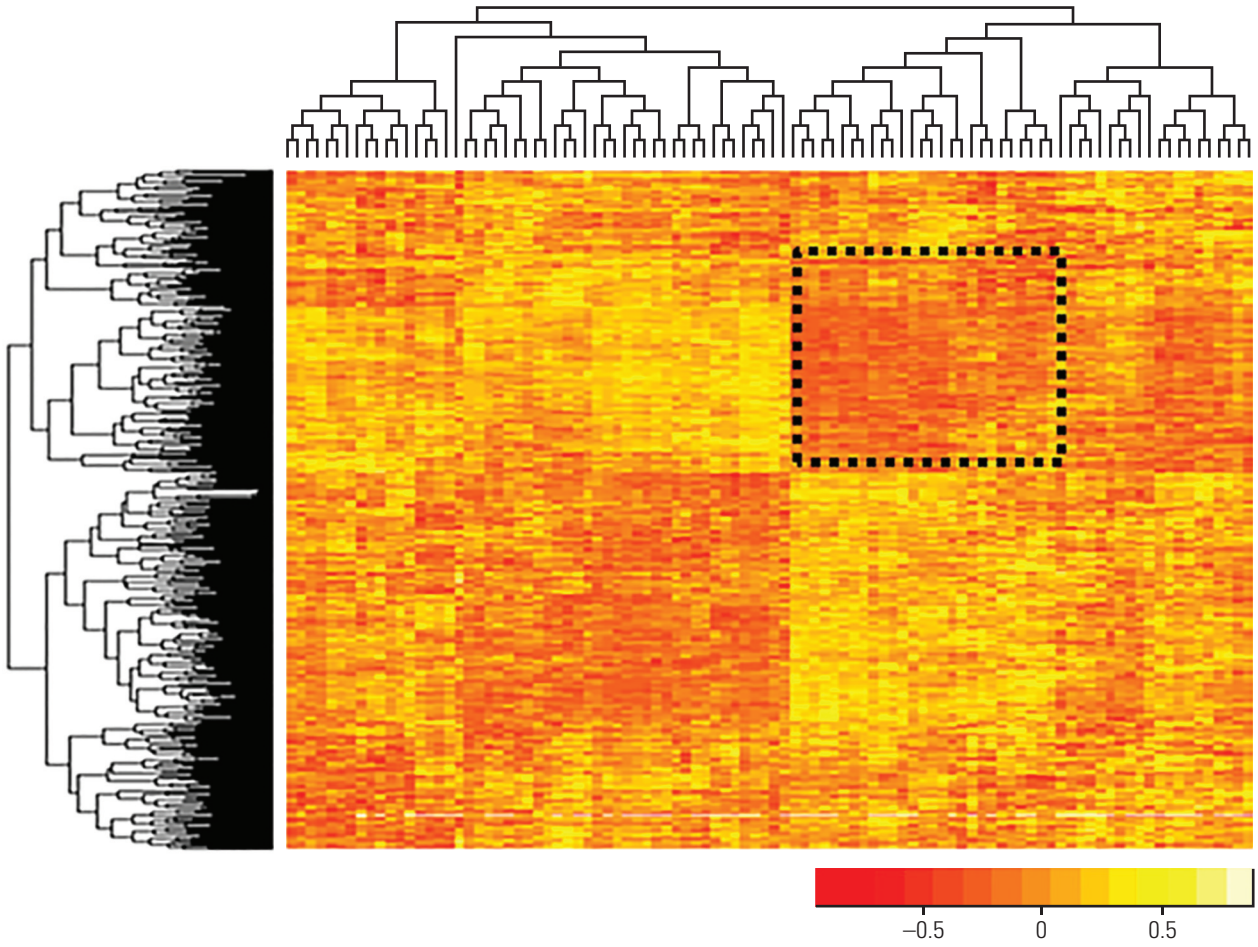

Fig. 5. DG-matrix. Correlation pattern of 638 gene expressions and standardized IC_{50} scores of 98 drugs. The vertical and horizontal axes represent 638 genes and 98 drugs, respectively. Correlation coefficients range from –1 to 1.

cancer cell lines are also related to drug sensitivity. Therefore, implementation of a drug sensitivity prediction model using gene expression data is reasonable, as gene expression differs between subtypes of breast cancer.

4. Selection of genes associated with drug-specific chemosensitivity

Correlation coefficients and their significance of both gene expression and IC_{50} scores in the four drugs were calculated. Drug sensitivity-related genes were identified as having correlation coefficients of $\geq |0.7|$ with drug-specific chemosensitivity. The results are listed in Table 2. Fifteen genes showed strong correlation with the IC_{50} score of each drug.

Genes showing strong correlation with specific drug sensitivity for implementing the drug sensitivity prediction model were selected.

5. Drug sensitivity prediction model using gene expression

In implementing our prediction model for sensitivity to doxorubicin, 10 genes were used as predictive variables. Five genes showed positive correlation with drug sensitivity and the other five genes showed negative correlation (Fig. 6).

As shown in Fig. 6, doxorubicin-related genes showed strong association with sensitivity to doxorubicin ($|r| > 0.77$, $p < 0.001$) but were not associated with sensitivity to mitomycin ($|r| < 0.45$, $p > 0.05$). This result indicates that the drug sensitivity model can be implemented for each specific drug and the genes related to it.

The optimal prediction model was identified by backward estimation in linear regression, and the final model included *PARVA*, *TBX21*, *SHF*, and *FAM158A* as predictors. The expression data of the four genes were combined to reduce the number of predictors. Combined gene expression was calculated by PCA.

Table 2. List of the top-ranked genes which showed positive or negative association with drug-specific chemosensitivity for four drugs

Mitomycin		Doxorubicin		Gemcitabine		Obatoclax mesylate	
Gene	r	Gene	r	Gene	r	Gene	r
FAM149A	0.8089	*LHFPL2*	0.8043	*IPW*	0.8351	*TAZ*	0.7864
GLRB	0.7851	*ARSB*	0.8026	*SNURF*	0.8122	*RHBDL2*	0.75747
PRDM5	0.7739	*GLRX*	0.7797	*NAGA*	0.7867	*YOD1*	0.73705
PPAP2A	0.7611	*GLRXL*	0.7787	*TNFAIP8L3*	0.7519	*KEL*	0.7330
PXK	0.7369	*PARVA*	0.7556	*COPS5*	0.7476	*LEMD2*	0.7227
ZNF525	–0.7583	*ACR*	–0.7751	*NKRF*	–0.8465	*FBXL19*	–0.8170
SCN7A	–0.7399	*DARS2*	–0.7712	*UBL4A*	–0.8294	*APOH*	–0.8033
PTPN4	–0.7359	*TBX21*	–0.7605	*USP29*	–0.7876	*TSEN2*	–0.7928
USP8	–0.7296	*SHF*	–0.7571	*CDH24*	–0.7829	*TUBD1*	–0.7908
FTO	–0.7291	*FAM158A*	–0.7537	*SPATA12*	–0.7791	*PFDN4*	–0.7777
EPB41	–0.7233	*RFC1*	–0.7429	*CC2D2B*	–0.7501	*BCAS3*	–0.7654
PPM1B	–0.7217	*KCTD15*	–0.7333	*KRT80*	–0.7433	*POU1F1*	–0.7502
WDR33	–0.7214	*PTP4A3*	–0.7299	*ANKRD39*	–0.7423	*CA4*	–0.7419
AIFM1	–0.7182	*MNX1*	–0.7296	*MFSD9*	–0.7409	*CTRB1*	–0.7411
BCL2L11	–0.7089	*KRT32*	–0.7270	*DXS542*	–0.7388	*GABPB1*	–0.7316

The p-values of the genes were less than 0.001. The standardized IC_{50} scores were used for calculating correlation coefficients. r represents correlation coefficient.

We compared the significances of two models, one using four independent genes and the other using the combined gene expression (Table 3).

A combined biomarker was calculated by linear, weighted combination of the four genes. The two models were similar in p-value and R^2, and the F value was significantly increased when a combined biomarker was used. Hence, the combined biomarker is reliable in our drug sensitivity prediction model.

Sensitivity to doxorubicin was associated with up-regulation of *TBX21*, *SHF*, and *FAM159A*. The combined factor showed a strong negative association with doxorubicin IC_{50} score. These results indicate that the more up-regulated the expression of these genes, the more sensitive the cells are to doxorubicin. Different color points represent subtypes of breast cancer cell lines. Breast cancer subtype was not found to be a predictive factor in this study (Fig. 7).

Discussion

Development of a method to predict what drug will be most effective in each case is needed in order to reduce the unpleasant side effects of chemotherapy for breast cancer patients. We suggest a prediction method using combined gene expression. This method is based on linear regression with IC_{50} scores of doxorubicin as the response variable. A statistical method was applied to reduce the number of predictors in the regression model so that the model is useful even in a small dataset.

The number of predictive variables has a significant influence on sample size and statitical power [21]. As the number of predictive variables is increased, the sample size must also be increased. Therefore, PCA was applied to reduce the number of predictive variables by combining several predictors.

By combining gene expression profiling with drug sensitivity data, we examined a large set of possible gene-drug relationships. These two datasets, which include gene expression and drug sensitivity, were derived from the public database (GEO database) and a previously published dataset [8]. Previous results showed that breast cancer is comprised of molecularly distinct subtypes that may respond differently to targeted therapies [22]. However, in this study drug sensitivity was related to gene expression rather than subtype of breast cancer cell lines.

Scherf et al. [8] reported that clustering of cell lines on the basis of gene expression yielded relationships that were very different from those obtained by clustering cell lines based on their response to drugs. This is in accordance with our study. However, a strong relationship was observed between gene expression and drug sensitivity, indicating that drug sensitivity can be predicted by gene expression data.

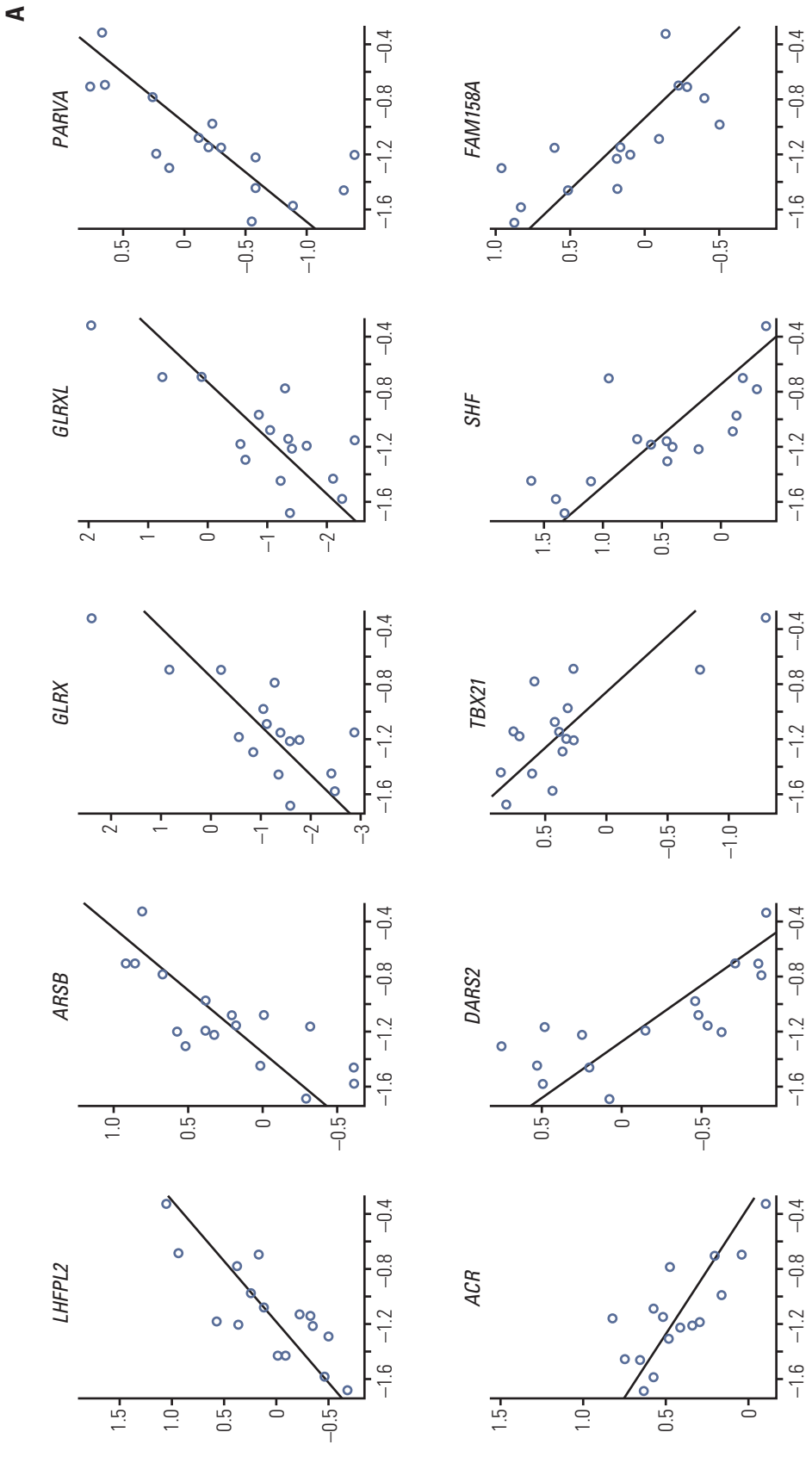

Fig. 6. The relationship between expressions of doxorubicin-related genes and drug sensitivity of doxorubicin (A) and mitomycin (B). The horizontal and vertical axes represent the standardized IC_{50} scores and gene expressions, respectively. *(Continued to the next page)*

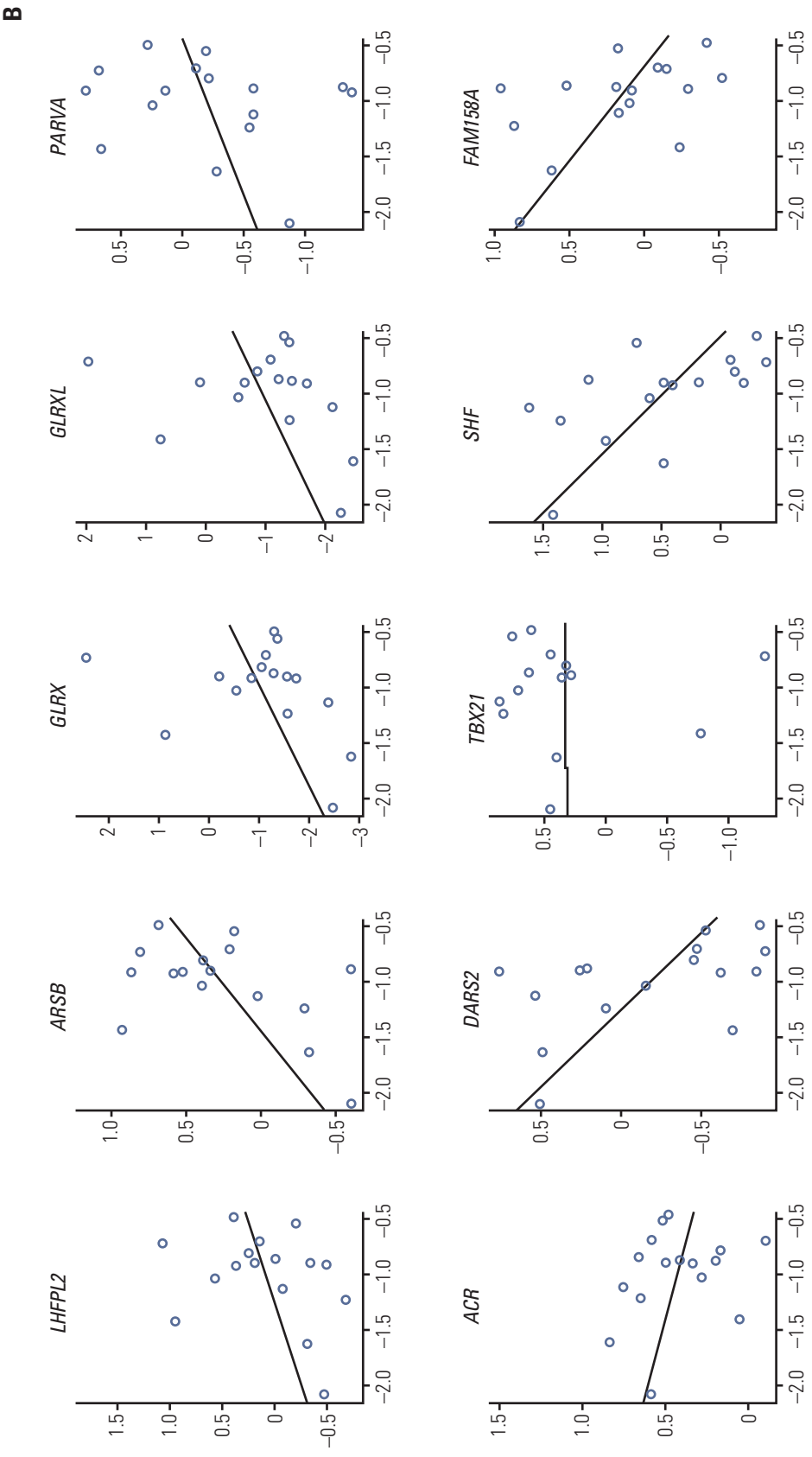

Fig. 6. *(Continued from the previous page)*

Table 3. Comparison of model significances in cases using four individual genes and a combined gene expression

Drug	Gene list in the model	Model significance	Weights of genes (combined biomarker)	Model significance
Doxorubicin	*PARVA*	F=58.238[a]	–0.812	F=198.309
	TBX21	$p < 0.001$[b]	0.772	$p < 0.001$
	SHF	R^2=0.939[c]	0.614	R^2=0.934
	FAM158A		0.552	

[a]F value represents model statistics, model significance is improved as F value is increased, [b]$p < 0.05$ is generally interpreted as statistically significant, [c]R^2 ranged from 0 to 1. When this value is 1, the model is perfectly predictive.

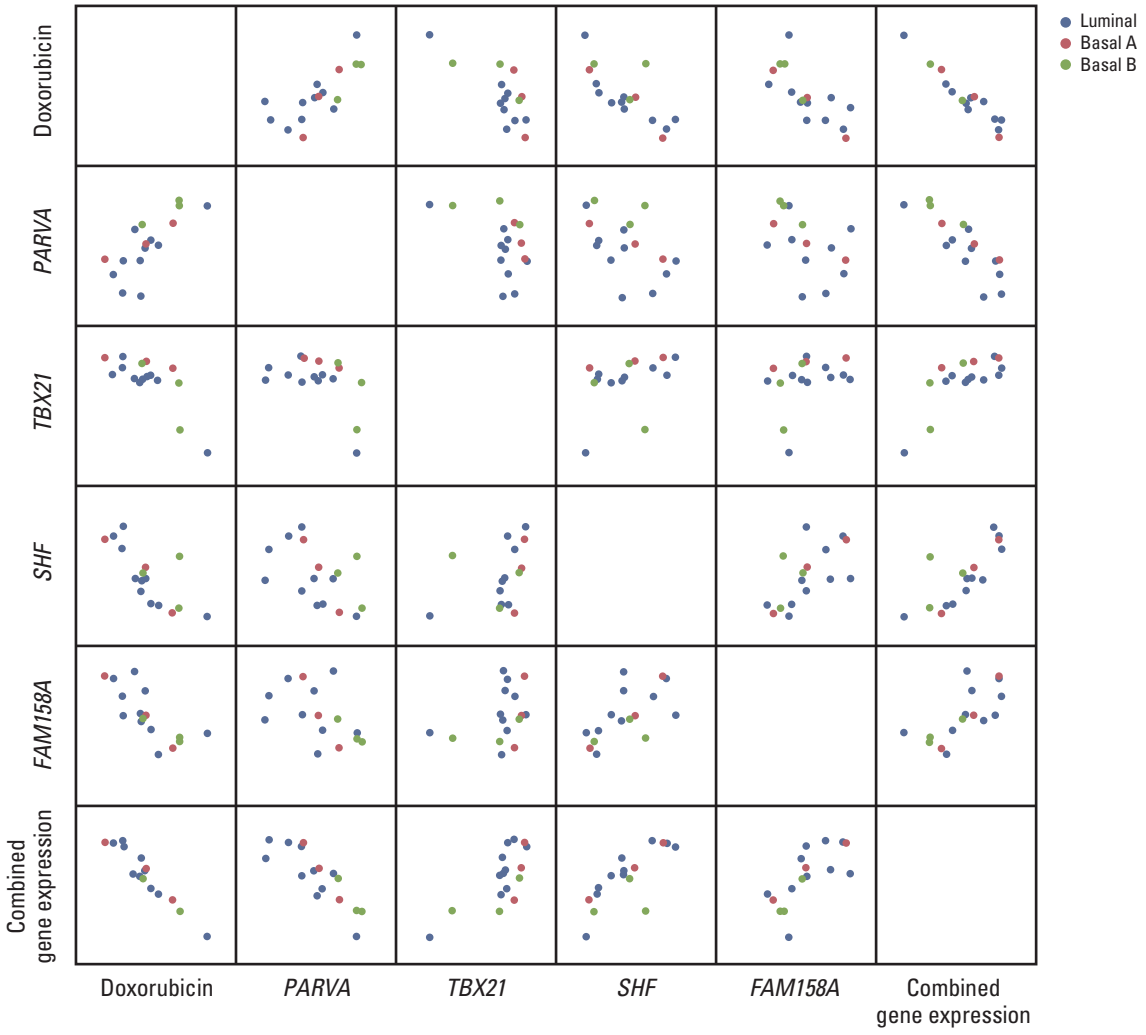

Fig. 7. Association of gene expressions and chemosensitivity to doxorubicin in breast cancer. Doxorubicin represents IC_{50} scores against doxorubicin and combined biomarker represents combined gene expression.

Various statistical and computational methods have been previously applied for prediction of IC_{50} scores [10-12]. When the statistical method is used for predicting chemosensitivity, sample size can be a very influential factor for achieving high statistical power, and sample size depends on the number of predictors in the linear regression method, which can be used for predicting numeric dependent variables (e.g., IC_{50}).

Therefore, our aim in this study was not to suggest a novel prediction method, but rather to show that the performance of a prediction model can be improved by reducing the number of predictors. PCA was used to reduce the number of predictors and to identify a combined gene expression profile.

The performance of the prediction model using combined gene expression was better than that of the model using several independent gene expressions. Model significance was improved when combined gene expression was used as a predictive variable.

The discovery of biomarkers of breast cancers has led to development of treatment strategies for breast cancer patients. Identification of combined gene expression profiles is expected to contribute to more personalized menagement of breast cancer patients and to improve existing therapies and it will be helpful in finding new therapies by identifying more predictive biomarkers. It can also be extended to other chemotherapeutic treatments and other cancer types.

Conclusion

The advantage of the proposed method is that the prediction model can be implemented with a relatively small sample dataset by reducing the number of predictive variables. A limitation of this study is that we did not consider the subtypes of breast cancer cell lines in our prediction model. With more available data, subtype of breast cancer should be considered as a predictive factor, even though it was not associated with drug sensitivity in this study. Our further study will include the validation of the result using an independent dataset and extension of the proposed method to other cancer types.

Conflicts of Interest

Conflict of interest relevant to this article was not reported.

Acknowledgments

This study was supported by a faculty research grant of Yonsei University College of Dentistry for 2015 (6-2015-0003).

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61:69-90.
2. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. CA Cancer J Clin. 2014;64:52-62.
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65:87-108.
4. Oh HS. Targeted therapy for breast cancer. Hanyang Med Rev. 2012;32:112-7.
5. Jung KW, Won YJ, Oh CM, Kong HJ, Cho H, Lee JK, et al. Prediction of cancer incidence and mortality in Korea, 2016. Cancer Res Treat. 2016;48:451-7.
6. BreastCancerTrials.org [Internet]. San Francisco, CA: BreastCancerTrials.org; 2016 [cited 2016 May 1]. Available from: https://www.breastcancertrials.org/.
7. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483:570-5.
8. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, et al. A gene expression database for the molecular pharmacology of cancer. Nat Genet. 2000;24:236-44.
9. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One. 2009;4:e6146.
10. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PLoS One. 2013;8:e61318.
11. Ferriss JS, Kim Y, Duska L, Birrer M, Levine DA, Moskaluk C, et al. Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: predicting platinum resistance. PLoS One. 2012;7:e30550.
12. Lee JK, Havaleshko DM, Cho H, Weinstein JN, Kaldjian EP, Karpovich J, et al. A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. Proc Natl Acad Sci U S A. 2007;104:13086-91.
13. De Abreu FB, Schwartz GN, Wells WA, Tsongalis GJ. Person-

alized therapy for breast cancer. Clin Genet. 2014;86:62-7.
14. Ring BZ, Chang S, Ring LW, Seitz RS, Ross DT. Gene expression patterns within cell lines are predictive of chemosensitivity. BMC Genomics. 2008;9:74.
15. Wu D, Pang Y, Wilkerson MD, Wang D, Hammerman PS, Liu JS. Gene-expression data integration to squamous cell lung cancer subtypes reveals drug sensitivity. Br J Cancer. 2013; 109:1599-608.
16. Knofczynski GT, Mundfrom D. Sample sizes when using multiple linear regression for prediction. Educ Psychol Meas. 2008;68:431-42.
17. Ku JM, Kim SR, Hong SH, Choi HS, Seo HS, Shin YC, et al. Cucurbitacin D induces cell cycle arrest and apoptosis by inhibiting STAT3 and NF-kappaB signaling in doxorubicin-resistant human breast carcinoma (MCF7/ADR) cells. Mol Cell Biochem. 2015;409:33-43.
18. Felding-Habermann B, O'Sullivan DM, Lorger M, MacDermed D, Fernandez-Santidrian A, Steele JB, et al. PD03-07: Breast cancer heterogeneity and treatment resistance: clues from metaplastic tumors. In: Thirty-Fourth Annual CTRC-AACR San Antonio Breast Cancer Symposium; 2011 Dec 6-10; San Antonio, TX, USA.
19. Fukunaga K. Introduction to statistical pattern recognition. Boston, MA: Academic Press; 1990.
20. R Foundation. The R Project for Statistical Computing [Internet]. Vienna: R Foundation; 2016 [cited 2016 May 1]. Available from: http://www.r-project.org/.
21. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
22. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S, et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci U S A. 2012;109:2724-9.