

# Identification of Small and Discriminative Gene Signatures for Chemosensitivity Prediction in Breast Cancer

Wei Hu

Department of Computer Science, Houghton College, Houghton, New York, USA.  
Email: [wei.hu@houghton.edu](mailto:wei.hu@houghton.edu)

Received January 22<sup>nd</sup>, 2011; revised April 5<sup>th</sup>, 2011; accepted April 15<sup>th</sup>, 2011.

## ABSTRACT

Various gene signatures of chemosensitivity in breast cancer have been discovered. One previous study employed *t*-test to find a signature of 31 probe sets (27 genes) from a group of patients who received weekly preoperative chemotherapy. Based on this signature, a 30-probe set diagonal linear discriminant analysis (DLDA-30) classifier of pathologic complete response (pCR) was constructed. In this study, we sought to uncover a signature that is much smaller than the 31 probe sets and yet has enhanced predictive performance. A signature of this nature could inform us what genes are essential in response prediction. Genetic algorithms (GAs) and sparse logistic regression (SLR) were employed to identify two such small signatures. The first had 13 probe sets (10 genes) selected from the 31 probe sets and was used to build a SLR predictor of pCR (SLR-13), and the second had 14 probe sets (14 genes) selected from the genes involved in Notch signaling pathway and was used to develop another SLR predictor of pCR (SLR-Notch-14). The SLR-13 and SLR-Notch-14 had a higher accuracy and a higher positive predictive value than the DLDA-30 with much lower *P* values, suggesting that our two signatures had their own discriminative power with high statistical significance. The SLR prediction model also suggested the dual role of gene *RNUX1* in promoting residual disease (RD) or pCR in breast cancer. Our results demonstrated that the multivariable techniques such as GAs and SLR are effective in finding significant genes in chemosensitivity prediction. They have the advantage of revealing the interacting genes, which might be missed by single variable techniques such as *t*-test.

**Keywords:** Genetic Algorithm, Gene Signature, Breast Cancer, Sparse Logistic Regression, Predictor, Chemosensitivity

## 1. Introduction

The molecular and pathological characteristics observed in breast cancer patients implies that breast cancer is a heterogeneous disease. Distinct molecular subtypes of breast cancer identified by patterns of gene expression could lead to different clinical outcomes. In current practice, chemotherapy is applied empirically, and does not benefit all patients, illustrating the imperative needs for a more personalized approach in cancer treatment. Therefore, the ability to predict whether an individual patient will benefit from a specific therapy is of great clinical significance. Single clinical or molecular indicators, such as tumor size, tumor grade, histology, hormone receptor or human epidermal growth factor receptor 2 (HER2) expression, does not always give reliable predictions of response to a treatment. With gene expression profiling, it is possible to find multiple genes that can be used to build

an enhanced predictor of response to chemotherapy in breast cancer [1-7].

In [8], *t*-test for unequal-variance was applied to find a signature of 31 probe sets (27 genes) with highest differentially expressed values between pCR and residual disease (RD). A 30-probe set diagonal linear discriminant analysis (DLDA-30) predictor of pCR was constructed based on this signature. The value of this predictor is the ability to identify those patients most likely to benefit from a particular treatment, the neoadjuvant chemotherapy, in this case. This predictor is able to recognize not all responsive patients but exclusively those that will benefit the most, as defined by attaining a pCR.

As a single variable technique, the *t*-test used in [8] analyzes one gene at a time and might miss the interactions between genes. We hypothesized that the signature of 31 probe sets in [8] could be optimized with help of

multivariable techniques such as genetic algorithms (GAs), which can explore multiple solutions concurrently to find interacting and informative genes. In this study, our aim was to identify a signature that is a small subset of the 31 probe sets in [8] and use it to develop a predictor of pCR that can achieve better predictions than the DLDA-30. A gene signature of this nature captures the essence of response prediction.

## 2. Patients and Methods

### 2.1. Patient Cohorts and Clinical Information

One breast cancer patient cohort was obtained from a previous publication [8] ( $n = 133$ ). Needle-biopsy samples were collected from 133 patients with stage I, II, or III breast cancer who received preoperative weekly paclitaxel and a combination of fluorouracil, doxorubicin, and cyclophosphamide (T/FAC). These 133 patients were divided into two subsets, one training set of size 81 and one validation set of size 52. These data contain clinical information including patient age, gender, race, histological classification, stage, nuclear grade, ER (estrogen receptor), PR (progesterone receptor), and HER2 (human epidermal growth factor 2) status, pathologic complete response, and residual disease. These data also contain each patient's genome-scale gene expression profiles generated using Affymetrix U133A chip (Santa Clara, CA). pCR was defined as no residual invasive cancer in the breast or lymph nodes. pCR is presently accepted as a reasonable early indicator for long-term survival.

### 2.2. Sparse Logistic Regression

A standard least squared linear regression solves the following problem:

Given data  $\{x_i, y_i\}_{i=1}^m$ , find  $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n)$  such that

$$\min_{\alpha} \sum_i (y_i - f(x_i))^2 \text{ where } f(x_i) = \sum_{j=1}^n \alpha_j x_{ij} + \alpha_0$$

The LASSO regression, studied by Tibshirani R [9], deals with the following problem:

$$\min_{\alpha} \sum_i (y_i - f(x_i))^2 \quad \text{with} \quad \sum_{j=1}^n |\alpha_j| \leq t,$$

where  $t$  controls the  $L_1$  norm of  $(\alpha_1, \alpha_2, \dots, \alpha_n)$ . This constraint on  $\alpha$  produces a sparse model, i.e., many components of  $\alpha$  can be zero. Following the idea of LASSO, Shevade *et al.* studied the following problem for sparse logistic regression, when  $y_i \in \{-1, +1\}$

$$\min_{\alpha} \sum_i g(-y_i f(x_i)) \quad \text{with} \quad \sum_{j=1}^n |\alpha_j| \leq t,$$

where  $g(\xi) = \log(1 + e^{\xi})$ , which is the negative log-likelihood function associated with the probability model

$$\text{Prob}(y|x) = \frac{1}{1 + e^{-y \cdot f(x)}}.$$

Cawley GC *et al.* [10] utilized a novel technique to solve this sparse logistic regression problem efficiently. In our study, we used +1 to label those cases of RD status, and used -1 to label those cases of pCR status.

### 2.3. Notch Signaling Pathway

Notch genes encode highly conserved cell surface receptors. The Notch signaling pathway consists of Notch receptors, ligands, negative and positive modifiers, and transcription factors. It plays a key role in the normal development of many tissues and cell types, through diverse effects on cell regulation, proliferation, and differentiation. Aberrant Notch signaling has been observed in several human cancers, including acute T-cell lymphoblastic leukemia, cervical cancer, and breast cancer [11,12]. The Oligo GEArray Human Notch Signaling Pathway Microarray [13] was designed for profiling expression of 113 genes (**Table 1**) involved in Notch signaling.

### 2.4. Prediction Accuracy Evaluation

In order to evaluate the significance of our predictions, we need to compare them with random guesses. For each dataset, a random-label permutation was conducted while keeping the number of instances in each group fixed. The matches between the permuted labels and the original ones were recorded. The standard P value was the percentage of 1000 random predictions with higher accuracy than the calculated predictions.

### 2.5. Genetic Algorithms

Genetic algorithms (GAs), a particular class of evolutionary algorithms, are search algorithms that adopt some common processes in genetics such as selection, mutation, and inheritance. The GAs outperform other traditional search algorithms in various applications.

Pseudo-code for genetic algorithms is as follows:

- Generate an initial population of individuals
- Evaluate initial population
- Repeat
  - Perform selection
  - Apply genetic operations such as mutation and crossover to generate a new generation of individuals
  - Evaluate individuals in the population
- Until some stopping criteria is satisfied

## 3. Results

To search for a subset of the 31 probe sets, we represented our solution, referred to as an individual in GAs, as a binary vector of size 31 to indicate the presence (1)

**Table 1. Genes involved in notch signaling pathway as described in [13].**

---

**Notch Signaling Pathway:**  
 Notch Binding: DLL1 (DELTA1), DTX1, JAG1, JAG2.  
 Notch Receptor Processing: ADAM10, PSEN1, PSEN2, PSENEN (PEN2).

**Notch Signaling Pathway Target Genes:**  
 Apoptosis Genes: CDKN1A, CFLAR (CASH), IL2RA, NFKB1.  
 Cell Cycle Regulators: CCND1 (Cyclin D1), CDKN1A (P21), IL2RA.  
 Cell Proliferation: CDKN1A (P21), ERBB2, FOSL1, IL2RA.  
 Genes Regulating Cell Differentiation: DTX1, PPARG.  
 Neurogenesis: HES1, HEY1.  
 Regulation of Transcription: DTX1, FOS, FOSL1, HES1, HEY1, NFKB1, NFKB2, NR4A2, PPARG, STAT6.  
 Other Target Genes with Unspecified Functions: CD44, CHUK, IFNG, IL17B, KRT1, LOR, MAP2K7, PDPK1, PTCRA.

**Other Genes Involved in the Notch Signaling Pathway:**  
 Apoptosis Genes: AXIN1, EP300, HDAC1, NOTCH2, PSEN1, PSEN2.  
 Cell Cycle Regulators: AXIN1, CCNE1, CDC16, EP300, FIGF, JAG2, NOTCH2, PCAF.  
 Cell Proliferation: CDC16, FIGF, FZD3, JAG1, JAG2, LRP5, NOTCH2, PCAF, STIL (SIL).  
 Genes Regulating Cell Differentiation: DLL1, JAG1, JAG2, NOTCH1, NOTCH2, NOTCH3, NOTCH4, PAX5, SHH.  
 Neurogenesis: DLL1, EP300, HEYL, JAG1, NEURL, NOTCH2, PAX5, RFNG, ZIC2 (HPE5).  
 Regulation of Transcription: AES, CBL, CTNNB1, EP300, GLI1, HDAC1, HEYL, HOXB4, HR, MYCL1, NCOR2, NOTCH1, NOTCH2, NOTCH3, NOTCH4, PAX5, PCAF, POFUT1, RUNX1, SNW1 (SKIIP), SUFU, TEAD1, TLE1.  
 Others Genes with Unspecified Functions: ADAM17, GBP2, LFNG, LMO2, MFNG, MMP7, NOTCH2NL, NUMB, SEL1L, SH2D1A.

**Other Signaling Pathways that Crosstalk with the Notch Signaling Pathway:**  
 Sonic Hedgehog (Shh) Pathway: GLI1, GSK3B, SHH, SMO, SUFU.  
 Wnt Receptor Signaling Pathway: AES, AXIN1, CTNNB1, FZD1, FZD2, FZD3, FZD4, FZD6, FZD7, GSK3B, LRP5, TLE1, WISPI1, WNT11.  
 Other Genes Involved in the Immune Response: CXCL9, FAS (TNFRSF6), G1P2, GBP1, IFNG, IL2RA, IL2RG, IL4, IL4R, IL6ST, IRF1, ISGF3G, OAS1, OSM, STAT5A, STUB1.

---

or absence (0) of each probe set in the 31 probe sets. Our fitness value was the prediction accuracy of SLR based on the training set. In each generation of GA, we randomly divided the training set ( $n = 82$ ) into five equal subsets for five-fold cross validation, where four subsets were used as a training set for SLR and one subset as a test set to get the accuracy of SLR on this test set. We ran the GA algorithm with population size 40, individual size

31, and 100 generations. In each generation, the top 50% of the individuals with highest fitness values were selected as parents to produce the next generation. A single point crossover and point mutations were applied to each individual in the population. As a result of GA selection, a collection of individuals with high prediction accuracy were discovered. The following individual has this property with training set accuracy equal to 0.89.

1	0	0	1	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

The above binary representation means the following 13 probe sets (10 genes) were selected (**Table 2**).

This 13 probe set signature highlighted the critical genes for response prediction, and therefore provided clear insight into the molecular mechanisms that regulate the chemosensitivity in breast cancer. Encouraged by the finding of the 13 probe sets from the 31 probe sets in [8], we applied the same selection procedure using GAs to the 113 genes involved in Notch signaling pathway (**Table 1**), resulting in a signature of 14 probe sets (SLR-Notch-14) (**Figure 1**). It is worth noting that our search for a smaller and yet more discriminative set of genes from the 31 probe sets was accomplished by two size reductions. First, the size of informative genes was reduced from 31 probe sets to 13 by GAs, and then SLR was able to further reduce this 13 probe sets to 10 probe sets (**Figure 1**).

We report in **Tables 2** and **3** the t-test statistic, P value, and gene names for the 13 probe sets from the 31 probe

sets in [8] and the 14 probe sets from the genes involved in Notch signaling pathway. The P values in these two tables showed that the most discriminative genes as a group measured by SLR were not necessarily those with the smallest P values as individual genes. In general, the genes that were highly expressed in the RD cases contributed positively to the prediction of RD and those highly expressed in the pCR cases contributed positively to the prediction of pCR.

In addition to testing the SLR-13 and SLR-Notch14 on the training set (**Table 4**), we also tested them on the validation set (**Tables 5** and **6**) for prediction performance and P values. **Table 5** suggests that the two predictors, SLR-13 and SLR-Notch-14, complemented each other in their specificity and sensitivity. On the other hand the confusion matrix in **Table 6** contains information about actual and predicted classifications obtained through the three predictors. The SLR-Notch-14 had three nonzero P values and they were all less than 0.05,

**Table 2. 13 informative probe sets selected by GA and SLR from the 31 probe sets in [8].**

SLR Weights	Probe Set ID	Gene Symbol	Gene Name	t-test	P value	Higher Expression in
1.0998	203929_s_at	MAPT	Microtubule-associated protein tau	6.6019	.00000001	RD
0.2047	203928_x_at	MAPT	Microtubule-associated protein tau	5.9166	.00000008	RD
0.14744	212207_at	THRAP2	Thyroid hormone receptor associated protein 2	6.0604	.00000010	RD
0.13463	217542_at	MGC5370	Hypothetical protein MGC5370	5.6999	.00000027	RD
0.26123	206401_s_at	MAPT	Microtubule-associated protein tau	5.6116	.00000028	RD
0	219741_x_at	ZNF552	Zinc finger protein 552	5.4961	.00000047	RD
0	205548_s_at	BTG3	BTG family, member 3	-5.9508	.00000056	pCR
-0.26046	202204_s_at	AMFR	Autocrine motility factor receptor	5.2823	.00000129	RD
0.46085	209617_s_at	CTNND2	Catenin (cadherin-associated protein), delta 2 (neural plakophilin-related armrepeat protein)	5.1606	.00000178	RD
0.14544	205354_at	GAMT	Guanidinoacetate N-methyltransferase	5.349	.00000250	RD
0.24103	204509_at	CA12	Carbonic anhydrase XII	5.0802	.00000282	RD
0.36339	214124_x_at	FGFR1OP	FGFR1 oncogene partner	5.1848	.00000311	RD
0.34214	215616_a_at	JMDJ2B	Jumonji domain containing 2B	5.0047	.00000559	RD

**Table 3. 14 informative probe sets selected by GA and SLR from Notch signaling pathway.**

SLR Weight	Probe Set ID	Gene Symbol	Gene Name	t-Test	P value	Higher Expression in
0.18233	202221_s_at	EP300	E1A binding protein p300	2.585026	0.011947	RD
0	204262_s_at	FZD1	frizzled homolog 1	-1.39285	0.174441	pCR
0	207535_s_at	SHH	sonic hedgehog homolog	-0.35197	0.726497	pCR
0	207539_s_at	NCOR2	nuclear receptor co-repressor 2	0.450195	0.654935	RD
0.16053	207586_at	RUNX1	runt-related transcription factor 1	1.130322	0.264823	RD
0	209835_x_at	HR	hairless homolog	-0.42314	0.675586	pCR
0	210086_at	CTNNB1	catenin (cadherin-associated protein), beta 1	-0.38441	0.7025	pCR
-0.08195	210930_s_at	RUNX1	runt-related transcription factor 1	-2.28195	0.031343	pCR
0.30715	211209_x_at	CFLAR	CASP8 and FADD-like apoptosis regulator	1.873074	0.066922	RD
0	212195_at	MFNG	Manic fringe homolog	-0.26482	0.79271	pCR
0.33756	212849_at	OSM	oncostatin M	3.355561	0.001316	RD
-0.29049	214058_at	SNW1	SNW domain containing 1	-2.47946	0.019481	pCR
0	214637_at	CD44	CD44 antigen	-1.12707	0.268386	pCR
0.14599	222183_x_at	SH2D1A	SH2 domain protein 1A	1.281554	0.207413	RD

**Table 4. Prediction measures of SLR-13 and SLR-Notch-14 on the training set (five-fold cross validation).**

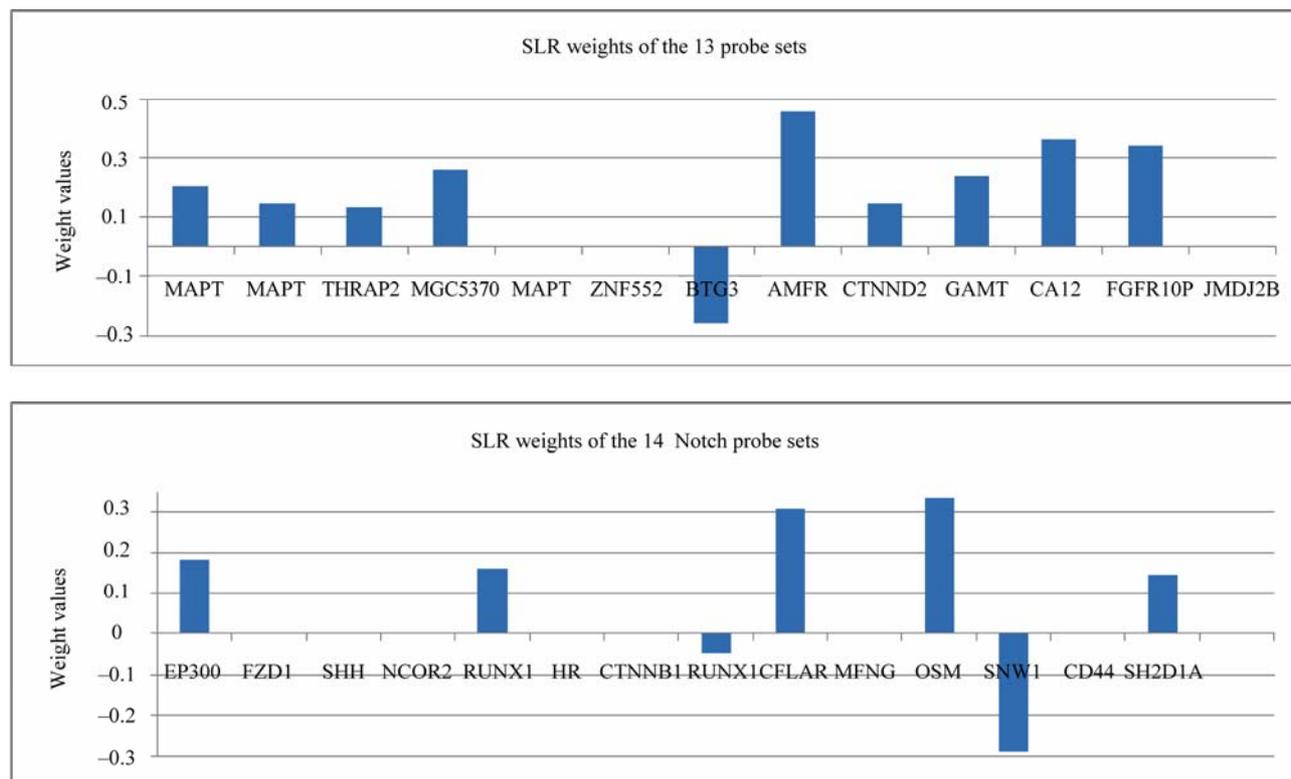
Measures	DLDA-30	SLR-13	SLR-Notch-14
Accuracy	0.73	0.83	0.78
Sensitivity	0.75	0.84	0.78
Specificity	0.73	0.83	0.78
PPV	0.50	0.62	0.56
NPV	0.90	0.94	0.91

**Table 5. Prediction measures of DLDA-30, SLR-13, and SLR-Notch-14 on the validation set.**

Measure	DLDA-30	P value	SLR-13	P value	SLR-Notch-14	P value
Accuracy	0.76	0.1900	0.88	0	0.82	0.0040
Sensitivity	0.92	0	0.77	0	0.77	0
Specificity	0.71	0.96	0.92	0	0.84	0.0190
PPV	0.52	0.0920	0.77	0	0.63	0.0040
NPV	0.96	0	0.92	0	0.91	0

**Table 6. Confusion matrices for DLDA-30, SLR-13, and SLR-Notch-14 on the validation set.**

DLDA-30	Predicted as pCR	Predicted as RD	SLR-13	Predicted as pCR	Predicted as RD	SLR-Notch-14	Predicted as pCR	Predicted as RD
Observed pCR	$n = 12$	$n = 1$	Observed pCR	$n = 10$	$n = 3$	Observed pCR	$n = 10$	$n = 3$
Observed RD	$n = 11$	$n = 27$	Observed RD	$n = 3$	$n = 35$	Observed RD	$n = 6$	$n = 32$

**Figure 1. The SLR weights of the 13 probe set and 14 probe set signature.**

implying that the prediction accuracies of SLR-13 and SLR-Notch14 were statistically significant. The DLDA-30 had three nonzero P values but they were all larger than 0.05, especially those for accuracy and specificity.

In **Table 2**, the genes were all highly expressed in the RD cases with one exception, whereas **Table 3** has more balanced gene numbers in the RD and pCR cases, which could partially contribute to the distinct predicting power of SLR-13 and SLR-Notch-14 (**Tables 5** and **6**). In the SLR model, genes with a positive weight contributed positively to the RD prediction and those with a negative weight contributed positively to the prediction of pCR. Obviously, genes of zero weight did not contribute to response prediction. The gene BTG3 [14] was highly expressed in the pCR cases, and so it was weighted positively toward the prediction of pCR in the SLR-13 signature (**Figure 1**). However, in the SLR-Notch-14 signature one UNIX1 probe set was expressed highly in the RD cases and another UNIX1 probe set was highly ex-

pressed in the pCR cases, which were also reflected by their opposite SLR weights (**Figure 1**). RUNX1 has a dual role in promoting cell cycle progression and differentiation [15] and can function as a transcription activator or as a repressor [16].

#### 4. Discussion

Predicting the treatment response for patients with breast cancer is a great challenge in clinics and is critical in personalized medicine. Various such genes signatures of different predicting power have been identified. The focus of the current study was to find a gene signature that is small but has improved predictions. Our strategy was to apply genetic algorithm, an efficient search algorithm, to refine the genes in a known signature so that only the most relevant genes were remained after this procedure. The findings of this report demonstrated the validity of this approach. We believed that genes in a small signature reveal the essence of association between gene ex-

pressions and clinical outcome.

## 5. Conclusions

Recent studies suggest that gene expression profile correlates with responses to neoadjuvant chemotherapy, with tumors displaying the ER-positive gene signatures being less likely to respond than other types of breast cancer [17]. In this study, we intended to discover a much smaller gene signature than the 31 probe sets in [8] that predicts whether a breast cancer patient will benefit from a preoperative treatment containing paclitaxel followed by fluorouracil, doxorubicin, and cyclophosphamide by achieving a pCR. With the ability to account for multiple gene interactions, two multivariable techniques, genetic algorithms and sparse logistic regression, were employed to identify a 13 probe sets from the 31 probe sets in [8] and a 14 probe set signatures from the genes involved Notch signaling pathway (**Table 1**). The response predictions of our two signatures had much lower P values than the 31 probe sets in [8], revealing the improved statistical significance of their predictions. The SLR prediction model also verified the dual role of gene RNUX1 in promoting RD or pCR in breast cancer.

## 6. Acknowledgements

We thank Houghton College for its financial support.

## REFERENCES

- [1] M. Chanrion, V. Negre, H. Fontaine, N. Salvetat, F. Bibeau, G. Mac Grogan, L. Mauriac, D. Katsaros, F. Molina, C. Theillet and J. M. Darbon, "A Gene Expression Signature That Can Predict the Recurrence of Tamoxifen-Treated Primary Breast Cancer," *Clinical Cancer Research*, Vol. 14, No. 6, 2008, pp. 1744-1752. [doi:10.1158/1078-0432.CCR-07-1833](https://doi.org/10.1158/1078-0432.CCR-07-1833)
- [2] S. P. Linke, T. M. Bremer, C. D. Herold, G. Sauter and C. Diamond, "A Multimarker Model to Predict Outcome in Tamoxifen-Treated Breast Cancer Patients," *Clinical Cancer Research*, Vol. 12, No. 4, 2006, pp. 1175-1183. [doi:10.1158/1078-0432.CCR-05-1562](https://doi.org/10.1158/1078-0432.CCR-05-1562)
- [3] P. E. Lønning, S. Knappskog, V. Staalesen, R. Chrisanthar and J. R. Lillehaug, "Breast Cancer Prognostication and Prediction in the Postgenomic Era," *Annals of Oncology*, Vol. 18, No. 8, 2007, pp. 1293-1306. [doi:10.1093/annonc/mdm013](https://doi.org/10.1093/annonc/mdm013)
- [4] M. A. Folgueira, D. M. Carraro, H. Brentani, D. F. Patrão, E.M. Barbosa, M. M. Netto, J. R. Caldeira, M. L. Katayama, F. A. Soares, C. T. Oliveira, L. F. Reis, J. H. Kaiano, L. P. Camargo, R. Z. Vêncio, I. M. Snitcovsky, F. B. Makdissi, P. J. e Silva, J. C. Góes and M. M. Brentani, "Gene Expression Profile Associated with Response to Doxorubicin-Based Therapy in Breast Cancer," *Clinical Cancer Research*, Vol. 11, No. 20, 2005, pp. 7434-7443. [doi:10.1158/1078-0432.CCR-04-0548](https://doi.org/10.1158/1078-0432.CCR-04-0548)
- [5] H. K. Dressman, C. Hans, A. Bild, J. A. Olson, E. Rosen, P. K. Marcom, V. B. Liotcheva, E. L. Jones, Z. Vujaskovic, J. Marks, M. W. Dewhirst, M. West, J. R. Nevins and K. Blackwell, "Gene Expression Profiles of Multiple Breast Cancer Phenotypes and Response to Neoadjuvant Chemotherapy," *Clinical Cancer Research*, Vol. 12, No. 3, 2006, pp. 819-826. [doi:10.1158/1078-0432.CCR-05-1447](https://doi.org/10.1158/1078-0432.CCR-05-1447)
- [6] O. Thuerigen, A. Schneeweiss, G. Toedt, P. Warnat, M. Hahn, H. Kramer, B. Brors, C. Rudlowski, A. Benner, F. Schuetz, B. Tews, R. Eils, H.-P. Sinn, C. Sohn and P. Lichter, "Gene Expression Signature Predicting Pathologic Complete Response with Gemcitabine, Epirubicin, and Docetaxel in Primary Breast Cancer," *Journal of Clinical Oncology*, Vol. 24, No. 12, 2006, pp. 1839-1845. [doi:10.1200/JCO.2005.04.7019](https://doi.org/10.1200/JCO.2005.04.7019)
- [7] S. Rathnagiriswaran, Y. W. Wan, J. Abraham, V. Castanova, Y. Qian and N. L. Guo, "A Population-Based Gene Signature is Predictive of Breast Cancer Survival and Chemoresponse," *International Journal of Oncology*, Vol. 36, No. 3, 2010, pp. 607-616.
- [8] K. R. Hess, K. Anderson, W. F. Symmans, *et al.*, "Pharmacogenomic Predictor of Sensitivity to Preoperative Chemotherapy with Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide in Breast Cancer," *Journal of Clinical Oncology*, Vol. 24, No. 26, 2006, pp. 4236-4244.
- [9] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society B*, Vol. 58, No. 1, 1996, pp. 267-288.
- [10] G. C. Cawley and L. C. Talbot, "Gene Selection in Cancer Classification Using Sparse Logistic Regression with Bayesian Regularization," *Bioinformatics*, Vol. 22, No. 19, 2006, pp. 2348-2355. [doi:10.1093/bioinformatics/btl386](https://doi.org/10.1093/bioinformatics/btl386)
- [11] S. Stylianou, R. B. Clarke, *et al.*, "Activation of Notch Signaling in Human Breast Cancer," *Cancer Research*, Vol. 66, No. 3, 2006, pp. 1517-1525. [doi:10.1158/0008-5472.CAN-05-3054](https://doi.org/10.1158/0008-5472.CAN-05-3054)
- [12] K. Brennan and A. M. C. Brown, "Is There a Role for Notch Signalling in Human Breast Cancer?" *Breast Cancer Research*, Vol. 5, No. 2, 2003, pp. 69-75. [doi:10.1186/bcr559](https://doi.org/10.1186/bcr559)
- [13] [http://www.sabiosciences.com/rt\\_pcr\\_product/HTML/PAHS-059A.html](http://www.sabiosciences.com/rt_pcr_product/HTML/PAHS-059A.html)
- [14] Y. H. Ou, P.-H. Chung, *et al.*, "The Candidate Tumor Suppressor BTG3 Is a Transcriptional Target of P53 That Inhibits E2F1," *The EMBO Journal*, Vol. 26, No. 17, 2007, pp. 3968-3980. [doi:10.1038/sj.emboj.7601825](https://doi.org/10.1038/sj.emboj.7601825)
- [15] J. Starkova, J. Madzo, G. Cario, T. Kalina, A. Ford, M. Zaliova, O. Hrusak and J. Trka, "The Identification of (ETV6)/RUNX1-Regulated Genes in Lymphopoiesis Using Histone Deacetylase Inhibitors In ETV6/RUNX1-Positive Lymphoid Leukemic Cells," *Clinical Cancer Research*, Vol. 13, No. 6, 2007, pp. 1726-1735. [doi:10.1158/1078-0432.CCR-06-2569](https://doi.org/10.1158/1078-0432.CCR-06-2569)
- [16] F. M. Mikhail, K. K. Sinha, Y. Sauntharajah and G. Nucifora, "Normal and Transforming Functions of RUNX1: A Perspective," *Journal of Cell Physiology*, Vol.

207, No. 3, 2006, pp. 582-593.  
[doi:10.1002/jcp.20538](https://doi.org/10.1002/jcp.20538)

- [17] N. S. Goldstein, D. Decker, D. Severson, *et al.*, "Molecular Classification System Identifies Invasive Breast Carcinoma Patients Who Are Most Likely and Those Who

Are Least Likely to Achieve a Complete Pathologic Response after Neoadjuvant Chemotherapy," *Cancer*, Vol. 110, No. 8, 2007, pp. 1687-1696.  
[doi:10.1002/cncr.22981](https://doi.org/10.1002/cncr.22981)