

# Screening of feature genes in distinguishing different types of breast cancer using support vector machine

Qi Wang  
Xudong Liu

Department of Emergency Surgery,  
Affiliated Hospital of Inner Mongolia  
Medical University, Hohhot, People's  
Republic of China

**Objective:** To screen the feature genes in estrogen receptor-positive (ER+) breast cancer in comparison with estrogen receptor-negative (ER-) breast cancer.

**Methods:** Nine microarray data of ER+ and ER- breast cancer samples were collected from Gene Expression Omnibus database. After preprocessing, data in five training sets were analyzed using significance analysis of microarrays to screen the differentially expressed genes (DEGs). The DEGs were further analyzed via support vector machine (SVM) function in e1071 package of R to construct a SVM classifier, the efficacy of which was verified by four testing sets and its combination with training sets using a leave-one-out cross-validation. Feature genes obtained by SVM classifier were subjected to function- and pathway-enrichment via the Database for Annotation, Visualization and Integrated Discovery and KEGG Orthology Based Annotation System, respectively.

**Results:** A total of 526 DEGs were screened between ER+ and ER- breast cancer. The SVM classifier demonstrated that these genes could distinguish different subtype samples with high accuracy of larger than 90%, and also showed good sensitivity, specificity, positive/negative predictive value, and area under receiver operating characteristic curve. The inflammatory and hormone biological processes were the common enriched results for two different function analyses, indicating that the inflammatory (ie, *IL8*) and hormone regulation (ie, *CGA*) genes may be the involved feature genes to distinguish ER+ and ER- types of breast cancer.

**Conclusion:** The gene-expression profile data can provide feature genes to distinguish ER+ and ER- samples, and the identified genes can be used for biomarkers for ER+ samples.

**Keywords:** classification, differentially expressed genes, biomarker

## Introduction

Breast cancer is the most common invasive cancer in females worldwide, with an estimated 232,670 newly diagnosed cases and approximately 40,000 deaths in 2014 in the USA.<sup>1</sup> Breast cancer is a hormone-dependent malignancy. At their primary diagnosis, approximately 75%–80% of breast cancer patients present as estrogen receptor-positive (ER+), while 20%–30% are estrogen-negative (ER-).<sup>2</sup> It is reported that ER- breast cancer is related with poor prognosis, whereas breast cancer patients who are ER+ have a favorable outcome.<sup>3</sup> This indicates the importance to distinguish between these two different subtypes of breast cancer, with the aim to provide prognosis and guide targeted treatment.

Traditional classification based on the histochemical analysis of ER expression is often limited, and does not have the ability to discern subtle differences in different

Correspondence: Xudong Liu  
Department of Emergency Surgery,  
Affiliated Hospital of Inner Mongolia  
Medical University, No. 1 Tongdaobei  
Street, Hohhot, Inner Mongolia Province  
010050, People's Republic of China  
Tel +86 0471 663 6963  
Email liuxuxuxusd@163.com

subtypes of breast cancer.<sup>4</sup> Thus, molecular identification is advocated. For example, Lim et al demonstrated that the lysine-specific demethylase 1 is highly expressed in ER– breast cancer,<sup>5</sup> while Mehta et al showed that the fork-head box protein A1 is an independent prognostic marker for ER+ breast cancer.<sup>6</sup> However, the research on the genes that could distinguish the different subtypes of breast cancer is so limited and needs further study.

Recent studies have shown that gene-expression profile generated by high-throughput platforms may provide comprehensive molecular characteristics of the tumors and may be informative for tumor classifications.<sup>4,7</sup> For example, Parker et al have identified a 50-gene transcriptional signature and demonstrated that they have a good prognosis performance for “intrinsic” subtypes of breast cancer (luminal A, luminal B, HER2-enriched, and basal-like).<sup>8</sup> Haibe-Kains et al reported a three-gene-expression model to classify tumors into four molecular entities (ER+/HER2–/low proliferative, ER+/HER2–/high proliferative, HER2+, and ER–/HER2–),<sup>9</sup> which displays relatively less prognosis ability compared to the 50-gene transcriptional signature.<sup>10</sup> However, a gene model to specifically distinguish ER+ and ER– breast cancer remains poorly investigated.

Several data-mining technologies have recently been developed to accomplish feature gene extraction and selection, among which the support vector machine (SVM) algorithm performs at a higher power in two categories of classification.<sup>11–13</sup> In the present study, we used the SVM to analyze the biomarkers for two subtypes of breast cancer: ER– and ER+ utilizing gene-expression profiling data.

## Materials and methods

### Microarray data and data preprocessing

Gene-expression data of breast cancer were downloaded from Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>).<sup>14</sup>

The following criteria were used to screen out the appropriate gene-expression data: 1) samples were not treated by any medicines; 2) all samples were breast cancer samples; and 3) samples were classified by ER status. Nine datasets were finally included, consisting of 1,289 samples, in which five expression profiles (626 samples, 425 ER+ and 201 ER–) were randomly assigned to training sets and the other four expression profiles (663 samples, 492 ER+ and 171 ER–) were used for testing sets (Table 1). The raw downloaded data first underwent background correction,<sup>15</sup> log2 transformation, and then quantiles normalization<sup>16</sup> using Affy package in R.

### Screening of differentially expressed genes

To select the differentially expressed genes (DEGs) in ER+ samples compared with ER– samples after data preprocessing, significance analysis of microarrays package in R ([www.r-project.org](http://www.r-project.org)) was utilized.<sup>17</sup> Genes with false discovery rates (FDRs) estimated by permutation method<sup>18</sup> to be less than 0.05 and log2 fold change (FC) >1 were considered as DEGs.

### Sample classification using SVM classifier

It was not certain whether the selected DEGs could distinguish the two types of breast cancers well; thus, SVM was used for this determination to build models based on “training” data and search for similar patterns in “testing” data. Based on the normalized expression values of 526 DEGs identified using training data, a SVM classifier was constructed via SVM function in e1071 package of R ([www.r-project.org](http://www.r-project.org)) with the nonlinear radial basis function as the kernel and penalty functions set at 1,000. The predictive results of the SVM model for the training set itself were evaluated by a leave-one-out cross-validation method,<sup>19</sup>

**Table 1** Summary of the nine included microarray data

Datasets ID	Number of samples	ER+	ER–	Average age (years)
Training sets				
E-GEOD-3494	241	209	32	62.004
E-GEOD-24185	94	56	38	49.03
E-GEOD-22597	82	37	45	51.23
E-GEOD-22093	82	41	41	48.51
E-GEOD-45255	127	82	45	–
Total	626	425	201	52.69
Testing sets				
E-GEOD-4922	245	211	34	62.12
E-GEOD-32518	71	40	31	47.69
E-GEOD-23988	61	32	29	48.69
E-GEOD-2034	286	209	77	–
Total	663	492	171	52.83

**Abbreviations:** ER+, estrogen receptor-positive; ER–, estrogen receptor-negative.

where one sample in  $n$  samples was randomly selected as the testing set, and the other  $n-1$  samples were regarded as the training set. The error rate ( $1 - \text{accuracy}$ ), when every single sample in the training sets has been used in the testing set, is the accuracy reference of the SVM classifier. The lower the error rate is, the more accurate the classifier is. Besides accuracy, another five indices were also utilized: sensitivity (Se), specificity (Sp), positive predictive value, negative predictive value, and area under receiver operating characteristic curve. For Se and Sp,  $P=0.5$  was the cutoff criteria; while area under the curve was a comprehensive assessment criteria. Subsequently, the accuracy of the SVM classifier was further verified using testing sets and the combined datasets according to the leave-one-out cross-validation method. If all of the earlier mentioned results suggest that the construed SVM classifier exhibits high reliability, DEGs collected from the training sets will be regarded as the feature genes to distinguish the two subtypes of breast cancer.

## Function- and pathway-enrichment of feature genes

The feature genes were subjected to enrichment analysis to identify their roles in breast cancer. The Database for

Annotation, Visualization and Integrated Discovery was used for function enrichment,<sup>20</sup> while KEGG Orthology Based Annotation System was applied for pathway-enrichment using hypergeometric distribution algorithm.<sup>21</sup>  $P < 0.05$  was the threshold for the enriched terms.

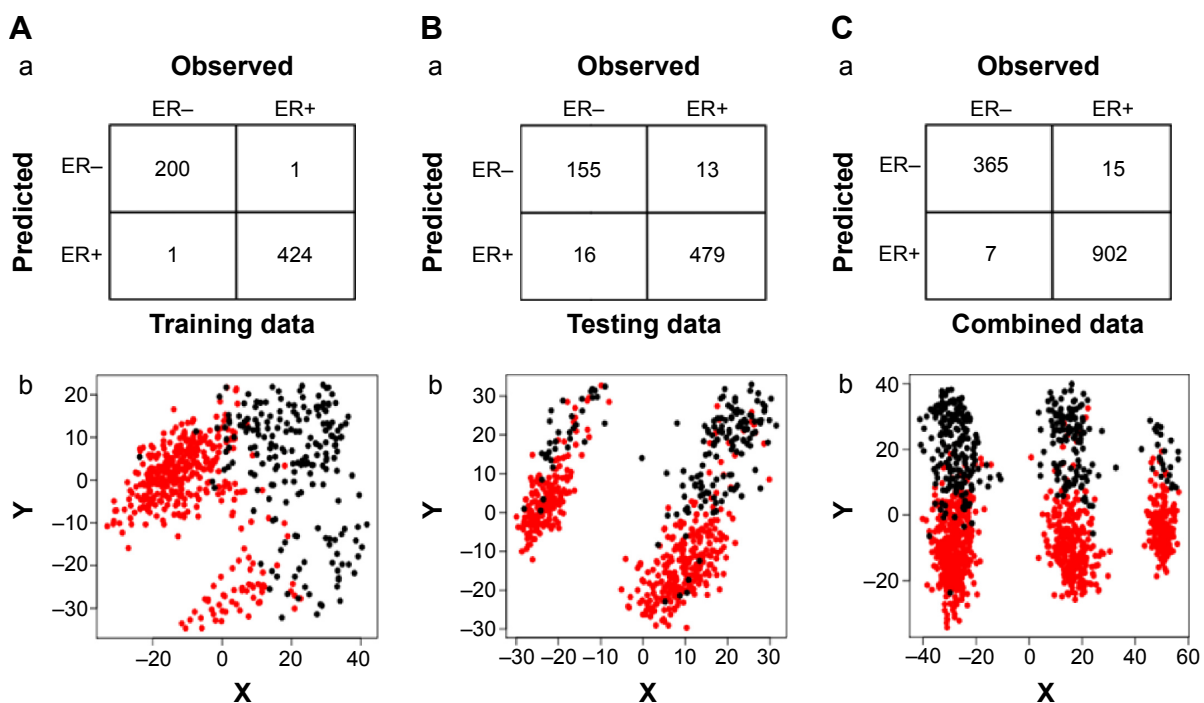
## Results

### Screening for DEGs

A total of 526 DEGs, consisting of 239 upregulated ones and 287 downregulated ones, were identified in ER+ samples, comparing with ER- samples in the five training datasets.

### Sample classification using SVM classifier

Using the normalized expression values of DEGs, a SVM classifier was constructed (Figure 1). After that, the accuracy of this classifier was detected. For training, testing, and the combined datasets, two (one ER- and one ER+), 29 (16 ER- and 13 ER+), and 22 (seven ER- and 15 ER+) samples were wrongly classified by the SVM classifier, respectively. However, the accuracies were all larger than 90% (99.7% [99.5% for ER- and 99.8% for ER+], 95.6% [90.6% for ER- and 97.3% for ER+], and 98.2% [98.1% for ER- and 98.4% for ER+], respectively), indicating the reliability of



**Figure 1** Classification of three sample datasets by constructed support vector machine classifier.

**Notes:** (A) Six hundred and twenty-six samples for training; (B) 663 samples for testing; (C) 1,289 combined samples for testing. (Aa, Ba, and Ca) indicate the sample distribution for ER+ and ER-. (Ab, Bb, and Cb) indicate the scatterplot of the classification, in which black dots represent ER- while red dots represent ER+ breast cancer samples.

**Abbreviations:** ER+, estrogen receptor-positive; ER-, estrogen receptor-negative.

**Table 2** Effect evaluation of the support vector machine classifier on training and testing datasets

	Number of samples	Correct rate	Sensitivity	Specificity	PPV	NPV	AUROC
Training	626	0.9968	0.9976	0.9950	0.9976	0.9950	0.999
Testing	663	0.9563	0.9736	0.9064	0.9677	0.9226	0.816
Combined	1,289	0.9829	0.9836	0.9812	0.9923	0.9605	0.890

**Abbreviations:** AUROC, area under receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

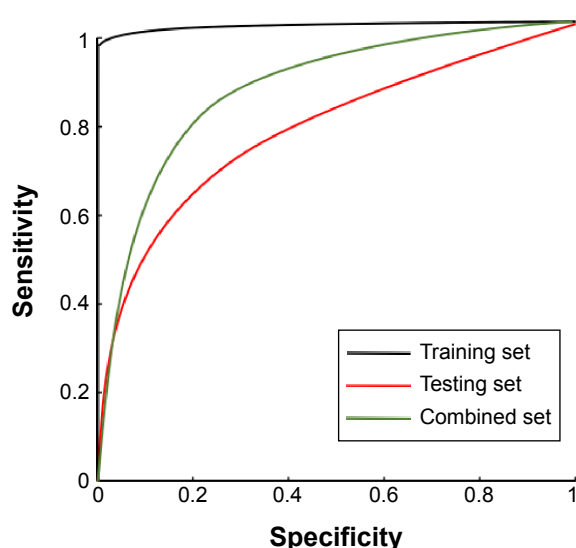
the classifier. Moreover, the results of Se, Sp, positive predictive value, negative predictive value, and area under receiver operating characteristic curve of the SVM classifier showed that it could not only distinguish training datasets, but also testing datasets well (Table 2; Figure 2).

### Significantly related functions of feature genes

A total of eight biological functions were enriched by the feature genes (Table 3), among which the response to inorganic substance was the most significant. Furthermore, the majority of features genes were collected in several biological functions, including response to organic substance, cell–cell signaling, response to wounding, behavior, and inflammatory response, each of which accounting for larger than 10%.

### Significantly related pathways of feature genes

Eight pathways of the feature genes were enriched, of which, cytokine–cytokine receptor interaction was the most significant one, involving 17 genes. The other pathways included drug metabolism, GnRH signaling pathway, etc (Table 4).



**Figure 2** Receiver operating characteristic curve used for training, testing, and combined datasets by support vector machine classifier.

## Discussion

In present study, ER+ and ER– types of breast cancer samples were investigated to screen the DEGs, which were further used for SVM classifier training. The SVM classifier could not only distinguish the training dataset, but could also distinguish the testing and combined datasets well, with accuracy higher than 90%. Thus, the DEGs could be considered as feature genes for ER+ and ER– types of breast cancer. Subsequently, function- and pathway-enrichment analyses were conducted, in which the inflammatory- and hormone-related biological process were the common results for these two different analyses, indicating these inflammatory (ie, *interleukin-8 [IL8]*) and hormone regulation (ie, *glycoprotein hormones, alpha polypeptide, CGA*) genes may be important for distinguishing ER+ and ER– types of breast cancer.

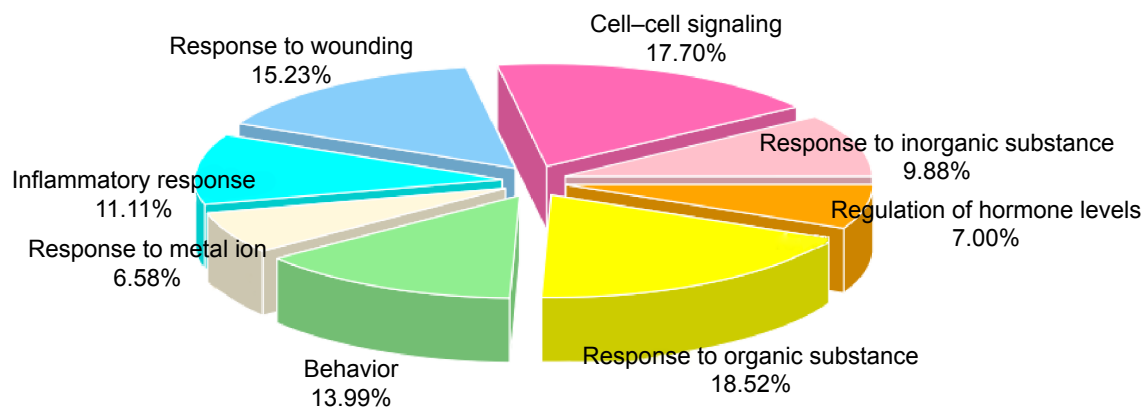
It is well reported that cytokine IL8 plays an important role in malignant tumor progression. *IL8* is highly expressed in breast cancer and is associated with an accelerated clinical course, a higher tumor load, and the presence of distant metastasis, ultimately leading to poor survival.<sup>22,23</sup> The depletion of *IL8* expression may promote the cell cycle arrest and inhibit migration and invasion in breast cancer cells, causing high response to chemotherapy.<sup>24,25</sup> Thus, *IL8* may be a biomarker for distinguishing subtypes of breast cancer because of the lower survival in ER– type of breast cancer. This hypothesis has been demonstrated by several studies.<sup>26,27</sup> For example, Lin et al showed that *IL8* is lowly expressed in ER+ whereas highly expressed in ER– cells.<sup>28</sup> Specifically, knockdown of *IL8* significantly reduces the cell invasion by suppressing the PI3K/Akt/NF-κB/integrin β3 pathway in ER– breast cancer cell lines,<sup>29,30</sup> and the microvessel density and neutrophil infiltration into the tumors in vivo.<sup>29</sup> Furthermore, exogenous addition of ERα in ER– cells may also achieve the goal of downregulating *IL8* expression.<sup>31</sup> These findings all suggest the negative relationship between *IL8* and ER status, which was also proved in our study (*IL8* was downregulated in ER+ samples, log<sub>2</sub> FC = –2.25, FDR = 1.04 × 10<sup>–70</sup>).

*CGA* codes for the common alpha subunit of four glycoprotein hormones (chorionic gonadotropin, luteinizing

**Table 3** Significantly enriched biological functions of feature genes

Term	Count	P-value	Enriched genes
GO:0010035, response to inorganic substance	24	$1.28 \times 10^{-7}$	SYT1, ERBB4, CRYAB, S100A7, EEF1A2, ALDOB, TRPA1, NR4A2, GGH, SOD2..... IGFBP2
GO:0007267, cell–cell signaling	43	$1.02 \times 10^{-6}$	LALBA, CGA, SYT1, CXCL5, NDP, S100A9, CXCL9, CCL8, CACNB2, GABBR2..... GDF15
GO:0009611, response to wounding	37	$1.2 \times 10^{-5}$	CXCL1, TF, S100A8, ERBB2, S100A9, CXCL9, CCL8, CXCL11, CDH3, CXCL10..... IGFBP4
GO:0006954, inflammatory response	27	$1.29 \times 10^{-5}$	CXCL1, TF, S100A8, S100A9, CXCL9, CCL8, CXCL11, CXCL10, FOS, IL17B..... IGFBP4
GO:0010038, response to metal ion	16	$1.3 \times 10^{-5}$	SYT1, ALDOB, GGH, PCSK1, FGG, CCND1, PLA2G4A, GRIA2, FGB, CYBRD1..... IGFBP2
GO:0007610, behavior	34	$1.38 \times 10^{-5}$	CXCL1, ADCY1, CXCL5, S100A9, UCHL1, CXCL9, CCL8, TRH, ZIC1, CXCL11..... CARTPT
GO:0010033, response to organic substance	45	$1.94 \times 10^{-5}$	CGA, TF, KYNU, ADCY1, CYP1B1, ERBB4, IL6ST, ERBB2, ARNT2, ALDOB..... IGFBP2
GO:0010817, regulation of hormone levels	17	$2.2 \times 10^{-5}$	KLK6, CGA, CYP1B1, TBX3, FOXA1, AFP, PCSK1, DHRS2, WNT4, SERPINA6..... SNAP25

**Abbreviation:** GO, gene ontology.

**Figure 3** Significantly enriched functions of feature genes.

**Note:** Different colors represent different biological functions and the percentages are the ratios of genes in a function.

**Table 4** Significantly enriched pathways of feature genes

ID	Pathway	P-value	Enriched genes
hsa04060	Cytokine–cytokine receptor interaction	$2.03 \times 10^{-2}$	CXCL1, IL1R2, IL2RA, IL8, CXCL5, IL6ST, CXCL9, CCL8, CXCL11, CCL18, CXCL10, IL17B, CXCL14, CCL20, IL20RA, BMPRI1, LTB
hsa00982	Drug metabolism	$2.05 \times 10^{-2}$	GSTA1, FMO5, FMO3, UGT2B4, CYP2A6, CYP2A7, GSTP1
hsa04912	GnRH signaling pathway	$2.07 \times 10^{-2}$	CGA, ADCY1, PLA2G4A, ADCY9, MAPK14, CAMK2B, CALML5, CACNA1D, ITPRI
hsa00232	Caffeine metabolism	$2.28 \times 10^{-2}$	NAT1, CYP2A6, CYP2A7
hsa04062	Chemokine signaling pathway	$3.00 \times 10^{-2}$	CXCL1, ADCY1, VAV3, IL8, CXCL5, CXCL9, CCL8, CXCL11, CCL18, CXCL10, CXCL14, CCL20, ADCY9
hsa04914	Progesterone-mediated oocyte maturation	$3.01 \times 10^{-2}$	PGR, IGF1R, ADCY1, ADCY9, MAPK14, BUB1, IGF2, CDC25A
hsa04110	Cell cycle	$3.03 \times 10^{-2}$	CCNE1, CCND1, CDC45, CDKN2A, BUB1, TTK, CDC20, SFN, MCM4, CDC25A
hsa00380	Tryptophan metabolism	$4.98 \times 10^{-2}$	WARS, KYNU, CYP1B1, IDO1, TPH1



hormone, follicle stimulating hormone, and thyroid stimulating hormone) that have a cystine knot motif formed by three of the five disulfide bonds.<sup>32,33</sup> This cystine knot motif is known to be a characteristic feature of growth factor and, thus, the expression of *CGA* may be related with the development of cancer, which was also demonstrated in previous studies.<sup>34,35</sup> However, compared with the beta subunit of glycoprotein hormones,<sup>36</sup> the alpha subunit may be a marker of tumors with low aggressiveness, eg, ER+ but not ER- cancer cells.<sup>37,38</sup> In this study, we also found the upregulated expression of *CGA* in ER+ breast cancer patients (log2 FC =2.63, FDR =2.82×10<sup>-34</sup>).

Despite the ideal classification of different types of breast cancer samples and satisfactory accuracy, the SVM classifier showed decreased recognizing ability. The potential reasons for this are: 1) breast cancer is one of the tumors which vary between individuals, and this variation among samples will affect gene expressions; and 2) samples used for training and testing were obtained from different experiments, which allows for some personal error. This kind of error can hardly be eliminated by normalization of the data. However, all the other indices showed reliability of our classifier by SVM method.

## Conclusion

Based on a set of gene-expression profiles, 526 DEGs were identified in ER+ samples in comparison with ER- samples, which were further used for SVM classifier construction. After being tested using the other microarray data, the SVM classifier showed satisfactory efficacy. The selected feature genes (such as *IL8* and *CGA*) could well distinguish those two subtypes of breast cancer. However, further experimental studies are needed to confirm the values of other involved genes.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9–29.
2. Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status – a metaanalysis. *Int J Cancer*. 2009;124(3):698–712.
3. Yamashita H, Iwase H, Toyama T, et al. Estrogen receptor-positive breast cancer in Japanese women: trends in incidence, characteristics, and prognosis. *Ann Oncol*. 2011;22(6):1318–1325.
4. West M, Blanchette C, Dressman H, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A*. 2001;98(20):11462–11467.
5. Lim S, Janzer A, Becker A, et al. Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis*. 2010;31(3):512–520.
6. Mehta RJ, Jain RK, Leung S, et al. FOXA1 is an independent prognostic marker for ER-positive breast cancer. *Breast Cancer Res Treat*. 2012;131(3):881–890.
7. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet*. 2011;378(9805):1812–1823.
8. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–1167.
9. Haibe-Kains B, Desmedt C, Loi S, et al. A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst*. 2012;104(4):311–325.
10. Prat A, Parker JS, Fan C, Perou CM. PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. *Breast Cancer Res Treat*. 2012;135(1):301–306.
11. Li T, Zhang C, Ogihara M. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*. 2004;20(15):2429–2437.
12. Fan XJ, Wan XB, Huang Y, et al. Epithelial-mesenchymal transition biomarkers and support vector machine guided model in preoperatively predicting regional lymph node metastasis for rectal cancer. *Br J Cancer*. 2012;106(11):1735–1741.
13. Upstill-Goddard R, Eccles D, Ennis S, et al. Support Vector Machine classifier for estrogen receptor positive and negative early-onset breast cancer. *PLoS One*. 2013;8(7):e68606.
14. Barrett T, Suzek TO, Troup DB, et al. NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res*. 2005;33(Database issue):D562–D566.
15. Liu WM, Mei R, Di X, et al. Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics*. 2002;18(12):1593–1599.
16. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods*. 2003;31(4):265–273.
17. Zhang S. A comprehensive evaluation of SAM, the SAM R-package and a simple modification to improve its performance. *BMC Bioinformatics*. 2007;8:230.
18. Xie Y, Pan W, Khodursky AB. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics*. 2005;21(23):4280–4288.
19. Cawley GC, Talbot NLC. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognit*. 2003;36(11):2585–2592.
20. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1–13.
21. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res*. 2006;34(Web Server issue):W720–W724.
22. Benoy IH, Salgado R, Van Dam P, et al. Increased serum interleukin-8 in patients with early and metastatic breast cancer correlates with early dissemination and survival. *Clin Cancer Res*. 2004;10(21):7157–7162.
23. Zuccari DA, Leonel C, Castro R, et al. An immunohistochemical study of interleukin-8 (IL-8) in breast cancer. *Acta Histochemica*. 2012;114(6):571–576.
24. Shao N, Chen LH, Ye RY, Lin Y, Wang SM. The depletion of interleukin-8 causes cell cycle arrest and increases the efficacy of docetaxel in breast cancer cells. *Biochem Biophys Res Commun*. 2013;431(3):535–541.
25. Wu K, Katiyar S, Li A, et al. Dachshund inhibits oncogene-induced breast cancer cellular migration and invasion through suppression of interleukin-8. *Proc Natl Acad Sci U S A*. 2008;105(19):6924–6929.
26. Ghoneim HM, Maher S, Abdel-Aty A, Saad A, Kazem A, Demian SR. Tumor-derived CCL-2 and CXCL-8 as possible prognostic markers of breast cancer: CORRELATION with estrogen and progesterone receptor phenotyping. *Egypt J Immunol*. 2009;16(2):37–48.
27. Todorović-Raković N, Milovanović J. Interleukin-8 in breast cancer progression. *J Interferon Cytokine Res*. 2013;33(10):563–570.

28. Lin Y, Huang R, Chen L, et al. Identification of interleukin-8 as estrogen receptor-regulated factor involved in breast cancer invasion and angiogenesis by protein arrays. *Int J Cancer*. 2004;109(4):507–515.
29. Yao C, Lin Y, Chua MS, et al. Interleukin-8 modulates growth and invasiveness of estrogen receptor-negative breast cancer cells. *Int J Cancer*. 2007;121(9):1949–1957.
30. Shao N, Lu Z, Zhang Y, et al. Interleukin-8 upregulates integrin  $\beta 3$  expression and promotes estrogen receptor-negative breast cancer cell invasion by activating the PI3K/Akt/NF- $\kappa$ B pathway. *Cancer Lett*. 2015;364(2):165–172.
31. Freund A, Chauveau C, Brouillet J-P, et al. IL-8 expression and its possible relationship with estrogen-receptor-negative status of breast cancer cells. *Oncogene*. 2003;22(2):256–265.
32. Darling RJ, Wilken JA, Ruddon RW, Bedows E. Intracellular folding pathway of the cystine knot-containing glycoprotein hormone alpha-subunit. *Biochemistry*. 2001;40(2):577–585.
33. Alvarez E, Cahoreau C, Combarnous Y. Comparative structure analyses of cystine knot-containing molecules with eight aminoacyl ring including glycoprotein hormones (GPH) alpha and beta subunits and GPH-related A2 (GPA2) and B5 (GPB5) molecules. *Reprod Biol Endocrinol*. 2009;7:90.
34. Abbud RA, Takumi I, Barker EM, et al. Early multipotential pituitary focal hyperplasia in the alpha-subunit of glycoprotein hormone-driven pituitary tumor-transforming gene transgenic mice. *Mol Endocrinol*. 2005;19(5):1383–1391.
35. Beck-Peccoz P, Persani L, Faglia G. Glycoprotein hormone alpha-subunit in pituitary adenomas. *Trends Endocrinol Metab*. 1992;3(2):41–45.
36. Iles RK. Ectopic hCGbeta expression by epithelial cancer: Malignant behaviour, metastasis and inhibition of tumor cell apoptosis. *Mol Cell Endocrinol*. 2007;260–262:264–270.
37. Bieche I, Parfait B, Le Doussal V, et al. Identification of CGA as a novel estrogen receptor-responsive gene in breast cancer: an outstanding candidate marker to predict the response to endocrine therapy. *Cancer Res*. 2001;61(4):1652–1658.
38. Bièche I, Latil A, Parfait B, et al. CGA gene (coding for the alpha subunit of glycoprotein hormones) overexpression in ER alpha-positive prostate tumors. *Eur Urol*. 2002;41(3):335–341.

## OncoTargets and Therapy

### Publish your work in this journal

OncoTargets and Therapy is an international, peer-reviewed, open access journal focusing on the pathological basis of all cancers, potential targets for therapy and treatment protocols employed to improve the management of cancer patients. The journal also focuses on the impact of management programs and new therapeutic agents and protocols on

Submit your manuscript here: <http://www.dovepress.com/oncotargets-and-therapy-journal>

patient perspectives such as quality of life, adherence and satisfaction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress