Published in final edited form as:

Breast Cancer Res Treat. 2010 February ; 119(3): 593–599. doi:10.1007/s10549-009-0365-6.

Derivation of molecular signatures for breast cancer recurrence prediction using a two-way validation approach

Yijun Sun,

Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL 32611, USA

Virginia Urquidi, and

Department of Medicine, University of Florida, Jacksonville, FL 32009, USA

Goodison Steve

Department of Surgery, Shands Health Science Center, University of Florida, 655 West 8th St., Jacksonville, FL 32009, USA

Goodison Steve: steve.goodison@jax.ufl.edu

Abstract

Previous studies have demonstrated the potential value of gene expression signatures in assessing the risk of post-surgical breast cancer recurrence, however, many of these predictive models have been derived using simple computational algorithms and validated internally or using one-way validation on a single dataset. We have recently developed a new feature selection algorithm that overcomes some limitations inherent to high-dimensional data analysis. In this study, we applied this algorithm to two publicly available gene expression datasets obtained from over 400 patients with breast cancer to investigate whether we could derive more accurate prognostic signatures and reveal common predictive factors across independent datasets. We compared the performance of three advanced computational algorithms using a robust two-way validation method, where one dataset was used for training and to establish a prediction model that was then blindly tested on the other dataset. The experiment was then repeated in the reverse direction. Analyses identified prognostic signatures that while comprised of only 10–13 genes, significantly outperformed previously reported signatures for breast cancer evaluation. The cross-validation approach revealed CEGP1 and PRAME as major candidates for breast cancer biomarker development.

Keywords

Microarray; Breast cancer prognosis; Predictive model; PRAME

Introduction

Breast cancer is the second most common cause of death from cancer among women in the United States. In 2009, it is estimated that about 182,480 new cases of breast cancer will be diagnosed, and 40,930 women are expected to die from this disease. The major clinical problem of breast cancer is the recurrence of disseminated disease. Adjuvant therapy (chemotherapy and hormonal therapy) reduces the risk of distant metastases by one-third, however, it is

Correspondence to: Goodison Steve, steve.goodison@jax.ufl.edu.

Electronic supplementary material: The online version of this article (doi:10.1007/s10549-009-0365-6) contains supplementary material, which is available to authorized users.

estimated that about 70% patients receiving treatment would have survived without it [1]. Being able to predict disease outcomes more accurately will aid physicians and patients to make more informed decisions regarding adjuvant treatment, and may lead to the development of individually tailored treatments.

For breast cancer, two currently used treatment guidelines are the St. Gallen [2] and NIH [3] consensus criteria, which determine whether a patient is at a high risk of distant metastases based on a panel of clinical markers. However, these criteria are less than precise in predicting primary therapy failure, with only 10% specificity at the 90% sensitivity level. The advent of microarray gene expression technology has greatly enabled the search for predictive disease biomarkers through multivariate data analyses. From expression information on thousands of genes per tissue sample, breast cancer prognostic signatures of 70-genes [4] and 76-genes [5] have been derived that achieve a higher specificity than the current clinical systems at the same sensitivity level. These results are encouraging and large-scale randomized clinical validation studies are currently being conducted in Europe to evaluate the prognostic value of these gene signatures [6]. However, it is not yet clear whether this is the best performance we can achieve for breast cancer prognosis using molecular profiling, and the issue of whether the existing gene signatures are ready for clinical validation trials is still under debate in the oncology community [6,7]. Key to optimizing molecular signature derivation is the development of advanced computational feature selection algorithms capable of identifying relevant genes from many thousands of genes on the basis of a limited number of tissue samples [8]. Many existing algorithms are a trade-off between computational efficiency and solution accuracy [9]. We have recently derived a feature selection algorithm that addresses several major issues with prior work. We have experimentally demonstrated that our algorithm is capable of achieving close-to-optimal solutions in problems with extremely large input data dimensionality, to a point far beyond that needed for gene expression data analysis of genetically complex organisms [10–12 and supplemental data]. The application of the algorithm to prostate tissue microarray datasets enabled us to derive highly accurate prognostic molecular signatures for prostate disease recurrence [13].

In this study, we applied our computational approach to two publicly available breast cancer gene expression microarray datasets obtained from over 400 patients. We used a robust two-way validation process, where one dataset was used for training to establish a prediction model which was then blindly tested on the other dataset, and then the experiment was repeated in the reverse direction. In contrast, most existing studies reported in the literature have used either leave-one-out cross validation or one-way validation on a single dataset. We demonstrated that our approach derived a shorter, yet more accurate prognostic signature for both datasets than has been previously reported. Our results show that more advanced computational studies on high-dimensional gene expression data are required for optimization of molecular signatures for potential clinical use.

Materials and methods

Datasets

This analysis used two publicly available breast cancer gene expression datasets. The dataset referred to as the Nature dataset [4] in the text was obtained from 97 patients. The other independent dataset containing data from 307 patients is referred to as the JNCI dataset in the text [14]. Raw microarray data and clinical information are available at the European Bioinformatics Institute Array-Express database (<http://www.ebi.ac.uk/arrayexpress/>).

Feature selection algorithm

We used our previously developed feature selection algorithm that addresses several major issues with prior work, including the capability to handle problems with large data dimensionality [10–13,15 and see our website <http://plaza.ufl.edu/sunyijun/>]. The key idea is to decompose an arbitrary complex model into a set of locally linear ones through local learning, and then estimate feature relevance globally within a large margin framework. The algorithm is a generic feature selection method that performs without making any assumptions about the underlying data distribution. For details of the computational algorithm see our previous publications [10–13 and supplemental data]. The Matlab implementation of the algorithm is available upon request for validating reported results and academic research.

Experimental procedure

Due to the small sample size compared to the number of genes, special care must be taken in experimental protocols to avoid possible over-fitting of a computational model to training data. In the Nature dataset, the ratio of metastatic to non-metastatic cases evaluated was close to equal. Conversely, for the JNCI dataset, the ratio between the metastatic and non-metastatic cases is about 1:4. Thus, in the training process, we tripled the metastatic samples to balance the two classes. The parameters of prediction models were estimated using tenfold cross validation based solely on training data. The training data was first partitioned randomly into ten groups, and then one group of samples was held out for testing and the remaining samples were used for training. The experiment was repeated until each group had been tested. We used the area under a receiver operating characteristic curve (AUC) as a criterion to tune the parameters. After a predictive gene subset was identified, we trained support vector machines with both linear and RBF kernels to predict the outcome of test samples (i.e., the independent dataset). We found that the performance obtained by using both kernels had no significant difference, and hence only the results of the linear kernel are reported. To compare the findings with conventional algorithms we also used SVM-RFE [16] and ℓ_1 regularized logistical regression [17] to obtain prognostic signatures. The parameters for both of these algorithms were also estimated using tenfold cross-validation based on training data.

Statistical analysis

Kaplan–Meier survival plots and log-rank tests [18] were used to assess the predictive values of different prognostic approaches. The Mantel–Cox estimation of hazard ratio was performed to quantify the relative risk of biochemical recurrence in the bad-prognosis group compared with the good-prognosis group. A hazard ratio above 1.0 indicates that the patients assigned to the bad-prognosis group have a higher probability to develop disease recurrence than those in the good-prognosis group. A receiver operating characteristic (ROC) curve obtained by varying a decision threshold provides a direct view on how a predictive approach performs at the different sensitivity and specificity levels. The specificity is defined as the probability that a patient who did not experience disease recurrence was assigned to the good-prognosis group, and the sensitivity is the probability that a patient who developed disease recurrence was in the bad-prognosis group. MedCalc version 8.0 (MedCalc Software, Mariakerke, Belgium) was used to perform the ROC curve analysis. A P value of 0.05 was considered statistically significant.

Results and discussion

Prediction models for assessing breast cancer recurrence

We applied our computational approach to two publicly available breast cancer gene expression datasets. The first one, referred to as the Nature data hereafter, has been used in [4] to derive the 70-gene prognosis signature. The dataset contains 24,481 probes that measure the gene

expression levels in tumor samples collected from 97 breast cancer patients. Among them, 46 developed distant metastases within 5 years, and 51 remained metastasis free for at least 5 years. The second independent dataset, referred to as the JNCI dataset, has been used to validate the prognostic value of the 70-gene signature [14]. This dataset consists of 1,145 gene expression values of 307 patient samples, including 64 that developed distant metastases within 5 years, and 243 who remained metastasis-free for at least 5 years. In order to perform a two-way validation, we were only able to use the 1,141 genes that were common to both datasets. The task was to construct a prediction model that would enable us to accurately predict the risk of distant recurrence of breast cancer within a 5-year post-surgery period.

We demonstrated the predictive values of our prognostic classifier models by comparing their performance with those of the clinical St. Gallen criterion, and to results we obtained using SVM-RFE [18] and ℓ_1 regularized logistical regression [17], two conventional algorithms often used in microarray data analysis. Figure 1 presents the receiver operating characteristic (ROC) curves of the three computational methods performed on the Nature and JNCI datasets. Following the study of [4], a threshold is set for each classifier so that the sensitivity of each classifier is equal to 90%. The corresponding specificities derived from the ROC plots are reported in Table 1. While both SVM-RFE and ℓ_1 regularized logistical regression significantly outperform the St. Gallen criterion, our method achieved by far the best specificities, 53 and 61% on the JNCI and Nature datasets, respectively. Our method also gave the highest odds ratios (OR) at 9.4 (95% CI: 3.3–27.1) for the JNCI data, and 16.3 (95% CI: 5.1–52.4) for the Nature data (Table 1). The St. Gallen criterion classified only a few samples into the good prognosis group, and the estimates of odds ratios are not reliable and hence were omitted.

To further demonstrate the performance of molecular signatures in assessing the risk of developing distant metastases, survival data analyses were performed. The Kaplan–Meier curves of our predictive models are plotted in Fig. 2, and shows a significant difference in the probability of remaining free of distant metastases in patients with good and bad prognosis (P value < 0.001). The calculated Mantel–Cox estimate of hazard ratios of distant metastases within 5 years for our model were 8.4 (95% CI: 3.0–23.6) for the JNCI data, and 10.2 (95% CI: 3.4–29.9) for the Nature data, which are much larger than those obtained using the SVM-RFE and the ℓ_1 regularized logistical regression models (Table 1).

Finally, we compared the performance of our predictive classifier derived from the Nature dataset with that of the 70-gene signature that was previously derived from the same data [4]. The comparison is somewhat in favor of the 70-gene signature, since it was derived using the full gene set while ours was derived using only the 1,141 genes common to both datasets. We note that the hazard ratios obtained for the 70-gene signature were relatively low (42%), which is consistent with the results reported in [14]. Our signature performed better than the 70-gene signature, in terms of specificity, AUC and hazard ratio (Table 2). Moreover, our high-performance signature consisted of only 13 genes (Table 3).

The gene signatures generated by training with the Nature and JNCI datasets contained 13 and 10 genes, respectively, with two genes present in both gene signatures. Of the thirteen genetic markers of the prognostic gene signature derived from the Nature data, eight (CEGP1, L2DTL, MMP9, IGFBP5, FLT1, AL080059, FLJ11354, and Contig24252) were listed in the 70-gene prognosis signature derived from the same dataset. Of the ten genes derived from the JNCI data, four genes (CEGP1, ESM1, FLJ11190, Contig20217) were present in the 70-gene signature. We also investigated the potential overlap between gene signatures generated when using our algorithm and the established algorithms SVM-RFE and ℓ_1 regularized logistical regression on the same training and test data. The prognostic signatures derived using these algorithms were substantially higher in gene content than ours. Except for the signature obtained using ℓ_1 regularized logistical regression trained on the Nature data, the other

signatures contained >50 genes. Because of the high number of genes in these other signatures, it was not surprising that the genes in our minimal signature were also present in the signatures generated by the other two methods. Of note, CEGP1 and PRAME were present in all signatures derived trained on the JNCI data, however, neither SVM-RFE nor ℓ_1 regularized logistical regression identified both CEGP1 and PRAME in signatures derived by training on the Nature dataset (see supplementary data).

The ascribed function of many of the genes in these signatures aligns well with mechanisms involved in breast cancer progression and dissemination. For example, cell cycle regulation and proliferation (L2DTL, ESM1), angiogenesis (FLT1), and extracellular matrix remodeling and invasion (MMP9, ESM1, IGFBP5). However, of particular interest are the two genes (CEGP1 and PRAME) that were included in the prognostic signatures derived from both the Nature and the JNCI datasets. PRAME was not included in either the 70-gene signature, nor a larger 231 geneset identified as being associated with disease outcome in the Nature study [4]. The CEGP1 gene (also known as SCUBE2) is located on human chromosome 11p15 and has homology to the basic helix-loop-helix (bHLH) family of transcription factors. The biological role of CEGP1/SCUBE2 is unknown, but the gene encodes a secreted and cell-surface protein containing EGF and CUB domains [19]. The expression of CEGP1/SCUBE2 has been reported to be associated with estrogen receptor status in breast cancer specimens [20]. In our previous work, we did find the CEGP1 to be significantly over-expressed in our studies of a breast tumor metastases model [21], and intriguingly, CEGP1 is one component of a 21 gene set that comprises a commercial breast cancer predictive test (OncotypeDX Recurrence Score from Genomic Health) currently under clinical trial [22].

The preferentially expressed antigen in melanoma (PRAME) gene has been linked to human disease, including cancer. PRAME is a cancer-testis antigen (CTA), a group of tumor-associated antigens that represent possible target proteins for immuno-therapeutic approaches. Expression of CTAs is high in a variety of malignancies but is negligible in healthy tissues [23]. Reports suggest that over-expression of PRAME in human cancers confers growth or survival advantages by antagonizing retinoic acid (RA) signaling, thus inhibiting RA-induced differentiation, growth arrest, and apoptosis [24]. Using quantitative PCR, Doolan et al. reported that PRAME was detected in the majority of tumor specimens and the minority of normal breast specimens, and multivariate analysis indicated PRAME expression to be an independent prognostic factor for shortened disease-free survival and overall survival [25]. In our own previous analyses of a breast cancer microarray dataset revealed PRAME to be useful in conjunction with clinical parameters in outcome prediction [15], and a subsequent targeted interrogation of PRAME expression in another microarray dataset also found an association with outcome [26]. In line with the findings reported in this study, these results suggest that PRAME expression may be a promising prognostic and predictive marker for breast cancer.

Previous studies have demonstrated the potential value of gene expression signatures in assessing the risk of post-surgical disease recurrence, and have laid a solid foundation for future studies on breast cancer prognosis. However, many existing predictive models have been derived using simple computational algorithms. Whether this is the best we can perform using genetic information is rarely addressed in the literature. The derivation of disease-associated molecular signatures is necessarily an ongoing, dynamic process, and in this report, we have demonstrated that an advanced feature selection algorithm can identify molecular signatures with dramatically reduced complexity, yet greater prognostic performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Cancer Institute under grant NIH/NCI RO1 CA108597 (SG), and Susan Komen Breast Cancer Foundation grant BCTR0707587 (YS). We thank those investigators that performed the original molecular tissue profiling and for sharing data through public gene expression databases.

References

1. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998;352:930–942.10.1016/S0140-6736(98)03301-7 [PubMed: 9752815]
2. Goldhirsch A, Wood WC, Gelber RD, et al. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 2003;21:3357–3365.10.1200/JCO.2003.04.576 [PubMed: 12847142]
3. National Institutes of Health Consensus Development Panel. National Institutes of Health consensus development conference statement: adjuvant therapy for breast cancer. *J Natl Cancer Inst* 2001;30:5–15.
4. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–536.10.1038/415530a [PubMed: 11823860]
5. Wang Y, Klijn J, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671–679. [PubMed: 15721472]
6. Loi S, Sotiriou C. Molecular forecasting of breast cancer: time to move forward with clinical testing. *J Clin Oncol* 2006;24:721–722.10.1200/JCO.2005.04.6524 [PubMed: 16446348]
7. Brenton JD, Carey LA, Ahmed AA, et al. Molecular classification and molecular forecasting of breast cancer: ready for clinical application. *J Clin Oncol* 2005;23:7350–7360.10.1200/JCO.2005.03.3845 [PubMed: 16145060]
8. Lafferty J, Wasserman L. Challenges in statistical machine learning. *Statist Sinica* 2006;16:307–322.
9. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007;23:2507–2517.10.1093/bioinformatics/btm344 [PubMed: 17720704]
10. Sun, Y.; Todorovic, S.; Goodison, S. A feature selection algorithm capable of handling extremely large data dimensionality. *Proceedings of 8th SIAM international conference on data mining*; 2008. p. 530-540.
11. Sun Y, Wu D. Feature extraction through local learning. *Stat Anal Data Min*. 2009 in press.
12. Sun Y. Iterative RELIEF for feature weighting: algorithms, theories, and applications. *IEEE Trans Pattern Anal Mach Intell* 2007;29:1035–1051.10.1109/TPAMI.2007.1093 [PubMed: 17431301]
13. Sun, Y.; Cai, Y.; Goodison, S. Combining nomogram and microarray data for predicting prostate cancer recurrence. *Proceedings of 8th IEEE international conference on bioinformatics and bioengineering*; 2008. p. 1-710.
14. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;98:1183–1192. [PubMed: 16954471]
15. Sun Y, Goodison S, Li J, Liu L, et al. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics* 2007;23:30–37.10.1093/bioinformatics/btl543 [PubMed: 17130137]
16. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;46:389–422.10.1023/A:1012487302797
17. Ng, AY. Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of 21st international conference on machine learning*; 2004. p. 78-86.
18. Sterne Kirkwood, B. *Essential medical statistics*. Blackwell Publishing; Oxford: 2003.
19. Yang RB, Ng CKD, Wasserman SM, et al. Identification of a novel family of cell-surface proteins expressed in human vascular endothelium. *J Biol Chem* 2002;277:46364–46373.10.1074/jbc.M207410200 [PubMed: 12270931]
20. Abba MC, Hu Y, Sun H, et al. Gene expression signature of estrogen receptor status in breast cancer. *BMC Genomics* 2005;6:37.10.1186/1471-2164-6-37 [PubMed: 15762987]

21. Goodison S, Yuan J, Sloan D, et al. The RhoGAP protein DLC-1 functions as a metastasis suppressor in breast cancer cells. *Cancer Res* 2005;65:6042–6053.10.1158/0008-5472.CAN-04-3043 [PubMed: 16024604]
22. Asad J, Jacobson AF, Estabrook A, et al. Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *Am J Surg* 2008;196(4):527–529.10.1016/j.amjsurg.2008.06.021 [PubMed: 18809056]
23. Juretic A, Spagnoli G, Schultz-Thater E, et al. Cancer/testis tumour-associated antigens: immunohistochemical detection with monoclonal antibodies. *Lancet Oncol* 2003;4:104–109.10.1016/S1470-2045(03)00982-3 [PubMed: 12573352]
24. Epping MT, Wang L, Edel MJ, et al. The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. *Cell* 2005;122:835–847.10.1016/j.cell.2005.07.003 [PubMed: 16179254]
25. Doolan P, Clynes M, Kennedy S, et al. Prevalence and prognostic and predictive relevance of PRAME in breast cancer. *Breast Cancer Res Treat* 2008;109:359–365.10.1007/s10549-007-9643-3 [PubMed: 17624586]
26. Epping MT, Hart AAM, Glas AM, et al. PRAME expression and clinical outcome of breast cancer. *Br J Cancer* 2008;99:98–403.10.1038/sj.bjc.6604494

Fig. 1.

Receiver operating characteristic plots comparing the predictive performance of three computational methods on two independent datasets (Nature and JNCI). Analysis performed using our computational approach (*top left panel*); L1-regularized logistical regression (*top right panel*); SVM-RFE (*lower panel*). AUC, area under the curve

Fig. 2.

Kaplan-Meier estimation of the probabilities of patients with a good or bad prognostic signature remaining metastasis free. Signatures were derived using a two-way validation process in which one of two independent datasets was used to train and establish a prediction model which was then blindly tested on the other dataset. In the upper panel (JNCI-Nature) we used the JNCI dataset to train and the Nature dataset to test. The P values were computed by log-rank test

Table 1

Comparison of prognostic results (at 90% sensitivity level) obtained using our computational method, SVM-RFE, ℓ_1 regularized logistic regression (ℓ_1 -LR), and the St. Gallen clinical criterion in a two-way validation of independent datasets (Nature and JNCI)

Methods	Train-Test	AUC	Specificity (%)	Odds ratio (95% CI)	Hazard ratio (HR)	HR (95% CI)	P value
Our method	Nature-JNCI	0.72	53	9.4 (3.3–27.1)		8.4 (3.0–23.6)	<0.001
	JNCI-Nature	0.80	61	16.3 (5.1–52.4)		10.2 (3.4–29.9)	<0.001
SVM-RFE	Nature-JNCI	0.66	40	7.6 (2.6–21.5)		6.6 (2.4–18.3)	<0.001
	JNCI-Nature	0.69	40	5.3 (1.8–15.7)		3.8 (1.5–9.6)	<0.001
ℓ_1 -LR	Nature-JNCI	0.65	33	6.2 (2.1–18.3)		4.3 (1.7–11.1)	<0.001
	JNCI-Nature	0.70	39	4.1 (1.6–10.8)		3.8 (1.5–9.7)	<0.001
St. Gallen	JNCI	–	10	–		–	–
	Nature	–	12	–		–	–

Both odds ratio and hazard ratio were computed using 5-year follow-up data

Table 2

Comparison of the predictive values of our 13-gene prognostic signature and the 70-gene signature, both derived from the Nature dataset, when tested on the JNCI data

Methods	Train-Test	AUC	Specificity (%)	Hazard ratio (HR)	P value
				HR (95% CI)	
Our method	Nature-JNCI	0.72	53	2.5 (1.5-4.3)	<0.001
70-Gene [12]	Nature-JNCI	0.68	42	2.3 (1.4-4.0)	0.002

Specificities are reported at the 90% sensitivity level

Table 3

Genes comprising the 13-gene prognostic signature derived from training on the Nature dataset and testing on the JNCI dataset, and the 10-gene signature derived in the reverse direction

Training data	Gene name	Sequence description
Nature data	L2DTL	L2DTL protein
	CEGP1 [*]	Homo sapiens CEGP1 protein (CEGP1), mRNA
	AL080059	Homo sapiens mRNA; cDNA DKFZp564H142
	PEX12	Peroxisomal biogenesis factor 12
	PRAME [*]	Preferentially expressed antigen in melanoma
	MMP9	Matrix metalloproteinase 9 (gelatinase B, 92kD gelatinase, 92kD type IV collagenase)
	DKFZP761L0424	Hypothetical protein DKFZp761L0424
	IGFBP5	Human insulin-like growth factor binding protein 5 (IGFBP5) mRNA
	FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
	NSP3	Novel SH2-containing protein 3
	FLJ11354	Homo sapiens cDNA FLJ11354 fis, clone HEMBA1000129, weakly similar to HYPOTHETICAL HELICASE C8A4.08C IN CHROMOSOME 1
	Contig24252-RC	ESTs
	Contig27312-RC	ESTs
JNCI data	KIAA0008	KIAA0008 gene product
	CEGP1 [*]	Homo sapiens CEGP1 protein (CEGP1), mRNA
	TRIP13	Thyroid hormone receptor interactor 13
	ACADS	Acyl-coenzyme A dehydrogenase, C-2 to C-3 short chain
	PRAME [*]	Preferentially expressed antigen in melanoma
	ESM1	Endothelial cell-specific molecule
	FLJ11190	Hypothetical protein FLJ11190
	SEC14L2	SEC14 (<i>S. cerevisiae</i>)-like 2
	Contig20217-RC	ESTs
	EPAC	Rap1 guanine-nucleotide-exchange factor directly activated by cAMP

^{*} Present in both signatures