

22-Gene Signature Predicts Response to Chemotherapy in a Broad Range of Breast Cancer Subtypes



Katherine J. Martin, David Magee and Marcia V. Fournier;
BioArray Therapeutics, Venture Development Center at UMass, Boston

Background

Currently only ~50% of breast cancer patients respond to chemotherapy with long term survival*

There is a need for predictive tests to determine ahead of time if an individual patient will respond to a particular treatment

Patients predicted to be non-responsive would benefit from immediately receiving an alternative treatment

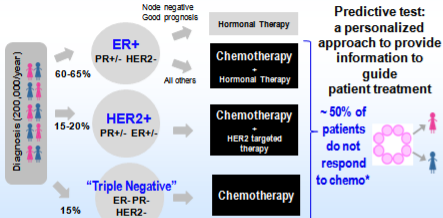


Figure 1. Outline of current treatment groups in breast cancer. *Citron, et al., 2003 J Clinical Oncology 21:1431-1439.

Approach

A novel approach was used to find genes with RNA expression levels that predict chemotherapy response

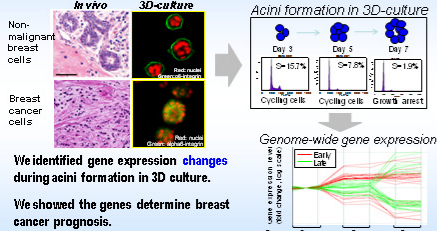


Figure 2. Approach used to identify 22-gene signature. *Fournier, et al., 2006, Cancer Research 66:7095-7102.

Methods

Microarray datasets. This study used 5 microarray datasets from a total of 610 patients.

Signature discovery. A time course of acini formation in 3D culture. Fournier, et al., 2006 Cancer Res, 66:7095.

Affymetrix HG-U133A microarray (GEO GSE2094).

Evaluation of response prediction. 3 overlapping datasets were used, all obtained at MD, Anderson Medical Center from fine needle tumor or aspirates from patients with stage III breast cancer obtained before reassignment combination treatment with paclitaxel, 5-fluorouracil, cyclophosphamide and doxorubicin (TFAC) followed by surgical resection. Response was categorized as pathological complete response (pCR), i.e. no residual invasive cancer in breast or nodes) or residual disease (RD). Affymetrix HG-U133A microarrays. Dataset of Hess, et al., 2006 J Clin Oncol, 24:2626 included 133 patients. Dataset of Popovici, et al., 2010 Breast Cancer Res 12:95 included 243 patients (GEO GSE20194). Dataset of Tabchy, et al., 2010, Clin Cancer Res 16: 5351-5361 included 79 patients (GEO GSE20271).

Evaluation of prognosis. A set of 286 lymph node negative patients with 5 year relapse as an endpoint. Wang et al., 2005, Lancet 365:671-679 (GEO GSE23049).

AUC (Area Under Curve). was determined from Receiver Operating Characteristic (ROC) curves with 3-fold cross validation. AUC is a value that incorporates both sensitivity and specificity into a single value that quantifies "how good is a test". AUC's range from 1.00 (perfect test) to 0.50 (random).

Molecular classes for tumors in dataset of Popovici 2010, were determined using the intrinsic gene set of 300 genes (Hu, et al., 2006). Expression values were organized by hierarchical clustering with Pearson metric. Clusters were identified as Luminal A = high ESR1, low AURKA; Luminal B = high ESR1, low AURKA; ERBB+ = high ERBB; ERBB-like = low ESR1, high KRT5.

Figure 4. Hierarchical cluster analysis to assign molecular class to 243 samples from dataset of Popovici et al., 2010.

Table 1. Numbers of samples for molecular class and response categories

	Actual number	Percentages	Total
Basal-like	42	27	60
ERBB+	8	11	19
LuminalA	55	1	56
LuminalB	43	7	50
Unclassified	43	5	48
Total	191	51	242

How does 22-gene signature predict chemotherapy response?

We mapped the 22 genes onto the signaling network that regulates the malignant phenotype. This network regulates tumor proliferation and survival in response to various agents.

We found that the 22 genes function at different sites across this network. Hence, we hypothesize the signature is like series of "surveillance cameras" that can readout signaling network status in tumor or cells. Given the proper algorithms, this information has the potential to predict a range of properties of tumor cells.

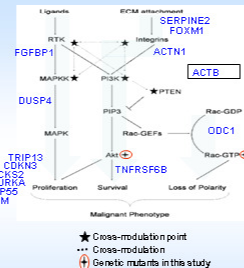


Figure 3. We mapped the 22 genes onto the breast cancer signaling network of Liu, et al., 2004, J. Cell Biology, 164:603-612.

Results

Table 2. Gene sets down-regulated during acini formation are enriched in genes associated with response to TFAC chemo.

Temporal expression pattern	Total genes (N)	Genes significantly associated with pCR (N)	(%)	Ability to stratify by response** (Chi ² coefficient)	(p-value)
Down early	6	3	50%	0.248	0.0005
Down late	22	12	55%	0.364	<0.000001
Up early	4	2	5%	-	-
Up late	11	2	18%	-	-
Down	28	15	54%	0.241	0.00059
Up	32	3	9%	-	-
Early	27	6	22%	-	-
Late	33	14	42%	0.344	<0.000001
All differentials	60	22	37%	0.283	<0.000001
All genome	22282	3766	17%	-	-
840 random	22	3.73	17%	-	-

**Test, p<0.05, was used to evaluate genes associated with response (pCR) in the TFAC response microarray dataset of Popovici et al. 2010 (243 patients).

**Hierarchical clustering was used to stratify patients from the TFAC response microarray dataset of Hess et al. 2006 (133 patients). Chi² coefficient and Fisher's Exact p-values are tabulated.

Table 3. Univariate analysis comparing genes associated with prognosis and TFAC response prediction in multiple subtypes.

Gene Symbol	PROGNOSIS				PREDICTION of TFAC Response			
	All	ER+	ER-	All	ER+	ER-	All	
1 CDKN2A	0.0002	0.0001	0.0001	0.0002	0.0001	0.0001	0.0001	
2 PCDH11	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
3 SLC12A1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
4 SLC12A1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
5 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
6 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
7 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
8 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
9 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
10 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
11 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
12 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
13 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
14 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
15 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
16 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
17 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
18 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
19 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
20 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
21 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
22 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
23 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
24 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
25 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
26 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
27 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
28 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
29 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
30 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
31 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
32 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
33 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
34 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
35 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
36 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
37 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
38 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
39 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
40 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
41 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
42 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
43 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
44 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
45 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
46 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
47 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
48 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
49 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	
50 CDKN2A	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	

Although the 22-gene signature includes both prognostic and predictive genes, results here show these features are not necessarily overlapping. Some genes predicted prognosis but not response and vice-versa, while some predicted both. Further, the 22-genes includes subsets that predicted response in different subtypes of breast cancer. Prognosis <0.05 for prognostic categories; see highlighted blue genes; values <0.10 for predictive categories; see highlighted pink. Association of gene expression with prognosis was assessed in the microarray dataset of Wang, et al., 2005, using log₂-linear analysis with 5 year follow up as an endpoint. Association of gene expression with TFAC response prediction was assessed in the microarray dataset of Popovici, et al., 2010, using discovery logistic regression analysis with pCR as an endpoint.

Results

Table 4. 22-gene signature stratified breast cancer subtypes by response to TFAC chemotherapy and outperformed clinical parameters.

Breast Cancer Subtype	22-genes	AUC Value* (n)			
		Node status	ER status	Tumor size	Tumor grade
ER Positive	0.723 (208)	0.400	-	0.475	0.680
ER Negative	0.744 (145)	0.481	-	0.525	0.689
HER2 Positive	0.772 (42)	0.513	-	0.525	0.316
Triple Negative (ER, PR, HER2 negative)	0.718 (95)	0.490	-	0.525	0.689
Luminal B	0.750 (50)	-	-	-	-
Basal-like	0.85 (69)	-	-	-	-
All subtypes	0.830 (353)	0.478	0.760	0.525	0.689

*AUC values for 22-gene signature test and clinical parameters were determined by logistic regression with 3-fold cross validation using the datasets of Popovici et al. 2010 and Tabchy et al. 2010.

Conclusions

Genes down regulated during acini formation in 3D culture were associated with tumor response to TFAC chemotherapy. The 22-gene signature (down-regulated late in acini formation) accurately predicted TFAC response across a broad range of breast cancer subtypes and outperformed clinical parameters. The signature includes both prognostic and predictive features, which are not necessarily overlapping. It also includes different genes that predict response in different subtypes of breast cancer.

Future Perspectives

- Our current work is directed to:
 - Develop RT-PCR version of test
 - Evaluate test parameters including sensitivity, specificity, PPV, etc, in RT-PCR version
 - Define ability of test to predict response to other chemo's
 - Bring test to patients - as soon as possible