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

BIOARRAY Therapeutics
www.bioarray.us

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

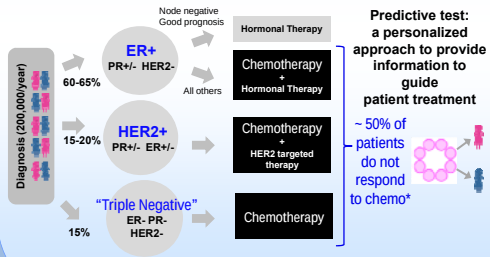


Figure 1. Outline of current treatment groups in breast cancer. *Citron, et al., 2003 J Clinical Oncology 21:1431-1439.

Methods

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

Gene discovery: A time course of acini formation in 3D culture. Fournier, et al., 2006 Cancer Res, 66:7095. Affymetrix HG-U133A microarrays (GEO GSE8096).

Evaluation of response prediction: 3 overlap-ping datasets were used, all obtained at MD Anderson Medical Center from fine-needle tumor aspirates from patients with stage I-III breast cancer obtained before neoadjuvant combination treatment with paclitaxel, 5-fluorouracil, cyclophosphamide and doxorub-icin (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, 24:4236 included 133 patients. Dataset of Popovici, et al., 2010 Breast Cancer Res 12:R5 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 GSE2034).

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 organi-zed 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; Basal-like = low ESR1, high KRT5.

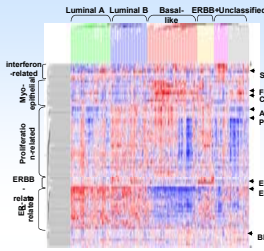


Figure 4. Hierarchical cluster analysis to assign molecular class to 243 samples from dataset of Popovici et al, 2010.

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

	Actual numbers			Percentages		
	no pCR	pCR	Total	no pCR	pCR	Total
Basal-like	42	27	69	17%	11%	29%
ERBB+	8	11	19	3%	5%	8%
Luminal A	55	1	56	23%	0%	23%
Luminal B	43	7	50	18%	3%	21%
Unclassified	43	5	48	18%	2%	20%
Total	191	51	242	79%	21%	100%

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** (Chi² coefficient)	(p-value)
Down early	6	3	50%	0.248	0.0005
Down late	22	12	55%	0.364	<0.000001
Up early	21		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 lists	22	3.73 (max 6, min 0)	17%	-	-

*t-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 subclasses.

		Gene Symbol	PROGNOSIS All	PROGNOSIS ER+	PROGNOSIS ER-	PREDICTION of TFAC Response All	PREDICTION of TFAC Response ER+	PREDICTION of TFAC Response ER-	PREDICTION of TFAC Response LumA	PREDICTION of TFAC Response Lum B	PREDICTION of TFAC Response Her2+	PREDICTION of TFAC Response Basal
Down Late Genes	5	CKS2	0.075	0.29	0.88	**0.0013**	**0.089**	0.17	**0.014**	0.77	**0.025**	**0.038**
	1	CDKN3	**0.0002**	**0.0014**	0.051	**0.0738**	0.65	0.72	0.20	0.95	0.52	0.58
	2	FOXM1	**< 0.0001**	**0.0001**	0.11	**0.0014**	**0.073**	0.61	0.90	0.68	**0.079**	0.34
	3	RRM2	**0.0014**	**0.0068**	0.15	**< 0.0001**	**0.020**	**0.050**	0.11	**0.023**	0.38	0.12
	4	VRK1	**0.0034**	**0.0081**	**0.016**	0.34	0.51	0.80	0.54	0.97	0.12	0.94
	6	TRIP13	**< 0.0001**	**< 0.0001**	0.68	**0.00018**	**0.025**	0.18	0.67	**0.069**	0.20	**0.096**
	7	ASPM	**0.019**	**0.0099**	0.81	**0.0058**	**0.017**	0.81	0.24	**0.036**	0.55	0.85
	9	CEP55	**< 0.0001**	**0.0015**	0.98	**0.010**	0.74	0.19	0.85	0.40	0.61	0.83
	11	ZWILCH	**0.012**	**0.0012**	0.65	0.20	0.80	0.42	0.28	0.65	0.14	0.23
	10	TUBG1	**0.049**	0.056	0.079	0.83	0.18	0.45	**0.017**	0.17	0.20	0.28
	8	AURKA	**< 0.0001**	0.91	0.55	**0.036**	0.16	0.15	0.56	0.90	0.74	0.35
	12	SERPINE2	0.98*	---	---	**0.00040**	0.38	0.33	1.00	0.45	**0.055**	**0.072**
	13	CAPRIN2	0.42*	---	---	0.17	0.42	0.53	0.65	0.87	0.95	0.35
	14	TNFRSF6B	0.23	0.18	0.90	0.35	0.60	0.14	0.45	0.20	0.18	**0.094**
	15	CAPG	0.23*	---	---	0.20	0.63	0.38	0.45	0.31	**0.093**	0.56
	16	ACTN1	**0.005**	0.11	0.065	0.65	0.58	0.37	0.57	0.74	0.20	0.79
	17	ACTB	**0.002**	0.89	**0.037**	0.62	0.44	0.14	0.56	0.38	**0.019**	0.31
	18	DUSP4	**< 0.0001**	**0.002**	0.51	0.60	0.90	**0.078**	0.57	**0.028**	**0.012**	0.37
	19	EPHA2	**0.074**	0.26	0.22	0.20	0.20	**0.088**	0.84	0.44	0.14	0.27
	20	FGFBP1	0.48	0.18	0.80	0.88	0.28	0.13	0.56	**0.055**	0.89	0.23
	21	EIF4A1	0.45	0.35	0.80	0.52	0.40	0.89	0.78	0.43	**0.040**	0.33
	22	ODC1	**0.046**	0.15	0.54	**< 0.0001**	0.43	**0.012**	**0.038**	**0.080**	0.67	**0.011**
Down Early Genes	23	AMIGO2	0.99	0.76	0.67	**0.010**	0.74	0.22	0.45	0.97	0.58	**0.0092**
	24	THBS1	0.29	0.12	**0.015**	**0.0078**	0.20	0.46	**<0.040**	0.11	0.17	0.93
	25	PHLDA	0.64	0.18	0.50	0.22	1.00	0.26	0.52	0.77	0.31	**0.029**
	26	MPRIP	0.47	0.64	0.98	0.49	0.46	0.34	0.80	0.42	0.42	0.57
	27	LRP8	**< 0.0001**	0.055	0.12	**0.0079**	0.13	0.28	0.74	0.87	**0.045**	0.78
	28	SLC20A1	0.66	0.47	0.95	0.37	0.70	0.47	0.12	0.90	0.12	0.81
	-	ER status	**< 0.0001**	---	---	**< 0.0001**	---	---	---	0.54	0.18	**0.052**
	-	Grade	**< 0.0001**	**< 0.0001**	0.20	**< 0.0001**	0.23	0.23	0.80	0.85	0.29	**0.036**
	-	Tumor size	**0.0003**	**0.0011**	1.00	0.75	0.27	0.17	0.63	0.37	0.91	0.74
	-	Age	**0.014**	0.37	**0.030**	0.12	0.69	**0.051**	0.94	0.55	0.79	**0.010**
	-	Node Status	0.47	0.21	0.64	0.20	0.51	0.72	0.18	0.82	0.64	0.52
	-	Race	---	---	---	0.49	0.32	0.70	0.36	0.89	0.39	0.64
	-	Chemo	0.30	0.34	0.64	---	---	---	---	---	---	---

Conclusions: 1) Some genes predicted prognosis but not response and vice-versa, while some predicted both,. Hence, prognosis and response prediction are overlapping but distinct. 2) Different genes predicted response in different subtypes.

P-values <0.05 for prognosis categories are highlighted blue-green; values <0.10 for prediction categories are highlighted purple. Association of gene expression with prognosis was assessed in the microarray dataset of Wang, et al., 2005, using Kaplan-Meier analysis with 5 year relapse 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

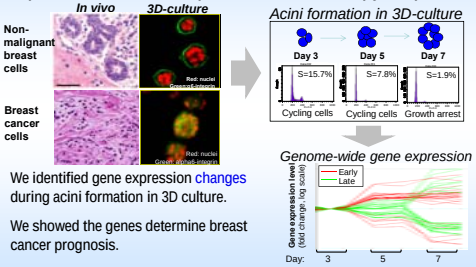
Table 4. 22-gene signature stratified breast cancer subtypes by response to TFAC chemotherapy and outperformed clinical parameters.

	AUC Value* (n)					
Breast Cancer Subtype	22-genes	Node status	ER status	Tumor size	Tumor grade	Ki67
ER Positive	0.723 (208)	0.490	-	0.475	0.689	0.650
ER Negative	0.744 (145)	0.481	-	0.525	0.689	0.635
HER2 Positive	0.772 (42)	0.513	-	0.525	0.316	0.350
Triple Negative (ER, PR, HER2 negative)	0.718 (95)	0.490	-	0.525	0.689	0.650
Luminal B	0.75 (50)	-	-	-	-	-
Basal-like	0.85 (69)	-	-	-	-	-
All subtypes	0.830 (353)	0.478	0.760	0.525	0.689	0.650

*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 .

Approach

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



We identified gene expression changes during acini formation in 3D culture.

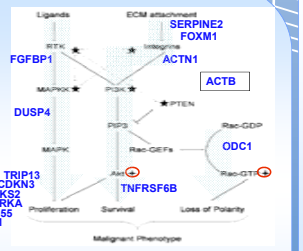
We showed the genes determine breast cancer prognosis.

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

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 cells. Given the proper algorithms, this information has the potential to predict a range of properties of tumor cells.



★ Cross-modulation point
Cross-modulation
Genetic mutants in this study

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

Conclusions

Genes down regulated during acini formation in 3D culture were associated with tumor response to TFAC chemotherapy.

The ability to predict response and prognosis in breast cancer are overlapping but not synonymous. Also, different genes predict response in different subtypes of breast cancer .

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.

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