

ONCOGENOMICS

An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen

AS Doane¹, M Danso², P Lal¹, M Donaton¹, L Zhang¹, C Hudis² and WL Gerald¹¹Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA and ²Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Little is known of the underlying biology of estrogen receptor-negative, progesterone receptor-negative (ER(-)/PR(-)) breast cancer (BC), and few targeted therapies are available. Clinical heterogeneity of ER(-)/PR(-) tumors suggests that molecular subsets exist. We performed genome-wide expression analysis of 99 primary BC samples and eight BC cell lines in an effort to reveal distinct subsets, provide insight into their biology and potentially identify new therapeutic targets. We identified a subset of ER(-)/PR(-) tumors with paradoxical expression of genes known to be either direct targets of ER, responsive to estrogen, or typically expressed in ER(+) BC. Differentially expressed genes included SPDEF, FOXA1, XBP1, CYB5, TFF3, NAT1, APOD, ALCAM and AR ($P < 0.001$). A classification model based on the expression signature of this tumor class identified molecularly similar BCs in an independent human BC data set and among BC cell lines (MDA-MB-453). This cell line demonstrated a proliferative response to androgen in an androgen receptor-dependent and ER-independent manner. In addition, the androgen-induced transcriptional program of MDA-MB-453 significantly overlapped the molecular signature of the unique ER(-)/PR(-) subclass of human tumors. This subset of BCs, characterized by a hormonally regulated transcriptional program and response to androgen, suggests the potential for therapeutic strategies targeting the androgen signaling pathway.

Oncogene (2006) 25, 3994–4008. doi:10.1038/sj.onc.1209415; published online 20 February 2006

Keywords: breast cancer; estrogen receptor-negative; androgen responsive; gene expression

Introduction

Breast cancer remains a major public health concern in the United States and is the second highest cause of

cancer death in women. It is estimated that in 2005 over 200 000 women will develop breast cancer and 40 400 will die of their disease (Jemal *et al.*, 2005). The estrogen receptor (ER) regulates growth and differentiation of the normal mammary gland and is important in the development and progression of about 70% of breast cancer. Like other steroid hormone receptors, the ER mediates its downstream effects by direct transcriptional regulation of target genes (Gruber *et al.*, 2002; McKenna and O'Malley, 2002). In recent years, alternative ER signaling via direct association with and activation of many signal transduction pathways has been described (Cato *et al.*, 2002; Losel *et al.*, 2003). For several decades, targeting the ER has been the cornerstone in treatment for ER-positive (ER(+)) breast cancer (Mouridsen *et al.*, 2001; Goss *et al.*, 2003; Smith and Dowsett, 2003; Osborne *et al.*, 2004; Tobias, 2004; Howell *et al.*, 2005).

ER-negative, progesterone receptor-negative (ER(-)/PR(-)) breast cancer represents approximately 25–30% of all breast cancers and generally has a more aggressive clinical course. In contrast to ER(+) breast cancer, patients with ER(-)/PR(-) tumors derive little or no benefit from anti-estrogen therapy (Howell *et al.*, 2005) and targeted therapies remain elusive (Slamon *et al.*, 2001). One notable exception has been the successful use of antibodies targeting the tyrosine kinase receptor HER-2-neu (ERBB2), which is disproportionably over-expressed in ER(-) breast cancer (Slamon *et al.*, 2001; Lal *et al.*, 2005). In addition to ER, breast cancer cells express other nuclear hormone receptors. For example, the androgen receptor (AR) is expressed in 60–80% of breast cancers and implicated in breast cancer biology (Isola, 1993). Recent studies have reported that among postmenopausal women, high androgen levels are associated with an increased risk of developing breast cancer (Agoff *et al.*, 2003). Furthermore, androgens can induce proliferation in breast tissue, and initiate tumor formation via the AR in animal models (Wong and Xie, 2001). The mechanisms by which the AR contributes to the initiation and progression of breast cancer and its functional relationship to the ER are unknown. It also remains to be determined if targeting the AR could extend the benefits of hormonal therapy to women with ER(-)/PR(-), AR-positive breast cancer.

Genome-wide transcript analysis using DNA microarray technology is an important and well-established

Correspondence: Dr WL Gerald, Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, Schwartz Building, 801, New York, NY 10021, USA.

E-mail: geraldw@mskcc.org

Received 25 October 2005; revised 15 December 2005; accepted 15 December 2005; published online 20 February 2006

new tool in the study of human disease. The technology allows the measurement of several thousands of mRNA species simultaneously. The resulting gene expression profiles have been used to distinguish tumor classes not evident by traditional methods (DeRisi *et al.*, 1996; Golub *et al.*, 1999). In breast cancer, DNA microarray analysis has demonstrated that ER(+) breast cancer and ER(-)/PR(-) disease have unique molecular profiles, has identified several distinct molecular subclasses and has been used to predict disease recurrence (Perou *et al.*, 2000; Gruvberger *et al.*, 2001; West *et al.*, 2001; van't Veer *et al.*, 2002; Pusztai *et al.*, 2003). Few reports specifically focus on gene expression analysis of ER(-)/PR(-) breast cancers and are limited by small sample size (Nagahata *et al.*, 2004). We carried out transcript profiling of human breast cancers and focused attention on ER(-) tumors in an effort to better define important biological subsets. We report the identification of a previously uncharacterized ER(-)/PR(-) breast cancer subset with a hormonally regulated gene expression signature and AR-dependent, androgen-induced cell growth in culture. This represents a clinically relevant subset of ER(-)/PR(-) breast cancer for which AR may provide a useful therapeutic target.

Results

Molecular heterogeneity of ER(-)/PR(-) breast cancers demonstrated by genome-wide expression analysis

In order to explore the molecular heterogeneity of breast cancers, we performed genome-wide transcript profiling for 99 primary breast carcinomas using oligonucleotide microarrays. In all cases, we performed immunohistochemical assessment of ER and PR to ensure the accuracy of receptor status and determine heterogeneity. Forty-one tumors were ER(-)/PR(-), two were ER(-)/PR(+) and 56 were ER(+). There was a close correspondence between the transcript level for ER determined by microarray and ER protein expression determined by a semiquantitative immunohistochemistry (IHC) score (% immunoreactive cells times intensity, spearman's $\rho = 0.834$, $P < 0.01$).

Unsupervised hierarchical clustering revealed a strong association between ER status and molecular profile as previously reported (Gruvberger *et al.*, 2001). However, nine ER(-)/PR(-) breast cancers were grouped with the ER(+) tumors and three ER(+) samples were grouped with ER(-)/PR(-) breast cancers (Figure 1). The finding of breast cancers molecularly discordant with ER status suggested heterogeneity within the major breast cancer subtypes and was further explored.

We focused on ER(-)/PR(-) breast cancers and performed unsupervised hierarchical clustering limited to the 41 ER(-)/PR(-) tumors. Of the major clusters in the dendrogram, it was of particular interest that the nine ER-discordant samples identified in the previous analysis were all closely correlated and contained in a single cluster with only one additional case (Figure 2a). To evaluate the reproducibility of these molecular

subgroups, we used a principal component analysis (PCA) to plot ER(-)/PR(-) samples in three dimensions. These same 10 samples were distinct from the other ER(-)/PR(-) samples, demonstrating relatively robust molecular phenotypes (Figure 2b). Therefore, within our sample set of ER(-)/PR(-) breast cancers, we detected two major molecular subdivisions: one composed of 10 samples with a molecular resemblance to ER(+) breast cancer (referred to hereafter as ER(-) class A) and another composed of the remaining 31 breast cancers (ER(-) class B).

Characterization of genes differentially expressed in ER(-)/PR(-) breast cancer subtypes

By visual inspection of two-dimensional cluster diagrams, it was evident that a number of genes differentially expressed in ER(-) class A relative to other ER(-)/PR(-) breast cancers are similarly expressed in ER(+) tumors (Figure 1). These initial observations suggested that ER(-) class A tumors expressed a molecular signature common to ER(+) breast cancers. To further evaluate this finding, we first identified 202 genes markedly differentially expressed according to ER status (at least threefold difference between the means of ER(+) and ER(-)/PR(-) cases and a Student's *t*-test $P < 0.0001$) (Supplementary Table 2). Not surprisingly, many of the differentially expressed genes have been identified in previous similar analyses (Gruvberger *et al.*, 2001; West *et al.*, 2001). We next identified 138 genes significantly differentially expressed between ER(-)/PR(-) class A and class B samples. Ninety-six genes were overexpressed and 42 genes were underexpressed in class A relative to class B. Of the 96 genes differentially overexpressed in class A, 15 have been reported to be experimentally valid direct targets of the ER (Tang *et al.*, 2004; Laganier *et al.*, 2005), 12 were responsive to estrogen in a previous genome-wide molecular study (Cunliffe *et al.*, 2003) and 24 genes were differentially overexpressed in ER(+) tumors compared to all ER(-)/PR(-) tumors in our data (Table 1). The number of genes common among the 96 overexpressed genes in class A and each gene set described above was much greater than expected by chance ($P < 0.0001$). Among the 42 genes underexpressed in ER(-) class A, three genes have been identified as experimental targets of the ER (Tang *et al.*, 2004; Laganier *et al.*, 2005) ($P = 0.065$), and five genes were differentially underexpressed in ER(+) tumors compared to all ER(-)/PR(-) tumors in our data ($P < 0.0001$). The data strongly suggest that ER(-) class A was distinguished from class B by expression of a transcriptional program that is associated with ER(+) breast cancer.

To explore the potential of the differentially expressed genes as direct targets for steroid hormone receptors, we searched 5 kb of DNA sequence 5' of the transcription start site for putative AR and ER response elements (ARE, ERE) in each of the 138 genes. Among the 96 genes overexpressed in class A, 24 had promoter regions containing at least one putative ERE, and 13 had promoter regions with at least one putative ARE.

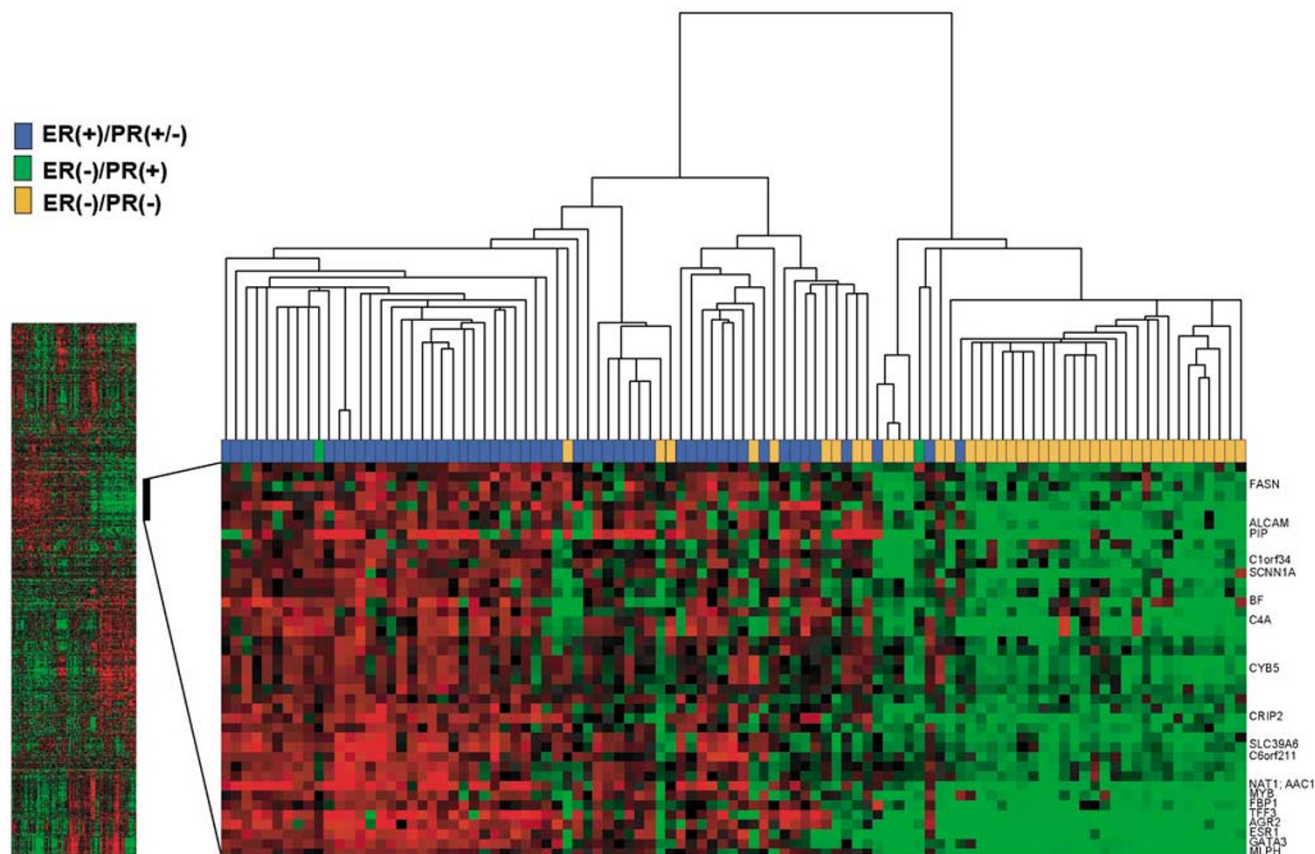


Figure 1 Molecular heterogeneity of breast cancers. Two-way hierarchical clustering was performed with 99 primary breast cancers based on 1961 genes with the greatest variance among samples. The dendrogram represents the relationship of samples. The length of the branches represents 1—the correlation coefficient between samples. A strongly differentially expressed gene cluster is enlarged and genes associated with ER status are labeled. Samples are arranged in columns and genes in rows. Expression levels are pseudocolored red to indicate transcript levels above the median for that gene across all samples and green below the median. Color saturation is proportional to the magnitude of expression.

Among the 42 genes underexpressed in class A, four genes had promoter regions containing at least one putative ERE, and two genes had promoter regions containing at least one putative ARE ($P < 0.0001$ for both ERE and ARE). These observations suggested that ER(−) class A breast cancer more closely resembled an ER(+) breast cancer molecular phenotype owing to expression of many hormonally regulated genes that are based on data from several lines of investigation.

In order to further evaluate this finding, we obtained unigene id numbers for 386 estrogen responsive genes identified in a previously published genome-wide expression analysis using an experimental platform different from that used in our study (Cunliffe *et al.*, 2003), and identified 508 corresponding Affymetrix probe sets. Cluster analysis limited to this set of genes tended to group ER(−) class A samples intermediately between ER(+) samples and the remaining ER(−)/PR(−) tumors (Supplementary Figure 1). This was also true for the ER(−) class A cell line, as described below. This provides an independent confirmation that ER(−) class A breast cancers were characterized by expression of gene profiles that are similar to those of ER(+) tumors.

Immunohistochemical analysis of gene transcript differences between ER(−)/PR(−) breast cancer subtypes

In order to further evaluate and validate the molecular differences identified by the genome-wide expression analysis, we performed IHC for proteins expected to be differentially expressed between ER(−) class A and ER(−) class B based on transcript analysis. Genes with a high magnitude of differential expression between ER(−) subtypes, for which useful antibody reagents were available, and those of known functional significance in breast cancer were studied (Table 1). Protein expressions of ALCAM and SPDEF were analysed on a continuous scale using an IHC score (percentage of cells staining times intensity). There was significant differential protein expression between ER(−) class A and ER(−) class B samples for both ($P = 0.023$ and $P < 0.0001$, respectively) (Figure 3). A significant proportion of ER(−) class A samples were immunoreactive for the AR and FOXA1 compared to ER(−) class B samples (50 vs 16%, $P = 0.045$ and 70 vs 9%, $P = 0.0005$, respectively) in concordance with transcript levels.

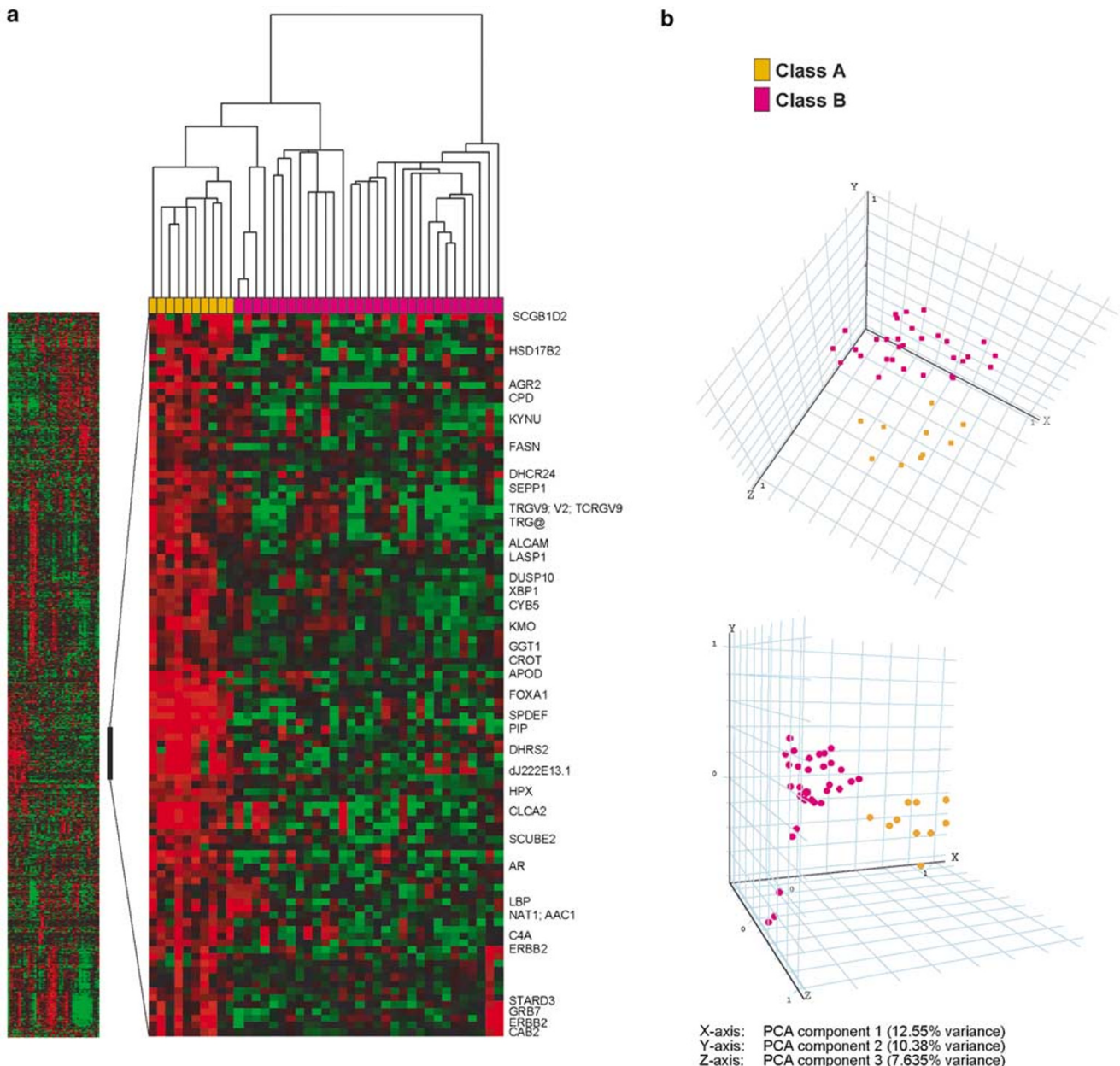


Figure 2 Molecular subclasses of ER(-)/PR(-) breast cancers. **(a)** Two-way hierarchical clustering was performed with 41 ER(-)/PR(-) breast cancers based on 1366 genes with greatest variance among samples. Samples with a molecular similarity to ER(+) breast cancers are labeled class A and the remaining as class B. A gene cluster highly differentially expressed between the two classes is enlarged and select characterized genes labeled. **(b)** Three-dimensional plot of ER(-)/PR(-) primary breast cancers based on the three principal components representing the greatest variance in gene expression across the 41 ER(-)/PR(-) samples identified by analysis of all 22,215 U133A probe sets.

We also evaluated two breast cancer-associated proteins that may participate in the regulation of hormonally responsive gene expression, ERBB2 and ER β . The proportion of ERBB2-positive samples for ER(-) class A and ER(-) class B was 0.30 and 0.15 respectively, and was in good agreement with the ERBB2 transcript levels. By IHC, luminal epithelial cells of non-neoplastic breast expressed moderate levels of ER β ; however, there was little or no ER β protein expression detected in any of the ER(-)/PR(-) samples.

Because several genes differentially expressed in ER(-) class A were identifiable at the protein level by IHC in formalin-fixed, paraffin-embedded (FFPE) tissue sections, we evaluated the feasibility of developing a combination of IHC markers for routine clinical identification of ER(-) class A breast cancers. A combination of ER, PR, SPDEF and ALCAM was estimated to predict ER(-)/PR(-) samples as class A with a sensitivity approaching 100% (95% CI 69–100%), and a specificity of 94% (95% CI 79–99%). It is

Table 1

				Gene Name	p-value	Class A v B Fold Change	Common Name	Genbank
				214451_at	5.04E-11	44.66	TFAP2B	NM_003221
				209173_at	7.88E-07	25.86	AGR2	AF088867
				217276_x_at	2.36E-08	22.61	dJ222E13.1	AL590118
				216623_x_at	1.14E-05	20.46	TNRC9	AK025084
				204607_at	1.81E-09	20.06	HMGCS2	NM_005518
				217284_x_at	2.56E-07	19.42	dJ222E13.1	AL589866
				206509_at	1.02E-08	17.82	PIP	NM_002652
				214774_x_at	2.29E-06	16.7	TNRC9	AK027006
				214243_s_at	5.46E-08	15.55	dJ222E13.1	AL450314
				207802_at	6.88E-06	14.47	CRISP3	NM_006061
				206463_s_at	1.37E-06	14.41	DHRS2	NM_005794
				220192_x_at	2.05E-10	13.99	SPDEF	NM_012391
				209309_at	1.78E-05	13.8	AZGP1	D90427
				214404_x_at	1.70E-13	12.31	SPDEF	AI435670
				215108_x_at	1.58E-05	11.79	TNRC9	U80736
				217528_at	7.75E-05	11.33	CLCA2	BF003134
				201525_at	1.80E-07	11.05	APOD	NM_001647
				218922_s_at	7.43E-08	9.676	LASS4	NM_024552
				213441_x_at	7.99E-11	9.494	SPDEF	AI745526
				204667_at	1.80E-12	7.782	FOXA1	NM_004496
				217562_at	5.78E-05	7.704	DBCCR1L	BF589529
				209813_x_at	5.71E-05	7.691	TRGV9; V2; T	M16768
				214079_at	4.21E-05	7.428	DHRS2	AK000345
				210576_at	6.64E-05	7.337	CYP4F8	AF133298
				205221_at	1.64E-06	6.974	HGD	NM_000187
				213884_s_at	1.04E-09	6.785	TRIM3	AA114843
				211657_at	2.95E-05	6.542		M18728
				211237_s_at	9.51E-07	6.317	FGFR4	AF202063
				204623_at	2.61E-07	6.007	TFF3	NM_003226
				218211_s_at	5.11E-12	5.895	MLPH	NM_024101
				39763_at	2.12E-10	5.599	HPX	M36803
				204719_at	7.17E-05	5.233	ABCA8	NM_007168
				220622_at	1.24E-05	5.222	FLJ23259	NM_024727
				210013_at	3.00E-05	5.165	HPX	BC005395
				211110_s_at	1.46E-05	4.951	AR	AF162704
				218313_s_at	2.45E-06	4.85	GALNT7	NM_017423
				204942_s_at	1.31E-07	4.721	ALDH3B2	NM_000695
				203722_at	1.06E-05	4.681	ALDH4A1	NM_003748
				210056_at	9.47E-09	4.538	RND1	U69563
				219734_at	5.96E-10	4.512	FLJ20174	NM_017699
				221584_s_at	7.57E-06	4.344	KCNMA1	U11058
				219197_s_at	8.36E-06	4.218	SCUBE2	AI424243

Table 1 (continued)

				Gene Name	p-value	Class A v B Fold Change	Common Name	Genbank
				204462_s_at	2.71E-05	4.158	SLC16A2	NM_006517
				211144_x_at	9.11E-05	4.091	TRG@	M30894
				205161_s_at	4.10E-07	4.083	PEX11A	NM_003847
				215465_at	2.10E-05	4.009	ABCA12	AL080207
				217014_s_at	3.44E-06	3.925		AC004522
				215806_x_at	5.00E-05	3.84	TRG@	M13231
				200670_at	1.22E-08	3.671	XBP1	NM_005080
				201952_at	3.35E-06	3.568	ALCAM	AA156721
				212218_s_at	1.11E-05	3.562	FASN	AI954041
				214295_at	5.13E-05	3.491		AW129056
				212510_at	8.81E-08	3.479	KIAA0089	AA135522
				205306_x_at	1.73E-05	3.441	KMO	AI074145
				207843_x_at	4.12E-07	3.426	CYB5	NM_001914
				216333_x_at	2.63E-05	3.305	TNXB	M25813
				215726_s_at	1.43E-06	3.224	CYB5	M22976
				205150_s_at	8.16E-06	3.204	KIAA0644	AV724192
				209366_x_at	1.17E-06	3.185	CYB5	M22865
				219956_at	5.72E-07	3.144	GALNT6	NM_007210
				219429_at	4.43E-05	3.124	FA2H	NM_024306
				218546_at	2.19E-05	3.095	FLJ14146	NM_024709
				209522_s_at	8.33E-05	3.082	CRAT	BC000723
				208284_x_at	1.74E-06	3.022	GGT1	NM_013421
				215559_at	6.47E-05	3.013	ABCC6	AI074459
				218776_s_at	4.25E-08	2.859	FLJ23375	NM_024956
				204579_at	3.21E-06	2.816	FGFR4	NM_002011
				207131_x_at	6.49E-07	2.764	GGT1	NM_013430
				217973_at	2.84E-06	2.74	DCXR	NM_016286
				206850_at	2.54E-05	2.71	RRP22	NM_006477
				212593_s_at	1.01E-06	2.706	PDCD4	N92498
				205160_at	2.72E-05	2.683	PEX11A	AL360141
				203740_at	3.34E-07	2.676	MPHOSPH6	NM_005792
				211417_x_at	1.79E-07	2.591	GGT1	L20493
				209919_x_at	5.16E-07	2.552	GGT1	L20490
				216638_s_at	2.70E-05	2.536	PRLR	S78505
				213557_at	4.36E-05	2.492		AW305119
				212594_at	1.10E-06	2.436	PDCD4	AI185160
				201941_at	5.57E-06	2.372	CPD	BE349147
				212736_at	8.29E-05	2.333	BC008967	BE299456
				218552_at	9.75E-06	2.306	FLJ10948	NM_018281
				51158_at	9.72E-07	2.24		AI801973
				212099_at	7.79E-05	2.189	ARHB	AI263909
				200810_s_at	2.67E-06	2.187	CIRBP	NM_001280
				212956_at	4.61E-06	2.166		AI348094
				200618_at	8.97E-05	2.151	LASP1	NM_006148

Table 1 (continued)

				Gene Name	p-value	Class A v B Fold Change	Common Name	Genbank
				211596_s_at	7.34E-05	2.146	LRIG1	AB050468
				213107_at	8.17E-05	2.139		R59093
				208872_s_at	6.53E-06	2.131	DP1	AA814140
				215603_x_at	3.43E-05	2.122	GGT2	AI344075
				219396_s_at	4.27E-05	2.116	NEIL1	NM_024608
				211621_at	9.37E-06	2.105	AR	M73069
				215299_x_at	2.29E-05	2.019	SULT1A1; PS	U37025
				219543_at	8.17E-05	2.014	MAWBP	NM_022129
				201940_at	5.50E-05	2.013	CPD	AA897514
				208873_s_at	7.34E-07	2.006	DP1	BC000232
				37966_at	4.40E-05	0.494	PARVB	AA187563
				200756_x_at	4.37E-05	0.489	CALU	U67280
				203167_at	7.48E-05	0.487	TIMP2	NM_003255
				219785_s_at	9.20E-05	0.485	MGC15419	NM_024735
				212650_at	9.18E-05	0.478	NACSIN	BF116032
				200757_s_at	2.99E-06	0.477	CALU	NM_001219
				211924_s_at	8.44E-06	0.463	PLAUR	AY029180
				205120_s_at	1.03E-05	0.458	SGCB	U29586
				209043_at	1.61E-05	0.457	PAPSS1	AF033026
				218629_at	4.29E-05	0.453	SMO	NM_005631
				200755_s_at	3.07E-05	0.443	CALU	BF939365
				209204_at	2.85E-05	0.429	LMO4	AI824831
				213003_s_at	9.43E-05	0.426		BF061054
				202990_at	2.27E-05	0.411	PYGL	NM_002863
				200934_at	6.04E-05	0.404	DEK	NM_003472
				221505_at	8.10E-06	0.378	ANP32E	AW612574
				210074_at	5.18E-06	0.377	CTSL2	AF070448
				214845_s_at	1.96E-05	0.375	CALU	AF257659
				60474_at	8.81E-05	0.371	C20orf42	AA469071
				202620_s_at	3.64E-05	0.37	PLOD2	NM_000935
				202236_s_at	2.99E-06	0.363	SLC16A1	NM_003051
				219944_at	1.86E-05	0.346	FLJ21069	NM_024692
				202134_s_at	3.20E-05	0.332	TAZ	NM_015472
				216488_s_at	2.33E-06	0.326	ATP11A	AL161996
				202619_s_at	6.84E-06	0.318	PLOD2	AI754404
				218851_s_at	8.81E-05	0.315	WDR33	NM_018383
				207675_x_at	5.89E-05	0.313	ARTN	NM_003976
				213256_at	1.53E-05	0.31	MGC48332	AW593996
				201564_s_at	3.22E-05	0.307	FSCN1	NM_003088
				219926_at	9.90E-05	0.306	POPCDC3	NM_022361
				202784_s_at	3.61E-05	0.293	NNT	NM_012343
				209900_s_at	5.16E-05	0.283	SLC16A1	AL162079
				209834_at	2.96E-05	0.249	CHST3	AB017915
				213260_at	4.56E-05	0.228	FOXCI	AU145890

Table 1 (continued)

				Gene Name	p-value	Class A v B Fold Change	Common Name	Genbank
				208103_s_at	6.99E-05	0.227	ANP32E	NM_030920
				204285_s_at	9.09E-09	0.222	PMAIP1	AI857639
				209875_s_at	7.49E-06	0.217	SPP1	M83248
				202235_at	1.95E-07	0.196	SLC16A1	BF511091
				204750_s_at	1.34E-05	0.144	DSC2	BF196457
				204286_s_at	9.42E-07	0.141	PMAIP1	NM_021127
				209800_at	1.03E-05	0.121	KRT16	AF061812
				204855_at	2.69E-05	0.0816	SERPINB5	NM_002639
				<div> <div>Up in ER(+) 3 fld p<0.0001</div> <div>Direct targets of ER (Tang <i>et al.</i>, 2004; Laganier <i>et al.</i>, 2003)</div> <div>Estrogen responsive genes (Cunliffe <i>et al.</i>, 2003)</div> <div>Putative ERE</div> <div>putative ARE</div> <div>Up in ER(-) >3 fold p<0.0001</div> </div>				

important to note that our analysis is limited by sample size and lack of independent validation, and therefore may overestimate the predictive value. The practical utility of an IHC assay for class distinction requires further study.

Class prediction and independent evaluation of ER(-) breast cancer subsets

In order to determine if the ER(-) class A subclass was a reproducible finding and identify appropriate breast cancer cell lines for further study, we developed a *k*-nearest neighbor classification model using genes that were differentially expressed (*P*-value <0.0001) between ER(-) class A and all other ER(-)/PR(-) tumors. We applied this classification method to an independent, publicly available breast cancer gene expression data set that used the same analytical platform (Wang *et al.*, 2005). A similar proportion of ER(-)/PR(-) samples was classified as ER(-) class A in this independent data set as in our samples (32 vs 24%). Many of the genes differentially expressed in the comparison of ER(-) class A and ER(-) class B in our original data were also differentially expressed in the independent predicted subsets. AR, CYB5, XBP1, FOXA1 and SPDEF, as well as the androgen responsive genes APOD (Hall *et al.*, 1996) and PIP (Carsol *et al.*, 2002), were among the top 50 significantly overexpressed genes in the predicted class A (*P*<1e-10). In addition, ERBB2 and FGFR4 were overexpressed in the predicted ER(-) class A of the independent data with a higher degree of significance than in our original sample set (*P* = 1.6e-14 and *P* = 1.4e-08, respectively).

Not only was the ER(-) class A clearly distinguishable in the independent data by supervised analysis, but unsupervised approaches indicated that these classes represent a primary distinction among ER(-)/PR(-) tumors. An unsupervised hierarchical clustering of the 77 ER(-)/PR(-) tumors yielded primary groups of

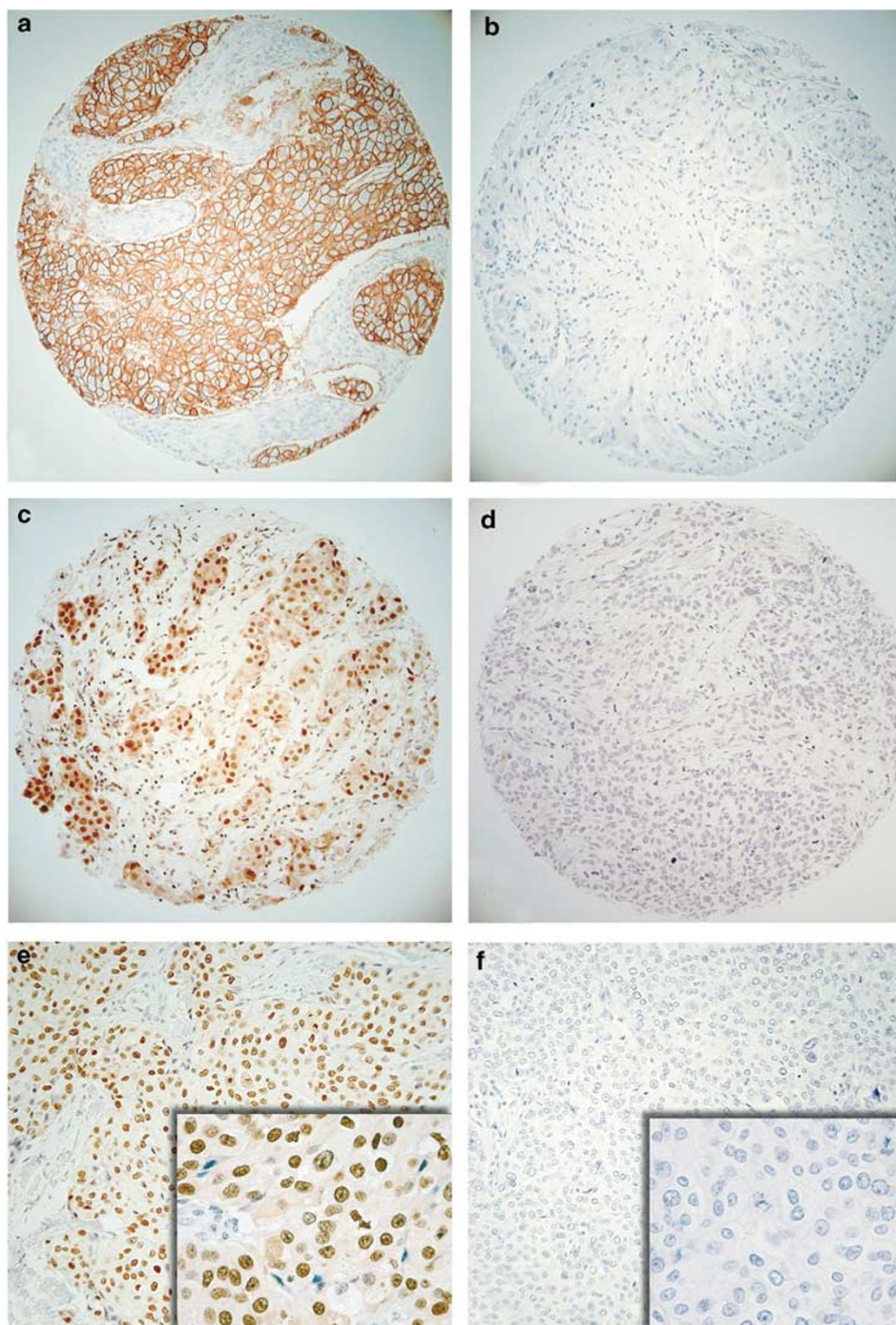


Figure 3 Immunohistochemical evaluation of differentially expressed genes. Representative photomicrographs of IHC studies for ALCAM in an ER(−) class A breast tumor (a) and an ER(−) class B breast tumor (b). SPDEF in a class A breast tumor (c) and class B breast tumor (d). AR in a class A breast tumor (e) and a class B breast tumor (f).

samples, which corresponded very closely to the class prediction assignments by the predictive model (Figure 4). Although the independent samples were not available for confirmation of ER status or further analyses, the data from class prediction and unsupervised clustering provided further evidence that the ER(−) class A and B distinction was reproducible and intrinsic to the primary molecular substructure of ER(−)/PR(−) tumors.

We then used the prediction model to evaluate breast cancer cell lines in order to identify cell lines corresponding to the ER(−) class A molecular phenotype. Expression profiles were generated for the ER(−)/PR(−) cell lines MDA-MB-231, MDA-MB-453, HCC-1937 and SKBR-3. These cell lines have been described to represent important distinctions within the spectrum of ER(−)/PR(−) disease (Lacroix and Leclercq, 2004). Our classification model identified the cell line MDA-MB-453 as

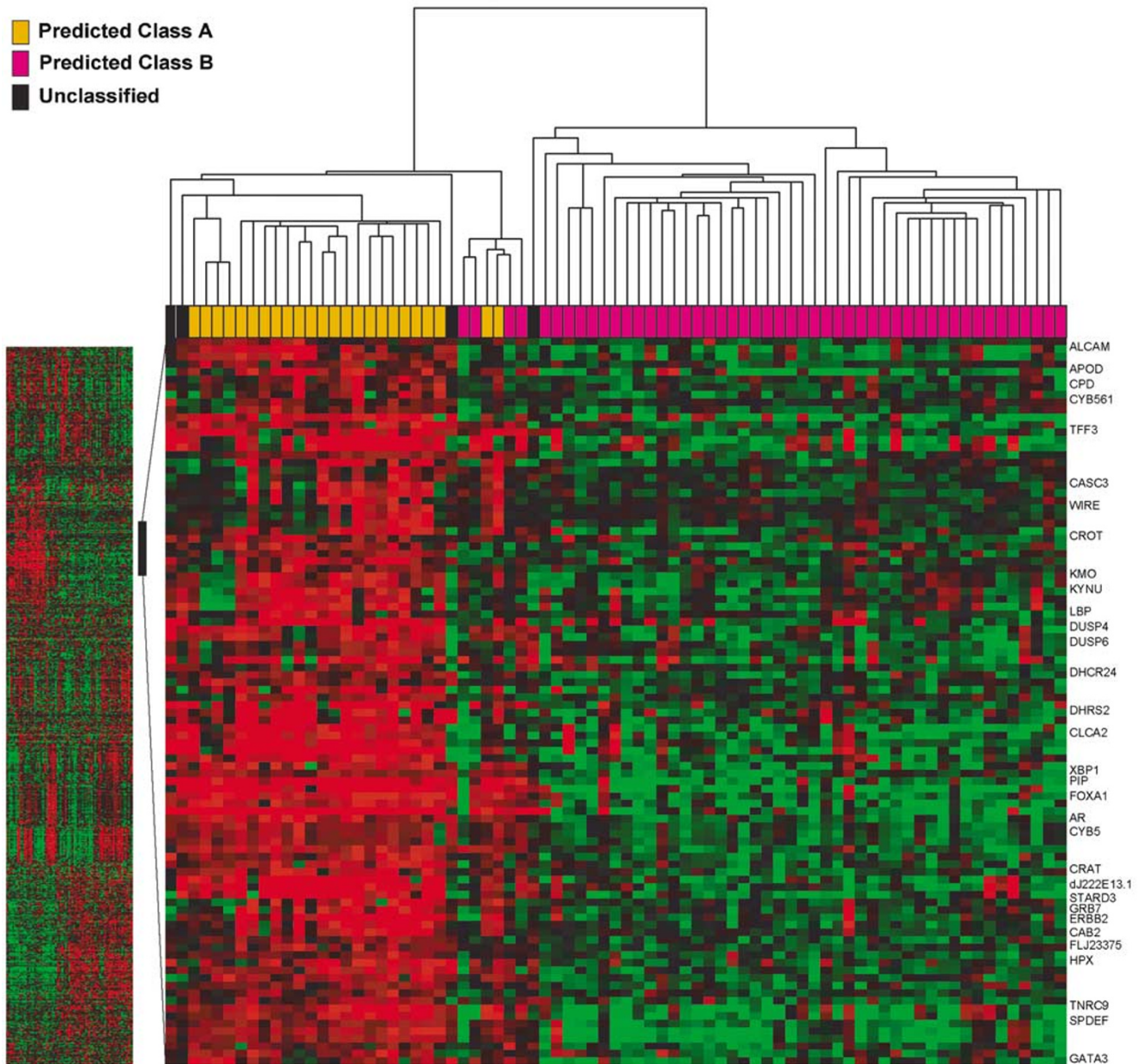


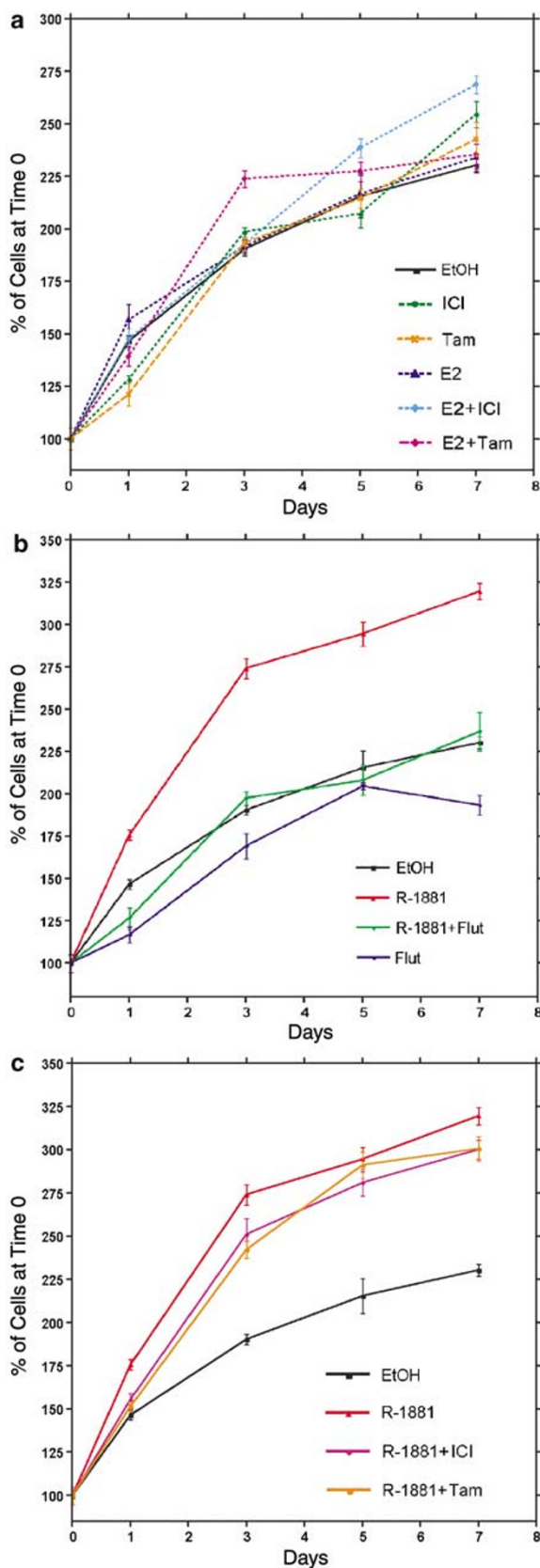
Figure 4 Reproducibility of ER(-) breast cancer subclasses. Two-way hierarchical clustering was performed with 77 ER(-) breast tumors from an independent data set using 1262 genes with greatest variance across samples. The resulting dendrogram revealed a tendency to group samples according to our class prediction assignments. A strongly differentially expressed gene cluster is enlarged and genes associated with ER status and class A are labeled.

ER(-) class A (P -value ratio = $5.75e-06$), and the remaining ER(-)/PR(-) cell lines as ER(-) class B. We have therefore used MDA-MB-453 as an *in vitro* model representing the ER(-) class A molecular phenotype.

The ER(-) class A cell line MDA-MB-453 shows a proliferative response to androgen that is AR dependent and ER independent

The identification of ER(-)/PR(-) breast tumors characterized by expression profiles including estrogen-regulated genes suggested an ER-independent mechanism

for activation of hormonally responsive transcription that contributed to tumor growth and survival. In order to define the mechanism for regulation of this transcriptional program, we first sought to determine whether low levels of active ER, below the limit of detection in clinical assays, might be contributing to growth of ER(-) class A tumors. In our group A model cell line MDA-MB-453, ER transcript levels were very low with an absolute adjusted intensity of 38.0 and Affymetrix MAS 5.0 call of absent. Incubation with 100 nM E2 had no effect on cell culture growth compared to vehicle control. Accordingly, incubation with either the pure anti-estrogen ICI or tamoxifen,



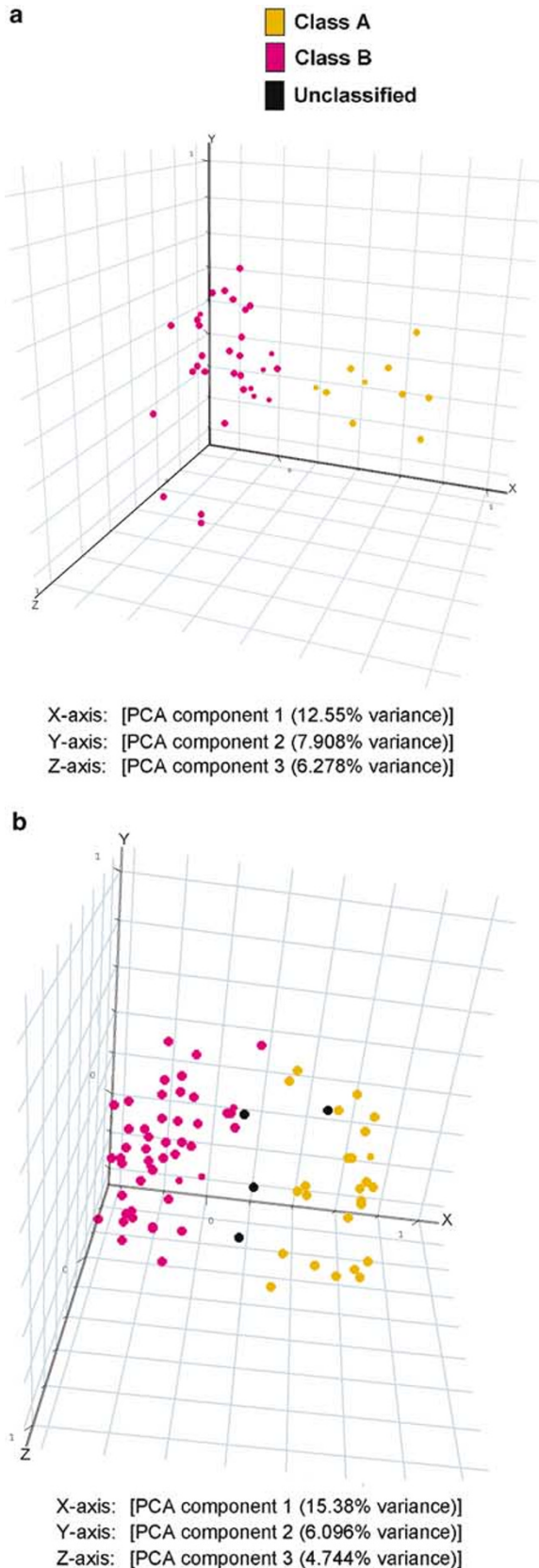
alone or in combination with 100 nM E2, had no effect on overall cell viability compared to vehicle control (Figure 5a). This is in contrast to the ER(+) cell line MCF-7 that was markedly growth stimulated by administration of 100 nM E2, and this effect was abrogated by the addition of the pure anti-estrogen ICI (data not shown). These results suggested that the ER was not playing an active role in the growth and survival of ER(-) class A breast cancer cells.

Because there is the potential for functional overlap of transcriptional regulation by steroid hormone receptors, we reasoned that other nuclear receptors might play a role in ER(-) class A breast cancers. We examined the expression of many known nuclear hormone receptors, including ER- β , ESRRA, PR, AR, RARA, RXRA, GCHR, PPAR and VDR and found that the AR was the only nuclear hormone receptor differentially over-expressed in ER(-) class A. The AR has been implicated in the pathogenesis of breast cancer (Wong and Xie, 2001), and it is known to activate a number of estrogen responsive genes (Nantermet *et al.*, 2005). Incubation with the synthetic non-metabolizable androgen R-1881 at concentrations between 0.1 and 10 nM stimulated growth in MDA-MB-453. This proliferative effect was abrogated by the addition of the AR antagonist flutamide, confirming that the response was AR dependent (Figure 5b). Again, we determined that the effects of androgen were not dependent on the ER, as MDA-MB-453 cells treated with androgens in combination with the anti-estrogens tamoxifen or ICI had minimal effect on the androgen-induced proliferation (Figure 5c). These observations indicated that AR signaling was intact in ER(-) class A breast cancer cells, and that cell growth and survival were responsive to androgen in an AR-dependent, ER-independent manner.

The ER(-) class A molecular phenotype is associated with androgen responsive genes

Because the ER(-) class A cell line MDA-MB-453 demonstrated a proliferative response to androgen, we set out to determine whether this was associated with the transcriptional program characteristic of ER(-) class A human breast cancers. We monitored gene expression changes after administration of androgens, androgen antagonists or vehicle control to the ER(-) class A cell line MDA-MB-453 under a variety of growth conditions (see Materials and methods). The results of the various experiments were concordant, but the most pronounced differences in gene expression were observed between

Figure 5 ER(-) class A breast cancer cells proliferate in response to androgen in an AR- dependent and ER-independent manner. MDA-MB-453 cells were treated with reagents as indicated and cell proliferation measured. All experiments were performed in triplicate. (a) Incubation with E2, the anti-estrogens tamoxifen and ICI with or without E2, and vehicle control. (b) Incubation with the androgen R-1881, R-1881 with the AR antagonist flutamide, flutamide alone and vehicle control. (c) Incubation with R-1881, R-1881 with tam, R-1881 with ICI and vehicle control.



those cells first incubated in steroid-deprived conditions for 48 h, and then treated with either R-1881 or vehicle for 48 h. After trimming to eliminate genes with very low-level expression (<200 in both conditions), 497 genes were differentially expressed by at least twofold between cells exposed to R-1881 or vehicle. The androgen-regulated gene *SARG* was upregulated by 247-fold, and has been previously shown to contain an experimentally verified, hormonally active androgen response element (Steketee *et al.*, 2004). Several other androgen responsive genes including *FASN*, *NDRG1* and *SORD* each contain putative androgen response elements in their promoters (Nelson *et al.*, 2002; Dhanasekaran *et al.*, 2005), and were highly upregulated after administration of R-1881. Furthermore, *FASN*, *NDRG1* and *SORD* are implicated in the androgen-mediated development of prostate cancer (Shurbaji *et al.*, 1996; Rossi *et al.*, 2003), and their expression levels are positively correlated with the expression of prostate-specific antigen in human prostate cancer cells (Nelson *et al.*, 2002). These observations provided indirect evidence that administration of R-1881 to MDA-MB-453 caused recruitment of an active AR transcription complex to highly specific AREs.

To evaluate the association between androgen responsive genes in MDA-MB-453 and the ER(-) class A molecular phenotype, we compared androgen-induced gene expression changes to genes differentially expressed between ER(-) classes A and B. Of the 497 differentially expressed genes between cells treated with R-1881 or vehicle control, 22 were common to our ER(-) class A expression signature, and this number of commonly expressed genes was highly significant ($P = 3e-8$) (Supplementary Table 3). In particular, *ALCAM* was overexpressed in cells treated with R-1881 compared to cells treated with vehicle control, and was also significantly overexpressed at the protein and transcript levels in class A breast cancers, as described above. Therefore, the genes that comprise the ER(-) class A molecular fingerprint were at least in part androgen responsive in the class A cell line.

To further explore the association between the ER(-) class A molecular phenotype and an androgen-dependent transcription program, we performed PCA of the 41 ER(-)/PR(-) breast tumors using the 497 androgen responsive genes and plotted samples based on three principal components. The ER(-) class A and B samples formed distinct clusters (Figure 6a). Furthermore, the same approach using the 77 ER(-)/PR(-) samples from the independent data set demonstrated clusters corresponding to our class predictions

Figure 6 Molecular subclasses of ER(-) breast cancer based on androgen responsive genes. Three-dimensional plot of the three principal components with the greatest variance across 41 ER(-)/PR(-) primary breast tumors using 497 genes responsive to androgen in the class A cell line MDA-MB-453. (b) Three-dimensional plot of the three principal components with the greatest variance among 77 ER(-) breast tumors from an independent data set using the 497 androgen responsive genes. Samples are colored according to class prediction assignments.

(Figure 6b). We next used a weighted Kolmogorov–Smirnov (KS) statistic (Subramanian *et al.*, 2005) to evaluate the degree to which the 497 androgen-regulated genes were correlated with the ER(–) class A and B distinction. The results indicated that among the original and independent data sets, tumors designated as ER(–) class A were significantly enriched in androgen-regulated genes ($P < 0.0001$ for both data sets) (Supplementary Figures 2A and B). These results suggested that the ER(–) class A molecular phenotype was partially recapitulated by the expression of genes regulated by androgen in ER(–) class A breast cancer cells.

We also determined whether genes induced by androgens in MDA-MB-453 corresponded to the transcriptional program activated by estrogens in ER(+) breast cancer cells and therefore could contribute to the molecular relationship between ER(–) class A and ER(+) breast cancers. Fifty of the 497 androgen responsive genes from our experiments were in common with the 386 estrogen responsive genes determined by an independent study using MCF7, T-47D and MDA-MB-436 breast cancer cells ($P = 4e-16$) (Cunliffe *et al.*, 2003). Therefore, androgen induced a transcriptional program in AR-positive ER(–) class A breast cancer that significantly overlapped with that induced by estrogen in ER(+) breast cancer.

Discussion

Clinicians have long recognized that the current classification of breast cancer based on histopathological grade, hormone receptor status and HER2 status does not sufficiently capture the clinical and biologic heterogeneity observed in practice. This has fueled efforts to develop more biologically and clinically meaningful classification based on molecular features. We applied unsupervised and supervised analyses to gene expression profiles of primary breast cancers, and identified a subset of ER(–)/PR(–) tumors with a molecular signature that suggests an active hormonally regulated transcriptional program. Gene expression signatures were used to develop a predictive model that identified this subset among independent tissue samples and breast cancer cell lines. The breast cancer cell line MDA-MB-453 recapitulated the molecular phenotype and was used to investigate the biological basis for this subclass. Several molecules that can initiate signal transduction contributing to tumor growth and survival were overexpressed in this tumor subset, including AR, ERBB2 and FGFR4. We found that androgen enhanced growth of MDA-MB-453 in an ER-independent but AR-dependent manner. In addition, the ER(–) class A molecular phenotype was at least partially androgen regulated. Taken together, our findings help to define a distinctive molecular subset of ER(–)/PR(–) breast cancer with the potential for novel targeted therapeutic strategies.

The potential for molecular subclassification of breast cancers based on genome-wide expression analysis has

been well documented in previous studies. Applying a class discovery approach using cDNA microarrays, Perou *et al.* (2000) identified at least five molecular subtypes of breast cancer (termed luminal subtypes A and B, ERBB2, basal and normal breast like) based on expression of an internally derived ‘intrinsic’ gene list. These subtypes have been repeatedly observed across various high-throughput platforms (Sorlie *et al.*, 2003; Wilson and Dering, 2004). The luminal subtype A and basal groups are the most robust in independent data analysis. This luminal subtype is primarily composed of ER(+) tumors, generally demonstrates a better prognosis and is characterized by relative overexpression of estrogen-related genes such as GATA3, XBP1, FOXA1, CCND1, TFF3 and ER α . The basal class is so named because its expression pattern resembles that of the basal epithelial cell component of the breast including lack of expression of ER and related genes, and expression of cytokeratins 5/6 and 17. Hierarchical clustering of our data using the intrinsic gene list revealed that the luminal A and basal subtypes were evident, whereas the remaining subtypes were not nearly as distinct (Supplementary Figure 3). This is similar to previous reports and suggests that other subtypes of breast cancer require refinement of their molecular definition (Sorlie *et al.*, 2003; Wilson and Dering, 2004). The ER(–) class A samples described here are distinct from the luminal A and basal subtypes. This subset is most distinct when clustering is not limited to the intrinsic gene set of Perou and even more so when the analysis is limited to ER(–) breast cancers. Our data suggest that ER(–) class A breast cancers bear a much closer molecular relationship to ER(+) breast cancers than to the basal subtype despite the shared ER(–) phenotype. This observation was recapitulated in a larger independent sample of 77 ER(–) breast cancers. This work serves to better define a previously unrecognized breast cancer subset, with important implications for the diagnosis and treatment of women with ER(–)/PR(–) breast cancer.

Our results suggest that the AR may play an important role in regulating the molecular events associated with ER(–) class A breast cancers. AR expression is well known in ER(+) breast cancer and has been associated with ER(–) breast cancer with apocrine histologic features (Gatalica, 1997; Matsuo *et al.*, 2002; Sapp *et al.*, 2003). Morphologic evaluation of our ER/PR–group of tumors demonstrated an increased number of cases with strong apocrine morphology in class A tumors compared to class B (70 vs 21%). Although the sample size is small, these results corroborate the findings of Farmer *et al.* (2005), who identified a group of breast tumors with expression of AR and AR-related genes and apocrine features. Nonetheless, the functional role of AR in breast cancer is poorly understood. Androgenic effects on the proliferation of breast cancer cell lines are highly variable (Birrell *et al.*, 1995), an observation not particularly surprising considering the heterogeneity of AR expression in breast cancer and the complexity of AR signaling. Although several breast cancer cell lines

appear to be growth inhibited by the addition of androgens, many are growth stimulated and may be androgen dependent (Lippman *et al.*, 1976; Greeve *et al.*, 2004). In MCF-7 ER(+) breast cancer cells, AR may interfere with ER signaling, leading to a partial G1-S cell cycle arrest and an associated decrease in cell proliferation (Greeve *et al.*, 2004; Buchanan *et al.*, 2005). In agreement with our results, previous studies have reported AR-dependent androgen-induced proliferation in ER(-) class A MDA-MB-453 cells (Hall *et al.*, 1994; Birrell *et al.*, 1995). Our data further suggest that this proliferative response is associated with a hormonally regulated transcriptional program that is common to ER(-) class A breast cancers and overlaps with ER-induced transcription in ER(+) tumors. However, the overlap is incomplete, and this may reflect the fact that an integrated network of signaling pathways regulates cell proliferation, which AR activation alone does not replicate. We speculate that AR may act in concert with other signal transduction pathways to contribute to the ER(-) class A molecular phenotype. For example, it is well known that receptor tyrosine kinase pathways function as modulators of nuclear hormone receptor activity (Shao and Lazar, 1999) and in this regard it is interesting that ERBB2 is differentially expressed in a proportion of ER(-) class A breast cancers. ERBB2 has been shown to stabilize AR protein levels and optimize binding of AR to promoters of androgen-regulated genes in prostate cancer cells (Mellinghoff *et al.*, 2004). In the ER(-) class A breast cancer line MDA-MB-453, blocking ERBB2 with PKI166 inhibits PI3K signaling, deactivates mTOR and decreases cell proliferation (Koziczak and Hynes, 2004). Given the proliferative effect of androgen on MDA-MB-453 that we have shown, the potential for cooperative crosstalk between ERBB2 signaling and AR deserves further study. In addition, the anti-proliferative effect of anti-androgens on MDA-MB-453 provides the rationale for study of anti-androgens to treat ER(-) class A breast cancer.

The expression of ERBB2 represents a biologically and clinically important feature of breast cancer, and a molecular subtype characterized by ERBB2 overexpression has been proposed (Perou *et al.*, 2000). Our observations suggest heterogeneity within the ERBB2 molecular subtype. Indeed, ERBB2 overexpression exists in estrogen responsive, ER(+) breast cancer as well as ER(-) breast cancer. Results of unsupervised hierarchical clustering of 99 primary tumors revealed expression of ERBB2 among several sample clusters, and suggested that the expression of ERBB2 alone does not capture the molecular phenotype of class A breast cancer. Indeed, among the class A samples in our data, only 30% were ERBB2 positive. Furthermore, SKBR-3 cells have ERBB2 gene amplification and protein overexpression (Lacroix and Leclercq, 2004), and were identified by our predictor as class B. Not surprisingly, the ERBB2 monoclonal antibody trastuzumab inhibits the growth of SKBR-3 cells (Mayfield *et al.*, 2001; Yakes *et al.*, 2002). MDA-MB-453 cells are not ERBB2-amplified and are resistant to the anti-proliferative

effects of trastuzumab (Yakes *et al.*, 2002). Further investigation into the diversity of ERBB2 signaling among various breast cancer subtypes is required.

FGFR4 is another signaling molecule that may cooperate with AR and ERBB2 to drive tumor growth in the ER(-) class A subtype of ER(-)/PR(-) breast cancer. FGFR4 is overexpressed in ER(-) class A tumors and gene amplification may exist in as many as 30% of all breast cancers (Dickson *et al.*, 2000). In MDA-MB-453 cells, FGFR4 and ERBB2 have been shown to work in concert to activate the mTOR translational pathway and regulate cyclin D1 levels (Koziczak and Hynes, 2004). Simultaneous inhibition of both pathways had a stronger anti-proliferative effect than either alone. In addition, FGFR4-dependent activation of the MAPK/ERK1/2 signaling cascade can drive cell proliferation via downstream initiation of cyclin D1 transcription (Koziczak *et al.*, 2004). This convergence of data suggests that further investigation into the role of FGFR4, ERBB2 and AR in ER(-) class A breast cancers is warranted and that this molecular complex may provide useful therapeutic targets for as many as 25% of ER(-)/PR(-) breast cancer patients.

Materials and methods

Samples and gene expression analysis

Tissue samples were obtained from therapeutic or diagnostic procedures performed as part of routine clinical management at Memorial Sloan-Kettering Cancer Center. All research procedures using human tissue were approved by the MSKCC institutional review board. Tissues were snap-frozen in liquid nitrogen and stored at -80°C. Each sample was examined histologically using hemotoxylin- and eosin-stained cryostat sections and enriched for areas of interest by manual trimming of tissue blocks. Total RNA was extracted from frozen tissue by homogenization in guanidinium isothiocyanate-based buffer (Trizol; Invitrogen, Carlsbad, CA, USA), purified using RNeasy (Qiagen, Valencia, CA, USA) and examined for quality using denaturing agarose gel. Complementary DNA was synthesized from RNA using a T7-promoter-tagged oligo-dT primer. RNA target was synthesized from cDNA by *in vitro* transcription, and labeled with biotinylated nucleotides (Enzo Biochem, Farmingdale, NY, USA) (Holzbeierlein *et al.*, 2004). Gene expression analysis was performed using HG-U133A oligonucleotide microarrays according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA). Data files for gene expression analysis are available at <http://caarraydb.nci.nih.gov/caarray/>. There were 99 primary breast tumors analysed (77 invasive ductal carcinoma, 10 invasive lobular carcinomas, seven mixed lobular and ductal carcinomas, four metaplastic carcinomas and one not specified). All ER(-)/PR(-) tumors were designated invasive ductal or invasive lobular type. Morphological examination was performed by an experienced surgical pathologist in a blinded manner with special attention to apocrine features. A score was assigned based on the presence of (1) abundant eosinophilic and granular cytoplasm, (2) vesicular nuclei with prominent nucleoli and (3) apical snouts (Miller *et al.*, 1985). A score of 0 represented no apocrine features, whereas 1 and 2 were reserved for moderate and marked degree of apocrine differentiation, respectively.

Data analysis

Signals were quantified using Affymetrix Microarray Suite 5.0 and expression values were scaled to have a mean expression of 500 across the central 96% of values for each array. Each sample was individually characterized by both probe set intensity values and associated clinical data. A master gene table was compiled, in which specific genes represented by GenBank accession numbers were identified for each probe set (<http://www.affymetrix.com>). Annotation information corresponding to the GenBank accession number for each probe set was retrieved from the GenBank, LocusLink, Unigene and Gene Ontology Consortium databases. All annotation information was downloaded through the Silicon Genetics Mirror server using the GeneSpider tool (GeneSpring, Silicon Genetics, Redwood City, CA, USA).

Before unsupervised analyses, the gene expression measurements were filtered and normalized using the following methods. We included probe sets that varied the most across samples. Additionally, a probe set was included only if > 10% of its measurements exceeded the per chip mean of 500. For each array, probe set values were \log_2 transformed and centered to median = 0. Normalization was performed so that all measurements for that array were multiplied by a scaling factor S such that the sum of the squares of the values equaled 1. Each probe set measurement was centered and normalized across samples according to the same procedure. Filtering and normalization were performed independently for each analysis.

Two-way unsupervised hierarchical clustering was performed using the software Cluster 3.0 (de Hoon *et al.*, 2004) and GeneSpring. To cluster data, we used an uncentered standard correlation (Pearson correlation around zero) as our measure of similarity. In constructing dendrograms, centroid linkage was used as the measure of proximity between clusters. PCA was performed using GeneSpring. Principal components were calculated for a designated set of genes and samples, and the three principal components representing the greatest variance in expression were plotted in order to visualize samples in three-dimensional gene expression space.

To identify differentially expressed genes between two groups, we used two different measures: fold change (ratio) between the normalized means of each group of samples and a Student's t -test. For gene expression data generated from cultured cells exposed to different treatments, the data were filtered to include only probe sets with an absolute expression value greater than 200 in at least one condition and differential expression was evaluated by fold change between different conditions.

To evaluate whether two gene lists contained a statistically significant number of overlapping genes, Fisher's exact test was used to calculate the hypergeometric probability of overlap between a specified list and a random list sampled from all genes. The resulting P -value was adjusted with a Bonferroni multiple testing correction.

To evaluate whether members of an *a priori* defined gene list were correlated with a specific tumor class, a modified KS analysis was performed as described (Subramanian *et al.*, 2005). Briefly, all 22 215 probe sets were first ranked according to the correlation between their expression and the class A designation. An enrichment score (ES) was then generated, which reflects the degree to which the members of the specified gene list were over-represented toward the top or bottom of the 22 215 ranked probe sets. The ES is analogous to a weighted KS statistic. To determine significance, the ES was recalculated with 1000 randomly permuted class labels to create a null distribution. A P -value for the observed ES was then calculated by comparing the observed ES to this null distribution.

Immunohistochemistry

Immunohistochemical detection was performed using streptavidin-biotin-peroxidase and microwave antigen retrieval methodology as described (Holzbeierlein *et al.*, 2004). Tissue blocks with multiple samples were prepared using a tissue arrayer (Beecher Instruments, Sun Prairie, WI, USA). For each sample, three 0.6 mm core sections of tissue were extracted from diagnostic areas of formalin-fixed, paraffin-embedded tissues. We defined Her2 positivity as 3+ by IHC, or 2+ by IHC with gene amplification of 2.1 or greater. Amplification was measured by fluorescence *in situ* hybridization as described (Lal *et al.*, 2005). For ER, PR, AR, ER β and FOXA1, samples were considered positive if greater than 10% of cell nuclei were immunoreactive. Fisher's exact test was used to compare the proportions of positive samples among independent groups. Semiquantitative analysis of ER expression was performed using whole sections obtained from the original paraffin-embedded tissue samples. Signal intensity was graded on a scale of 0–3. A final IHC score was computed by multiplying the percent of positive nuclei by the intensity. The magnitude of the difference in IHC score between groups was evaluated using the Mann–Whitney test. For a given gene, the correlation between IHC score and probe set intensity was calculated using Spearman's rank correlation coefficient. Estimates of sensitivity and specificity for predicting ER(–)/PR(–) class using ALCAM and SPDEF IHC scores were calculated with a receiver-operator characteristic curve using Graphpad Prism (Graphpad Software, San Diego, CA, USA).

Cell culture

The breast cancer cell lines MDA-MB-453, MDA-MB-231, SKBR-3, HCC-1937, ZR75-1, MCF7, BT-474 and T-47D were obtained from American Type Culture Collection (Rockville, MD, USA) (<http://www.atcc.org>). Cells were maintained at 37°C in a humidified atmosphere containing 5% CO₂ in 75 cm² flasks containing minimal essential medium (MEM) supplemented with 10% fetal bovine serum, 2% L-glutamine, non-essential amino acids (NEAA), 1 mM sodium pyruvate, 1.5 g/l sodium bicarbonate, 100 IU/ml penicillin and 100 μ g/ml streptomycin. Cells were passaged every 3–4 days when they reached 80% confluence, and harvested with 0.25% trypsin/EDTA.

For cell proliferation studies, cells were pelleted by centrifugation and resuspended in medium containing phenol red-free MEM supplemented with 10% charcoal-stripped fetal bovine serum (Hyclone, Logan, UT, USA), 2% L-glutamine, NEAA, 1 mM sodium pyruvate and 1.5 g/l sodium bicarbonate. Cells were plated in replicates of six at a density of 1×10^4 cells/well in 96-well microtiter plates. At 24 h after seeding, cells were treated with various reagents and media and reagents were replenished every 3 days. Reagents used were 10 nM E2 (Sigma-Aldrich, St Louis, MO, USA), 0.1–10 nM R-1881 (Sigma-Aldrich, St Louis, MO, USA), 10 μ M flutamide (Sigma-Aldrich, St Louis, MO, USA), 100 nM 4-OHT (tamoxifen) (Sigma-Aldrich, St Louis, MO, USA) and 100 nM anti-estrogen ICI 162,780 (fulvestrant, ICI) (Tocris, Ellisville, MO, USA). Cell viability and proliferation were measured using the 3-(4,5 dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide colorimetric assay (American Type Culture Collection) (Mosmann, 1983) and quantified by measuring absorbance at 570 nm (Victor V7 microplate reader, Perkin-Elmer, Wellesley, MA, USA).

Genome-wide expression profiling was performed for MDA-MB-453 cells in six experimental conditions that included incubation with combinations of androgen, AR antagonist and vehicle control. The six expression time-course

experiments, referred to as experiments I–VI, were performed simultaneously. Cells were grown to confluence in one 125 cm² flask, trypsinized, resuspended and seeded in six 75 cm² flasks at a density of 1×10^6 cells per flask. Cells were then incubated in media containing 10% FBS until 60% confluence, washed with ice-cold PBS and treated with media and reagents according to the six experimental conditions. Experiment I incubated cells in media containing 10% FBS; experiment II used charcoal-stripped media supplemented with vehicle control; experiment III used stripped media with 1 nM R-1881; experiment IV used stripped media with 1 nM R-1881 and 10 μ M flutamide. For experiments I–IV, RNA was extracted after 48 h. In experiments V and VI, cells were incubated in stripped media for 48 h and then exposed to either 1 nM R-1881 (V) or vehicle control (VI) for 48 h followed by RNA extraction.

Identification of ERE and ARE motifs

For each probe set, GenBank accession numbers identified specific genes. We retrieved 9999 bp of sequence 5' to the start of the transcription site for all genes from the ENSEMBL database using build NCBI 34 (Version 2), updated February 2004, from the Silicon Genetics website (<http://www.silicon-genetics.com/Downloads/HumanGenome9999.zip>). For genes of interest, sequence within 1–5000 bp upstream of the transcription site was analysed for homology to the ERE consensus 5'-GGTCAnnnTGACC-3' and the ARE consensus 5'-AGAACAAnnnTGTTCT-3'. We allowed for two single point discrepancies in each sequence homology analysis. For genes identified as having putative regulatory sequences, a false positive probability was estimated by observing both the frequency of the regulatory sequence upstream of all other genes and the frequency of the regulatory sequence within a random distribution of bases. In the latter case, the percent occurrence of each base in the random distribution was set to equal the percent occurrence of each base within the sequence in question. Genes with homologous response elements were

reported if the higher *P*-value obtained from these two observations was less than 0.0001.

Class prediction

A prediction algorithm was developed in order to identify samples that expressed a relevant gene signature. Tissue samples were assigned to a subclass based on our unsupervised hierarchical clustering of ER(–)/PR(–) tumors. Differentially expressed genes between the two clusters (designated classes A and B) were ranked by Student's *t*-test and those with a *P*-value <0.0001 were selected for use in the prediction model. The expression of each predictor gene was used to classify unknown samples using the *k*-nearest neighbors method (Golub *et al.*, 1999). Based on normalized expression values, we examined 11 samples near (as measured in Euclidian distance) the unclassified samples, and for each class, computed a *P*-value of the likelihood of finding the observed number of this class among the identified neighborhood members by chance, given the proportion of class membership in the training set. The class with the lowest *P*-value was assigned to the unclassified sample. We specified a *P*-value cutoff of 0.15, so that if there was not sufficient evidence in favor of a particular class, no prediction was made. The *P*-value cutoff is a ratio of the *P*-value of the predicted class to the alternate class.

Acknowledgements

We thank Dr Dennis Watson for generously providing SPDEF antibody, Yixin Wang and John A Foekens for providing unpublished data, Adam Olshen and Gaorav Gupta for advice and manuscript review, Louis Vargus, Yonghong Xiao and Lishi Chen for technical assistance, Andea Aleodor for histological analysis and Faye Taylor for data management. We are indebted to the Pathology and Genomics Core Facilities at MSKCC for technical support. WG was supported by DAMD BC010104.

References

- Agoff SN, Swanson PE, Linden H, Hawes SE, Lawton TJ. (2003). *Am J Clin Pathol* **120**: 725–731.
- Birrell SN, Bentel JM, Hickey TE, Ricciardelli C, Weger MA, Horsfall DJ *et al.* (1995). *J Steroid Biochem Mol Biol* **52**: 459–467.
- Buchanan G, Birrell SN, Peters AA, Bianco-Miotto T, Ramsay K, Cops EJ *et al.* (2005). *Cancer Res* **65**: 8487–8496.
- Carsol JL, Gingras S, Simard J. (2002). *Mol Endocrinol* **16**: 1696–1710.
- Cato AC, Nestl A, Mink S. (2002). *Sci STKE* **2002**: RE9.
- Cunliffe HE, Ringner M, Bilke S, Walker RL, Cheung JM, Chen Y *et al.* (2003). *Cancer Res* **63**: 7158–7166.
- de Hoon MJ, Imoto S, Nolan J, Miyano S. (2004). *Bioinformatics* **20**: 1453–1454.
- DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M *et al.* (1996). *Nat Genet* **14**: 457–460.
- Dhanasekaran SM, Dash A, Yu J, Maine IP, Laxman B, Tomlins SA *et al.* (2005). *FASEB J* **19**: 243–245.
- Dickson C, Spencer-Dene B, Dillon C, Fantl V. (2000). *Breast Cancer Res* **2**: 191–196.
- Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D *et al.* (2005). *Oncogene* **24**: 4660–4671.
- Gatalica Z. (1997). *Pathol Res Pract* **193**: 753–758.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP *et al.* (1999). *Science* **286**: 531–537.
- Goss PE, Ingle JN, Martino S, Robert NJ, Muss HB, Piccart MJ *et al.* (2003). *N Engl J Med* **349**: 1793–1802.
- Greeve MA, Allan RK, Harvey JM, Bentel JM. (2004). *J Mol Endocrinol* **32**: 793–810.
- Gruber CJ, Tschugguel W, Schneeberger C, Huber JC. (2002). *N Engl J Med* **346**: 340–352.
- Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A *et al.* (2001). *Cancer Res* **61**: 5979–5984.
- Hall RE, Aspinall JO, Horsfall DJ, Birrell SN, Bentel JM, Sutherland RL *et al.* (1996). *Br J Cancer* **74**: 1175–1180.
- Hall RE, Birrell SN, Tilley WD, Sutherland RL. (1994). *Eur J Cancer* **30A**: 484–490.
- Holzbeierlein J, Lal P, LaTulippe E, Smith A, Satagopan J, Zhang L *et al.* (2004). *Am J Pathol* **164**: 217–227.
- Howell A, Cuzick J, Baum M, Buzdar A, Dowsett M, Forbes JF *et al.* (2005). *Lancet* **365**: 60–62.
- Isola JJ. (1993). *J Pathol* **170**: 31–35.
- Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A *et al.* (2005). *CA Cancer J Clin* **55**: 10–30.
- Koziczak M, Hynes NE. (2004). *J Biol Chem* **279**: 50004–50011.
- Koziczak M, Holbro T, Hynes NE. (2004). *Oncogene* **23**: 3501–3508.
- Lacroix M, Leclercq G. (2004). *Breast Cancer Res Treat* **83**: 249–289.

- Laganiere J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguere V. (2005). *Proc Natl Acad Sci USA* **102**: 11651–11656.
- Lal P, Tan LK, Chen B. (2005). *Am J Clin Pathol* **123**: 541–546.
- Lippman M, Bolan G, Huff K. (1976). *Cancer Res* **36**: 4610–4618.
- Losel RM, Falkenstein E, Feuring M, Schultz A, Tillmann HC, Rossol-Haseroth K et al. (2003). *Physiol Rev* **83**: 965–1016.
- Matsuo K, Fukutomi T, Hasegawa T, Akashi-Tanaka S, Nanasawa T, Tsuda H. (2002). *Breast Cancer* **9**: 43–49.
- Mayfield S, Vaughn JP, Kute TE. (2001). *Breast Cancer Res Treat* **70**: 123–129.
- McKenna NJ, O'Malley BW. (2002). *Endocrinology* **143**: 2461–2465.
- Mellinghoff IK, Vivanco I, Kwon A, Tran C, Wongvipat J, Sawyers CL. (2004). *Cancer Cell* **6**: 517–527.
- Miller WR, Telford J, Dixon JM, Shivas AA. (1985). *Breast Cancer Res Treat* **5**: 67–73.
- Mosmann T. (1983). *J Immunol Methods* **65**: 55–63.
- Mouridsen H, Gershanovich M, Sun Y, Perez-Carrion R, Boni C, Monnier A et al. (2001). *J Clin Oncol* **19**: 2596–2606.
- Nagahata T, Onda M, Emi M, Nagai H, Tsumagari K, Fujimoto T et al. (2004). *Cancer Sci* **95**: 218–225.
- Nantermet PV, Masarachia P, Gentile MA, Pennypacker B, Xu J, Holder D et al. (2005). *Endocrinology* **146**: 564–578.
- Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, White J et al. (2002). *Proc Natl Acad Sci USA* **99**: 11890–11895.
- Osborne CK, Wakeling A, Nicholson RI. (2004). *Br J Cancer* **90**(Suppl 1): S2–S6.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. (2000). *Nature* **406**: 747–752.
- Pusztai L, Ayers M, Stec J, Clark E, Hess K, Stivers D et al. (2003). *Clin Cancer Res* **9**: 2406–2415.
- Rossi S, Graner E, Febbo P, Weinstein L, Bhattacharya N, Onody T et al. (2003). *Mol Cancer Res* **1**: 707–715.
- Sapp M, Malik A, Hanna W. (2003). *Breast J* **9**: 335–336.
- Shao D, Lazar MA. (1999). *J Clin Invest* **103**: 1617–1618.
- Shurbaji MS, Kalbfleisch JH, Thurmond TS. (1996). *Hum Pathol* **27**: 917–921.
- Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A et al. (2001). *N Engl J Med* **344**: 783–792.
- Smith IE, Dowsett M. (2003). *N Engl J Med* **348**: 2431–2442.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A et al. (2003). *Proc Natl Acad Sci USA* **100**: 8418–8423.
- Steketee K, Ziel-van der Made AC, van der Korput HA, Houtsmuller AB, Trapman J. (2004). *J Mol Endocrinol* **33**: 477–491.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. (2005). *Proc Natl Acad Sci USA* **102**: 15545–15550.
- Tang S, Han H, Bajic VB. (2004). *Nucleic Acids Res* **32**: D533–D536.
- Tobias JS. (2004). *Ann Oncol* **15**: 1738–1747.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al. (2002). *Nature* **415**: 530–536.
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F et al. (2005). *Lancet* **365**: 671–679.
- West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R et al. (2001). *Proc Natl Acad Sci USA* **98**: 11462–11467.
- Wilson CA, Dering J. (2004). *Breast Cancer Res* **6**: 192–200.
- Wong YC, Xie B. (2001). *Ital J Anat Embryol* **106**: 111–125.
- Yakes FM, Chinratanalab W, Ritter CA, King W, Seelig S, Arteaga CL. (2002). *Cancer Res* **62**: 4132–4141.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>)