Biomarkers in Cancer

ORIGINAL RESEARCH

Identification of Gene Expression Signature in Estrogen Receptor Positive Breast Carcinoma

Arvind D. Thakkar[1], Hemanth Raj[2], Debarshi Chakrabarti[1], Ravishankar[2], N. Saravanan[2], Bhaskaran Muthuvelan[3], Arun Balakrishnan[1] and Muralidhara Padigaru[1]

[1]Department of Biomarker Discovery, Piramal Life Sciences Limited, Mumbai, Maharashtra, India. [2]Apollo Specialty Hospital, Chennai, Tamil Nadu, India. [3]School of Bio Sciences and Technology, Vellore Institute of Technology University, Vellore, Tamil Nadu, India. Email: arvind.thakkar@piramal.com; muralidhara.padigaru@piramal.com

Abstract: A significant group of patient with estrogen receptor (ER) α positive breast tumors fails to appreciably respond to endocrine therapy. An increased understanding of the molecular basis of estrogen-mediated signal transduction and resultant gene expression may lead to novel strategies for treating breast cancer. In this study, we sought to identify the dysregulated genes in breast tumors related to ERα status. Microarray analyses of 31 tumor samples showed 108 genes differentially expressed in ERα (+) and ERα (–) primary breast tumors. Further analyses of gene lists indicated that a significant number of dysregulated genes were involved in mRNA transcription and cellular differentiation. The majority of these genes were found to have promoter-binding sites for E74-like factor 5 (ELF5; 54.6% genes), E2F transcription factor 1 (E2F1; 22.2% genes), and nuclear transcription factor Y alpha (NFYA; 32.4% genes). Six candidate genes (*NTN4, SLC7A8, MLPH, ENPP1, LAMB2*, and *PLAT*) with differential expression were selected for further validation studies using RT-qPCR (76 clinical specimen) and immunohistochemistry (48 clinical specimen). Our studies indicate significant overexpression of all the six genes in ERα (+) breast tumors as compared to ERα (–) breast tumors. *In vitro* studies using T-47D breast cancer cell line confirmed the estrogen dependant expression of four of the above six genes (*SLC7A8, ENPP1, LAMB2*, and *PLAT*). Collectively, our study provides further insights into the molecular basis of estrogen-dependent breast cancer and identifies "candidate biomarkers" that could be useful for predicting endocrine responsiveness.

Keywords: estrogen receptor, breast cancer, gene expression, estrogen response element

Biomarkers in Cancer 2010:2 1–15

This article is available from http://www.la-press.com.

Introduction

Breast cancer is a serious life threatening condition observed in women worldwide. It ranks second (after lung cancer) as a cause of cancer death in women. In US, from 1975 through 2003, 394,891 invasive and 59,837 in situ breast cancer cases were diagnosed in women.[1,2] An increased understanding of the pathogenesis of this disease is imperative in the pursuit of innovative therapies for treatment of and/or diagnosis of patients. Extensive research has been conducted to unravel the molecular basis of breast cancer. *In vitro, in vivo*, and, most importantly, clinically relevant studies have established that naturally occurring estrogens play a critical role in the initiation, progression and maintenance of breast cancers.[3] Under physiological conditions, estrogens are vital for the normal development, growth control, differentiation and function of breast.[4] However, under a variety of abnormal settings, increased sensitivity or longer exposures to estrogens predisposes an individual to breast cancer.

Estrogen stimulates proliferation of breast cancer cells, in large part, via estrogen receptors (ERs; members of the superfamily of nuclear receptors).[5] There are two main subtypes of ERs: ERα and ERβ. Both the ERs functionally act as transcription factors to initiate target gene expression.[6] Much is already known about the ER-mediated signal transduction pathway. The ERs regulate the transcription activity by two different activation domains: an activation function (AF)-1 and AF-2. The transcription of ERα involves serine phosphorylation events in AF-1 domain which influence both DNA binding and recruitment of cofactors.[7] In classical mechanism, upon binding of estrogen to ER, resulting ligand-receptor complex bind to DNA at estrogen response elements (ERE) in the promoters of target gene.[8] In nonclassical mechanism, the estrogen-ER complex can promote transcription via activator protein (AP)-1 and specificity protein (SP)-1 complexes.[6] Alternatively, it has been reported that E2F transcription factor 1 (E2F1) target genes are differentially expressed during the bimodal regulation of estrogen in ER α breast cancer cells.[9] The transcription factor E74-like factor 5 (ELF5) is member of the ETS (E-twenty six) transcription factor family.[10] As transcription factors, ETS proteins regulate the normal biological processes,[11] and also have oncogenic and tumor suppressive activity.[12,13]

Clinically, an ERα (+) status correlates with improved prognosis, lower risk of relapse and better overall survival.[14] Furthermore, ERα status is essential in making decisions about endocrine therapy with anti-estrogens. However, it is observed that approximately half of all ERα (+) patients fail to respond to anti-estrogen therapy.[15] The molecular basis for this paradoxical observation is currently under extensive investigation. Given these findings, in recent times, the progesterone receptor (*PgR*) positivity is determined along with ERα status.[16] An alternative and emerging trend is to adopt gene expression-based approaches to identify a set of genes that would determine hormone responsive-breast cancer phenotype. Indeed, several studies have shown differential expression of genes (e.g. *GATA3, TFF1*) in ERα (+) and ERα (–) breast cancers.[17,18]

To further elucidate the molecular basis of estrogen dependent breast carcinoma, we used oligo-based microarray to identify dysregulated gene signature that can discriminate between ERα (+) and ERα (–), followed-by validation of gene expression by RT-qPCR (reverse transcription quantitative real-time polymerase chain reaction) and tissue microarray based immunohistochemistry for protein levels. The gene expression analyses were combined with promoter sequence analysis for genes of interest.

Materials and Methods

Tumor specimens

All procedures involving human cells/specimen were approved by appropriate institutional human ethics committee. Breast tumor samples (Table 1) were obtained from patients undergoing surgery after informed written consent (Apollo Hospital; Chennai, India). The excised tumor specimen were immediately preserved in RNAlater (Sigma-Aldrich, MO, USA) and stored at 4 °C until shipment. All tumor samples utilized in this study were invasive ductal carcinoma. The receptor status of the tumors was determined at the hospital using immunohistochemistry (Supplementary Table 1). Thirty one tumor samples [15 ERα (+) and 16 ERα (–)], for which sufficient total RNA could be obtained (~20 µg total RNA), were used in the microarray analysis and additional 45 tumor samples [31 ERα (+) and 14 ERα (–)] were utilized in RT-qPCR studies.

Table 1. Clinicopathological characteristics of breast cancer patients.

Patient group	No. of patients
All patients	100
Age (years)	
Average	53.5
Less than 50	45
More than 50	55
Estrogen receptor	
Positive	53
Negative	40
Unknown	7
Progesterone receptor	
Positive	46
Negative	47
Unknown	7
HER2	
Positive	26
Negative	67
Unknown	7
Grade	
1	8
2	48
3	33
Unknown	11
Lymph node	
Negative	37
Positive	53
Unknown	10

RNA isolation and reverse transcription

The procedure followed was as described by others.[19] Briefly, 30 to 50 mg tissues were homogenized in 1 ml of TRIzol (Invitrogen Corporation, CA, USA). The samples were spun at 13,000 rpm for 15 mins at 4 °C. Subsequently, the upper aqueous layer was collected and purified using RNeasy Mini kit (Qiagen GmbH, Hilden, Germany). The concentration and purity of RNA samples was determined using NanoDrop ND-1000 spectrophotometer (NanoDrop products, DE, USA). RNA quality was determined using RNA 6000 Nano Lab-on-a-Chip kit (Agilent Technologies, Santa Clara, CA) on the Bioanalyzer 2100 (Agilent Technologies). RNA samples were reverse-transcribed in a total volume of 20 µL using 200 units of reverse transcriptase, 50 pmol of random hexamer, and 10 mM of deoxynucleotide triphosphates (Invitrogen).

Gene expression profiling

Microarray based gene expression data was generated by using 35 K human oligo chip based on 70-mer oligonucleotides (Operon, AL, USA). Fifteen µg of each tumor RNA and human universal reference RNA (Stratagene, CA, USA) was labeled with Cy5 and Cy3 (GE Healthcare, UK), respectively, according to the protocol established by Pronto Indirect Labeling kit (Promega Corporation, WI, USA). The labeled cDNA concentration and Cy5, Cy3 incorporation were assessed with NanoDrop spectrophotometer. The spotted aminosilane slides were pre-hybridized and post-hybridized according to Universal Microarray Hybridization kit (Corning Incorporated, NY, USA). Hybridization was run for 6 hrs at 42 °C, 6 hrs at 35 °C, and 6 hrs at 30 °C. Hybridized slides were scanned using the GeneTAC GT UC scanner (Genomic Solutions Inc., USA). GeneTAC Integrator software was used to analyze the image followed by global normalization. The microarray data were evaluated using GeneSpring Software (Agilent Technologies, Santa Clara, California).

RT-qPCR

Template cDNA were synthesized from total RNA isolated from tumor samples. All the PCR reactions were performed using the QuantiFast SYBR Green PCR Kit (Qiagen GmbH) as described by others.[20] Briefly, each PCR reaction contained 1X master mix, 1 µL of the diluted cDNA, and 250 nM of forward and reverse primers designed to yield 80 to 125-bp amplicons. PCR was carried out through 40 cycles (95 °C for 10 secs, 60 °C for 30 secs) following initial 3 mins enzyme activation at 95 °C. Reactions were carried out on an Eppendorf Realplex 4 (Eppendorf AG, Hamburg, Germany). The primers used in this study for RT-qPCR validation are listed in Table 2. Experiments were performed in duplicate for each data point and *GAPDH* gene was used as endogenous reference control. Results are expressed as mean ± 2 standard error based on Log2 transformation of normalized RT-qPCR values of the assayed genes. The fold change in expression of each gene

Table 2. Primer sequences used in RT-qPCR study.

Gene symbol	Forward primer	Reverse primer
GATA3	GTCCTGTGCGAACTGTCAGA	TTTCTGGTCTGGATGCCTTC
NTN4	ATTTTCCGAGGAAAGCGAAC	CCTCATGTCCTGCTACAAGGT
SLC7A8	CTAACCCTGGTGAGCCAGAA	CTGCTCCTCCATGTCCTCAT
MLPH	CAATGGCTGTGCCCTATCTT	CGAGCCTCGGTACACTGATT
ENPP1	ATCTCAGACGCCTTTGCACT	CTGTGATCCGTGCTCTGTGT
LAMB2	AGGCAAGGGCAGAACAACT	GGTGCCTTCCAATTCCTGTA
PLAT	AACAGTCACCGACAACATGC	CCATCGTTCAGACACACCAG

was calculated using the ΔΔ Ct method as described by Livak et al.[21]

Immunohistochemistry

Tissue microarray (TMA) slides were purchased from US biomax (US Biomax, MD, USA). These slides contained 48 cases (in duplicates) of common types of breast carcinoma, of which 24 cases (in duplicate) were invasive ductal carcinoma (IDC). The TMA slide manufacturer provided patient's AR/ER/PgR/Her-2 (neu) status and clinicopathological information. Slides were baked at 60 °C for 2 hrs and de-paraffinized in xylene and rehydrated through a graded alcohol series. The antigen was retrieved using antigen retrieval solution (Vector Labs, CA, USA). Subsequently, slides were placed in water bath at 95 °C for 20 mins before being immunostained using Vectastain® ABC Elite kits, in accordance with the manufacturer's protocol (Vector Labs). Briefly, sections were blocked by either 10% normal goat/rabbit serum for 1 hr followed by overnight incubation at 4 °C with primary antibody. The primary antibodies utilized included: anti-*NTN4* (1:100 dilution; R&D systems, Minneapolis, MN, USA), anti-*SLC7A8* (1:25 dilution; SantaCruz Biotechnology, CA, USA), anti-*MLPH* (1:25 dilution; SantaCruz Biotechnology), and anti-*ENPP1* (1:25 dilution; SantaCruz Biotechnology). Sections were then incubated with either biotinylated anti-rabbit/mouse antibody for 30 mins followed by Vectastain® Elite ABC reagent for 30 mins. Liquid diaminobenzidine (Vector labs; Burlingame, USA) was used as chromogenic agent and counterstained with Mayer's hematoxylin. Negative controls included slides incubated only with blocking buffer. Antibody stained tissues were assessed using scoring system based on the quickscore method (Detre et al, 1995). Briefly, the proportion of positive cells were estimated and given a score on a scale of 1 to 6 (1 = 0% to 4%; 2 = 5% to 19%; 3 = 20% to 39%; 4 = 40% to 59%; 5 = 60% to 79%; and 6 = 80% to 100%). The intensity of the staining was estimated and given a score from 0 to 3 (0 = no staining; 1 = weak; 2 = intermediate; and 3 = strong staining). A score was then calculated by multiplying the percentage of cells staining score by the intensity score, to yield a minimum value of 0 and a maximum value of 18.

Cell culture and treatments

T-47D cells were obtained from ATCC (Mannasas, VA) and cultured as described by others.[22] T-47D cells were maintained in DMEM medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Hyclone, USA), 1% penicillin/streptomycin. For estrogen treatments, cells were washed with PBS and pre-cultured in phenol-red-free DMEM medium supplemented with 4% charcoal-treated FBS (Hyclone, USA) for 48 hrs. Subsequently, T-47D cells were treated with varying concentrations of 17β-estradiol (E2; Sigma-Aldrich, MO, USA) or ICI 182780 (Sigma-Aldrich) for 96 hrs. Cells treated only with 0.1% ethanol were used as vehicle control. Following the completion of incubation period, the cells were washed and processed for gene expression studies as described above.

Statistical analyses

Mann-Whitney t-test was used to evaluate the difference between gene expression levels in ERα (+) and ERα (–) breast tumors. *P* values less than 0.05 were considered statistically significant. The over-represented transcription factor sites in the distal promoters of the differentially expressed genes were

Table 3. Differentially expressed genes in ERα (+) breast tumors identified by oligo microarray analyses. Mann-Whitney t-test was utilized to evaluate the statistical significance.

Gene name	Refseq accession	*P* value	Regulation in ERα (+)
mRNA transcription			
GATA3 (GATA binding protein 3)	NM_002051	0.0002	Up
ENPP1 (Ectonucleotide pyrophosphatase/phosphodiesterase 1)	NM_006208	0.0047	Up
CDC2 (Cell division cycle 2)	NM_033379	0.032	Down
FOXA1 (Forkhead box A1)	NM_004496	0.0015	Up
ESR1 (Estrogen receptor a)	NM_000125	0.0144	Up
Proteolysis			
NTN4 (Netrin-4 precursor)	NM_021229	0.0050	Up
NAT1 (Arylamine N-acetyltransferase 1)	NM_000662	0.0330	Up
Signal transduction			
LAMB2 (Laminin, beta 2)	NM_002292	1.977E-11	Up
PCSK6 (Proprotein convertase subtilisin/kexin type 6)	NM_138319	0.016	Up
SCUBE2 (Signal peptide, CUB domain, EGF-like 2)	NM_020974	0.037	Up
IL1R2 (Interleukin 1 receptor, type II)	NM_004633	0.038	Down
DNA repair			
KPNA2 (Karyopherin alpha-2)	NM_002266	0.0425	Down
TM4SF1 (Transmembrane 4 L6 family member 1)	NM_014220	0.0178	Down
DNA binding/Transcription factors			
HIST1H1B (histone cluster 1, H1b)	NM_005322	0.0133	Down
AFF3 (AF4/FMR2 family member 3)	NM_002285	0.0009	Up
RRM2 (Ribonucleoside-diphosphate reductase M2 subunit)	NM_001034	0.0459	Down
Other G-protein modulator			
KPNA2 (Karyopherin alpha-2)	NM_002266	0.0425	Down
TCERG1 (Transcription elongation regulator 1)	NM_006706	0.0045	Down
YWHAQ (Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide)	NM_006826	0.012	Down
Small GTPase			
YWHAZ (Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide)	NM_145690	0.0220	Down
SLC2A3 (solute carrier family 2 (facilitated glucose transporter), member 3)	NM_006931	0.0389	Down
RNA helicase			
PFKP (6-phosphofructokinase type C)	NM_002627	0.0407	Down
HAS2 (Hyaluronan synthase 2)	NM_005328	0.0159	Down
CALU (Calumenin precursor)	NM_005173	0.0488	Down
Cell adhesion molecules			
THBS2 (Thrombospondin-2 precursor)	NM_003247	0.0088	Down
MAGED2 (Melanoma-associated antigen D2)	NM_201222	0.0037	Up

carried out using oPOSSUM[23] analysis. The gene symbols were used as input and all classes of vertebrate transcription factors were screened for over-represented start sites using the JASPAR core database. The top 30% conservation with 5000bp sequences upstream and downstream of the TSS was used in the analysis. The significance for selecting the appropriate transcription factor was maintained with Z score (>5) or Fisher exact score (<0.05).

Results

Identification of a discriminating 108-gene signature associated with ERα status

Initially, we sought to determine the gene signature of breast tumor samples depending upon the expression status of ERα. Accordingly, microarray analyses of 15 ERα (+) and 16 ERα (–) breast tumors were conducted. Mann-Whitney t-test was performed to identify genes

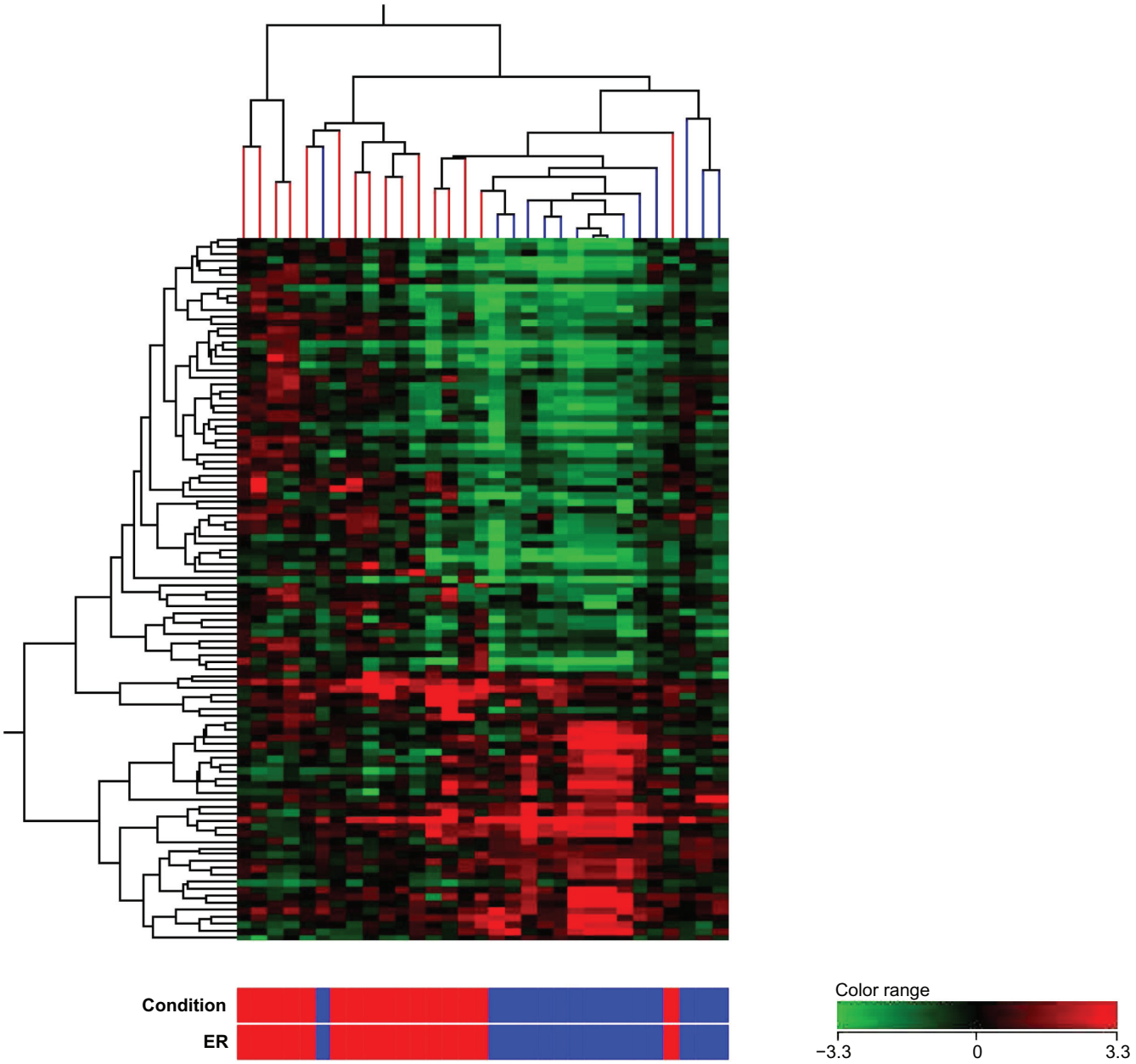

Figure 1. Dendrogram of 31 breast tumor samples (15 ERα (+) and 16 ERα (–)). Unsupervised hierarchical, uncentered Pearson distance correlation clustering was performed to classify the 108 genes into homogeneous clusters. The columns in the dendrogram represent the patient's tumor samples, while the rows represent the genes classified into clusters based on similar expression patterns. The expression color bar demonstrates the limits of regulation on either direction. The ERα (+) tumor samples are colored blue and ERα (–) samples colored red.

Table 4. Transcription factor binding sites over represented in the promoter region of ERα dysregulated genes.

Transcription factor	Percentage	Z-score	Target gene list
ERE (one ERE binding site)	14.0%	NA	*XBP1, IL1R2, KPNA2, S100A6, SLC2A3, THBS2*
ERE (multiple ERE binding site)	6.5%	NA	*PLAT, ICA1, LDHA, CSTB, DIAPH1 and UQCRH*
NFYA	32.4%	11.19	*GATA3, NTN4, SLC7A8, XBP1, AFF3, FADS2*
ELF5	54.6%	5.95	*ESR1, GATA3, NTN4, SLC7A8, ENPP1, MLPH, LAMB2, PLAT, XBP1, NAT1*
E2F1	22.2%	5.25	*ESR1, GATA3, NTN4, ENPP1, LAMB2, MAGED2, CALU, FOXA1*

differentially expressed in ERα (+) and ERα (–) groups (fold change ≥1.8 and $P \leq 0.05$). Subsequently, unsupervised clustering was carried out using a hierarchical algorithm and Pearson-based distance approach. These analyses discriminated 108 genes based on ERα status of breast tumor specimen (Fig. 1). Among the 108 genes, 41 genes were up regulated and 67 genes were down regulated in ERα (+) tumors as compared to ERα (–) (Supplementary Table 2). We performed a robust cross-platform validation of ERα-associated genes. Meta-analysis showed that 20% of the genes identified in our study was confirmed as having statistically up- or down-regulated in other studies[18,24–27] related to ERα (*ESR1, GATA3, XBP1, NAT1, FOXA1, IL1R2, SLC39A6, CALU, ID1, ICA1, PFKP, SCUBE2, PLAT, CDC2, S100A6, SLPI, SLC2A3*).

We next classified the 108 discriminating genes based on biological and molecular function. Expression Analysis Systematic Explorer software (EASE)[28] was used to annotate these genes according to the information provided by the Gene Ontology(GO) consortium.[29] The GO database provided annotation for 67% of the genes identified by our study (Supplementary Table 3). We observed that 54% of dysregulated genes were related to mRNA transcription regulation (e.g. *GATA3, ENPP1, CDC2, FOXA1, ESR1, MAGED2*), 29% were related to proteolysis (e.g. *NTN4, NAT1*), 26% were related to signal transduction (e.g. *LAMB2, PCSK6, CALU, SCUBE2, IL1R2*), and 11.1% were related to DNA repair (e.g. *KPNA2, YWHAZ, TM4SF1*). Further, molecular classification revealed that 41.7% of dysregulated genes were related to DNA binding (e.g. *HIST1H1B, AFF3, RRM2*), 25% were related to other G-protein modulator (e.g. *KPNA2, TCERG1, YWHAQ*), 15.3% were related to small GTPase (e.g. *YWHAZ, SLC2A3, MAGED2*), 13.9% were related to RNA helicase (e.g. *PFKP, HAS2, CALU*), and 8% were associated with cell adhesion molecules (e.g. *THBS2, MAGED2*) (Fig. 2 and Table 3). Interestingly, using the enrichment GO terms analysis, we identified statistical significant over-representation of specific groups of proteins including mRNA transcription and cellular differentiation. The observation of functionally related group of genes over representation analysis allows the identification of distinct biological pathways directly or indirectly associated to estrogen response related processes. Accordingly, we next utilized genome-wide high-affinity estrogen response elements (ERE) database[22] to identify putative EREs in the promoter region of the discriminating 108 genes. Interestingly, only a small fraction of the dysregulated genes contained high-affinity EREs (22 out of 108 genes; 20.3%). Sixty eight percent of these genes (15 out of 22) have one high affinity ERE and 32% of these genes (7 out of 22) contain two or more high affinity EREs. The transcriptional control of dysregulated genes were also investigated using in-silico approaches for mining the transcription factor binding sites (TFBS) across the 5' distal promoter region of the reported genes. However, the genes demonstrated high affinity binding sites for ELF5 (54.6% genes; Z-score = 5.95), E2F1 (22.2% genes; Z-score = 5.25), and NFYA (32.4% genes; Z-score = 11.19) among the significant selections (Table 4, see supplementary Table 4 for complete list of genes).

RT-qPCR validation of ERα (+)-associated transcripts in primary breast tumors

To validate microarray data, we chose a set of genes which were showing high statistical significance and relatively unexplored with reference to ER status in breast cancer. This set of genes included: GATA binding protein 3 (*GATA3*), Netrin-4 (*NTN4*), Solute carrier

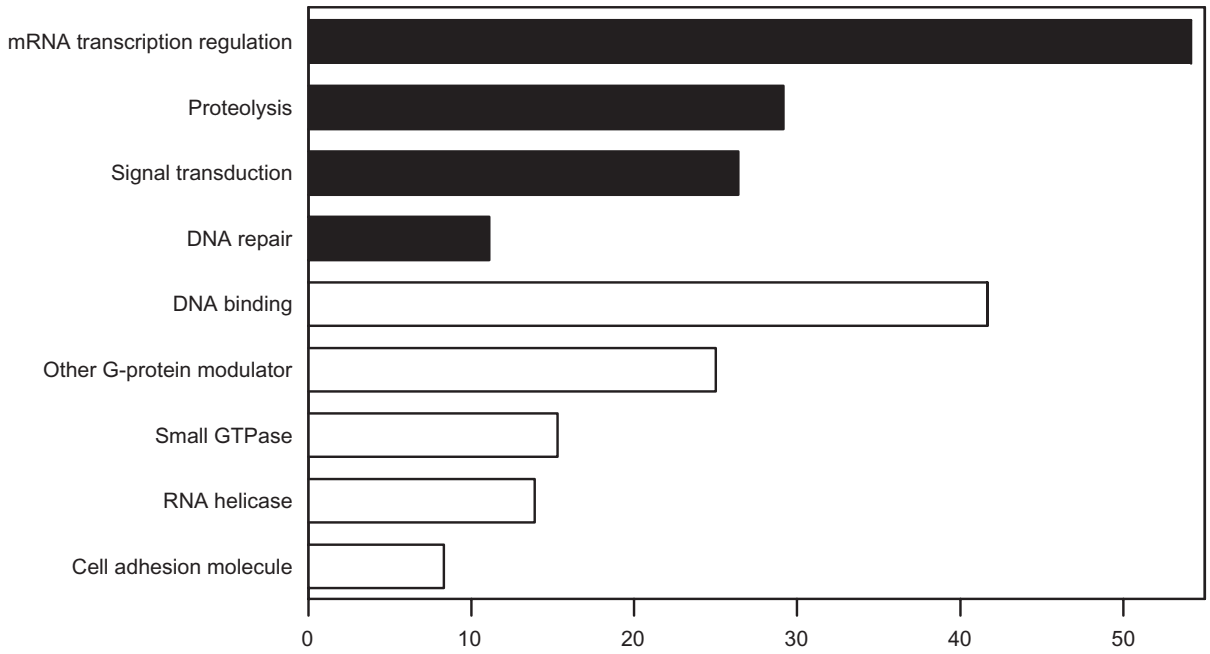

Figure 2. GO classification of the ERα associated genes. Percentage of genes annotated with a specific GO term related to biological process (black bars) and molecular function (white bars).

family 7 member 8 (*SLC7A8*), Melanophilin (*MLPH*), Ectonucleotide pyrophosphates/phosphodiesterase 1 (*ENPP1*), Laminin beta-2 chain precursor (*LAMB2*), and Plasminogen activator, tissue (*PLAT*). Since the dysregulation of *GATA3* to ERα (+) status has been extensively studied,[17,18] it was chosen as an "experimental validation control". Two complementary approaches were adopted to validate the data obtained using microarray analysis: RT-qPCR (mRNA expression) and immunohistochemistry (protein expression).

RT-qPCR analysis of the above 7 transcripts was performed in 31 tumors (for which microarray analyses were conducted) and in an independent set of 45 invasive ductal breast carcinomas. In agreement with our microarray analysis, we detected statistically significant over-expression of all the 7 genes in ERα (+) breast tumors as compared to ERα (−) breast tumors: *GATA3* ($P < 0.0001$), *NTN4* ($P < 0.0001$), *SLC7A8* ($P < 0.0001$), *MLPH* ($P < 0.0001$), *ENPP1* ($P < 0.0001$), *LAMB2* ($P = 0.0006$), and *PLAT* ($P = 0.003$) (Fig. 3).

Immunohistochemistry analysis of *NTN4, SLC7A8, MLPH* and *ENPP1*

Immunohistochemistry analysis was carried out to validate our data using tumor tissues from an independent set of patient cohort, obtained as tissue microarray slides from US biomax due to non-availability of paraffin blocks for our patient cohorts. We studied the protein expression for *NTN4, SLC7A8, MLPH, ENPP1* and their ability to discriminate ERα (+) and ERα (−) breast tumors. As shown in Figure 4, tumor tissues from ERα (+) group showed strong staining for our selected proteins as compared to weak staining associated with ERα (−) group. Furthermore, unpaired t-test demonstrated statistically significant difference in the expression levels of these proteins in ERα (+) and ERα (−) breast cancers: *NTN4* ($P < 0.0009$), *SLC7A8* ($P < 0.007$), *MLPH* ($P < 0.0001$), and *ENPP1* ($P < 0.0147$).

Estrogen regulates the mRNA expression of *SLC7A8, ENPP1, LAMB2* and *PLAT* in *T-47D* breast cancer cells

Further we investigated the estrogen dependant expression *of NTN4, SLC7A8, MLPH, ENPP1, LAMB2,* and *PLAT* in ERα (+) breast cancer. We analyzed the estrogen (17β-estradiol) induced mRNA expression in T-47D breast cancer cell line in the presence or absence of inhibitors of estrogen signaling pathway (ICI 182780 and tamoxifen). Since *GATA3* is not regulated by estrogen under *in-vitro* conditions,[30] *TFF1*, a well-known ERα induced gene under *in vitro* conditions, was used as a positive control.[31] The relative

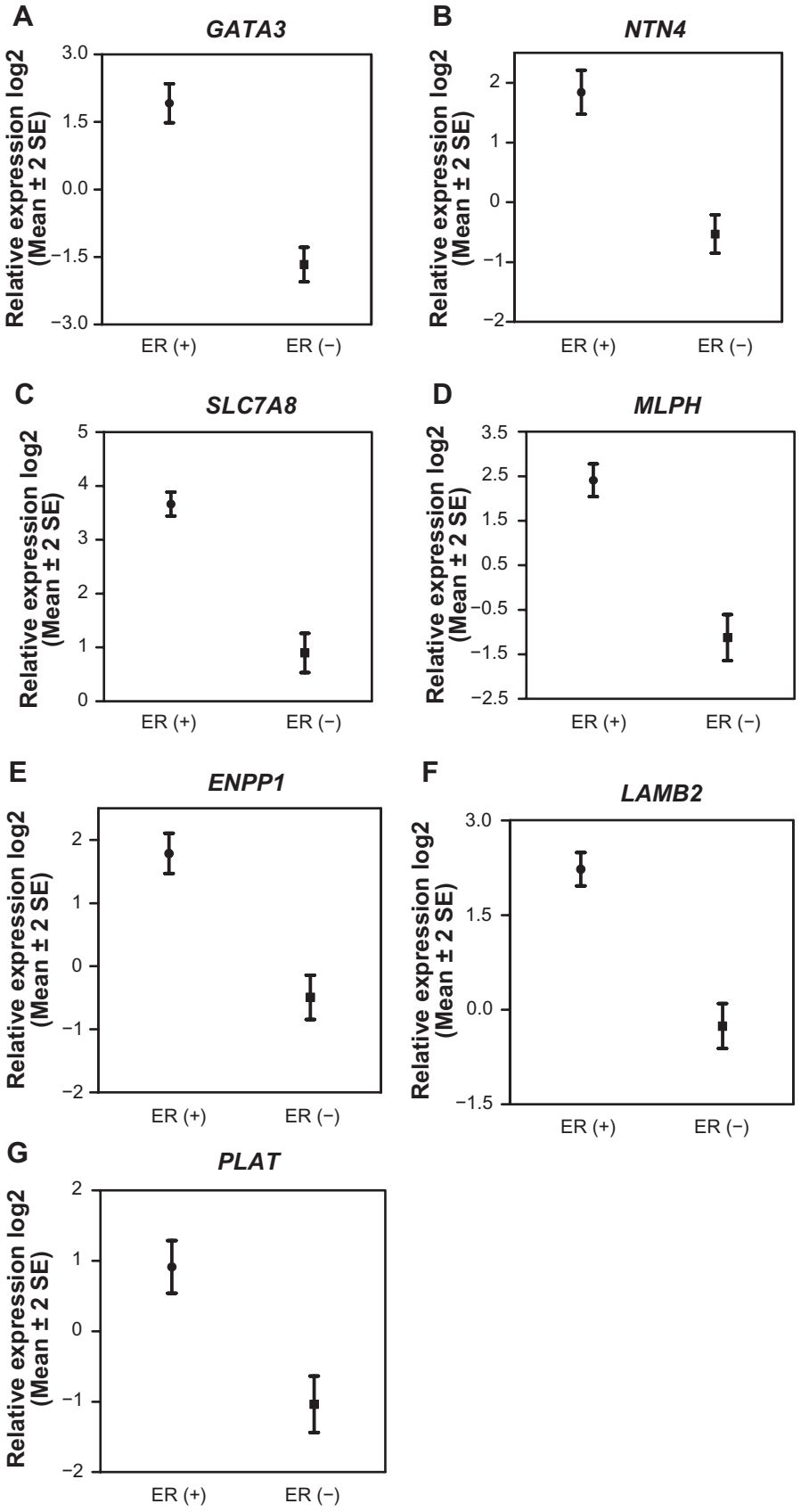

Figure 3. RT-qPCR validation of seven over expressed genes in 76 invasive breast carcinomas. Results were expressed as mean ± 2 standard error based on Log2 transformation of normalized RT-qPCR values of the assayed genes. *GAPDH* gene was used as normalization control. **A**) *GATA3* ($P < 0.0001$); **B**) *NTN4* ($P < 0.0001$); **C**) *SLC7A8* ($P < 0.0001$); **D**) *MLPH* ($P < 0.0001$); **E**) *ENPP1* ($P < 0.0001$); **F**) *LAMB2* ($P = 0.0006$); **G**) *PLAT* ($P = 0.003$).

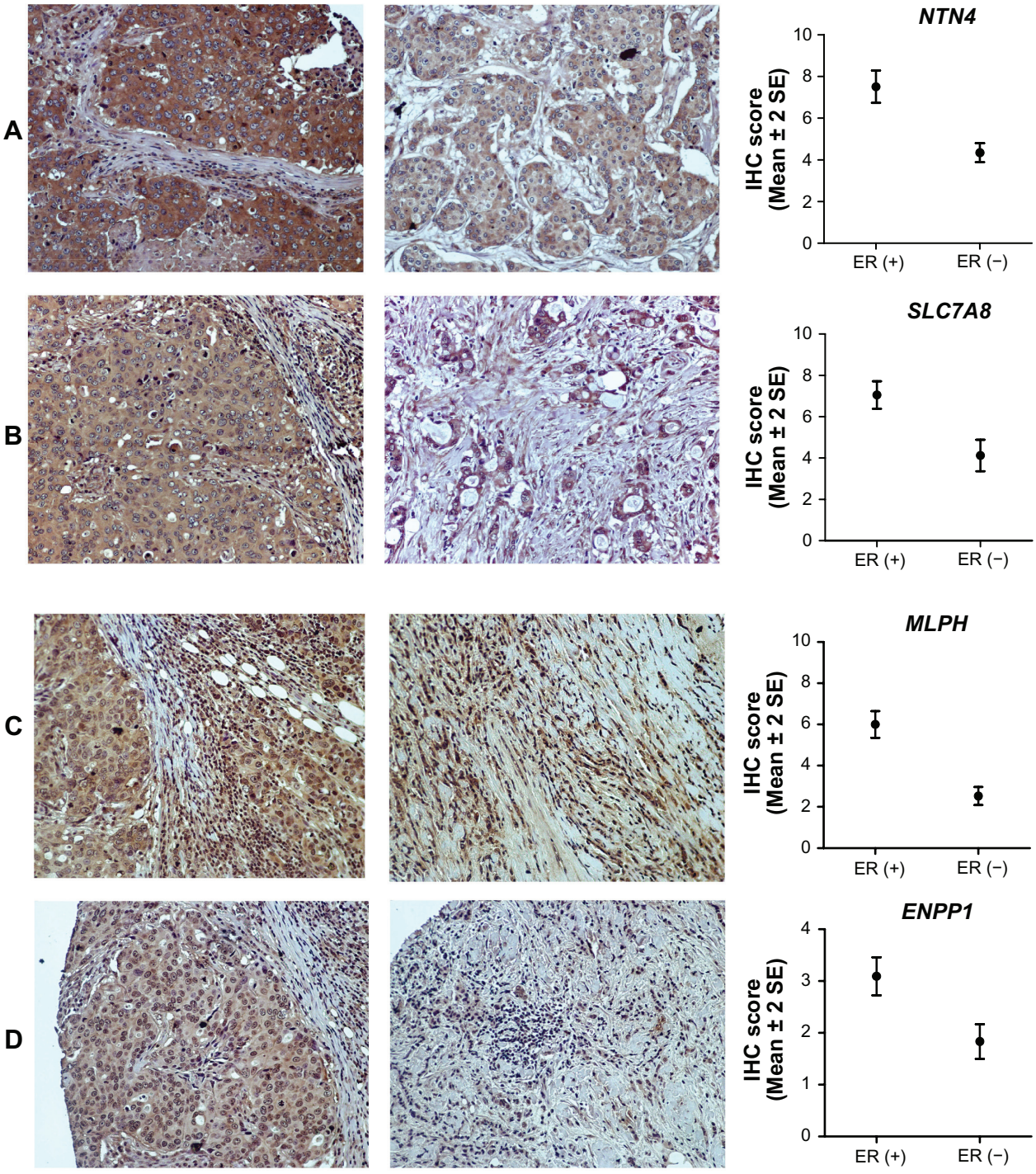

Figure 4. IHC staining for the expression of A) *NTN4* ($P < 0.0009$); B) *SLC7A8* ($P < 0.007$); C) *MLPH* ($P < 0.0001$); and D) *ENPP1* ($P < 0.0147$) in breast invasive ductal carcinoma. The left panel shows ERα positive tissue and right panel shows ERα negative tissue. Antibody stained tissues were assessed using scoring system based on the quickscore method. Results were expressed as mean ± 2 standard error based on IHC scores. Representative results are shown (magnification, 200X).

expression of the genes was determined by RT-qPCR and the results were expressed as fold change as compared to control cells (vehicle control). As shown in Figure 5, estrogen up-regulated the mRNA expression of *SLC7A8, ENPP1, LAMB2*, and *PLAT* (≥1.8 fold expression as compared to the control cells). ICI 182780 and tamoxifen abrogated the estrogen-induced upregulation of these genes (Fig. 5). Interestingly, we

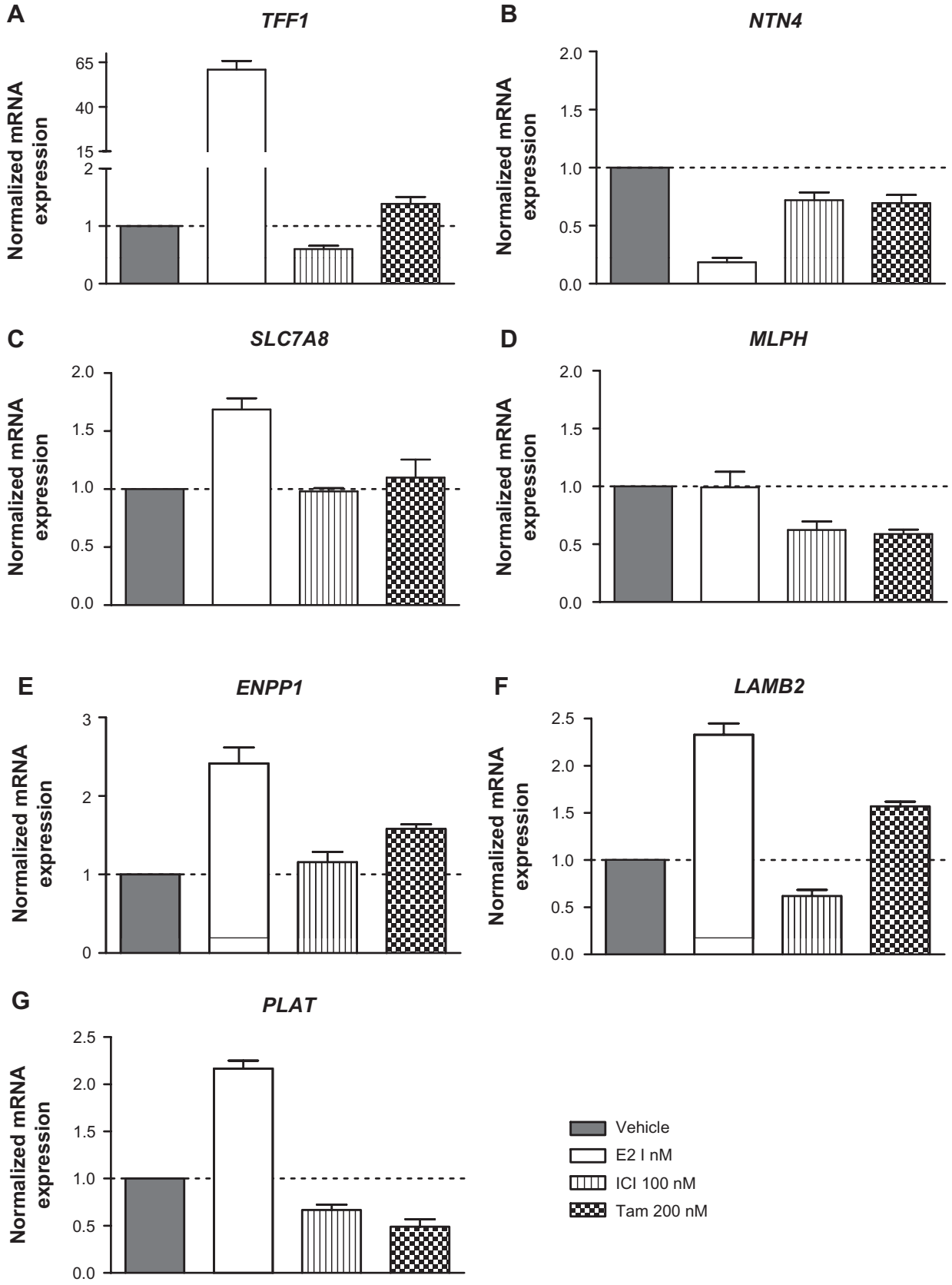

Figure 5. mRNA levels of A) *TFF1*, B) *NTN4*, C) *SLC7A8*, D) *MLPH*, E) *ENPP1*, F) *LAMB2*, and G) *PLAT* in T-47D cells after 4 days of culture in steroid-depleted medium or with E2, ICI 182780 (ICI) and tamoxifen (Tam). T-47D cells were treated with 1nM of 17β-estradiol (E2), 100 nM ICI 182, 780 and 200 nM of tamoxifen for 96 hrs. As a control, cells were treated with 0.1% ethanol (vehicle control) for 96 hrs. For each gene, mRNA levels were normalized with *GAPDH* and such that the value of the vehicle samples was 1.

observed that estrogen treatment down-regulated the mRNA expression of *NTN4* and had no effect on mRNA expression of *MLPH*.

Discussion

Half of all patients with ERα (+) breast tumor fail to respond favorably to anti-estrogen therapy. Identification of novel "molecular or biological" markers may lead to better understanding of the role of estrogen in breast tumorigenesis. Identification of genes that co-cluster with ER status is a first step towards identifying reliable markers to predict ER status and response to endocrine therapy. The current study has attempted to identify signatures that could be used as potential classifiers for ERα status in breast cancer patients in addition to globally accepted list of ERα classifiers. Accordingly, we utilized an oligo microarray approach to measure the expression of large number of genes (approx. 35,000) in 31 breast tumor samples. These analyses discriminated 108 genes based on ERα status of breast tumor specimen. Confirming data sets generated on different gene expression platforms increases the confidence of specific gene expression classifier data sets.[32] We did not observe 100% overlap of findings between various studies[18,24–27,33] and our study. This is not entirely surprising given that these studies have been done with different platforms, different number of genes in the various platforms and heterogeneous patient populations (with regard to age, tumor staging and treatment).

Classification of genes based on Gene Ontology (GO) terms is a powerful bioinformatics tool suited for the analysis of DNA microarray data. Analysis of GO annotation allows one to identify families of genes that may play significant roles related to specific molecular or biological processes in expression profiles.[29] Ontology analysis on data for biological function revealed genes belong to functional categories such as mRNA transcription regulation (59%), proteolysis (29%), signal transduction (26%), and DNA repair (11.1%). Similarly the molecular function categories include those involved in DNA binding (41.7%), G-protein modulators (25%), small GTPases (15.3%), RNA helicases (13.9%), and cell adhesion (8%). Gene enrichment analysis indicated that majority of dysregulated genes were involved in mRNA transcription and cellular differentiation. The observation of functionally related groups of genes identified via GO over representation analysis helps in understanding of distinct biological pathways associated to estrogen response related processes.

Accordingly, we used genome-wide high affinity estrogen response elements (ERE) database to search for ERE binding sites. Fourteen percent genes showed one ERE binding site and 6.5% genes showed two or more ERE binding sites. In our efforts to identify EREs in the promoter region of the dysregulated genes, only a small fraction of the dysregulated genes contained high affinity EREs. These observations are in line with earlier reports.[33] The possibility exists that many of these genes are transcriptionally regulated by non-ERE mediated mechanisms. The transcriptional factor binding sites (TFBS) analyses using oPPOSUM led to identification of ELF5 binding sites in 54.6% genes, E2F1 binding sites in 22.2% genes, and NFYA binding sites in 32.4% genes. Stender *et al* reported over representation of E2F in the promoter region of many cell cycle related genes stimulated by estrogen in MCF-7.[9] Another study reported RNA interference mediated knockdown of E2F1 blocked estrogen regulation resulted in loss of estrogen regulation of proliferation. The ELF5 transcription factor is a member of the ETS subfamily.[10] ETS proteins regulate biological processes including development, differentiation, proliferation and apoptosis and have oncogenic and tumor suppressive activity.[34,35] The T47D breast cancer cell line was observed to be express ELF5 transcription factor.[13] This is the first evidence to demonstrate that these transcription factors play an important role in expression of ERα dysregulated genes of patient samples. However, these genes need to be validated in larger patient cohort to further establish the regulatory role of these transcription factors in breast cancer biology.

It is noteworthy that some of these dysregulated genes that code for secreted proteins such as *NTN4, SLC7A8* and *PLAT* could potentially be used in development of plasma/serum based predictive biomarkers. However additional studies are required to investigate the clinical utility of these markers. The set of genes selected based on high statistical significance in ERα (+) tumors include: *NTN4, SLC7A8, MLPH, ENPP1, LAMB2*, and *PLAT*. These six genes of interest were then investigated in independent set of 46 ERα (+) and 30 ER (–) patient cohort including 31 tumor samples used for microarray analysis. All six genes showed mRNA over expression in ERα (+) patients compared with ERα (–) patients, making them putative

ERα-responsive genes. To validate these signatures at protein level, we performed immunohistochemistry on tissue microarray slide containing 24 tissue cores in duplicate. Intense staining of tissues indicated the higher expression of these proteins associated with ERα (+) breast tumors.

To further investigate the potential role of these signatures in ERα (+) breast cancer, we analysed the expression of these genes in estrogen responsive and tamoxifen sensitive T47D cell line.[36] We found that four out of six genes (*SLC7A8, ENPP1, LAMB2*, and *PLAT*) were regulated by estrogen. Moreover, the estrogen response was abolished by ICI 182780 treatment for all four genes but tamoxifen could only reduce the expression of *PLAT*. Our findings that only few genes are estrogen responsive in cell culture are in line with earlier reports.[25] There are several possible explanations for these findings. The existence of other ER-signaling pathways, independent of estrogen has been postulated.[37,38] The observation that these transcriptional activities are manifested in a tissue selective manner suggests that the receptor does not function in isolation, but rather, requires specific cellular factors for maximal responses. The complex network of coactivators and corepressors provide balanced, and sensitive control of ER target gene expression.[39]

NTN4 is a secreted molecule with roles in axon guidance and angiogenesis. *NTN4* acts as an antiangiogenic factor through binding to neogenin and recruitment of *UNC5B*.[40] The *NTN4* expression is associated with longer disease-free survival and overall survival in breast cancer patients.[41] The Shennan's study confirms that MCF-7 cells express *LAT*1 and *SLC7A8* (*LAT2*) mRNA but MDA-MB-231 cells express only *LAT*1 mRNA. A *SLC7A8* expression and activity in MCF-7 cells is also up-regulated by 17β-estradiol and could contribute to the proliferative capacity by increasing amino acid uptake via systems A and L.[42,43] Our study confirms the above observation with higher expression of *SLC7A8* in ERα (+) than ERα (–) breast cancer patients. The *MLPH* gene encodes a member of the exophilin subfamily of *RAB* effector proteins.[44] The low expression of *ER, PgR, HER2* and *MLPH* genes expression was reported in basal-like subtypes in high risk breast cancer patients.[45] Another study reported down regulation of *MLPH* gene in lymph node positive breast cancer patients.[24] Our study confirms the lower expression of *MLPH* in ERα (–) breast cancer patients.

A study revealed that *ENPP1* was a downstream target of AR (Androgen receptor) and expression of *ENPP1* might play a potential role in the development of androgen-refractory prostate cancer. *ENPP1* over expression also promoted the tumorigenic phenotype *in vitro* and *in vivo* during androgen-depleted condition.[46] The other study showed expression of *ENPP1* as a positive regulator in the Akt signaling pathway.[47] *LAMB2* belongs to the laminin family and are secretory proteins localized to the extra-cellular matrix and basement membrane in breast tissues. Laminins have been reported in a variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis.[48] *PLAT* is the core protein involved in physiological plasminogen action in the tissue. *PLAT* is also involved in cell migration of epithelial/myo-epithelial cells in the human breast.[49]

In conclusion, our study identified and validated estrogen-regulated genes (*NTN4, SLC7A8, ENPP1, MLPH, LAMB2* and *PLAT*). The reliability of the genes identified in this study was reinforced by validation at the mRNA and at protein level. This work provides potential candidates for understanding the pharmacological effects of estrogen and their consequences in estrogen-dependent diseases. We are now studying relevance of these signatures in predicting prognosis/risk classification and would depend upon the use and validation of these signatures in meta-analysis of breast cancer studies.

Abbreviations

AF-1, activation function 1; AP-1, activation protein 1; ER, estrogen receptor; ERE, estrogen response element; E2, 17β-estradiol; GO, gene ontology; PBS, phosphate buffer saline; PgR, progesterone receptor; RT-qPCR, reverse transcription quantitative real-time polymerase chain reaction; TFBS, transcription factor binding site; ELF5, E74-like factor 5; E2F1, E2F transcription factor 1; TMA, tissue microarray; NFYA, nuclear transcription factor Y alpha; AR, androgen receptor.

Acknowledgements

We thank Dr. Nilesh M. Dagia for critically reviewing this manuscript.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors report no conflicts of interest.

References

1. Hausauer AK, Keegan TH, Chang ET, Clarke CA. Recent breast cancer trends among Asian/Pacific Islander, Hispanic, and African-American women in the US: changes by tumor subtype. *Breast Cancer Res*. 2007;9(6):R90.
2. Jemal A, Ward E, Thun MJ. Recent trends in breast cancer incidence rates by age and tumor characteristics among US women. *Breast Cancer Res*. 2007;9(3):R28.
3. Couse JF, Korach KS. Estrogen receptor null mice: what have we learned and where will they lead us? *Endocr Rev*. 1999 Jun;20(3):358–417.
4. Shekhar MP, Werdell J, Santner SJ, Pauley RJ, Tait L. Breast stroma plays a dominant regulatory role in breast epithelial growth and differentiation: implications for tumor development and progression. *Cancer Res*. 2001 Feb 15; 61(4):1320–6.
5. Simpson ER. Sources of estrogen and their importance. *J Steroid Biochem Mol Biol*. 2003 Sep;86(3–5):225–30.
6. Heldring N, Pike A, Andersson S, et al. Estrogen receptors: how do they signal and what are their targets. *Physiol Rev*. 2007 Jul;87(3):905–31.
7. Hall JM, Couse JF, Korach KS. The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J Biol Chem*. 2001 Oct 5;276(40):36869–72.
8. O'Lone R, Frith MC, Karlsson EK, Hansen U. Genomic targets of nuclear estrogen receptors. *Mol Endocrinol*. 2004 Aug;18(8):1859–75.
9. Stender JD, Frasor J, Komm B, Chang KC, Kraus WL, Katzenellenbogen BS. Estrogen-regulated gene networks in human breast cancer cells: involvement of E2F1 in the regulation of cell proliferation. *Mol Endocrinol*. 2007 Sep;21(9):2112–23.
10. Sharrocks AD. The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol*. 2001 Nov;2(11):827–37.
11. Sharrocks AD, Brown AL, Ling Y, Yates PR. The ETS-domain transcription factor family. *Int J Biochem Cell Biol*. 1997 Dec;29(12):1371–87.
12. Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A*. 2003 May 13;100(10):5974–9.
13. Zhou J, Ng AY, Tymms MJ, et al. A novel transcription factor, ELF5, belongs to the ELF subfamily of ETS genes and maps to human chromosome 11p13-15, a region subject to LOH and rearrangement in human carcinoma cell lines. *Oncogene*. 1998 Nov 26;17(21):2719–32.
14. Ring BZ, Seitz RS, Beck R, et al. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006 Jul 1;24(19):3039–47.
15. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol*. 2006 Apr 10; 24(11):1656–64.
16. Stendahl M, Ryden L, Nordenskjold B, Jonsson PE, Landberg G, Jirstrom K. High progesterone receptor expression correlates to the effect of adjuvant tamoxifen in premenopausal breast cancer patients. *Clin Cancer Res*. 2006 Aug 1;12(15):4614–8.
17. Mehra R, Varambally S, Ding L, et al. Identification of *GATA3* as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res*. 2005 Dec 15;65(24):11259–64.
18. West M, Blanchette C, Dressman H, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A*. 2001 Sep 25;98(20):11462–7.
19. Bouras T, Southey MC, Chang AC, et al. Stanniocalcin 2 is an estrogen-responsive gene coexpressed with the estrogen receptor in human breast cancer. *Cancer Res*. 2002 Mar 1;62(5):1289–95.
20. Tozlu S, Girault I, Vacher S, et al. Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. *Endocr Relat Cancer*. 2006 Dec;13(4):1109–20.
21. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001 Dec;25(4):402–8.
22. Bourdeau V, Deschenes J, Metivier R, et al. Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol Endocrinol*. 2004 Jun;18(6):1411–27.
23. Ho Sui SJ, Mortimer JR, Arenillas DJ, et al. oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res*. 2005;33(10):3154–64.
24. Abba MC, Sun H, Hawkins KA, et al. Breast cancer molecular signatures as determined by SAGE: correlation with lymph node status. *Mol Cancer Res*. 2007 Sep;5(9):881–90.
25. Gruvberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res*. 2001 Aug 15;61(16):5979–84.
26. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A*. 2003 Sep 2;100(18):10393–8.
27. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002 Jan 31;415(6871): 530–6.
28. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol*. 2003;4(10):R70.
29. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000 May;25(1):25–9.
30. Hoch RV, Thompson DA, Baker RJ, Weigel RJ. *GATA-3* is expressed in association with estrogen receptor in breast cancer. *Int J Cancer*. 1999 Apr 20; 84(2):122–8.
31. Masiakowski P, Breathnach R, Bloch J, Gannon F, Krust A, Chambon P. Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line. *Nucleic Acids Res*. 1982 Dec 20; 10(24):7895–903.
32. Detours V, Dumont JE, Bersini H, Maenhaut C. Integration and cross-validation of high-throughput gene expression data: comparing heterogeneous data sets. *FEBS Lett*. 2003 Jul 3;546(1):98–102.
33. Abba MC, Hu Y, Sun H, et al. Gene expression signature of estrogen receptor alpha status in breast cancer. *BMC Genomics*. 2005;6(1):37.
34. Desmaze C, Brizard F, Turc-Carel C, et al. Multiple chromosomal mechanisms generate an EWS/FLI1 or an EWS/ERG fusion gene in Ewing tumors. *Cancer Genet Cytogenet*. 1997 Aug;97(1):12–9.
35. Peter M, Couturier J, Pacquement H, et al. A new member of the ETS family fused to EWS in Ewing tumors. *Oncogene*. 1997 Mar 13;14(10):1159–64.
36. Karey KP, Sirbasku DA. Differential responsiveness of human breast cancer cell lines MCF-7 and T47D to growth factors and 17 beta-estradiol. *Cancer Res*. 1988 Jul 15;48(14):4083–92.
37. Ding L, Yan J, Zhu J, et al. Ligand-independent activation of estrogen receptor alpha by XBP-1. *Nucleic Acids Res*. 2003 Sep 15;31(18):5266–74.
38. Zwijsen RM, Buckle RS, Hijmans EM, Loomans CJ, Bernards R. Ligand-independent recruitment of steroid receptor coactivators to estrogen receptor by cyclin D1. *Genes Dev*. 1998 Nov 15;12(22):3488–98.
39. Hall JM, McDonnell DP. Coregulators in nuclear estrogen receptor action: from concept to therapeutic targeting. *Mol Interv*. 2005 Dec;5(6):343–57.
40. Lejmi E, Leconte L, Pedron-Mazoyer S, et al. Netrin-4 inhibits angiogenesis via binding to neogenin and recruitment of Unc5B. *Proc Natl Acad Sci U S A*. 2008 Aug 26;105(34):12491–6.
41. Esseghir S, Kennedy A, Seedhar P, et al. Identification of *NTN*4, TRA1, and STC2 as prognostic markers in breast cancer in a screen for signal sequence encoding proteins. *Clin Cancer Res*. 2007 Jun 1;13(11):3164–73.
42. del Amo EM, Urtti A, Yliperttula M. Pharmacokinetic role of L-type amino acid transporters LAT1 and LAT2. *Eur J Pharm Sci*. 2008 Oct 2;35(3): 161–74.
43. Shennan DB, Thomson J, Gow IF, Travers MT, Barber MC. L-leucine transport in human breast cancer cells (MCF-7 and MDA-MB-231): kinetics, regulation by estrogen and molecular identity of the transporter. *Biochim Biophys Acta*. 2004 Aug 30;1664(2):206–16.

44. Fukuda M, Kuroda TS, Mikoshiba K. Slac2-a/melanophilin, the missing link between Rab27 and myosin Va: implications of a tripartite protein complex for melanosome transport. *J Biol Chem*. 2002 Apr 5;277(14):12432–6.
45. Fountzilas G, Wirtz RM, Skarlos P, et al. Molecular classification based on only four genes identifies basal-like subtypes in high-risk breast cancer patients. *J Clin Oncol* 26. 2008;(May 20 suppl; abstr 11103).
46. Wang R, Xu J, Saramaki O, et al. PrLZ, a novel prostate-specific and androgen-responsive gene of the TPD52 family, amplified in chromosome 8q21.1 and overexpressed in human prostate cancer. *Cancer Res*. 2004 Mar 1; 64(5):1589–94.
47. Zhang H, Wang J, Pang B, et al. PC-1/PrLZ contributes to malignant progression in prostate cancer. *Cancer Res*. 2007 Sep 15;67(18):8906–13.
48. Patarroyo M, Tryggvason K, Virtanen I. Laminin isoforms in tumor invasion, angiogenesis and metastasis. *Semin Cancer Biol*. 2002 Jun;12(3):197–207.
49. Vassalli JD, Sappino AP, Belin D. The plasminogen activator/plasmin system. *J Clin Invest*. 1991 Oct;88(4):1067–72.