Microarray analysis of suppression subtracted hybridisation libraries identifies genes associated with breast cancer progression

Dong Liu Barraclough [a,b,*], Susan Sewart [a,b], Philip S. Rudland [a,b], Balvinder S. Shoker [c],
D. Ross Sibson [d,e], Roger Barraclough [b,**] and Michael P.A. Davies [d,e]

[a] *Cancer Tissue Bank Research Centre, University of Liverpool, Liverpool, UK*

[b] *School of Biological Sciences, University of Liverpool, Liverpool, UK*

[c] *Department of Cellular Pathology, Royal Hampshire County Hospital, Winchester, Hampshire, UK*

[d] *Clatterbridge Cancer Research Trust, Clatterbridge Hospital, Bebington, Wirral, UK*

[e] *School of Cancer Studies, University of Liverpool, Liverpool, UK*

Abstract. *Background:* A major challenge of cancer research is to identify key molecules which are responsible for the development of the malignant metastatic phenotype, the major cause of cancer death.

Methods: Four subtracted cDNA libraries were constructed representing mRNAs differentially expressed between benign and malignant human breast tumour cells and between micro-dissected breast carcinoma *in situ* and invasive carcinoma. Hundreds of differentially expressed cDNAs from the libraries were micro-arrayed and screened with mRNAs from human breast tumor cell lines and clinical specimens. Gene products were further examined by RT-PCR and correlated with clinical data.

Results: The combination of subtractive hybridisation and microarray analysis has identified a panel of 15 cDNAs which shows strong correlations with estrogen receptor status, malignancy or relapse. This panel included S100P, which was associated with aneuploidy in cell lines and relapse/death in patients, and AGR2 which was associated with estrogen receptor and with patient relapse. X-box binding protein-1 is also an estrogen-dependent gene and is associated with better survival for breast cancer patients.

Conclusions: The combination of subtracted cDNA libraries and microarray analysis has thus identified potential diagnostic/prognostic biomarkers and targets for cancer therapy, which have not been identified from common prognostic gene signatures.

Keywords: Suppression subtracted hybridisation, cDNA microarray, patient survival, breast cancer, quantitative RT-PCR

Abbreviations

AGR2 anterior gradient-2;
ERα estrogen receptor α;
DCIS ductal carcinoma *in situ*;
GAPDH glyceraldehyde-3-phosphate dehydrogenase;
HPRT hypoxanthine phosphoribosyl transferase;
IDC invasive ductal carcinoma;
PCR polymerase chain reaction;
ROC receiver operating characteristic;
RT reverse transcription;
XBP-1 X-box binding protein-1.

*Present address: School of Clinical Sciences, University of Liverpool, Liverpool, L69 3GA, UK.

**Corresponding author: Dr. Roger Barraclough, School of Biological Sciences, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK. Tel.: +44 151 795 4469; Fax: +44 151 795 4406; E-mail: brb@liverpool.ac.uk.

1. Introduction

The behaviour of breast and other cancers probably reflects the pattern of expression of genes which contribute to the malignant phenotype [34]. Thus, genes which are differentially expressed between benign and malignant lesions have been sought using a variety of molecular techniques. Foremost amongst these techniques has been micro-array hybridisation using micro-arrays bearing oligonucleotides or cDNAs which represent large collections of human genes [38].

The result of these array experiments has been the identification of gene signatures which are thought to indicate good or poor prognosis for the patients [47]. Of key interest, however, is the identification of those individual genes and their protein products which drive the progression of cancers and which might therefore be therapeutic targets. The identity of such genes could be derived from global micro-arrays, however, there is a severe bioinformatics burden due to the large amount of data generated by these global arrays. In contrast to prognostic correlations, it is necessary to isolate the full length cDNAs/genes in order to demonstrate the functional link between the effect of the expression of the protein products and cancer progression [13,14].

An alternative approach to global microarray analysis is the isolation of differentially-expressed cDNAs from human cancer specimens or cell lines, using suppression subtractive hybridisation, a technique which can identify differentially expressed genes that yield rarer gene products [15,50]. A custom microarray containing the resulting cloned cDNAs can be produced and hybridised to cDNA from a variety of tumour-derived cell lines and carcinoma specimens exhibiting defined properties. This approach has now been used to identify mRNAs which encode proteins that are both potential biomarkers and also key mediators of breast cancer progression.

2. Materials and methods

2.1. Human tumour specimens

Total cellular RNA from 13 ductal carcinoma specimens of the breast was obtained with full informed patient consent and ethical approval from the Liverpool Cancer Tissue Bank Research Centre. All specimens were from postmenopausal women who received adjuvant tamoxifen treatment, but no chemotherapy or primary endocrine therapy. All were histological grade 2 or 3. These cancers were selected as representative of cases with either early breast cancer relapse/death or prolonged survival and estrogen receptor alpha (ERα) positive or negative status (Table 1). RNA from twelve specimens was used for screening the arrays and from 13 specimens for the reverse transcript (RT)-Polymerase Chain Reaction (PCR) assay. A panel of 97 breast carcinoma specimens (including the above 13) was also used, and this was a subset of the 100 specimens described previously [12]. All patient data was anonymised and the study was conducted with approval of the local Liverpool Adult Research Ethics Committee [12].

Table 1
Clinical characteristics of human breast carcinoma specimens

No.	Age of patient	Grade	Nodal status	ERα status	p53 mutation	Relapsed	Relapse-free follow-up (months)
1	61	3	−	−	No	No	11.0
2	59	3	−	−	Yes	No	67.4
3	65	3	+	−	Yes	No	94.6
4	87	3	+	−	No	Yes	6.8
5	71	3	N/A	−	Yes	Yes	12.4
6	81	3	N/A	−	Yes	Yes	14.7
7	74	3	−	+	No	No	69.5
8	66	2	+	+	No	No	46.5
9	69	3	+	+	Yes	No	83.4
10	62	2	+	+	No	No	93.5
11	58	2	+	+	No	Yes	15.1
12	56	3	+	+	Yes	Yes	23.5
13[a]	63	3	+	+	No	No	51.2

Notes: N/A: not available; −: negative; +: positive.
[a]Cancer specimen used for quantitative RT-PCR but not for screening of arrays.

2.2. Cell lines and cell culture

RNA from 8 human breast epithelial cell lines was used. The normal derived human mammary epithelial cell lines, Huma 7 and SVE3 (diploid, ERα negative) were subcloned from primary cultures of reduction mammoplasty specimens of normal breast tissue immortalized with SV40 virus [41]; the benign human mammary epithelial cell lines, Huma 123 and Huma 109 [25] (diploid, ERα negative), were derived from HMT-3522, itself obtained from a primary cell culture of human benign breast disease displaying prominent epithelial hyperplasia [7]. Malignant human mammary epithelial cell lines, MCF-7, T47D, ZR-75 [44] (aneuploid, ERα positive) and MDA-MB-231 [8] (aneuploid, ERα negative) were derived from pleural effusions of breast cancer patients and were cultured as described previously [25]. All cell lines were passaged on reaching 70% confluency. Cell culture in medium depleted of steroid hormones was carried out as described previously [17].

2.3. Construction of subtractive hybridisation cDNA libraries

Benign and malignant subtractive hybridisation libraries consisting of cDNAs representing mRNAs up- or down-regulated between the ERα negative, benign human mammary cell line, Huma 123 [25] and the

ERα positive malignant cell line, MCF-7 [44] were constructed as described previously [30,31]. Two further carcinoma subtracted libraries of cDNAs representing mRNAs which were up or down regulated between a histological section of an invasive ductal carcinoma (IDC) and a ductal carcinoma *in situ* (DCIS) specimen [28] were constructed as follows. Poly(A)-containing RNA was isolated from 250,000 cells microdissected from 20 frozen, 20 μm sections of a ductal carcinoma *in situ* (DCIS) specimen and from non-microdissected frozen sections of an invasive ductal carcinoma (IDC), which contained >90% carcinoma cells, using a DYNAbead poly(A) RNA isolation kit (DYNAbead, UK). First strand cDNA of the DCIS and IDC RNAs were synthesized and amplified using the Smart PCR cDNA synthesis system (Clontech, Mountainview, CA, USA) to provide material for the subtractive hybridisations, which were carried out as described previously [28,31].

Randomly isolated cDNA clones from the 4 subtracted cDNA libraries were PCR amplified using Taq DNA polymerase (Clontech, Mountainview, CA, USA), and the resulting PCR products purified using a PCR purification kit (Qiagen, Crawley, UK). The DNA was analyzed by DNA sequencing from both directions using a DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE (GE Healthcare, Amersham, UK) and analysed on a MegaBACE 1000 DNA Analysis System (GE Healthcare, Amersham, UK). Individual cDNA clones were identified by searching for matches in the GenBank public database. Differential expression of a number of clones was confirmed by reverse Northern hybridization carried out as described previously [30,31]. The redundancy of the clone set was minimised by removal of duplicate clones (i.e. the same cDNA fragment cloned multiple times) and alternative cDNA fragments (i.e. different fragments of the same cDNA identified by database searching).

2.4. In-house human breast cancer cDNA microarrays

Poly-lysine coated slides for cDNA microarrays were prepared as follows. New 'Goldseal' microscope slides (3×1 inch, Fisher Scientific Ltd., Loughborough, UK) were soaked in a solution of 10% (w/v) NaOH in 60% (v/v) ethanol for 2 h with gentle rocking at room temperature and rinsed 6 times with >2 l of sterile water. The cleaned slides were soaked in fresh 0.02% (w/v) poly-L-lysine solution containing 10% (v/v) phosphate-buffered saline (PBS) for 30 min with gentle rocking and washed 6 times with a total of 2 l of sterile water. The slides were immediately centrifuged at 600 rpm at 20°C for 10 min and stored in a clean dust-free slide box until used for DNA spotting.

Cloned cDNAs from the subtracted libraries (151 from the Huma 123/MCF-7 libraries and 178 from the DCIS/IDC libraries), cDNAs of 33 cancer genes and 14 controls were PCR amplified, purified and robotically spotted onto the poly-L-lysine-coated glass slides as described previously [21]. Briefly, the amplified PCR products were purified and diluted 1 in 100 and re-amplified, purified, quantified, precipitated with 1/10 volume of 5 M sodium acetate and 1 volume of isopropanol overnight, centrifuged and the pellet was rinsed with 70% (v/v) ethanol. The DNA pellets were air dried and dissolved in 75 mM sodium phosphate, pH 8.5, 0.005% (w/v) SDS and the final concentration of the cDNA was adjusted to 100–200 ng/ml. The PCR products, along with Cot1 and Salmon Sperm DNA hybridisation controls were transferred to 384-well plates before being robotically spotted in quadruplicate onto glass slides using a Biorobotics plc Microgrid II robot.

The spotted slides were hydrated with 1×SSC for 4 min using a hydration chamber (Sigma, St. Louis, MO, USA) and were snap dried on a heated block at 100°C for 10 s. The hydrated slides were soaked in 157 mM succinic anhydride, 43 mM sodium borate, pH 8.0 and 1-methyl-2-pyrrolidinone for 15 min at room temperature with vigorous stirring under the surface and transferred into a 95°C water bath for 1.5 min to denature the spotted double-stranded DNA. The slides were rinsed with 95% (v/v) ethanol 5 times and centrifuged at 600 rpm for 5 min at room temperature to dry the slides.

2.5. Fluorescent labelling of probes and microarray hybridisation

cDNA probes were synthesized from high quality total RNA and labelled with Cy3 or Cy5 fluors (GE Healthcare, Amersham, UK) with dye swaps from each cancer specimen or cell line, pair-wise against a Universal human reference RNA (Cat No. 740000, Stratagene, La Jolla, CA, USA).

Total cellular RNA was prepared from cell lines and carcinoma specimens using the guanidinium isothiocyanate/caesium chloride method [4] and digested with DNase I, using Qiagen DNase I digestion and RNeasy mini kits to remove any contamination of cellular DNA from the RNA. The qualities of the RNAs were checked by running samples on denaturing RNA

gels and carrying out Northern blot hybridisation procedures, as described previously [30].

Total RNA for each probe was coupled with amino allyl dUTP (Sigma, St. Louis, MO, USA) and labelled separately with Cy3 and Cy5 dyes. Twenty μg of total RNA for each probe and 5 μg of oligo dT_{16} were denatured at 70°C for 10 min, placed on ice for 10 min and reverse transcribed at 42°C for 2 h in a final volume of 30 μl containing 0.5 mM each of dATP, dGTP, dCTP, 0.25 mM dTTP and 0.25 mM aminoallyl-dUTP, 10 mM DTT and 150 units of SuperScript II RNase H Reverse Transcriptase (InVitrogen, Paisley, UK). The resulting first strand cDNAs coupled with amino allyl dUTP were denatured at 65°C for 15 min by adding 10 μl of 1 M NaOH and 10 μl 500 mM EDTA and neutralised by adding 25 μl of 1 M Tris-HCl, pH 7.4. Prior to probe labelling, each cDNA was concentrated to a final volume of 18 μl using a Microcon YM-30 column (Fisher Scientific Ltd., UK), the 18 μl of each cDNA probe were divided into two equal volumes and each brought to a final concentration of 50 mM sodium bicarbonate buffer and labelled with 1.25 μl of freshly-prepared solutions of either Cy3 or Cy5 in DMSO (GE Healthcare, Amersham, UK), for 1 h in the dark. The labelling reactions with Cy3 and Cy5 were quenched by adding 4.5 μl of 4 M hydroxylamine to each tube and incubating for 15 min in the dark at room temperature. Before hybridisation, the probe labelled with Cy3 from one sample was combined with the probe labelled with Cy5 from another, purified using a Qiagen PCR purification kit and concentrated on a Microcon YM-30 column to give a final volume of 15 μl in sterile water.

Cy3- and Cy5-labelled probes, each from 20 μg total RNA, were mixed with a hybridization buffer (3 × SSC, 750 ng/ml of oligo $(dA)_{16}$, 0.2% (w/v) SDS and 25 mM HEPES buffer, pH 7.0), denatured at 100°C for 2 min and cooled at room temperature for 5 min before hybridization.

Microarray hybridizations were carried out overnight at 63°C in a humid environment. After being washed twice with 0.57 × SSC and 0.03% (w/v) SDS and once with 0.057 × SSC for 20 min at room temperature with stirring, the arrays were air-dried by being centrifuged at 600 rpm for 5 min at room temperature. Bound fluorescent probes were detected by scanning the slide in an Axon 4000A scanner and the image data were extracted and visualised using GENEPIX 3.0 software (Axon Instruments, Union City, CA, USA), as described previously [21]. The array hybridisation results were analysed using GeneSpring GX 7.3.1 software (Silicon Genetics, Palo Alto, CA, USA), as described for the individual results.

2.6. Quantitative RT-PCR

1.5 μg high quality total cellular RNA for each sample was digested with DNase I (InVitrogen, Paisley, UK) and divided equally into three tubes for reverse transcription (RT) reactions (duplicate tubes and one lacking reverse transcriptase as a control for genomic DNA contamination). First strand cDNA was synthesized from the 0.5 μg aliquot of total RNA with 0.5 μg oligo $(dT)_{18}$ primer and 0.5 μl of ribonuclease inhibitor (Promega, Madison, WI, USA) in a final volume of 10 μl with 200 units of SuperScript II RNase H Reverse Transcriptase (InVitrogen, Paisley, UK) in the manufacturer's buffer at 42°C for 90 min. Five μl of a 1:50 dilution of the RT reaction was amplified by quantitative PCR using BioRad Icycler reaction mix on a BioRad Icycler (BioRad, Hercules, CA, USA), as described previously [23].

The PCR primers (Table 2) were designed to cross intron–exon boundaries (to allow detection of products of contaminating genomic DNA) for each particular cDNA. The quantitative PCR results for cDNAs were corrected for quantitative PCR-detected levels of housekeeping mRNAs, human ribosomal phosphoprotein PO [26], glyceraldehyde-3-phosphate dehydrogenase (GAPDH) [46], and hypoxanthine phosphoribosyl transferase (HPRT) [42], and either compared with a standard and expressed as relative level of mRNA, or fully quantified by being compared to cloned PCR products of known concentration and corrected for known mean input amount of total RNA per PCR reaction.

2.7. Statistical analyses

Statistical analyses were carried out using Stats Direct Software and analyses of outcomes were carried out using SPSS 11.0 software. X-Box Binding Protein-1 (XBP-1) quantitative RT-PCR amounts were dichotomised for outcome analysis by applying receiver operating characteristic (ROC) plots (see Section 3) for breast cancer relapse-free survival and breast cancer overall survival at 5 years after surgery. Curves for breast cancer relapse-free survival and breast cancer overall survival of the ERα protein-positive, Tamoxifen-treated cohort were produced using the Kaplan–Meier method for censored data and compared using log-rank tests. Unadjusted hazard ratios ±95% confidence intervals were obtained using Cox's univariate analysis. Cox's regression model was used for multivariate survival analysis (all parameters

Table 2
PCR primers used for quantitative RT-PCR

Identity	GenBank accession number	Forward primer	Reverse primer
Human acidic ribosomal phosphoprotein PO (36B4)	NM_001002	5′-CATTGCCCCA TGTGAAGT-3′	5′-AGAAGGGGGAGA TGTTGA-3′
Anterior gradient 2 (AGR2)	NM_006408	5′-GAGCCGATATCA CTGGAAGA-3′	5′-CAAGGCCTGACA GACAGAAG-3′
Estrogen receptor α	NM_000125	5′-CCACCAACCAGT GCA CCA TT-3′	5′-GGT CTT TTC GTA TCC CAC CTT TC-3′
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	NM_002046	5′-GCATCCTGGGCT ACACTGAG-3′	5′-TCCACCACCCTG TTGCTGTA-3′
Hypoxanthine phosphoribosyl transferase (HPRT)	NM_000194	5′-GTGTTGGATATA AGCCAGACTTTGTT-3′	5′-AACTCAACTTGAA CTCTCATCTTAGGC-3′
X-box binding protein-1 (XBP-1)	NM_005080	5′-GATTCTGGCGGT ATTGAC-3′	5′-ATGACGTCCCCAC TGACA-3′

entered, or backward conditional with removal limit of 0.01).

The times of disease-specific, patient relapse and patient death were obtained from hospital records.

3. Results

3.1. Microarray analysis of subtractive hybridisation cDNA libraries from breast cancer

Microarrays constructed from 376 cDNAs including 329 from the subtractive hybridisation experiments (see Section 2) were hybridised with cDNA from 12 invasive ductal breast carcinoma specimens from postmenopausal, Tamoxifen-treated patients (Table 1) and with cDNA from 8 breast tumour cell lines (details in Section 2). The results of the array hybridisations were analysed by class prediction on multiple analyses to identify a small group of differentially-expressed genes which clustered carcinoma and/or cell lines according to selected parameters, namely, ERα status (cell lines and carcinoma specimens), relapse (carcinoma specimens), benign versus malignant (cell lines) or diploid vs aneuploid (cell lines) (Table 3). The analyses yielded a panel of 8 genes, expression of which clustered for ERα and 7 genes, which clustered for malignant or aneuploidy (cell lines) and/or relapse (carcinomas) (Fig. 1).

The array results were also analysed by filtering cell line and tumour-derived data on level of expression, namely, genes selected as greater than 2-fold difference in level between groups of tumours (Student t-test, $p < 0.01$). Nine cDNAs were identified for immuno-cytochemical ERα positivity and 4 cDNAs were identified for relationship with relapse (Fig. 2). Amongst these two sets of cDNAs were three from the panel of 15 cDNAs from class prediction, namely S100P, Anterior gradient-2 (AGR2) and X-box binding protein, XBP-1.

3.2. Quantitative RT-PCR analysis of ERα responsive genes in breast cancer specimens

To validate the array results for these cDNAs, RNA from the cell lines and carcinoma specimens were subjected to quantitative RT-PCR using gene-specific primers (Section 2) and the results were normalised for the expression of three independent 'constitutive' cDNAs, GAPDH, HPRT and ribosomal protein, PO. The level of AGR2 mRNA was significantly less in immunocytochemically ERα negative cell lines and carcinomas than in immunocytochemically ERα positive carcinomas (Mann Witney U-test, two sided $p = 0.0002$). Furthermore, there was a strong positive correlation between the level of ERα mRNA and AGR2 mRNA by quantitative RT-PCR in both cell lines (Pearson coefficient $r = 0.951$, $p = 0.0003$) and carcinomas (Pearson coefficient $r = 0.828$, $p = 0.0005$) (Fig. 3A).

There was a positive correlation between the quantitative RT-PCR-determined levels of AGR2 mRNA and XBP-1 mRNA in cell lines and carcinoma spec-

Table 3

Panel of 8 expressed genes, which clustered for ERα and 7 expressed genes, which clustered for malignant or aneuploidy (cell lines) and/or relapse (carcinomas)

| ER (tumours)	Relapse or dead (tumours)	ER (cell lines)	Malignant (cell lines)	Aneuploid (cell lines)	GenBank accession number	Identity	Chromosome	7 genes for malignant/relapse	8 genes for ER
		Yes			AF376725	Lung seven transmembrane receptor 1 (LUSTR1) mRNA	9q34.11		Yes
					NM_020960	G protein-coupled receptor 107 (GPR107)			
Yes		Yes			FJ462787	No match	Unknown		Yes
Yes		Yes			NM_005080	X-box binding protein 1 (XBP-1)	22q12.1		Yes
Yes	Yes	Yes			AF465980 or BQ778361	Mitochondrial genome EST match to Cytochrome C oxidase	Mitochondrial		Yes
Yes	Yes	Yes			NM_006408	Anterior gradient 2 homologue (*Xenopus laevis*) (AGR2)	7p21.3		Yes
Yes	Yes	Yes			BC000724/ NM_001183	Human ATPase, H+ transporting, lysosomal accessory protein 1 (ATP6AP1)	Xq28		Yes
Yes	Yes	Yes			NM_000982	Human ribosomal protein L21 (RPL21)	13q12.1		Yes
	Yes	Yes			NM_000034	Fructose-bisphosphate aldolase A (ALDOA) transcript variant 1	16p11.2		Yes
Yes	Yes		Yes		BC034328/ NM_02433	Human iroquois homeobox protein 3 (IRX3)	16q12.2	Yes	
	Yes		Yes	Yes	AK091065	Human septin 4 nucleotide binding protein	17q22-q23	Yes	
	Yes		Yes		AF131851	Human family with sequence similarity 19 (chemokine (C-C motif)-like), member A5 (TAFA5)	22q13.32	Yes	
	Yes		Yes	Yes	NM_006456	Human ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2	17q25.1	Yes	
	Yes		Yes		NM_005980	S100 calcium binding protein P (S100P)	4p16	Yes	
			Yes	Yes	BC009195	Human heat shock protein 90 kDa β (Grp94), member 1 (HSP90B1)	12q24.2-q24.3	Yes	
				Yes	AC022083	Human chromosome 15 clone CTD-2137J4	15q21.3	Yes	
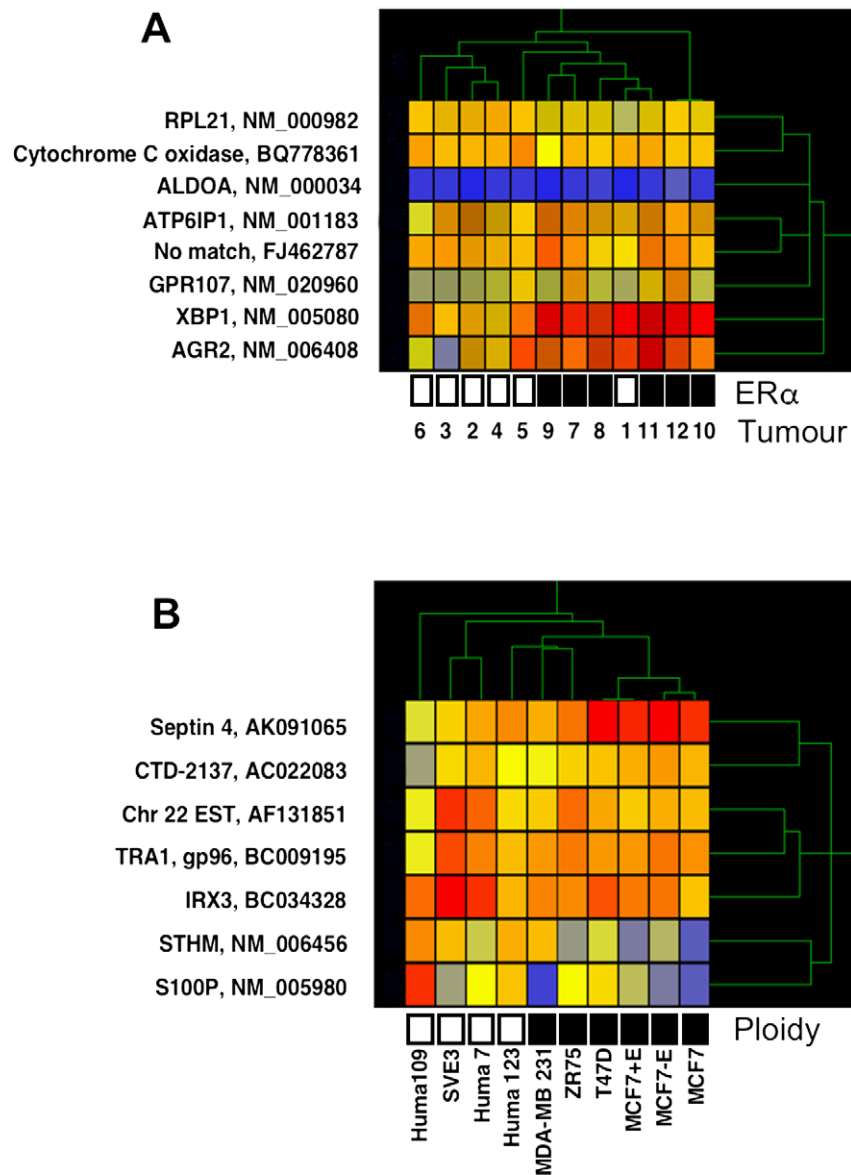
Fig. 1. Clustering of tumour samples (A) and cell-lines (B) based on gene lists from class-prediction for either ERα status (A, 8 genes) or ploidy (B, 7 genes): dendograms represent hierarchical clustering of genes (right) and samples (top) using Pearson correlation as a similarity measure. ERα: black boxes are ERα protein positive tumour samples; white boxes are ERα protein negative tumour samples. Ploidy: black boxes are aneuploid cell-lines; white boxes are diploid cell-lines. The listed genes are: (A) NM_000982, Human ribosomal protein L21 (RPL21); BQ778361, EST match to Cytochrome C oxidase; NM_000034, fructose bisphosphate aldolase A (ALDOA); NM_001183, Human ATPase, H+ transporting, lysosomal accessory protein 1 (ATP6AP1); FJ462787, novel clone no match in database; NM_020960, G protein-coupled receptor 107 (GPR107); NM_005080, X-box binding protein (XBP-1); NM_006408, anterior gradient 2 (AGR2); (B) AK091065, Human septin 4 nucleotide binding protein; AC022083, Human chromosome 15 clone CTD-2137J4; AF131851, Chr 22 EST, Human family with sequence similarity 19 (chemokine (C–C motif)-like), member A5 (TAFA5); BC009195, Human heat shock protein 90 kDa beta (Grp94), member 1 (HSP90B1); BC034328/NM_02433, Human iroquois homeobox protein 3 (IRX3); NM_006456, Human ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2; NM_005980, S100 calcium binding protein P (S100P). Colouring according to relative expression: red – higher, blue – lower, yellow – intermediate.
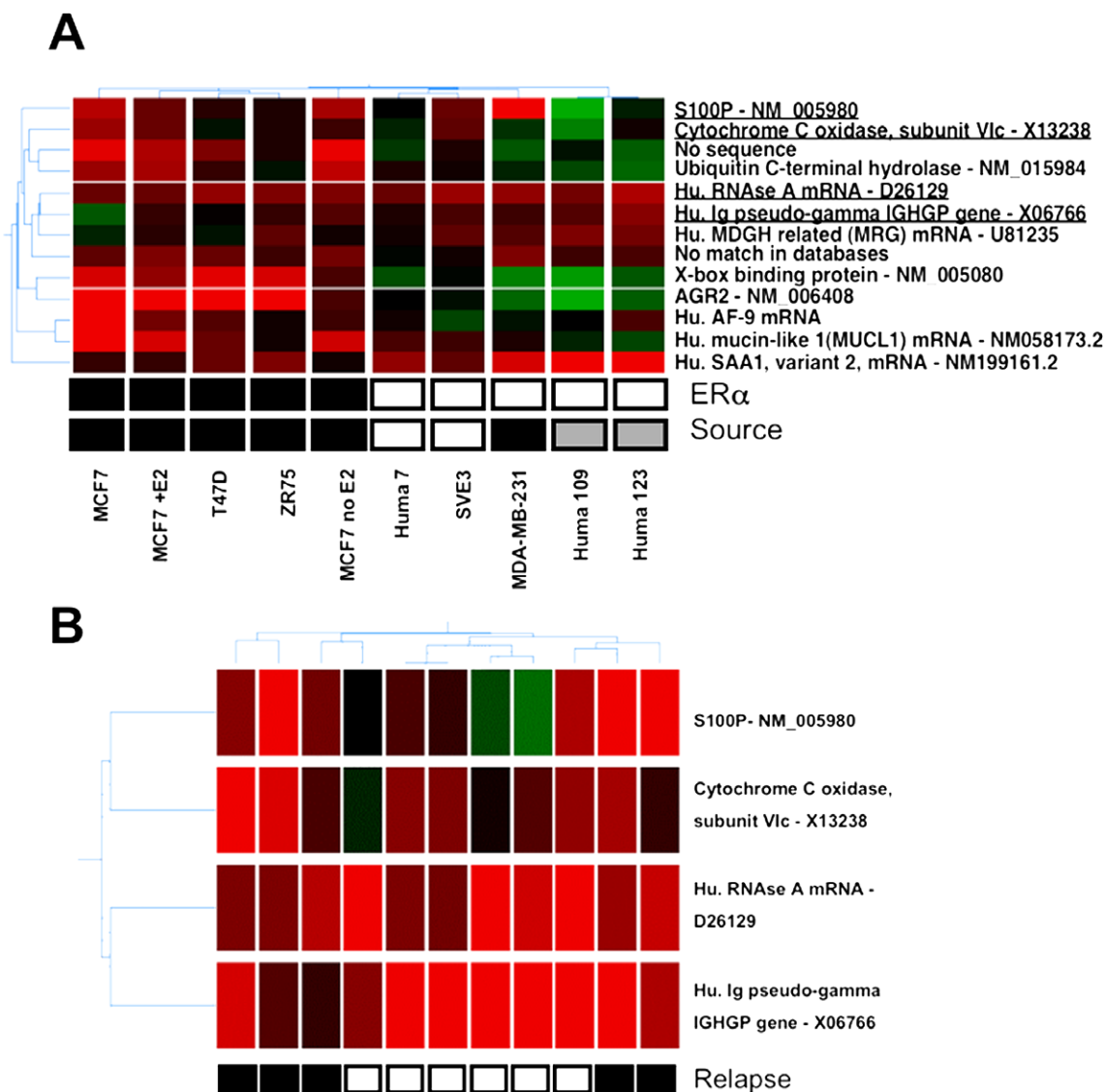
A

B

Fig. 2. Cell-line and tumour data clustered on tumour-selected genes as a two fold expression difference. Dendograms represent hierarchical clustering of genes (left) and samples (top) using Pearson correlation as a similarity measure. (A) Genes identified from tumour samples by expression filtering are meaningful for cell-lines; four relapse-associated genes are underlined, the remaining 9 genes were positively associated with immunohistochemical ERα protein positivity. ERα: black boxes are cell lines which are ERα protein-positive; white boxes are cell lines which are ERα protein-negative. Source: black boxes are cell-lines from malignant tumours; grey boxes are cell-lines from benign tumours; white boxes are cell-lines derived from normal tissue. Hu. – Human; MDGH – Mammary-derived growth inhibitor; SAA1 – Serum amyloid A1. (B) Tumour data clustered on 4 genes selected as 2-fold expression difference between relapsed and non-relapsed cases. Relapse: black boxes are tumours from cases with earlier relapse (mean time to relapse was 15 months); white boxes are tumours from cases without relapse at follow-up (mean relapse-free follow-up was 67 months). Colouring according to relative expression: red – higher, green – lower, black – intermediate.

imens that was stronger in cell lines than in carcinomas (cell lines, Pearson coefficient $r = 0.93$, $p = 0.0008$; carcinomas, Pearson coefficient $r = 0.579$, $p = 0.038$) (Fig. 3B). The level of X-box binding protein mRNA correlated positively and strongly with the levels of ERα mRNA in cell lines (Pearson coefficient $r = 0.938$, $p = 0.0006$) but did not reach significance in carcinoma specimens as determined by quantitative PCR (Pearson coefficient $r = 0.469$, $p = 0.106$) (Fig. 3C).

In order to analyse further the relationship between XBP-1 mRNA levels and ERα mRNA levels, quanti-
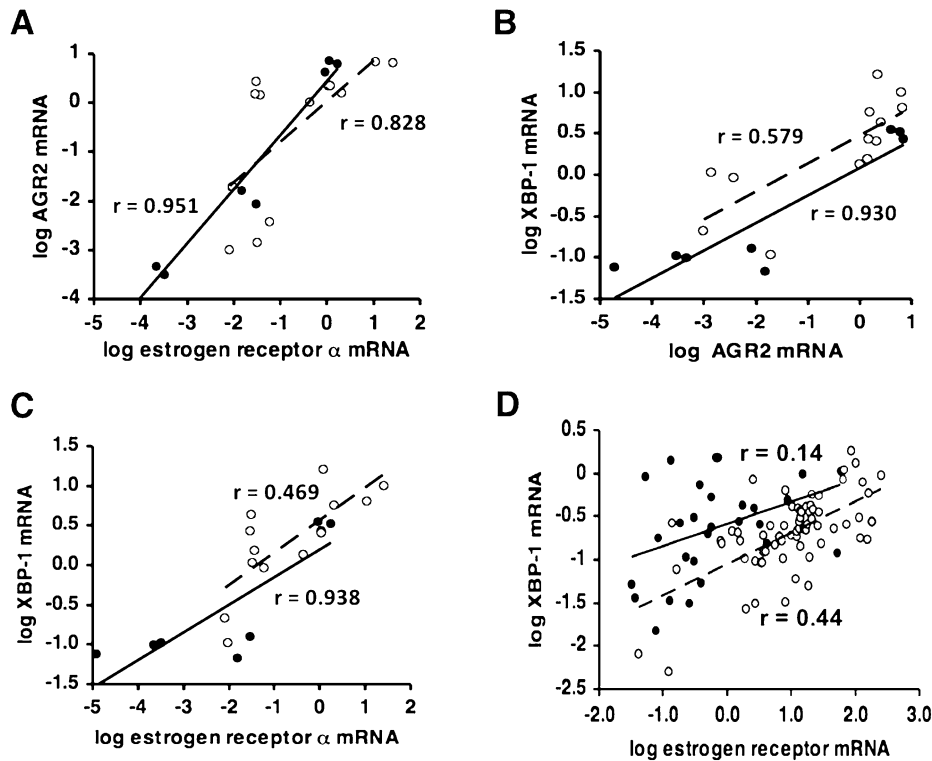
Fig. 3. Correlations between estrogen receptor α mRNA, AGR2 mRNA and XBP-1 mRNA in breast tumour cell lines and carcinoma specimens. The levels of estrogen receptor α mRNA, AGR2 mRNA and X-box binding protein-1 (XBP-1) mRNA were determined by quantitative RT-PCR in either 8 breast tumour cell lines and 13 carcinoma specimens (A, B, C), or 97 breast carcinoma specimens (D). The results were normalised against three independent constitutive mRNAs as described in Section 2. The results are plotted after logarithmic transformation for estrogen receptor α mRNA against AGR2 mRNA (A), for AGR2 against X-box binding protein (XBP-1) (B) and ERα mRNA against X-box binding protein-1 (XBP-1) mRNA (C, D). In (A, B and C), data from cell lines are shown as black circles with a solid trendline and data from carcinoma specimens is shown as white circles with a broken trend-line. In (D), the data for specimens that were ERα protein positive or negative by immunohistochemistry are shown as white circles (with broken trendline) or black circles (with solid trendline), respectively. Statistical analyses are given in the text, but Pearson coefficients (r) are shown for each trendline.

tative RT-PCR was carried out on a panel of 97 carcinoma specimens, with the results being normalised to the three constitutive cDNAs as above. In this larger group of specimens, there was a significant positive correlation between XBP-1 mRNA and ERα mRNA (Pearson coefficient $r = 0.27$, $p = 0.007$). When the 97 carcinoma specimens were divided into two groups, immunohistochemically-detected ERα protein positive and ERα protein negative, there was a strong positive correlation between log XBP-1 mRNA and ERα mRNA in the immunohistochemically ERα-positive group (Pearson coefficient $r = 0.44$, two-sided $p = 0.0002$), but not in the immunohistochemically ERα negative group (Pearson coefficient $r = 0.14$, $p = 0.47$). This strong link between ERα mRNA expression and XBP-1 mRNA expression is consistent with XBP-1 being an estrogen-responsive gene, at least in

cases defined as ERα immunohistochemically positive (Fig. 3D).

3.3. Association of XBP-1 mRNA with patient survival

The biological relevance of XBP-1 mRNA was therefore assessed by examining the association between the level of XBP-1 mRNA determined by quantitative RT-PCR and patient survival using Kaplan–Meier plots in a cohort of immunohistochemically ERα protein-positive primary breast cancers from patients receiving adjuvant Tamoxifen [12]. Using ROC analysis, there was a significant positive relationship between XBP-1 mRNA level and patient survival time (area under curve 0.64; 95% confidence interval 0.53–0.76, $p = 0.017$). A cut-off of 0.2 attomoles per μg total RNA (at the apex of the ROC plot giving
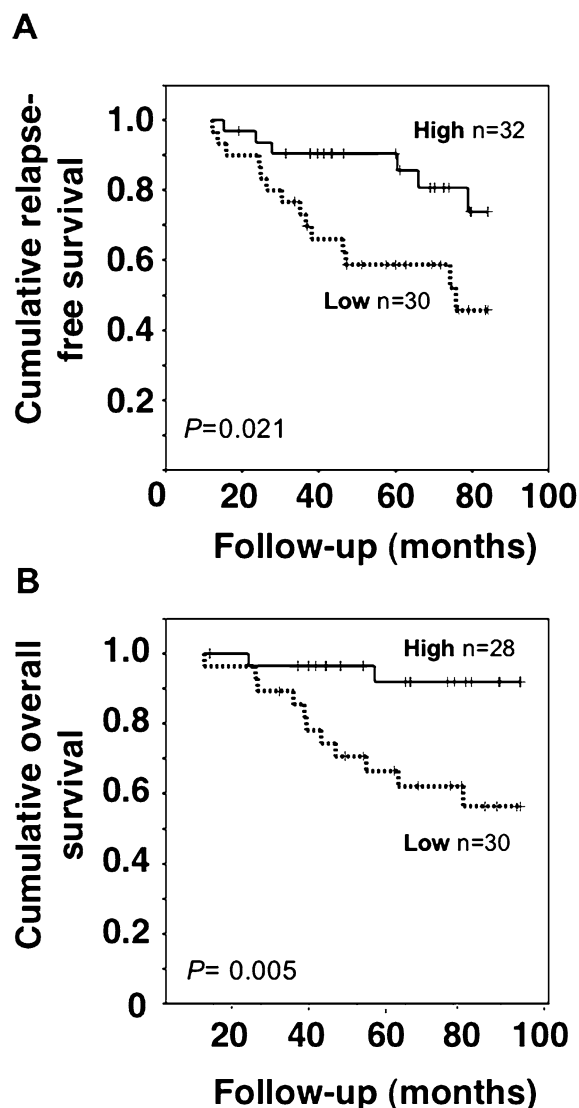
A

B

Fig. 4. XBP-1 mRNA expression and outcome. Kaplan–Meier plots are shown for breast cancer relapse-free (A) and overall survival (B) in a cohort of postmenopausal women with ERα-positive, primary breast cancer, receiving adjuvant tamoxifen with up to 7 years follow-up. XBP-1 quantitative RT-PCR amounts were dichotomised for outcome analysis by applying Receiver Operating Characteristic plots (see text) for breast cancer relapse and breast cancer survival at 5 years after surgery; cut-off 0.2 attomoles/μg total RNA. Unbroken lines represent cases with high levels of XBP-1, dotted lines represent cases with low levels of XBP-1. In all cases crosses represent censored data and p-values are given for log-rank tests.

the largest product of specificity and sensitivity) divided cases into two groups, containing high XBP-1 mRNA ($n = 32$) and low XBP-1 mRNA ($n = 30$). The high XBP-1 mRNA group exhibited increased relapse-free (log-rank $p = 0.021$) or overall survival

(log-rank $p = 0.005$) relative to the low XBP-1 subgroup (Fig. 4A and B respectively). For breast cancer relapse, the hazard ratio was 0.34 (Cox regression, 95% confidence interval 0.13–0.89, $p = 0.028$) and for breast cancer death, the hazard ratio was 0.15 (Cox regression, 95% confidence interval 0.03–0.69, $p = 0.01$). Furthermore, XBP-1 was the only significant marker for breast cancer survival in multivariate analysis including grade, size, nodal status and PgR status (Cox regression, hazard ratio 0.17, 95% confidence interval 0.03–0.98, $p = 0.048$). In a backward conditional model of these same markers, only XBP-1 (hazard ratio 0.16, 95% confidence interval 0.03–0.74, $p = 0.019$) and nodal status (hazard ratio 3.9, 95% confidence interval 1.2–13.0, $p = 0.028$) remain independently associated with breast cancer survival.

In the clinically interesting subgroup of ERα protein-positive, lymph node negative primary breast cancers ($n = 33$), higher XBP-1 mRNA expression ($n = 16$) was positively associated with longer breast cancer relapse-free survival (Kaplan–Meier log-rank $p = 0.024$) and better breast cancer overall survival (Kaplan–Meier log-rank $p = 0.048$).

4. Discussion

A combination of suppression subtractive hybridisation and microarray analysis has been used to identify cDNAs with differential expression associated with breast cancer progression. Suppression subtractive hybridisation libraries of cDNAs up- or down-differentially expressed in malignant/benign, ERα protein-positive/ERα protein-negative, microdissected DCIS/carcinoma specimens were arrayed onto glass slides and hybridised to denatured cDNA preparations from 8 benign and malignant breast cell lines and 12 breast carcinoma specimens. The array hybridisation experiments were analysed using Gene Spring Software by class prediction or by expression level. The class prediction yielded a panel of eight cDNAs associated with positive immunohistochemical ERα protein status and 7 cDNAs associated with relapse, whilst analysis by expression level yielded 9 cDNAs associated with positive immunohistochemical ERα protein status and 4 cDNAs associated with relapse. This analytical approach is different from previous approaches using sequential suppression subtractive and micro-array hybridisation. Whilst others have used these two techniques in parallel [39],

a sequential suppression subtractive hybridisation and micro-array approach has been used previously to identify differentially-expressed cDNAs in calcium-induced, terminally-differentiating keratinocytes using hierarchical clustering [43]. However, unsupervised hierarchical clustering was not found to be very informative in the current study, in relation to immunohistochemical ERα protein status or relapse. Recently, a modified suppression subtractive hybridisation method has been developed to provide RNA probes representing low abundance mRNAs which were used to identify differentially-expressed cDNAs in breast, hepatocellular and nasopharyngeal carcinomas using an Affimetrix array [29], however, the involvement of the differentially-expressed cDNAs in the carcinomas was not determined.

In the present experiments, the class prediction and expression level analyses predicted some genes/proteins which are already known to have a significant association with cancer progression. Thus, increased levels of S100P, a member of the S100 protein family of EF-hand containing, calcium-binding proteins [5] was found in the early relapse group of both array analyses. Increased levels are shown here to be positively associated with relapse in carcinomas and aneuploidy in cell lines. S100P levels have been shown previously to be up-regulated in pancreatic cancer cells by array hybridisation [11], to be androgen responsive in prostate cancers [2], to be negatively associated with reduced patient survival and positively associated with metastasis in breast cancers [49] and to be linked with cancer progression in a number of independent systems [3,16, 22,36].

AGR2 is the human homologue of the *Xenopus* developmental protein, XAG-2, which is associated with the development of the *Xenopus* tadpole cement gland [1]. AGR2 mRNA was associated with ERα status in both the class prediction and expression analyses and elevated levels of AGR2 mRNA in the present experiments were positively associated with immunohistochemically detected ERα-positive cancers, as described previously [19,23,32,45]. Elevated levels of AGR2 protein/mRNA have been shown also to confer upon benign mammary tumour cells a metastatic phenotype in an *in vivo* assay of metastasis [32] and to be associated with reduced survival in Tamoxifen-treated breast cancer patients [23]. These results show that cDNAs representing mRNAs that are relevant to cancer progression in model systems and in cancer patients have been selected by our present suppression subtractive hybridisation/array hybridisation strategy and they serve to validate the analysis of the array data for the

discovery of mRNAs/proteins which have relevance to both the progression of cancer and patient survival.

The cDNA for XBP-1 mRNA, which encodes a transcription factor associated with the unfolded protein response of the endoplasmic reticulum [40], was also identified as a strong candidate from amongst the immunohistochemical ERα-positive-associated genes/gene products in the analysis of the present array data. XBP-1 mRNA levels correlated with ERα mRNA levels in immunohistochemically ERα positive carcinomas, a group of patients for whom better markers of anti-estrogen therapy are required. Higher levels of total XBP-1 mRNA, as determined by quantitative RT-PCR, are here shown to be associated with a better relapse-free survival and with better overall survival from breast cancer than lower levels, suggesting that XBP-1 mRNA may be a marker of a good response to hormone treatment in this group of patients. XBP-1 mRNA exists in spliced and unspliced forms encoding different proteins with opposing effects on apoptosis [20] and cancer patient survival [12]. In the present experiments, the quantitative RT-PCR did not distinguish between the mRNAs encoding these different forms. However, the result that higher levels of total XBP-1 mRNA are associated with better survival reflect those obtained for the unspliced XBP-1 mRNA variant for relapse-free survival [12].

XBP-1 and AGR2 were selected using similar criteria on analysis of the microarray data and both AGR2 [23] and XBP-1 (here and previous report [12]) have been shown to have potential use as markers of breast cancer outcome. However, AGR2 does not appear in existing predictive microarray-based cancer gene signatures which correlate with generally reduced breast cancer patient survival. Furthermore S100P and XBP-1 are only evident in the intrinsic molecular portrait of human breast tumours [38], and do not appear in other gene signatures, including good/poor prognosis [47], wound response [9,10], breast vs lung metastasis [27, 35], ER+/ER$-$ [48], invasiveness [33], 10-year survival in node negative patients [24] and breast vs lymph node metastasis [18], which show positive or negative correlations with survival of breast cancer patients. Thus, the present experiments have identified individual cDNAs, the protein products of which are positively associated with reduced patient survival (AGR2 [32] and S100P [49]) or negatively associated with reduced survival (XBP-1) of breast cancer patients, but which are not generally identified in common cancer related gene signatures.

A combination of subtractive hybridisation and micro-array analysis of the resulting cloned cDNAs

has yielded not just the mRNAs discussed above, but a panel of 15 mRNAs with altered levels of expression that are related to immunohistochemical ERα protein status, quantitative RT-PCR-detected levels of ERα mRNA and/or patient outcome. Other members of this panel include mRNAs encoding enzymes which are elevated in level in ERα-protein positive cell lines/carcinomas; these include aldolase A, cytochrome C oxidase and the mRNA encoding the G protein-coupled receptor, GPR107, a member of a family of chemokine receptors, related members of which have been reported to be associated with cancer metastasis [51]. The malignancy/relapse-associated members of the panel, in addition to S100P, included a cDNA that was similar to Iroquois homeobox protein 3 (IRX 3), a member of a family of developmental proteins [6], one member of which has recently been implicated as a regulator of apoptosis in prostate cancer cells [37]. Thus, these panels of differentially-regulated cDNAs have identified key elements of cancer progression and also potential diagnostic/prognostic biomarkers and targets for therapy.

Acknowledgements

References

[1] F. Aberger, G. Weidinger, H. Grunz et al., Anterior specification of embryonic ectoderm: the role of the Xenopus cement gland-specific gene XAG-2, *Mech. Dev.* **72** (1998), 115–130.

[2] L.C. Amler, D.B. Agus, C. LeDuc et al., Dysregulated expression of androgen-responsive and nonresponsive genes in the androgen-independent prostate cancer xenograft model CWR22-R1, *Cancer Res.* **60** (2000), 6134–6141.

[3] T. Arumugam, D.M. Simeone, K. Van Golen et al., S100P promotes pancreatic cancer growth, survival and invasion, *Clin. Cancer Res.* **11** (2005), 5356–5364.

[4] R. Barraclough, R. Kimbell and P.S. Rudland, Differential control of mRNA levels for Thy-1 antigen and laminin in rat mammary epithelial and myoepithelial-like cells in culture, *J. Cell Physiol.* **131** (1987), 393–401.

[5] T. Becker, V. Gerke, E. Kube et al., S100P, a novel Ca(2+)-binding protein from human placenta. cDNA cloning, recombinant protein expression and Ca^{2+} binding properties, *Eur. J. Biochem.* **207** (1992), 541–547.

[6] A. Bosse, A. Zulch, M.B. Becker et al., Identification of the vertebrate Iroquois homeobox gene family with overlapping expression during early development of the nervous system, *Mech. Dev.* **69** (1997), 169–181.

[7] P. Briand, O.W. Petersen and B. Van Deurs, A new diploid non-tumorigenic human breast epithelial cell line isolated and propagated in chemically-defined medium, *In Vitro Cell Dev. Biol.* **23** (1987), 181–188.

[8] R. Cailleau, R. Young, M. Olivé et al., Breast tumour lines from pleural effusions, *J. Natl. Cancer Inst.* **53** (1974), 661–674.

[9] H.Y. Chang, D.S. Nuyten, J.B. Sneddon et al., Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival, *Proc. Natl. Acad. Sci. USA* **102** (2005), 3738–3743.

[10] H.Y. Chang, J.B. Sneddon, A.A. Alizadeh et al., Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds, *PLoS Biol.* **2** (2004), E7.

[11] T. Crnogorac-Jurcevic, E. Missiaglia, E. Blaveri et al., Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent, *J. Pathol.* **201** (2003), 63–74.

[12] M.P.A. Davies, D.L. Barraclough, C. Stewart et al., Expression and splicing of the unfolded protein response gene XBP-1 are significantly associated with clinical outcome of endocrine-treated breast cancer, *Int. J. Cancer* **123** (2008), 85–88.

[13] B.R. Davies, M.P.A. Davies, F.E.M. Gibbs et al., Induction of the metastatic phenotype by transfection of a benign rat mammary epithelial cell line with the gene for p9Ka, a rat calcium-binding protein but not with the oncogene EJ ras-1, *Oncogene* **8** (1993), 999–1008.

[14] M.P.A. Davies, P.S. Rudland, L. Robertson et al., Expression of the calcium-binding protein S100A4 (p9Ka) in MMTV-neu transgenic mice induces metastasis of mammary tumours, *Oncogene* **13** (1996), 1631–1637.

[15] L. Diatchenko, Y.-F.C. Lau, A. Campbell et al., Suppression subtractive hybridisation: A method for generating differentially-regulated or tissue-specific cDNA probes and libraries, *Proc. Natl. Acad. Sci. USA* **93** (1996), 6025–6030.

[16] S.E. Dowen, T. Crnogorac-Jurcevic, R. Gangeswaran et al., Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer, *Am. J. Pathol.* **166** (2005), 81–92.

[17] M. El-Tanani and C. Green, Estrogen-induced genes, pLiv-1 and pS2, respond divergently to other steroid-hormones in MCF-7 cells, *Mol. Cell Endocrinol.* **111** (1995), 75–81.

[18] R.E. Ellsworth, J. Seebach, L.A. Field et al., A gene expression signature that defines breast cancer metastases, *Clin. Exp. Metastasis* **26** (2009), 205–213.

[19] F.R. Fritzsche, E. Dahl, S. Pahl et al., Prognostic relevance of AGR2 expression in breast cancer, *Clin. Cancer Res.* **12** (2006), 1728–1734.

[20] B.P. Gomez, R.B. Riggins, A.N. Shajahan et al., Human X-box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines, *FASEB J.* **21** (2007), 4013–4027.

[21] A.Y. Gracey, E.J. Fraser, W. Li et al., Coping with cold: An integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate, *Proc. Natl. Acad. Sci. USA* **101** (2004), 16970–16975.

[22] I.D. Guerreiro Da Silva, Y.F. Hu, I.H. Russo et al., S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells *in vitro* and of breast cancer development *in vivo*, *Int. J. Oncol.* **16** (2000), 231–240.

[23] H.E. Innes, D. Liu, R. Barraclough et al., Significance of the metastasis-inducing protein AGR2 for outcome in hormonally-treated breast cancer patients, *Br. J. Cancer* **94** (2006), 1057–1065.

[24] E. Karlsson, U. Delle, A. Danielsson et al., Gene expression variation to predict 10-year survival in lymph-node-negative breast cancer, *BMC Cancer* **8** (2008), 254, doi:10.1186/1471-2407-8-254.

[25] Y. Ke, D.G. Fernig, M.C. Wilkinson et al., The expression of basic fibroblast growth factor and its receptor in cell lines derived from normal human mammary gland and a benign mammary lesion, *J. Cell Sci.* **106** (1993), 135–143.

[26] J. Laborda, 36B4 cDNA used as an estradiol-independent mRNA control is the cDNA for human acidic ribosomal phosphoprotein PO, *Nucleic Acids Res.* **19** (1991), 3998.

[27] T. Landemaine, A. Jackson, A. Bellahcene et al., A six-gene signature predicting breast cancer lung metastasis, *Cancer Res.* **68** (2008), 6092–6099.

[28] D. Liu, The identification of genes differentially expressed in human breast lesions, Ph.D. dissertation, University of Liverpool, Liverpool, 2001.

[29] B.H. Liu, C.H. Goh, L.L. Ooi et al., Identification of unique and common low abundance tumour-specific transcripts by suppression subtractive hybridization and oligonucleotide probe array analysis, *Oncogene* **27** (2008), 4128–4136.

[30] D. Liu, P.S. Rudland, D.R. Sibson et al., Expression of calcium-binding protein S100A2 in breast lesions, *Br. J. Cancer* **83** (2000), 1473–1479.

[31] D. Liu, P.S. Rudland, D.R. Sibson et al., Identification of mRNAs differentially-expressed between benign and malignant breast tumour cells, *Br. J. Cancer* **87** (2002), 423–431.

[32] D. Liu, P.S. Rudland, D.R. Sibson et al., Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas, *Cancer Res.* **65** (2005), 3796–3805.

[33] R. Liu, X. Wang, G.Y. Chen et al., The prognostic role of a gene signature from tumorigenic breast-cancer cells, *N. Engl. J. Med.* **356** (2007), 217–226.

[34] K. Martin, D.M. Kritzman, L.M. Price et al., Linking gene expression patterns to therapeutic groups in breast cancer, *Cancer Res.* **60** (2000), 2232–2238.

[35] A.J. Minn, G.P. Gupta, P.M. Siegel et al., Genes that mediate breast cancer metastasis to lung, *Nature* **436** (2005), 518–524.

[36] S. Mousses, L. Bubendorf, U. Wagner et al., Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays, *Cancer Res.* **62** (2002), 1256–1260.

[37] A. Myrthue, B.L. Rademacher, J. Pittsenbarger et al., The iroquois homeobox gene 5 is regulated by 1,25-dihydroxyvitamin D3 in human prostate cancer and regulates apoptosis and the cell cycle in LNCaP prostate cancer cells, *Clin. Cancer Res.* **14** (2008), 3562–3570.

[38] C.M. Perou, T. Sorlie, M.B. Eisen et al., Molecular portraits of human breast tumours, *Nature* **406** (2000), 747–752.

[39] C. Quan and S.J. Lu, Identification of genes preferentially expressed in mammary epithelial cells of Copenhagen rat using subtractive hybridization and microarrays, *Carcinogenesis* **24** (2003), 1593–1599.

[40] D. Ron and P. Walter, Signal integration in the endoplasmic reticulum unfolded protein response, *Nat. Rev. Mol. Cell Biol.* **8** (2007), 519–529.

[41] P.S. Rudland, G. Ollerhead and R. Barraclough, Isolation of simian virus 40 transformed human mammary epithelial stem cell line that can differentiate to myoepithelial-like cells in culture and *in vivo*, *Develop. Biol.* **136** (1989), 167–180.

[42] A.R. Santos and C.B. Duarte, Validation of internal control genes for expression studies: Effects of the neurotrophin BDNF on hippocampal neurons, *J. Neurosci. Res.* **16** (2008), 3684–3692.

[43] E.Y. Seo, J.H. Namkung, K.M. Lee et al., Analysis of calcium-inducible genes in keratinocytes using suppression subtractive hybridization and cDNA microarray, *Genomics* **86** (2005), 528–538.

[44] H.D. Soule, A. Vazquez, A. Long et al., Human cell line from a pleural effusion derived from a breast carcinoma, *J. Natl. Cancer Inst.* **51** (1973), 1409–1413.

[45] D.A. Thompson and R.J. Weigel, hAG-2, the human homologue of the *Xenopus laevis* cement gland gene XAG-2, is co-expressed with estrogen receptor in breast cancer cell lines, *Biochem. Biophys. Res. Commun.* **251** (1998), 111–116.

[46] J.Y. Tso, X.-H. Sun, T.-H. Kao et al., Isolation and characterisation of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs: genomic complexity and molecular evolution of the gene, *Nucleic Acids Res.* **13** (1985), 2485–2502.

[47] L.J. van't Veer, H. Dai, M.J. van de Vijver et al., Gene expression profiling predicts clinical outcome of breast cancer, *Nature* **415** (2002), 530–536.

[48] Y. Wang, J.G. Klijn, Y. Zhang et al., Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, *Lancet* **365** (2005), 671–679.

[49] G. Wang, A. Platt-Higgins, J. Carroll et al., Induction of metastasis by S100P in a rat mammary model and its association with poor survival of breast cancer patients, *Cancer Res.* **66** (2006), 1199–1207.

[50] X.Y. Yang, C.P. Ren, L. Wang et al., Identification of differentially expressed genes in metastatic and non-metastatic nasopharyngeal carcinoma cells by suppression subtractive hybridization, *Cell Oncol.* **27** (2005), 215–223.

[51] A. Zlotnik, New insights on the role of CXCR4 in cancer metastasis, *J. Pathol.* **215** (2008), 211–213.