Effects of oestrogen on gene expression in epithelium and stroma of normal human breast tissue

C L Wilson[1], A H Sims[2*], A Howell[2], C J Miller[1] and R B Clarke[2]*

[1]Cancer Research UK Bioinformatics Group, Paterson Institute for Cancer Research, Wilmslow Road, Withington, Manchester M20 4BX, UK
[2]Breast Biology Group, Cancer Research UK Department of Medical Oncology, University of Manchester, Christie Hospital (NHS) Trust, Wilmslow Road, Withington, Manchester M20 4BX, UK

(Requests for offprints should be addressed to R B Clarke; Email: robert.clarke@manchester.ac.uk)
*(C L Wilson and A H Sims contributed equally to this work)

Abstract

Oestrogen (E) is essential for normal and cancer development in the breast, while anti-oestrogens have been shown to reduce the risk of the disease. However, little is known about the effect of E on gene expression in the normal human breast, particularly when the epithelium and stroma are intact. Previous expression profiles of the response to E have been performed on tumour cell lines, in the absence of stroma. We investigated gene expression in normal human breast tissue transplanted into 9–10-week-old female athymic nude (Balb/c *nu*/*nu*) mice. After 2 weeks, when epithelial proliferation is minimal, one-third of the mice were treated with 17β-oestradiol (E2) to give human luteal-phase levels in the mouse, which we have previously shown to induce maximal epithelial cell proliferation. RNA was isolated from treated and untreated mice, labelled and hybridized to Affymetrix HG-U133A (human) GeneChips. Gene expression levels were generated using BioConductor implementations of the RMA and MAS5 algorithms. E2 treatment was found to represent the largest source of variation in gene expression and cross-species hybridization of mouse RNA from xenograft samples was demonstrated to be negligible. Known E2-responsive genes (such as TFF1 and AREG), and genes thought to be involved in breast cancer metastasis (including mammoglobin, KRT19 and AGR2), were upregulated in response to E treatment. Genes known to be co-expressed with E receptor α in breast cancer cell lines and tumours were both upregulated (XBP-1 and GREB1) and downregulated (RARRES1 and GATA3). In addition, genes that are normally expressed in the myoepithelium and extracellular matrix that maintain the tissue microenvironment were also differentially expressed. This suggests that the response to oestrogen in normal breast is highly dependent upon epithelial–stromal/myoepithelial interactions which maintain the tissue microenvironment during epithelial cell proliferation.

Endocrine-Related Cancer (2006) **13** 617–628

Introduction

Ovarian steroids, acting through nuclear receptors, are important for the development of the normal breast and breast cancer. An early menarche and a late menopause have been shown to increase breast cancer risk, whereas an early menopause is protective; these data suggest that breast cancer risk is related to a cumulative exposure to ovarian hormones. Oestrogen (E) is the focus of breast cancer therapies because tumours are often dependent on this steroid for growth (Clarke *et al*. 2004). Oestrogen is central to breast cancer development and interventions that alter exposure, such as oophorectomy or selective E receptor modulators (SERMs) and aromatase inhibitors has been shown to significantly reduce breast cancer risk (reviewed in Howell *et al*. 2005). Despite this, little is known about the effect of E on the normal human breast, particularly when the epithelium and stroma are intact.

Improved understanding of the molecular and cell biology of the breast is important for future risk-prediction and -reduction strategies. Direct effects

DOI:10.1677/erc.1.01165
1351-0088/06/013–617 © 2006 Society for Endocrinology *Printed in Great Britain*
Online version via http://www.endocrinology-journals.org

on the human breast are difficult to study because of problems of access, and breast heterogeneity. Genetic experiments on the rodent breast, especially during puberty, have highlighted the importance not only of the epithelium but interactions between these cells and other cell types within the breast, such as fibroblast, adipocytes and macrophages (Howell *et al*. 2005, Wiseman & Werb 2002). The stroma represents the major component of the adult human breast; during early reproductive life approximately 20% of breast tissue comprises epithelium, 20% fat and 60% connective tissue; the amounts of epithelium and connective tissue decline with age (Hutson *et al*. 1985). It has been suggested that despite the lack of E receptors in human stromal cells, the stroma plays an important role in the control of hormone-mediated epithelial cell proliferation (Anderson *et al*. 1998), through paracrine mechanisms (Shekhar *et al*. 2001).

Tumour cell lines or mouse models are extensively relied upon in cancer studies; however, there are significant differences between the cellular composition and behaviour of cultured cells in an *in vitro* environment and those of normal human tissue *in vivo*. It is likely that these differences will have a big impact on intracellular signalling and proliferation in response to environmental stimuli. Harvell *et al*. (2006) recently demonstrated that 17β-oestradiol (E2) regulates different genes in human breast tumour xenografts compared with the identical cells in culture. To date, a number of gene-expression profiling studies have investigated the response to E stimulation in tumour cell lines (Coser *et al*. 2003, Frasor *et al*. 2003, 2004, Inoue *et al*. 2002, Vendrell *et al*. 2004), in the absence of stroma, which does not accurately portray the breast tissue microenvironment. Some of these studies (Frasor *et al*. 2003, 2004, Inoue *et al*. 2002, Seth *et al*. 2002) employed high doses (10 nmol/l) of E2, equivalent to those observed during pregnancy and considerably higher than the fluctuating levels normally seen during the menstrual cycle. Normal breast tissue cell lines, such as MCF10As, are inappropriate for studying the response to E stimulation as they have few or no classical E receptors (ERα/ESR1). Mouse models are also unsatisfactory for studying the effects of E on normal breast as mouse mammary fat pad stromal cells are ESR1-positive, whereas their human counterparts are ESR1-negative (Anderson *et al*. 1998). Although expression profiling of tumour cell line xenografts has been previously performed (Armes & Venter 2002, Creighton *et al*. 2003, 2005), we report the first attempt to characterize the effects of E on normal human breast tissue using xenografts.

In this study, we transplanted normal human breast tissue into athymic nude mice in order to obviate the effects of variations in the stage of the menstrual cycle between patients, which is known to affect the proliferative activity of the normal breast (McManus & Welsch 1984, Potten *et al*. 1988) and has previously been shown to provide a baseline for analysis of the effects of E2 stimulation (Clarke *et al*. 1997, Laidlaw *et al*. 1995). Subcutaneous implantation of whole pieces of normal breast tissue into the athymic nude mice preserves the normal tissue architecture and, presumably, normal epithelial–mesenchymal interactions. We have investigated genes whose E regulation is persistent and thus would reflect the physiological exposure to endogenous E, whereas previous studies have generally focused on short-term effects (Frasor *et al*. 2003, Inoue *et al*. 2002, Xu *et al*. 2005). Administration of a 2 mg E2 slow-release pellet for 7 days produces similar serum concentrations of E2 to those published for the luteal phase of the menstrual cycle and which we have previously shown to be optimal for inducing proliferation of normal breast epithelium (Laidlaw *et al*. 1995). Here we describe the first attempts to study the gene expression profile in epithelium and stroma of normal human breast tissue in response to E stimulation at the level of the luteal phase of the menstrual cycle.

Materials and methods

Patient samples

Following approval of the local research ethics committee and with patient consent, normal human breast tissue was obtained from six pre-menopausal women, either at least 1 cm away from benign lesions or from women undergoing reduction mammoplasty (median age, 36.5 years; range, 30–42 years; three parous, three nulliparous; median menarche age, 13.5 years; range, 11–15 years). The tissue was confirmed to contain no abnormalities by histology.

Implantation of breast tissue and E2 administration

All of the animals used were 9–10-week-old intact female athymic nude (Balb/c *nu*/*nu*) mice (Clarke *et al*. 1997; Harlan, Bicester, Oxon, UK). All surgical procedures were performed as described previously (Laidlaw *et al*. 1995) and were carried

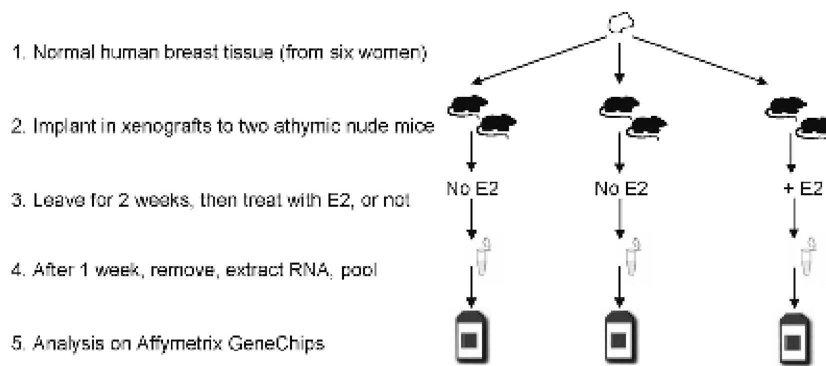

Figure 1 Overview of the experimental protocol for each sample. Normal human breast tissue was implanted into athymic nude mice. After 2 weeks one-third of the mice were treated with E for 1 week. Xenografts were removed and pooled, and RNA was extracted and hybridized to Affymetrix Genechips.

out under the Animals (Scientific Procedures) Act of 1986 (UK). Intact mice were used since they have low median levels of -E2 (94 pmol/l) with short-lasting (<24 h) peak serum levels during oestrous that are equivalent to the human follicular phase. As we have previously demonstrated, human breast tissue implanted into intact mice shows little variation in E-dependent proliferation and progesterone receptor expression (Clarke *et al.* 1997, Laidlaw *et al.* 1995).

The normal breast tissue samples were divided into approximately $2 \times 2 \times 1$ mm pieces and implanted subcutaneously by making two small incisions across the midline dorsal skin as previously described (Clarke *et al.* 1997). Eight pieces of breast tissue were implanted into six mice for each breast tissue sample. Two weeks after implantation of breast tissue, exogenous E2 was administered to two of the mice by inserting a 2 mg subcutaneous silastic pellet at the base of the tail away from the site of tissue implantation. At 1 week post-treatment or 3 weeks with no treatment, the xenografts were retrieved. In order to compensate for possible issues of breast tissue heterogeneity, we compared 12 pools of RNA extracted from untreated xenografts from six women with six pools of RNA from E2-treated tissue from the same women (see Fig. 1). Portions of breast tissue both prior to and following implantation were fixed in formalin (3.7% formaldehyde in PBS) for 1 h before transfer into paraffin for sectioning and histological examination of the tissue.

Array processing and analysis

Tissue samples were ground to a fine powder under liquid nitrogen and RNA was isolated using Trizol (Ambion) according to the manufacturer's instructions. RNA was purified using Qiagen RNeasy columns (Qiagen, Valencia, CA, USA) and quantified using a Nanodrop spectrophotometer (Labtech). The quality and amount of starting RNA were confirmed with an Agilent Bioanalyzer 2100 (Agilent) prior to labelling and hybridization to Affymetrix HG-U133A GeneChips (using standard protocols available at http://bioinformatics.picr.man.ac.uk/mbcf/protocols.shtml). After scanning, array images were assessed by eye to confirm scanner alignment and the absence of significant bubbles or scratches on the chip surface. Ratios (3′/5′) for GAPDH and β-actin were confirmed to be within acceptable limits (0.86–1.38 and 0.70–1.24, respectively), and BioB spike controls were found to be present on all chips, with BioC, BioD and CreX also present in increasing intensity. When scaled to a target intensity of 100 (using Microarray Analysis Suite^{TM} (MAS) version 5.0 array analysis software; Affymetrix) scaling factors for all arrays were within acceptable limits (0.93–2.60), as were background, Q value and mean intensities. The raw spot readings were imported into BioConductor (http://www.bioconductor.org) using R (Ihaka & Gentleman 1996) and normalized using either MAS or RMA algorithms (Irizarry *et al.* 2003) as implemented in the Simpleaffy package (Wilson & Miller 2005). The MAS data were filtered to remove probesets that were not called present on either four out of the six treated chips or eight out of the 12 untreated chips. Due to the low number of samples, a rank product analysis (Breitling *et al.* 2004) method was used to determine the significance of differentially expressed genes using the RankProd BioConductor module within R. Using the rank product analysis method, a false discovery rate

of 10% was employed and fold changes were calculated by comparing the mean expression level of probesets from the two untreated samples with those of the E2-treated samples. Unsupervised principle components analysis was performed using the MaxD View software available from the University of Manchester Bioinformatics Department (http://bioinf. man.ac.uk/microarray/maxd/). The microarray data are Minimum Information About a Microarray Experiment (MIAME) compliant and accessible via MIAME VICE (http://bioinformatics.picr.man. ac.uk/mvice/index.jsp).

Possible cross-species hybridization

Recovery of the human tissue from the mice introduces the potential for skewing of the data due to contamination of the human RNA with traces of mouse RNA. To investigate potential cross-species hybridization of mouse sequences with probesets designed to detect human sequences, varying concentrations of mouse RNA (100, 50 or 20% of the normal protocol amounts, in duplicate) were hybridized to six Affymetrix human U95A GeneChips (these experiments were performed prior to later versions of GeneChips becoming available).

Morphometry

Before and after implantation, the normal human breast tissue was examined by morphometry. The cellular percentage composition of the tissue samples was determined using an 11×11 grid at $\times 200$ magnification and by counting the squares filled with epithelial cells, stromal cells (including fibroblasts and endothelium) and fat.

Immunohistochemical quantification

Proliferative activity was assessed by immunohistochemistry using the mouse monoclonal antibody MIB-1 (Coulter) raised against the Ki67 proliferation-associated antigen. Expression of PgR and pS2 was determined by immunohistochemistry using a rat monoclonal anti-PR antibody (clone KD68; Abbott Laboratories) and a rabbit anti-pS2 polyclonal (NCL-pS2; Novocastra) respectively. Microwave antigen-retrieval methods and dilutions were as described previously (Clarke *et al.* 1997). Antibody binding was detected indirectly using the appropriate biotinylated second antibodies, a peroxidase-conjugated avidin–biotin complex (ABC Elite; Vector Laboratories) and diaminobenzidine as the chromogen. Quantitation of immunostaining was carried out with a light microscope and was restricted to the epithelial cells of the terminal-duct lobulo-alveolar units. Areas to be counted were selected out of focus at low power and then complete high-power fields were scored. At least 1000 epithelial cells were scored per sample and the number of labelled cells was expressed as a percentage of the total cells counted. The intensity of staining was not assessed.

Results

Differential expression in response to E

In order to find statistically significant changes in gene expression with our relatively small data set we utilized a rank products method which reliably and consistently outperforms a non-parametric *t* test, even with noisy data (Breitling *et al.* 2004). Analysis of gene expression in normal human breast tissue identified that a greater number of probesets are significantly differentially expressed in response to luteal phase levels of E2 when data are analysed using RMA (365; Irizarry *et al.* 2003), compared with MAS (93; Sasik *et al.* 2002). Comparing these two lists identified a total of 61 probesets (Table 1) commonly differentially expressed with both methods. A consensus approach that only considers probesets identified as changing by a number of different normalization methods has the advantage of providing a smaller but consistent gene list at the expense of missing some differentially expressed genes (Choe *et al.* 2005). However, the proportion of up- and downregulated genes was different between the two methods; 79% (289/365) of probesets were downregulated in response to E using RMA, compared with 52% (48/93) of probesets downregulated using MAS.

The list of significantly differentially expressed genes in response to E2 in normal human breast tissue (Table 1) can be assigned into three major groups. Firstly, a number of established E2-responsive genes indicated in Table 1 have been identified in previous studies of breast cancer cell lines (Coser *et al.* 2003, Frasor *et al.* 2003, 2004, Inoue *et al.* 2002, Vendrell *et al.* 2004), including Trefoil factor 1 (TFF1; formerly pS2), amphiregulin (AREG), myosin-binding protein C (MYBPC1) and transcription factor *b*-Myc (MYB). TFF1 has previously been shown to stimulate cell migration of breast cancer cells (Crosier *et al.* 2001, Prest *et al.* 2002) and a paracrine role for TFF1 has been postulated in gastrointestinal epithelial cells in

Table 1 E2-response probesets identified as differentially expressed in normal human breast tissue. FC, fold change; pfp is the estimated false discovery rate value for the list up to each gene if that gene was the cutoff point. Italics represent a mouse sequence present on the human GeneChip, which should be discounted (see Table 2).

Gene	Description	Probeset ID	FC MAS	pfp MAS	FC RMA	pfp RMA
TFF1[*]	Trefoil factor 1 (pS2)	205009_at	12.2	0.000	10.9	0.000
MYBPC1[*]	Myosin-binding protein C, slow type	214087_s_at	4.4	0.000	2.9	0.000
AREG[*]	Amphiregulin (schwannoma-derived growth factor)	205239_at	4.2	0.000	2.5	0.000
SCGB1D2†	Secretoglobin, family 1D, member 2 (lipophilin B)	206799_at	3.6	0.000	2.3	0.000
TFF3[*]	Trefoil factor 3 (intestinal)	204623_at	3.3	0.000	2.6	0.000
SCGB2A2†	Secretoglobin, family 2A, member 2 (mammaglobin)	206378_at	2.9	0.000	4.2	0.000
GREB1[*]	GREB1 protein	205862_at	2.8	0.000	1.9	0.000
SERPINA1‡	Serine (or cysteine) proteinase inhibitor, clade A	202833_s_at	2.5	0.000	2.0	0.000
SERPINA1‡	Serine (or cysteine) proteinase inhibitor, clade A	211429_s_at	2.3	0.000	2.5	0.000
C1orf34[*]	Chromosome 1 open reading frame 34 (DEME-6)	210652_s_at	1.9	0.002	1.6	0.002
PIP†	Prolactin-induced protein	206509_at	1.9	0.004	1.6	0.002
AGR2†	Anterior gradient 2 homologue	209173_at	1.9	0.003	1.9	0.000
–	CES hBr3 mRNA for brain carboxylesterase hBr3	208007_at	1.8	0.002	2.0	0.000
SERPINA3‡	Serine (or cysteine) proteinase inhibitor, clade A	202376_at	1.8	0.003	1.9	0.000
PRR4	Proline-rich 4 (lacrimal)	204919_at	1.7	0.012	1.4	0.037
–†	Human glomerular epithelial protein 1 (GLEPP1)	211600_at	1.6	0.006	1.7	0.001
HBE1	Haemoglobin, ε 1	217683_at	1.6	0.011	1.8	0.000
EEF1A2[*]	Eukaryotic translation elongation factor 1 α2	204540_at	1.6	0.011	1.5	0.002
DSU	Likely orthologue of mouse dilute suppressor	219648_at	1.6	0.013	1.7	0.001
MYB[*]	*v*-Myb myeloblastosis viral oncogene homologue (avian)	204798_at	1.6	0.020	1.6	0.003
AZGP1	α-2-Glycoprotein 1, zinc	209309_at	1.5	0.026	1.7	0.000
TACSTD1	Tumour-associated calcium signal transducer 1	201839_s_at	1.5	0.031	1.6	0.003
KRT19†	Keratin 19	201650_at	1.5	0.033	1.8	0.000
–	*Mouse anti-colorectal carcinoma light chain*	*215906_at*	*1.4*	*0.053*	*1.6*	*0.001*
CELSR2	Cadherin, EGF LAG seven-pass G-type receptor 2	204029_at	1.4	0.075	1.5	0.005
FXYD3	FXYD domain containing ion transport regulator 3	202489_s_at	1.4	0.076	1.5	0.006
XBP1[*]	X-box binding protein 1	200670_at	1.4	0.083	1.6	0.001
PRG4	Proteoglycan 4	206007_at	1.3	0.012	1.4	0.003
MMP1‡	Matrix metalloproteinase 1 (interstitial collagenase)	204475_at	1.2	0.025	0.9	0.008
ZNF236	Zinc finger protein 236	222227_at	1.0	0.013	1.1	0.001
–	Moderately similar to HBA_human haemoglobin α	217572_at	0.7	0.015	0.9	0.000
MMP12‡	Matrix metalloproteinase 12 (macrophage elastase)	204580_at	0.7	0.047	0.9	0.000
ME1	Malic enzyme 1, NADP(+)-dependent, cytosolic	211204_at	0.6	0.005	0.7	0.000
CXCL11‡	Chemokine (C-X-C motif) ligand 11	210163_at	0.6	0.006	0.6	0.000
COL6A1‡	Collagen, type VI, α 1	212938_at	0.6	0.004	0.7	0.000
CXCL11‡	Chemokine (C-X-C motif) ligand 11	211122_s_at	0.6	0.009	0.7	0.000
GATA3[*]	GATA binding protein 3	209604_s_at	0.6	0.014	0.7	0.000
HSPG2‡	Heparan sulphate proteoglycan 2 (perlecan)	201654_s_at	0.6	0.013	0.8	0.002
CCL2‡	Chemokine (C-C motif) ligand 2	216598_s_at	0.6	0.014	0.7	0.001
IGHM	Immunoglobulin heavy constant γ1	217022_s_at	0.6	0.021	0.9	0.008
DPT	Dermatopontin	207977_s_at	0.6	0.037	0.7	0.001
EMILIN1‡	Elastin microfibril interfacer 1	204163_at	0.6	0.047	0.8	0.052
CXCL10‡	Chemokine (C-X-C motif) ligand 10	204533_at	0.6	0.058	0.7	0.000
RARRES1	Retinoic acid receptor responder (tazarotene induced) 1	206392_s_at	0.6	0.068	0.8	0.003
S100A8	S100 calcium-binding protein A8 (calgranulin A)	202917_s_at	0.6	0.085	0.8	0.019
IGJ	Immunoglobulin J polypeptide,	212592_at	0.5	0.004	0.8	0.000
CXCL9‡	Chemokine (C-X-C motif) ligand 9	203915_at	0.5	0.003	0.6	0.000
FN1‡	Fibronectin 1	214701_s_at	0.5	0.005	0.7	0.000
–	Endomucin-1	220833_at	0.5	0.014	0.6	0.000
DPT	Dermatopontin	213068_at	0.5	0.014	0.6	0.000
RGS5	Regulator of G-protein signalling 5	209070_s_at	0.5	0.014	0.7	0.000
CCL19‡	Chemokine (C-C motif) ligand 19	210072_at	0.5	0.021	0.7	0.001
DPT	Dermatopontin	217430_x_at	0.5	0.038	0.7	0.000
TPSAB1	Tryptase α/β 1	215382_x_at	0.5	0.048	0.6	0.000
CSPG2‡	Chondroitin sulphate proteoglycan 2 (versican)	215646_s_at	0.5	0.053	0.7	0.001
ENPEP‡	Glutamyl aminopeptidase (aminopeptidase A)	204845_s_at	0.5	0.055	0.9	0.036

Table 1 *continued*

Gene	Description	Probeset ID	FC MAS	pfp MAS	FC RMA	pfp RMA
SERPINH1‡	Serine (or cysteine) proteinase inhibitor, clade H (hsp 47)	207714_s_at	0.5	0.073	0.8	0.061
RGS5	Regulator of G-protein signalling 5	209071_s_at	0.5	0.074	0.7	0.002
IL8‡	Interleukin 8 (CXCL8)	202859_x_at	0.4	0.003	0.7	0.001
IGHA1	Immunoglobulin heavy constant á 1	211430_s_at	0.4	0.005	0.8	0.000
BGN‡	Biglycan	201262_s_at	0.3	0.000	0.7	0.001

*Known E2-responsive genes from previous studies of breast cancer cell lines; †Known prognostic or metastatic genes; ‡Putative microenvironment maintenance genes; see text.

Table 2 Possible cross-species hybridization of mouse RNA to human GeneChips. Probesets detected when mouse RNA was hybridized to human chips that were also identified as significantly differentially expressed in response to E (Table 2). H, human; M, mouse.

Affymetrix GeneChip				No. of exact matches, probes to human/mouse sequence				
U133	U95	Gene ID	Nucleotide similarity, H/M	U95A versus mouse	U133 versus mouse	U95 versus human	U133 versus human	Transcript identified in E2 experiment
215906_at	31565_at	(S65921)	*	16/16	10/11	4/16	3/11	Mouse
211204_at	33644_at	ME1	1654/1832 (90%)	16/16	11/11	16/16	11/11	Unclear
204540_at	35174_i_at†	EEF1A2	1241/1392 (89%)	0/16	0/11	4/16	11/11	Human

*The full target consensus sequence is derived from the mouse nucleotide sequence; †Many non-specific nucleotides in the consensus/target sequence, which includes the poly-A tail and 3′ untranslated sequence.

inflammatory bowel disease (Rodrigues *et al.* 2001, Wright *et al.* 1993). Amphiregulin is a secreted growth factor related to epithelial growth factor (EGF) and transforming growth factor α (TGFα). TFF1 or AREG are likely candidates for stimulating the paracrine secretion of growth factors which induce proliferation of ER-positive cells in response to systemic E. Indeed, Sternlicht *et al.* (2005) recently used $AREG^{-/-}$ mice to suggest that AREG may have an essential role in the epithelial/stromal crosstalk that drives mammary development (Sternlicht *et al.* 2005). Some genes that are thought to be co-expressed with ESR1 were downregulated in E-treated xenogafts, including RARRES1 and GATA3 (Lacroix & Leclercq 2004). In contrast, other genes that are thought to be co-expressed with ESR1 in breast cancer cell lines and tumours were upregulated, including XBP-1 (Lacroix & Leclercq 2004), GREB1 (Lin *et al.* 2004), elongation factor 1 α2 (EEF1A2) and the little-studied C1orf34 (DEME-6), expression of which has been previously detected in primary breast carcinomas, but not in normal breast tissue (Kuang *et al.* 1998).

A second group of genes with increased expression due to E treatment (Table 1) were known diagnostic or prognostic markers, including lipophilin B (SCGB1D2), mammaglobin (SCGB2A2), prolactin-inducible peptide (PIP) and cytokeratin 19 (KRT19), which were recently used for a multigene reverse transcriptase PCR assay to detect circulating epithelial cells in the blood of patients with breast cancer (Ring *et al.* 2005). KRT19 is also a putative stem cell marker in the breast (Clarke *et al.* 2005, Gudjonsson *et al.* 2002). Anterior gradient 2 homologue (AGR2) was significantly upregulated in response to E2; the presence of detectable AGR2 mRNA in breast carcinoma cells is known to significantly correlate with carcinoma in preference to normal tissue, and ERα-positive in preference to ERα-negative carcinomas, suggesting that the metastasis-related properties of AGR2 may contribute, in some way, to the malignant progression of some ER-positive breast cancers (Liu *et al.* 2005).

A third category of differentially expressed genes was identified that appear to be involved with breast tissue microenvironment maintenance (Table 1). Serpins A1 and A3 were highly upregulated; serpin A1 encodes an anti-trypsin proteinase inhibitor which plays a key role in the control of homeostasis by neutralizing the deleterious effects of neutrophil elastase. Inhibition of serine proteinases

may explain the action of serpin A1 in the suppression of cell growth in breast cancer cells, because it prevents the proteolytic release of membrane-bound TGFα (Kang *et al.* 2005). This model would predict that serpin A1 can act as a tumour suppressor in inhibiting the growth of breast cancer cells (Yavelow *et al.* 1997). Genes involved with the generation of vascular stroma, which is thought to be essential for solid tumour growth, were downregulated. These genes include collagen (COL6A1), perlecan (HSPG2, heparan sulphate proteoglycan), versican (CSPG2, chondroitin sulphate proteoglycan 2), biglycan (BGN) and fibronectin (FN1). These genes encode components of the extracellular matrix and basement membrane, are involved with stabilization of other molecules that are associated with fibroblasts and are over-expressed in breast cancer (Brown *et al.* 1999). A number of chemokines (CCL2, interleukin-8, CXCL9, CXCL11, CCL19) were also down-regulated by E2 in normal breast tissue. In mammary tissues, SDF1 staining is primarily seen in stromal cells and weakly in epithelial cells (Kang *et al.* 2005). CCL2 and interleukin-8 have previously been suggested to have promalignant activity (Ben-Baruch 2003).

Tissue heterogeneity

Unsupervised principle component analysis identified that there were greater differences in the overall gene expression profile between the E2-treated and untreated tissues, than between the two untreated samples (data not shown). Up to seven different tissue samples from before and after implantation were examined by morphometry for cellular content to look at the variability in breast tissue heterogeneity. The median breast tissue composition from individual samples was 87% connective tissue and stroma (interquartile (IQ) range, 80–90%), 11% epithelium (IQ range, 8–14%) and 2% fat (IQ range, 0–5%). There were smaller differences in cellular composition after combining the xenograft samples for each woman (median, 86% stroma (IQ range, 83–87%); 11% epithelium (IQ range, 10–13%) and 3% fat (IQ range, 2–5%), as was performed to generate the RNA pools for gene-expression analysis. There was no significant difference between the median cellular compositions, before and after implantation or between treated and untreated samples (Mann–Whitney U test). Tissue blocks were not available for every single xenograft, but, based on the samples we did look at, we feel we can be confident that each pool of RNA extracted from the multiple xenografts is derived from comparable proportions of stroma, epithelium and fat. The proportions of different cell types in the xenografts is in the range previously published for pre-menopausal women (Hutson *et al.* 1985).

Immunohistochemistry

Proliferation of epithelial cells before and after treatment with E2 was measured by immunohistochemistry. The percentage of epithelial cells expressing Ki67 following E2 treatment (median, 3%; IQ range, 2–6%) was significantly higher ($P < 0.0001$; Mann–Whitney U test; Fig. 2A) than in untreated xenografts (median, 1%; IQ range, 0.5–2%), confirming that the current study was in accordance with the previously validated xenograft model (Clarke *et al.* 1997, Laidlaw *et al.* 1995). Immunohistochemical analysis also demonstrated that, as previously shown (Clarke *et al.* 1997), progesterone receptor (PgR) levels in epithelial cells were significantly ($P < 0.0001$ Mann–Whitney U test; Fig. 2B), increased in the E2-treated xenograft tissue (median, 19%; IQ range, 10–31%) compared with the untreated tissue (median, 6%; IQ range, 2–13%). Gene-expression levels of PgR were significantly increased following addition of E2; however, the representative probeset (208305_at) was not called present on a sufficient number of GeneChips, thus eliminating it from the analysis using our strict criteria. The reason for probesets not being called present in many of the untreated samples is likely to be a direct result of very low levels of PgR transcripts. Neither protein nor transcript levels of E receptor α (ESR1) were observed to be significantly different in the E2-treated xenografts. Only one of the previous studies on the effect of E on breast cancer cell lines reported a significant reduction in ESR1 transcript levels (Inoue *et al.* 2002). In the current study, genes known to be co-expressed with ESR1 in breast cancer were both upregulated (XBP-1 and GREB1) and downregulated (RARRES1 and GATA3) in E2-treated normal tissue. The protein level of TFF1 (pS2) was seen to be significantly increased, in agreement with its transcript levels (Fig. 2C).

Cross-hybridization of mouse RNA to human GeneChips

The potential for skewing the data due to contamination of human RNA with traces of mouse RNA

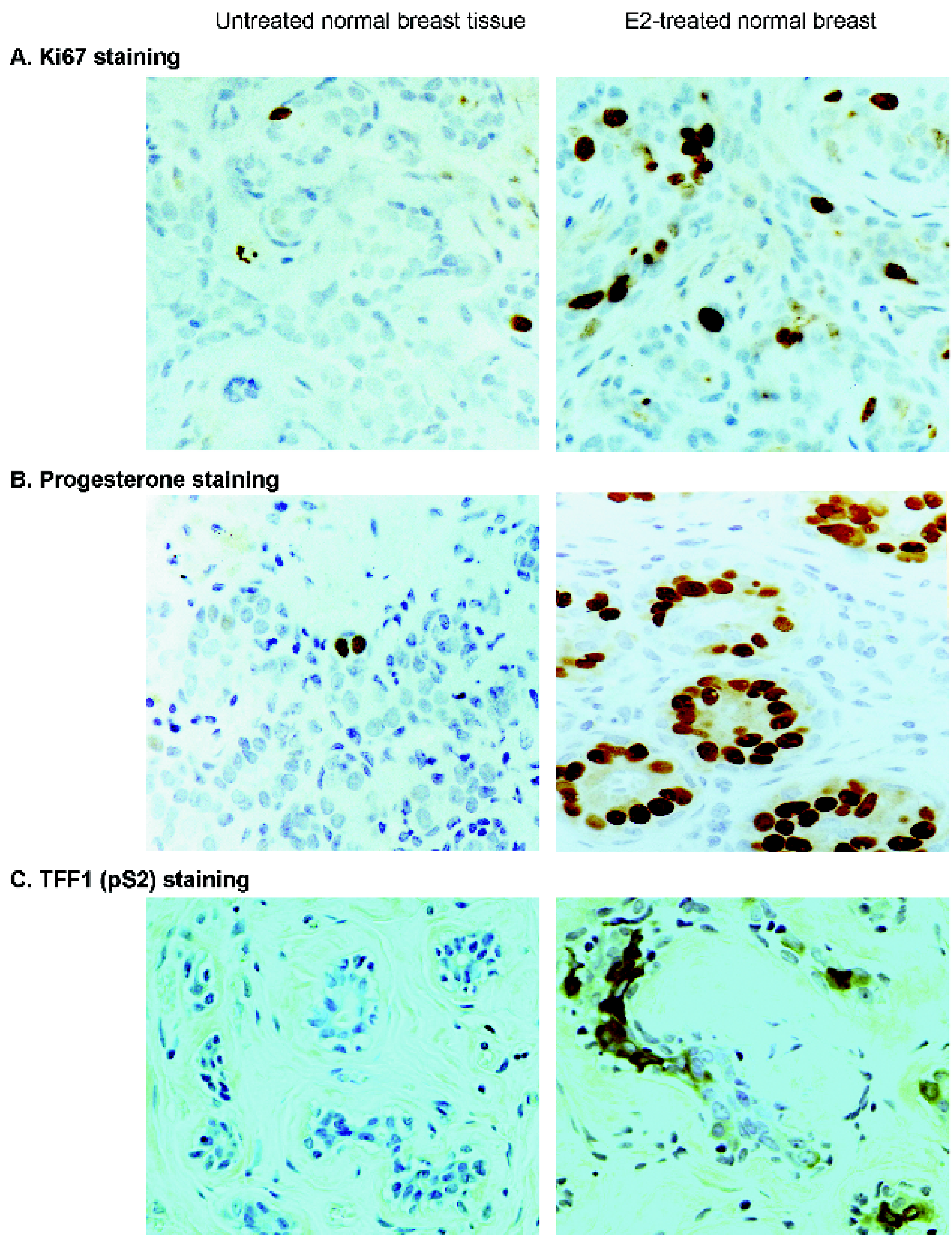

Figure 2 Immunohistochemisty analysis of E2-treated and untreated normal breast tissue. Proliferation of epithelial cells following treatment with E2. (A) Proliferation using Ki67 staining; (B) progesterone receptor staining; (C) TFF1 (pS2) staining.

was investigated by hybridizing mouse RNA to human GeneChips. Of the 12 265 full-length human genes represented on the human U95 GeneChips, fewer than 6% of the probesets (685) were called present on three or more arrays. Affymetrix array comparison spreadsheets (http:// www. affymetrix.com) were used to reconcile the 685 probesets from the U95 array with their counterparts on the U133A array (containing 22 823 probesets). Matches were found for 559 of the 685 probesets, which represents just 2.5% of the total number of probesets on the U133A GeneChips, compared with an average 55% (range, 43–61%) of probesets called present on the 18 GeneChips used to investigate the response to E2. Just three probesets (215906_at, 211204_at and 204540_at) were identified as having the potential to be affected by cross-species hybridization and significantly differentially expressed in response to E2 (see Table 2 and Fig. 3). Probeset 215906_at was designed against a consensus sequence thought to be homologous to mouse anti-colorectal carcinoma light-chain mRNA (S65921); however, the mouse transcript only partially overlaps with relatively short human cDNAs (AW605031, AW605037, AW605049 and AW934907) with the rest of the sequence made up of the mouse transcript. Consequently, only three out of the 11 probes exactly match the known human cDNAs, whereas 10 out of 11 probes match the mouse transcript, suggesting that the probeset more reliably detects the mouse transcript (Table 1). All probes from 204540_at exactly match human EEF1A2; hybridization of mouse RNA can be explained due to four out of 16 probes on the U95A GeneChip matching the orthologous human sequence for this gene. However, none of the probes on the U133A GeneChip match mouse EEF1A2, so it is highly likely that this probeset hybridizes to human transcripts (Table 2). All of the 211204_at probes exactly match human and mouse ME1 sequences. It was therefore not possible to determine whether the resulting signal for this probeset is due to the presence of human or mouse RNA (Table 2).

Discussion

While steroid receptor expression and epithelial cell proliferation occur in separate cells in the normal breast (Anderson *et al.* 1998), proliferating breast

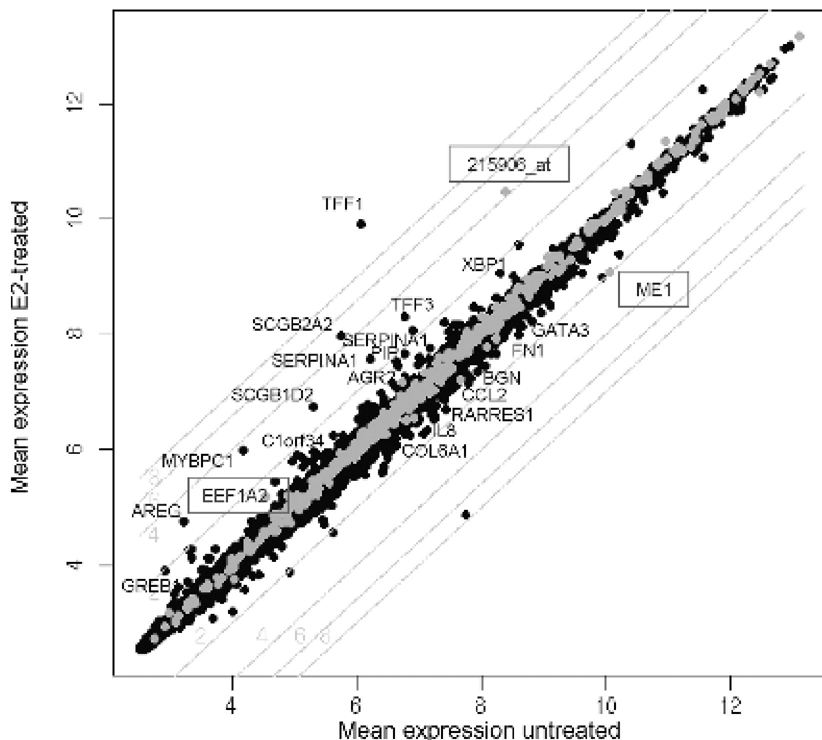

Figure 3 Comparing expression levels in the presence (*y* axis) and absence (*x* axis) of E2. Data were normalized using the RMA method. Highlighted probesets were significant using the Rankprod method (see Table 2 for a full list). Probesets shown in grey were identified as having some signal when mouse RNA was hybridized to human GeneChips.

tumour epithelial cells often express ESR1. This distinction in the biology between normal and tumour epithelium is apparent at an early point in breast tumourigenesis and it can be detected in premalignant lesions (Shoker *et al.* 1999). The differences in gene expression that we have observed in response to E2 in normal human breast tissue compared with the response previously described in tumours or cell lines may reflect these physiological changes.

Over 60% of the probesets identified as E2-responsive in (Table 2) with known Gene Ontology cellular component classifications are assigned as extracellular, suggesting that a major element of the response to E2 in normal tissue is devoted to signalling. These proteins include growth factors, matrix metaloproteases and protease inhibitors. Allinen *et al.* (2004) showed that the most dramatic and consistent changes in breast cancer progression occur in myoepithelial cells and myofibroblasts and the majority of differentially expressed genes encode secreted and cell-surface proteins. The breast microenvironment is increasingly being studied in the context of its effect on tumourigenesis and metastasis. Inflammatory cells, cytokines and chemokines have been suggested to play a key role in breast carcinoma (Ben-Baruch 2003). The present study suggests that growth factors, proteases and chemokines are also important in the response to E in normal breast. Two recent studies have used different methods to determine breast-tissue cell-type-specific genes. Jones *et al.* (2004) combined an immunomagnetic sorting protocol selecting with the luminal epithelial marker (EMA) and myoepithilial membrane antigen (CD10) with cDNA microarray hybridizations, whereas Allinen *et al.* (2004) combined cell sorting with serial analysis of gene expression. The majority of genes identified in this study as being involved with proliferation and metastasis (TFF1, AGR2, KRT19, MYB1) appear to be most associated with the luminal epithelium in the cell-type-specific studies; by contrast, the genes identified as being involved with tissue maintenance are most specifically expressed in the myoepithelium and stroma (SEPINA1, BGN, COL6A1, FN1, CCL2). Normal mammary tissue has an intact epithelial basement membrane, as shown by a continuous linear staining for collagen, laminin, perlecan and fibronectin. This staining is widely lost in the invasive carcinomas (Brown *et al.* 1999). While putative proliferation and metastasis genes appear to demonstrate increased expression in both breast cancer and normal epithelium in response to E, the tissue-maintenance genes have reduced expression in normal tissue, but have been shown to have increased expression in DCIS myoepithelium compared with normal myoepithelium (Allinen *et al.* 2004). Similarly, SEPINA1 was upregulated in normal tissue but downregulated in breast cancer (Yavelow *et al.* 1997). These changes in the expression of tissue maintenance genes may account for differences in the response to E by normal breast tissue during menstrual cycles and its role in uncontrolled epithelial proliferation and carcinogenesis.

Although every effort was made to only isolate RNA from the recovered implanted human tissue, we also looked at the possibility for contaminating traces of mouse RNA to hybridize with probesets designed to recognize human sequences. Reassuringly, our analysis is consistent with previous cross-species hybridization studies (Creighton *et al.* 2003, 2005) showing there is a minimal contribution to the observed expression levels from contaminating RNA.

We report the first global gene expression study to examine the effects of E on the epithelium and stroma of normal human breast tissue *in vivo*. Previously, we showed that treatment of xenografts with 2 mg E2 for 1 week induces epithelial proliferation equivalent to that seen during the luteal phase of the menstrual cycle (Clarke *et al.* 1997). The majority of genes that we have identified as E2-responsive in normal human tissue appear to be either connected to proliferation/migration/invasion or are involved with maintaining tissue architecture. However, at this point it is unclear whether the differences in the gene expression profiles observed between normal breast tissue and tumour cell lines in response to E are due to the difference between tumour and normal cells or between tissue and cell lines. As with the studies of E2 response in tumour cell lines (Coser *et al.* 2003, Frasor *et al.* 2003, 2004, Inoue *et al.* 2002, Vendrell *et al.* 2004), we acknowledge that our list of E2-responsive genes is not absolute, but dependent upon the specific approach (including regimens of E2 treatment, microarray platforms and analysis methods). However, we believe that the genes we have identified in this study are representative of the true response to E in normal tissue. Further studies will be required to characterize the true differences and similarities in the response to E2 in normal and tumour tissues, which may provide much-needed clues as to how E as a risk factor leads to breast cancer development.

Acknowledgments

We thank our colleagues in the University of Manchester Department of Surgery for breast tissue sample collection and are grateful for the excellent technical assistance of Stuart Pepper, and Yvonne Hey of the Cancer Research UK Affymetrix core facility and Caroline Chadwick of the Biological Resource Unit at the Paterson Institute for Cancer Research. R B C, C J M and C L W are supported by Cancer Research UK. A H S is supported by the Genetic Innovation network and Breakthrough Breast Cancer. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References

Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H, Porter D, Hu M, Chin L, Richardson A *et al.* 2004 Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* **6** 17–32.

Anderson E, Clarke RB & Howell A 1998 Estrogen responsiveness and control of normal human breast proliferation. *Journal of Mammary Gland Biology Neoplasia* **3** 23–35.

Armes JE & Venter DJ 2002 The pathology of inherited breast cancer. *Pathology* **34** 309–314.

Ben-Baruch A 2003 Host microenvironment in breast cancer development: inflammatory cells, cytokines and chemokines in breast cancer progression: reciprocal tumor-microenvironment interactions. *Breast Cancer Research* **5** 31–36.

Breitling R, Armengaud P, Amtmann A & Herzyk P 2004 Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters* **573** 83–92.

Brown LF, Guidi AJ, Schnitt SJ, Van De Water L, Iruela-Arispe ML, Yeo TK, Tognazzi K & Dvorak HF 1999 Vascular stroma formation in carcinoma in situ, invasive carcinoma, and metastatic carcinoma of the breast. *Clinical Cancer Research* **5** 1041–1056.

Choe SE, Boutros M, Michelson AM, Church GM & Halfon MS 2005 Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biology* **6** R16.

Clarke RB, Howell A & Anderson E 1997 Estrogen sensitivity of normal human breast tissue in vivo and implanted into athymic nude mice: analysis of the relationship between estrogen-induced proliferation and progesterone receptor expression. *Breast Cancer Research Treatment* **45** 121–133.

Clarke RB, Anderson E & Howell A 2004 Steroid receptors in human breast cancer. *Trends in Endocrinology Metabolism* **15** 316–323.

Clarke RB, Spence K, Anderson E, Howell A, Okano H & Potten CS 2005 A putative human breast stem cell population is enriched for steroid receptor-positive cells. *Developmental Biology* **277** 443–456.

Coser KR, Chesnes J, Hur J, Ray S, Isselbacher KJ & Shioda T 2003 Global analysis of ligand sensitivity of estrogen inducible and suppressible genes in MCF7/BUS breast cancer cells by DNA microarray. *PNAS* **100** 13994–13999.

Creighton C, Kuick R, Misek DE, Rickman DS, Brichory FM, Rouillard JM, Omenn GS & Hanash S 2003 Profiling of pathway-specific changes in gene expression following growth of human cancer cell lines transplanted into mice. *Genome Biology* **4** R46.

Creighton CJ, Bromberg-White JL, Misek DE, Monsma DJ, Brichory F, Kuick R, Giordano TJ, Gao W, Omenn GS, Webb CP *et al.* 2005 Analysis of tumor-host interactions by gene expression profiling of lung adenocarcinoma xenografts identifies genes involved in tumor formation. *Molecular Cancer Research* **3** 119–129.

Crosier M, Scott D, Wilson RG, Griffiths CD, May FE & Westley BR 2001 High expression of the trefoil protein TFF1 in interval breast cancers. *American Journal of Pathology* **159** 215–221.

Frasor J, Danes JM, Komm B, Chang KC, Lyttle CR & Katzenellenbogen BS 2003 Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144** 4562–4574.

Frasor J, Stossi F, Danes JM, Komm B, Lyttle CR & Katzenellenbogen BS 2004 Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. *Cancer Research* **64** 1522–1533.

Gudjonsson T, Villadsen R, Nielsen HL, Ronnov-Jessen L, Bissell MJ & Petersen OW 2002 Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes and Development* **16** 693–706.

Harvell DM, Richer JK, Allred DC, Sartorius CA & Horwitz KB 2006 Estradiol regulates different genes in human breast tumor xenografts compared to the identical cells in culture. *Endocrinology* **147** 700–713.

Howell A, Sims AH, Ong KR, Harvie M, Evans DG & Clarke RB 2005 Mechanisms of Disease: prediction and prevention of breast cancer – cellular and molecular interactions. *Nature Clinical Practice Oncology* **2** 635–646.

Hutson SW, Cowen PN & Bird CC 1985 Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer. *Journal of Clinical Pathology* **38** 281–287.

Ihaka R & Gentleman R 1996 R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5** 299–314.

Inoue A, Yoshida N, Omoto Y, Oguchi S, Yamori T, Kiyama R & Hayashi S 2002 Development of cDNA microarray for expression profiling of estrogen-responsive genes. *Journal of Molecular Endocrinology* **29** 175–192.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U & Speed TP 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4** 249–264.

Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E *et al*. 2004 Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Research* **64** 3037–3045.

Kang H, Watkins G, Parr C, Douglas-Jones A, Mansel RE & Jiang WG 2005 Stromal cell derived factor-1: its influence on invasiveness and migration of breast cancer cells in vitro, and its association with prognosis and survival in human breast cancer. *Breast Cancer Research* **7** R402–R410.

Kuang WW, Thompson DA, Hoch RV & Weigel RJ 1998 Differential screening and suppression subtractive hybridization identified genes differentially expressed in an estrogen receptor-positive breast carcinoma cell line. *Nucleic Acids Research* **26** 1116–1123.

Lacroix M & Leclercq G 2004 About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. *Molecular and Cellular Endocrinology* **219** 1–7.

Laidlaw IJ, Clarke RB, Howell A, Owen AW, Potten CS & Anderson E 1995 The proliferation of normal human breast tissue implanted into athymic nude mice is stimulated by estrogen but not progesterone. *Endocrinology* **136** 164–171.

Lin CY, Strom A, Vega VB, Kong SL, Yeo AL, Thomsen JS, Chan WC, Doray B, Bangarusamy DK, Ramasamy A *et al*. 2004 Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biology* **5** R66.

Liu D, Rudland PS, Sibson DR, Platt-Higgins A & Barraclough R 2005 Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. *Cancer Research* **65** 3796–3805.

McManus MJ & Welsch CW 1984 The effect of estrogen, progesterone, thyroxine, and human placental lactogen on DNA synthesis of human breast ductal epithelium maintained in athymic nude mice. *Cancer* **54** 1920–1927.

Potten CS, Watson RJ, Williams GT, Tickle S, Roberts SA, Harris M & Howell A 1988 The effect of age and menstrual cycle upon proliferative activity of the normal human breast. *British Journal of Cancer* **58** 163–170.

Prest SJ, May FE & Westley BR 2002 The estrogen-regulated protein, TFF1, stimulates migration of human breast cancer cells. *FASEB Journal* **16** 592–594.

Ring AE, Zabaglo L, Ormerod MG, Smith IE & Dowsett M 2005 Detection of circulating epithelial cells in the blood of patients with breast cancer: comparison of three techniques. *British Journal of Cancer* **92** 906–912.

Rodrigues S, Nguyen QD, Faivre S, Bruyneel E, Thim L, Westley B, May F, Flatau G, Mareel M, Gespach C *et al*. 2001 Activation of cellular invasion by trefoil peptides and src is mediated by cyclooxygenase- and thromboxane A2 receptor-dependent signaling pathways. *FASEB Journal* **15** 1517–1528.

Sasik R, Calvo E & Corbeil J 2002 Statistical analysis of high-density oligonucleotide arrays: a multiplicative noise model. *Bioinformatics* **18** 1633–1640.

Seth P, Porter D, Lahti-Domenici J, Geng Y, Richardson A & Polyak K 2002 Cellular and molecular targets of estrogen in normal human breast tissue. *Cancer Research* **62** 4540–4544.

Shekhar MP, Werdell J, Santner SJ, Pauley RJ & Tait L 2001 Breast stroma plays a dominant regulatory role in breast epithelial growth and differentiation: implications for tumor development and progression. *Cancer Research* **61** 1320–1326.

Shoker BS, Jarvis C, Clarke RB, Anderson E, Hewlett J, Davies MP, Sibson DR & Sloane JP 1999 Estrogen receptor-positive proliferating cells in the normal and precancerous breast. *American Journal of Pathology* **155** 1811–1815.

Sternlicht MD, Sunnarborg SW, Kouros-Mehr H, Yu Y, Lee DC & Werb Z 2005 Mammary ductal morphogenesis requires paracrine activation of stromal EGFR via ADAM17-dependent shedding of epithelial amphiregulin. *Development* **132** 3923–3933.

Vendrell JA, Magnino F, Danis E, Duchesne MJ, Pinloche S, Pons M, Birnbaum D, Nguyen C, Theillet C & Cohen PA 2004 Estrogen regulation in human breast cancer cells of new downstream gene targets involved in estrogen metabolism, cell proliferation and cell transformation. *Journal of Molecular Endocrinology* **32** 397–414.

Wilson CL & Miller CJ 2005 Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics* **21** 3683–3685.

Wiseman BS & Werb Z 2002 Stromal effects on mammary gland development and breast cancer. *Science* **296** 1046–1049.

Wright NA, Poulsom R, Stamp G, Van Noorden S, Sarraf C, Elia G, Ahnen D, Jeffery R, Longcroft J, Pike C *et al*. 1993 Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology* **104** 12–20.

Xu J, Fan S & Rosen EM 2005 Regulation of the estrogen-inducible gene expression profile by the breast cancer susceptibility gene BRCA1. *Endocrinology* **146** 2031–2047.

Yavelow J, Tuccillo A, Kadner SS, Katz J & Finlay TH 1997 Alpha 1-antitrypsin blocks the release of transforming growth factor-alpha from MCF-7 human breast cancer cells. *Journal of Clinical Endocrinology and Metabolism* **82** 745–752.