# Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach

*S Tozlu[1,2], I Girault[1,2], S Vacher[1,2], J Vendrell[3], C Andrieu[1,2], F Spyratos[1,2], P Cohen[3], R Lidereau[1,2] and I Bieche[1,2]*

[1]Centre René Huguenin, FNCLCC, St-Cloud F-92210, France
[2]INSERM, U735, St-Cloud F-92210, France
[3]CNRS UMR 5160, Faculté de Pharmacie, Centre de Biotechnologie et de Pharmacologie pour la Santé, Montpellier, France

(Requests for offprints should be addressed to I Bieche, Centre René Huguenin, FNCLCC, St-Cloud F-92210, France; Email: i.bieche@stcloud-huguenin.org)

## Abstract

The estrogen receptor alpha (ERα) plays a critical role in the pathogenesis and clinical behavior of breast cancer. To obtain further insights into the molecular basis of estrogen-dependent forms of this malignancy, we used real-time quantitative reverse transcription (RT)-PCR to compare the mRNA expression of 560 selected genes in ERα-positive and ERα-negative breast tumors. Fifty-one (9.1%) of the 560 genes were significantly upregulated in ERα-positive breast tumors compared with ERα-negative breast tumors. In addition to well-known ERα-induced genes (*PGR*, *TFF1*/*PS2*, *BCL2*, *ERBB4*, *CCND1*, etc.) and genes recently identified by cDNA microarray-based approaches (*GATA3*, *TFF3*, *MYB*, *STC2*, *HPN*/*HEPSIN*, *FOXA1*, *XBP1*, *SLC39A6*/*LIV-1*, etc.), an appreciable number of novel genes were identified, many of, which were weakly expressed. This validates the use of large-scale real-time RT-PCR as a method complementary to cDNA microarrays for molecular tumor profiling. Most of the new genes identified here encoded secreted proteins (*SEMA3B* and *CLU*), growth factors (*BDNF*, *FGF2* and *EGF*), growth factor receptors (*IL6ST*, *PTPRT*, *RET*, *VEGFR1* and *FGFR2*) or metabolic enzymes (*CYP2B6, CA12*, *ACADSB*, *NAT1*, *LRBA*, *SLC7A2* and *SULT2B1*). Importantly, we also identified a large number of genes encoding proteins with either pro-apoptotic (*PUMA*, *NOXA* and *TATP73*) or anti-apoptotic properties (*BCL2*, *DNTP73* and *TRAILR3*). Surprisingly, only a small proportion of the 51 genes identified in breast tumor biopsy specimens were confirmed to be ERα-regulated and/or E2-regulated *in vitro* (cultured cell lines). Therefore, this study identified a limited number of genes and signaling pathways, which better delineate the role of ERα in breast cancer. Some of the genes identified here could be useful for diagnosis or for predicting endocrine responsiveness, and could form the basis for novel therapeutic strategies.

*Endocrine-Related Cancer* (2006) **13** 1109–1120

## Introduction

Estrogens are important regulators of growth and differentiation in the normal mammary gland, and also play a major role in the onset and progression of breast cancer (Pike *et al.* 1993). Estrogens act via their receptors (estrogen receptors; ERs), which belong to the nuclear receptor superfamily of ligand-activated transcription factors that control physiological and pathological processes, largely by regulating gene transcription (McDonnell & Norris 2002).

The mitogenic effects of estrogens are largely attributed to their ability to increase the expression of key cell-cycle regulatory genes (Prall *et al.* 1997). However, regulation of cell proliferation is only one


1351–0088/06/013–1109 
DOI:10.1677/erc.1.01120
*Online version via http://www.endocrinology-journals.org*

aspect of estrogen action, and there is a pressing need to identify the full set of estrogen-responsive genes. The existence of other ER signaling pathways that are independent of estrogen has also been postulated (Zwijsen *et al.* 1998, Ding *et al.* 2003). Thus, to investigate the full range of ER signaling, gene expression profiling studies should compare ER+ and ER− tumors rather than focus solely on ERα-positive breast tumor cell lines, regulated or not by estrogens (MCF7, T-47D, etc.), which may not accurately reflect the physiological and pathological effects of ER signaling *in vivo*.

The recent advent of efficient tools for large-scale gene expression analysis has already provided new insights into the involvement of gene networks and regulatory pathways in various tumoral processes (DeRisi *et al.* 1996). cDNA microarrays can be used to test the expression of thousands of genes at a time, while real-time RT-PCR offers more accurate and quantitative information on smaller numbers of selected candidate genes (Latil *et al.* 2003, Bieche *et al.* 2004*a*).

Here, to identify new estrogen-responsive (or estrogen receptor-responsive) genes, we used real-time RT-PCR to quantify the mRNA expression of a large number of selected genes in pooled ERα-positive breast tumors, in comparison with pooled ERα-negative breast tumors (screening set). Thus we determined the expression level of 560 genes known to be involved in various cellular and molecular mechanisms associated with tumorigenesis. We particularly focused on the expression of genes found, by means of microarray analysis of breast tumor biopsies, to co-cluster with ERα, such as *TFF3*, *GATA3*, *FOXA1/HNF3A*, *SLC39A6/LIV-1*, *XBP1*, *STC2*, *HPN/HEPSIN* and *MYB* (Perou *et al.* 2000, Gruvberger *et al.* 2001, Sorlie *et al.* 2001, West *et al.* 2001, Bertucci *et al.* 2002, van't Veer *et al.* 2002).

Genes of interest were further investigated in an independent well-characterized series of 36 individual breast tumor samples, including 24 ERα-positive and 12 ERα-negative samples (validation set), as well as in five breast tumor cell lines and in the MCF7 cell line treated with E2.

## Materials and methods

### Patients and samples

We analyzed tissue samples from primary breast tumors excised from 48 women at Centre René Huguenin. Tumor samples containing more than 70% of tumor cells were considered suitable for the study.

Twelve tumors (six ERα-positive and six ERα-negative breast tumors) were used for the initial pooled sample analysis (screening set).

Thirty-six tumors constituted the validation set: all originated from two previous published studies (Bieche *et al.* 2001*a*, 2001*b*) and were selected so that two-third (24) were ERα-positive, whereas the remaining 12 were ERα-negative.

The 36 patients from the validation set met the following criteria: primary unilateral non-metastatic breast carcinoma; complete clinical, histological, and biological information available; no radiotherapy or chemotherapy before surgery; and full follow-up at Centre René Huguenin.

Patients underwent physical examinations and routine chest radiography every 3 months for 2 years and then annually. Mammograms were also done annually. Estrogen receptor status was determined at the protein level by biochemical enzymatic immuno-assay (EIA) method and confirmed by ERα real-time quantitative RT-PCR assay (Bieche *et al.* 2001*c*). The mRNA level median of *ERα* gene was 1 (range, 0.2–5.1) in the ERα-negative breast tumor group and 711 (range, 70.8–1938) in the ERα-positive breast tumor group.

All the 24 ERα-positive breast tumor patients received post-operative adjuvant endocrine therapy (tamoxifen, 20 mg daily for 3–5 years), and no other treatment. The median follow-up was 7.3 years (range 3.2–12.5 years). Twelve of the 24 ERα-positive breast tumor patients relapsed.

The tumor samples were flash-frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

We also analyzed five breast tumor cell lines obtained from the American Type Culture Collection (ATCC), including two ERα-positive cell lines (MCF7 and T-47D) and 3 REα-negative cell lines (MDA-MB-231, MDA-MB-435 and SK-BR-3).

### MCF7 cell line treated with E2

Prior to treatment, MCF7 cells were purged for four days in Dulbecco's Modified Eagle Medium without phenol red supplemented with 3% of steroid-depleted, dextran-coated charcoal-treated fetal calf serum. Cells were then treated for 4 days (with one media change) under the following pharmacological conditions: steroid-depleted medium (vehicle) and 1nM E2 (17β-estradiol).

### Real-time RT-PCR

#### *Theoretical basis*

Reactions are characterized by the point during cycling when amplification of the PCR product is first detected,

rather than the amount of PCR product accumulated after a fixed number of cycles. The larger the starting quantity of the target molecule, the earlier a significant increase in fluorescence will be observed. The parameter $C_t$ (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of a TaqMan probe (or by SYBR green dye–amplicon complex formation) passes a fixed threshold above baseline. The increase in fluorescent signal associated with exponential growth of PCR products is detected by the laser detector of the ABI Prism 7700 Sequence Detection System (Perkin–Elmer Applied Biosystems, Foster City, CA, USA), using PE Biosystems analysis software according to the manufacturer's manuals.

The precise amount of total RNA added to each reaction mix (based on optical density) and its quality (i.e., lack of extensive degradation) are both difficult to assess. Therefore, we also quantified transcripts of two endogenous RNA control genes involved in two cellular metabolic pathways, namely *TBP* (Genbank accession NM_003194), which encodes the TATA box-binding protein (a component of the DNA-binding protein complex TFIID), and *RPLP0* (also known as 36B4; NM_001002), which encodes human acidic ribosomal phosphoprotein P0. Each sample was normalized on the basis of its *TBP* (or *RPLPO*) content.

Results, expressed as *N*-fold differences in target gene expression relative to the *TBP* (or *RPLPO*) gene, and termed '$N_{target}$', were determined as

$$N_{target} = 2^{\Delta C_{t\ sample}}$$

where the $\Delta C_t$ value of the sample was determined by subtracting the average $C_t$ value of the target gene from the average $C_t$ value of the *TBP* (or *RPLP0*) gene (Bieche *et al.* 1999, 2001*a*).

The $N_{target}$ values of the samples were subsequently normalized such that the median of the ERα-negative breast tumor values was 1.

### Primers and controls

Primers for *TBP*, *RPLP0* and the 560 target genes (list in Supplemental data) were chosen with the assistance of the Oligo 5.0 computer program (National Biosciences, Plymouth, MN, USA).

We conducted searches in dbEST, htgs and nr databases to confirm the total gene specificity of the nucleotide sequences chosen as primers, and the absence of single nucleotide polymorphisms. In particular, the primer pairs were selected to be unique relative to the sequences of closely related family member genes or of the corresponding retropseudogenes. To avoid amplification of contaminating genomic DNA, one of the two primers was placed at the junction between two exons, if possible. In general, amplicons were between 70 and 120 nucleotides long. Gel electrophoresis was used to verify the specificity of PCR amplicons.

For each primer pair, we performed no-template control (NTC) and no-reverse-transcriptase control (RT negative) assays, which produced negligible signals (usually $>40$ in $C_t$ value), suggesting that primer–dimer formation and genomic DNA contamination effects were negligible.

### RNA extraction

Total RNA was extracted from frozen tumor samples by using the acid–phenol guanidinium method. The quality of the RNA samples was determined by electrophoresis through agarose gels and staining with ethidium bromide, the 18S and 28S RNA bands being visualized under u.v. light.

### cDNA Synthesis

Total RNA was reverse transcribed in a final volume of 20 µl containing 1× RT buffer (5 µM each dNTP, 3 mM MgCl2, 75 mM KCl, 50 mM Tris–HCl pH 8.3), 20 units RNasin RNase inhibitor (Promega), 10 mM DTT, 100 units Superscript II RNase H-reverse transcriptase (Invitrogen), 3 µM random hexamers (Pharmacia) and 1 µg total RNA. The samples were incubated at 25 °C for 10 min and 42 °C for 30 min, and reverse transcriptase was inactivated by heating at 99 °C for 5 min and cooling at 4 °C for 5 min.

### PCR amplification

All PCR were performed using an ABI Prism 7700 Sequence Detection System (Perkin–Elmer Applied Biosystems) and either the TaqMan® PCR Core REAGENTS Kit or the SYBR® Green PCR Core Reagents kit (Perkin–Elmer Applied Biosystems). A 5 µl diluted sample of cDNA (produced from 2 ng total RNA) was added to 20 µl of the PCR master-mix.

The thermal cycling conditions comprised an initial denaturation step at 95 °C for 10 min, and 50 cycles at 95 °C for 15 s and 65 °C for 1 min.

## Statistical analysis

As the mRNA levels did not fit a Gaussian distribution, (a) the mRNA levels in each subgroup of samples were characterized by their median values and ranges, rather than their mean values and coefficients of variation, and (b) relationships between the molecular markers and clinical and biological parameters were tested

using the non-parametric Mann–Whitney *U*-test (Mann & Whitney 1947). Differences between two populations were judged significant at confidence levels greater than 95% ($P<0.05$).

To visualize the capacity of a given molecular marker to discriminate between two populations (in the absence of an arbitrary cutoff value), we summarized the data in a receiver operating characteristic (ROC) curve (Hanley & McNeil 1982). This curve plots sensitivity (true positives) on the *Y* axis against 1-specificity (false positives) on the *X* axis, considering each value as a possible cutoff. The area under curve (AUC) was calculated as a single measure for the discriminatory capacity of each molecular marker. When a molecular marker had no discriminatory value, the ROC curve lies close to the diagonal and the AUC is close to 0.5. In contrast, when a molecular marker has strong discriminatory value, the ROC curve moves to the upper left-hand corner and the AUC is close to 1.0.

## Results

We first determined the mRNA expression level of the 560 selected genes in an ERα-positive and an ERα-negative breast tumor pools (screening set). These pools were each prepared by mixing identical amounts of tumor RNA from six patients. The mean *TBP* gene $C_t$ (threshold cycle) values for the six individual tumor samples were $25.63\pm0.28$ (ERα-positive pool) and $25.82\pm0.34$ (ERα-negative pool).

Genes, whose expression in the ERα-positive breast tumor pool was at least three times higher than in the ERα-negative breast tumor pool were then examined for their mRNA expression in an independent well-characterized series of 24 individual ERα-positive breast tumors and 12 ERα-negative breast tumors (validation set).

This robust selection criterion ensures the identification of genes of marked interest.

### Expression of the 560 genes in the ERα-positive and ERα-negative breast tumor pools (screening set)

mRNA levels of 45 (8.0%) of the 560 genes were detectable but not reliably quantifiable by means of real-time quantitative RT-PCR ($C_t>35$), in both the ERα-positive and ERα-negative breast tumor pools.

Fifty-six (10.8%) of the remaining 517 genes were upregulated (>3-fold) in the ERα-positive pool compared with the ERα-negative pool.

In contrast, 25 (4.8%) of the 517 genes were downregulated (>3-fold) in the ERα-positive pool compared with the ERα-negative pool. It is probable that these 25 latter genes are not estrogen-regulated, but correspond rather to genes that are mainly upregulated in undifferentiated tumors (i.e., ERα-negative breast tumors), independently of ERα status.

### mRNA expression of *ESR1*/*ERα*, *ESR2*/*ERβ* and 56 candidate genes in 24 individual ERα-positive breast tumors and 12 ERα-negative breast tumors (validation set)

The expression level of the 56 upregulated genes identified by pooled sample analysis was then determined individually in an independent series of 24 ERα-positive breast tumors and 12 ERα-negative breast tumors. Fifty-one (91.1%) of the 56 upregulated genes identified by pooled sample analysis were significantly upregulated in the 24 individual ERα-positive breast tumors relative to the 12 ERα-negative breast tumors ($P<0.05$; Table 1).

The 51 upregulated genes mainly encoded growth factors and secreted proteins, (*STC2*, *TFF1*/*PS2*, *SEMA3B*, *IGFBP4*, *BDNF*, *CLU*, *IGFBP5*, *FGF2*, *EGF* and *CGA*) growth factor receptors (*IL6ST*, *ERBB4*, *PTPRT*, *RET* and *FGFR2*), transcription factor (*FOXA1*, *PGR*, *BLU*, *GATA3*, *XBP1*, *MYB*, *AR* and *PAX3*), metabolic enzymes (*CYP2B6*, *CA12*, *ACADSB*, *NAT1*, *LRBA*, *SLC7A2* and *SULT2B1*), and proteins involved in cell proliferation (*p27/CDKN1B* and *CCND1*) and apoptosis (*BCL2*, *TNFRSF10C*/*TRAILR3*, *PUMA*, *NOXA*, *DNTP73*, *TATP73*).

The capacity of each of these 51 genes to discriminate between ERα-positive and ERα-negative breast tumors was then tested by ROC curve analysis. The overall diagnostic values of the 51 molecular markers were assessed in terms of their AUC values (Table 1). Three genes perfectly discriminated between the ERα-positive and ERα-negative breast tumors (AUC-ROC, 1.000), namely *CYP2B6*, *CA12* and *IL6ST*. Fig. 1 shows the mRNA levels of these three genes in each of the 24 ERα-positive breast tumors and the 12 ERα-negative breast tumors.

In the same set of 36 tumors, we also examined the expression of the *ESR2*/*ERβ* gene and found that it was similar in the ERα-positive and ERα-negative breast tumors (AUC-ROC, 0.502).

The mRNA levels indicated in Table 1 (calculated as described in Materials and methods) show the abundance of the target relative to the endogenous control (*TBP*), used to normalize the starting amount and quality of total RNA. Similar results were obtained with a second endogenous control, *RPLP0* (also known as 36B4).

**Table 1** List of the significantly upregulated genes in the 24 ERα-positive breast tumors relative to the 12 ERα-negative breast tumors

| Genes | Gene definition | Gene characterization | REα-negative ($n=12$) | REα-positive ($n=24$) | $P$[a] | ROC-AUC[b] |
|---|---|---|---|---|---|---|
| *CYP2B6* | Cytochrome P450 CYP2B6 | Metabolic enzyme | 1.0 (0–9.2)[c] | 2105 (12.4–13 896) | 0.0000014 | 1.000 |
| *CA12* | Carbonic anhydrase XII | Metabolic enzyme | 1.0 (0.1–8.5) | 96.1 (12.2–504) | 0.0000014 | 1.000 |
| *IL6ST* | Interleukin 6 signal transducer (gp130) | Growth factor receptor | 1.0 (0.3–2.2) | 16.5 (6.1–128) | 0.0000014 | 1.000 |
| *STC2* | Stanniocalcin 2 | Growth factor | 1.0 (0.1–2.8) | 84.3 (1.8–750) | 0.0000016 | 0.997 |
| *ACADSB* | Acyl-coenzyme A dehydrogenase, short/branched chain | Metabolic enzyme | 1.0 (0.2–4.5) | 16.2 (4.2–63,1) | 0.0000019 | 0.993 |
| *FOXA1* | Forkhead box A1 | Transcription factor | 1.0 (0.1–49.5) | 101 (36.6–213) | 0.0000019 | 0.993 |
| *SLC39A6* | Solute carrier family 39, member 6 | Unkown function | 1.0 (0.2–7.5) | 39.2 (6.3–311) | 0.0000019 | 0.993 |
| *RERG* | RAS-like, estrogen-regulated, growth inhibitor | Signal transduction | 1.0 (0.1–7.4) | 25.3 (1.8–146) | 0.0000026 | 0.986 |
| *PGR* | Progesterone receptor | Nuclear receptor | 1.0 (0–3.9) | 138 (2.3–1634) | 0.0000026 | 0.986 |
| *RABEP1* | Rabaptin, RAB GTPase binding effector protein 1 | Signal transduction | 1.0 (0.3–3.3) | 7.3 (1.7–26.7) | 0.0000034 | 0.981 |
| *NAT1* | *N*-acetyltransferase 1 | Metabolic enzyme | 1.0 (0.3–7.9) | 110 (3.9–368) | 0.0000036 | 0.979 |
| *ZMYND10*/*BLU* | Candidate tumor suppressor gene *BLU* | Transcription factor | 1.0 (0.1–5.4) | 26.2 (3.1–96.2) | 0.0000036 | 0.979 |
| *GATA3* | GATA binding protein 3 | Transcription factor | 1.0 (0.1–43.3) | 77.1 (18.7–212) | 0.0000043 | 0.976 |
| *XBP1* | X-box binding protein 1 | Transcription factor | 1.0 (0.1–15.0) | 22.7 (5.6–96.9) | 0.0000050 | 0.972 |
| *TFF1*/*PS2* | Trefoil factor 1 (pS2) | Secreted protein | 1.0 (0–23.0) | 1059 (6.6–19 498) | 0.0000050 | 0.972 |
| *ERBB4* | c-erbb-4 | Growth factor receptor | 1.0 (0.1–22.5) | 54.6 (9.3–1135) | 0.0000050 | 0.972 |
| *SEMA3B* | Semaphorin 3B | Secreted protein | 1.0 (0.3–14.1) | 26.4 (2.1–145) | 0.0000059 | 0.969 |
| *IGFBP4* | Insulin-like growth factor binding protein 4 | Growth factor | 1.0 (0.1–3.3) | 8.4 (1.9–41.6) | 0.0000069 | 0.965 |
| *DNAJC12* | DnaJ (Hsp40) homolog, subfamily C, member 12 | Unkown function | 1.0 (0.1–38.0) | 105 (6.1–1091) | 0.0000081 | 0.962 |
| *BDNF* | Brain-derived neurotrophic factor | Growth factor | 1.0 (0.2–3.3) | 6.9 (0.9–152) | 0.0000081 | 0.962 |
| *BCL2* | B-cell leukemia 2 oncogene | Apoptosis | 1.0 (0.2–5.7) | 9.4 (1.1–54.6) | 0.000011 | 0.955 |
| *RARRES3* | Retinoic acid receptor responder (tazarotene induced) 3 | Unkown function | 1.0 (0.1–5.4) | 10.5 (1.9–160) | 0.000015 | 0.948 |
| *HPN* | Hepsin (transmembrane protease, serine 1) | Unkown function | 1.0 (0–7.5) | 18.6 (0.4–54.1) | 0.000020 | 0.941 |
| *MYB* | v-myb myeloblastosis viral oncogene homolog (avian) | Transcription factor | 1.0 (0.1–6.6) | 9.4 (2.7–28.3) | 0.000020 | 0.941 |
| *TNFRSF10C* | TRAILR3 | Apoptosis | 1.0 (0.2–4.4) | 5.9 (0.7–29.8) | 0.000046 | 0.922 |
| *CLU* | Clusterin | Secreted protein | 1.0 (0.1–3.2) | 9.7 (0.7–51.2) | 0.000049 | 0.920 |
| *KRT18* | Keratin 18 | Cytoskeletal | 1.0 (0.2–21.7) | 13.8 (3.2–41.4) | 0.000057 | 0.917 |
| *BBC3*/*PUMA* | BCL2 binding component 3 (PUMA) | Apoptosis | 1.0 (0.4–2.6) | 3.1 (1.2–8.6) | 0.000065 | 0.913 |
| *PMAIP1*/*NOXA* | Phorbol-12-myristate-13-acetate-induced protein 1 | Apoptosis | 1.0 (0.3–4.9) | 3.9 (0.7–29.9) | 0.000065 | 0.913 |
| *LRBA* | LPS-responsive vesicle trafficking, beach and anchor containing | Metabolic enzyme | 1.0 (0.2–4.2) | 3.5 (1.1–9.7) | 0.000086 | 0.906 |
| *PTPRT* | Protein tyrosine phosphatase, receptor type, T | Growth factor receptor | 1.0 (0.2–53.0) | 201 (0.4–1015) | 0.000086 | 0.906 |
| *AR* | Androgen receptor | Nuclear receptor | 1.0 (0.1–115) | 102 (14.7–307) | 0.000086 | 0.906 |
| *TFF3* | Trefoil factor 3 (intestinal) | Unkown function | 1.0 (0.3–388) | 149 (5.6–5737) | 0.00020 | 0.885 |
| *SLC7A2* | Solute carrier family 7, member 2 | Metabolic enzyme | 1.0 (0.2–79.4) | 38.4 (0.9–821) | 0.00025 | 0.878 |
| *CDKN1B* | Cyclin-dependent kinase inhibitor 1B (p27, Kip1) | Cell cycle regulation | 1.0 (0.6–2.1) | 3.7 (0.5–25.6) | 0.00025 | 0.878 |
| *LOC255743* | Hypothetical protein LOC255743 | Unkown function | 1.0 (0.1–13.6) | 5.5 (0.6–19.9) | 0.00029 | 0.875 |
| *PAX3* | Paired box gene 3 (Waardenburg syndrome 1) | Transcription factor | 1.0 (0.3–15.1) | 14.2 (0.5–78.8) | 0.00048 | 0.861 |
| *DNTP73* | Tumor protein p73, isoform DeltaNp73 | Apoptosis | 1.0 (0–163) | 4.9 (0.7–56.2) | 0.00062 | 0.854 |
| *TIM14* | Homolog of yeast TIM14 | Unkown function | 1.0 (0.3–3.3) | 2.3 (0.9–6.8) | 0.0013 | 0.833 |
| *CCND1* | Cyclin D1 | Cell cycle regulation | 1.0 (0.1–11.2) | 4.3 (0.2–111) | 0.0016 | 0.826 |
| *SULT2B1* | Sulfotransferase family, cytosolic, 2B, member 1 | Metabolic enzyme | 1.0 (0.1–13.3) | 4.8 (0.1–52.2) | 0.0018 | 0.823 |

Table 1 *continued*

| Genes | Gene definition | Gene characterization | REα-negative ($n=12$) | REα-positive ($n=24$) | $P$[a] | ROC-AUC[b] |
|---|---|---|---|---|---|---|
| *IGFBP5* | Insulin-like growth factor binding protein 5 | Growth factor | 1.0 (0.2–31.0) | 4.9 (0.9–68.6) | 0.0035 | 0.802 |
| *TATP73* | Tumor protein p73, isoform Tp73 | Apoptosis | 1.0 (0.1–49.2) | 4.8 (0.8–26.7) | 0.0042 | 0.793 |
| *EMS-1* | Cortactin | Signal transduction | 1.0 (0.2–6.6) | 3.4 (0.1–16.8) | 0.0043 | 0.795 |
| *RET* | Ret proto-oncogene | Growth factor receptor | 1.0 (0.1–35.8) | 10.6 (0.2–180) | 0.0073 | 0.778 |
| *FGF2* | Fibroblast growth factor 2 (basic) | Growth factor | 1.0 (0.2–3.4) | 3.1 (0.4–51.6) | 0.014 | 0.755 |
| *VEGFR1* | fms-related tyrosine kinase 1 (VEGF receptor 1) | Angiogenesis | 1.0 (0.2–4.7) | 2.7 (1.0–4.7) | 0.015 | 0.750 |
| *EGF* | Epidermal growth factor | Growth factor | 1.0 (0–13.3) | 6.1 (0.9–48.5) | 0.017 | 0.745 |
| *CGA* | Glycoprotein hormones, alpha polypeptide | Growth factor | 1.0 (0–41.1) | 6.9 (0–847) | 0.018 | 0.744 |
| *GJA1* | Gap junction protein, alpha 1 (connexin 43) | Cell adhesion and cell junction | 1.0 (0.2–5.0) | 2.9 (0.5–60.4) | 0.019 | 0.743 |
| *FGFR2* | Fibroblast growth factor receptor 2 | Growth factor receptor | 1.0 (0.1–3.9) | 2.2 (0.3–75.1) | 0.029 | 0.726 |
| *ESR2/ERβ* | Estrogen receptor 2 (beta) | Nuclear receptor | 1.0 (0.1–3.8) | 0.6 (0.1–3.7) | NS | 0.502 |
| *ESR1/ERα* | Estrogen receptor 1 (alpha) | Nuclear receptor | 1.0 (0.2–5.1) | 711 (70.8–1938) | 0.0000014 | 1.000 |

[a]Mann and Whitney's *U*-test.
[b]ROC (receiver operating characteristics)–AUC (area under curve) analysis.
[c]Median (range) of gene mRNA levels.
NS, not significant.

### mRNA expression of the 51 upregulated genes in ERα-positive breast tumors, according to relapse

Twelve (50%) of the 24 patients with ERα-positive breast tumors relapsed. Comparison of the median mRNA levels of the 51 genes between patients, who relapsed ($n=12$) and those who did not relapse ($n=12$) identified only *NAT1* as having significantly different expression ($P=0.024$).

### mRNA expression of the 51 genes in five breast tumor cell lines

The expression level of the 51 genes upregulated in the ERα-positive breast tumors was then determined in five well-characterized breast tumor cell lines, including two ERα-positive cell lines (MCF7 and T-47D) and three ERα-negative cell lines (MDA-MB-231, MDA-MB-435 and SK-BR-3) (Table 2). Fourteen genes (*TFF1/PS2*, *PGR*, *FOXA1*, *GATA3*, *TATP73*, *TFF3*, *KRT18*, *CA12*, *ERBB4*, *TNFRSF10C/TRAILR3*, *SULT2B1*, *AR*, *STC2* and *CGA*) were upregulated (>3-fold the median value for the ERα-negative breast tumors) in both ERα-positive cell lines (MCF7 and T-47D). Seven genes (*SLC7A2*, *SEMA3B*, *RET*, *CLU*, *DNTP73*, *CCND1* and *NAT1*) were upregulated only in the ERα-positive cell line MCF7, and four other genes (*CYP2B6*, *RERG*, *BLU* and *EGF*) were upregulated only in the ERα-positive cell line T-47D. Surprisingly, 9 of these 25 putative ERα-responsive genes (*FOXA1*, *TFF3*, *KRT18*, *CA12*, *CGA*, *SEMA3B*, *CLU*, *CYP2B6* and *EGF*), were also upregulated in the ERα-negative cell line SK-BR-3. Likewise, 26 genes, whose expression was tightly linked to ERα-positivity of the breast tumor biopsies (Table 1) were not upregulated in any of the cell lines (*SLC39A6*, *p27/CDKN1B*, *LRBA*, *EMS-1*, *PTPRT*, *RABEP1*, *LOC255743*, *IL6ST*, *TIM14*, *HPN*, *BCL2*, *FGFR2*, *MYB*, *IGFBP4*, *IGFBP5*, *GJA1*, *VEGFR1*, and *RARRES3*) or were upregulated in the ERα-negative cell lines (*BDNF* and *NOXA* in MDA-231, *PAX23*, and *FGF2* in MDA-435, *PUMA* and XBP1 in SK-BR-3, and *DNAJC12* and *ACADSB* in both MDA-435 and SK-BR-3; Table 2).

### mRNA expression of the 51 genes in MCF7 cells treated with E2 (17β-estradiol)

Only 8 of the 51 genes (*PGR*, *TFF1/PS2*, *MYB*, *IGFBP4*, *RET*, *NOXA*, *SEMA3B* and *CA12*) were upregulated (>3-fold) in E2-treated MCF7 cells relative to untreated MCF7 cells (Table 3). Surprisingly, 9 genes were downregulated (>3-fold) by E2 treatment, namely *FOXA1*, *GATA3*, *SLC7A2*, *PUMA*, *CLU*, *ERBB4*, *LOC255743*, *PAX3* and *CGA*. It is also

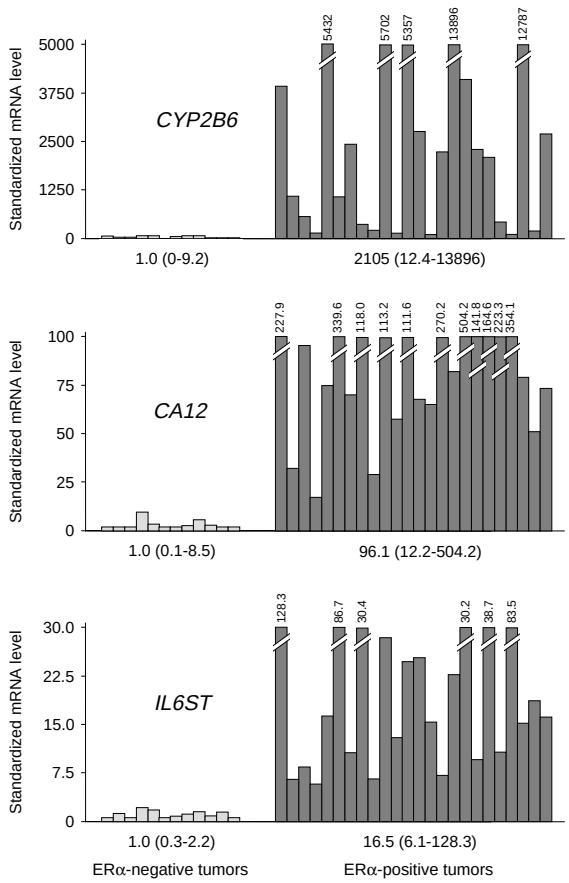


**Figure 1** mRNA levels of *CYP2B6*, *CA12* and *IL6ST* in 22 individual 12 ERα-negative breast tumors (gray bars) and 24 ERα-positive breast tumors (black bars). Median values (and ranges) are indicated for each tumor subgroup.

noteworthy that the *ERα* mRNA level was 2.3-fold lower in MCF7 cells treated with E2 than in untreated MCF7 cells, suggesting that E2 might act, via a negative feedback loop, on ER transcription. Finally, the ERβ mRNA level in MCF7 cells was not modified by E2 treatment.

## Discussion

We first used real-time quantitative RT-PCR to compare the mRNA expression of 560 selected genes in an ERα-positive breast tumor pool and an ERα-negative breast tumor pool (screening set). Thus the 56 genes of interest identified were then investigated in an independent well-characterized series of 24 individual ERα-positive breast tumors and in 12 ERα-negative breast tumors (validation set). Comparison of the pool values with the mean values of the individual samples showed that RNA pooling was an appropriate initial screening approach, significantly limiting the required number of PCR experiments.

Fifty-one (91.1%) of the 56 upregulated genes identified by the pooled sample analysis were significantly upregulated in the individual ERα-positive breast tumors relative to the ERα-negative breast tumors (Table 1). Using the same approach, we have previously shown the involvement of several altered molecular pathways in the genesis of prostate and liver cancer (Latil *et al.* 2003, Paradis *et al.* 2003).

Real-time quantitative RT-PCR is a promising complementary methodology to cDNA microarrays for molecular tumor profiling. In particular, real-time RT-PCR is far more precise, reproducible and quantitative than cDNA microarrays. Real-time RT-PCR is also more useful for analyzing weakly expressed genes, such as *CGA*, *BDNF*, *DNTP73*, *TATP73* and *NOXA* in the present study. Finally, real-time RT-PCR requires smaller amounts of total RNA (about 2 ng per target gene), and is therefore suitable for analyzing small or microdissected tumor samples.

We studied a number of genes involved in various cellular and molecular mechanisms that are associated with tumorigenesis and are known to be altered (mainly at the transcriptional level) in various cancers. These genes encode proteins involved in cell cycle control, cell–cell interactions, signal transduction pathways, apoptosis and angiogenesis, etc. (about 10–20 genes were selected per pathway; see list in Supplemental Data which can be viewed online at http://erc.endocrinology-journals.org/supplemental/). After scrutinizing the literature, we also included the well-known ERα-induced genes in breast cancer (*PGR*, *TFF1*/*PS2*, *BCL2*, *CCND1*) and a large number of genes that were found to co-cluster with ERα in microarray studies of breast tumor biopsies (Perou *et al.* 2000, Gruvberger *et al.* 2001, Sorlie *et al.* 2001, West *et al.* 2001, Bertucci *et al.* 2002, van't Veer *et al.* 2002). In consequence, it was not surprising in this present study to identify a large number of genes (51 of the 560 genes tested) significantly upregulated in ERα-positive breast tumors as compared with ERα-negative breast tumors.

This analysis was by no means exhaustive, and many possibly relevant genes were certainly missed, but it nevertheless demonstrates the ability of real-time RT-PCR to identify several potentially useful marker genes.

The first important result obtained in this study is that, in total agreement with Gruvberger *et al.* (2001), only a small proportion of the 51 genes that co-clustered with ERα status in our breast tumor series were confirmed *in vitro* to be ERα-regulated (i.e., upregulated in ERα-positive cell lines compared

**Table 2** mRNA expression of the 51 identified genes in five breast tumor cell lines

| | *REα*-positive cell lines | | *REα*-negative cell lines | | |
|---|---|---|---|---|---|
| **Genes** | MCF7 | T-47D | MDA-231 | MDA-435 | SK-BR-3 |
| *Genes upregulated in both the two REα-positive breast tumor cell lines: MCF7 and T-47D* | | | | | |
| *TFF1/PS2* | **3066**[a] | **4.9** | 0[b] | 0 | 2.8 |
| *PGR* | **75.3** | **399** | 0 | 0 | 0 |
| *FOXA1* | **27.2** | **12.2** | 0.11 | 0.05 | **9.67** |
| *GATA3* | **22.9** | **15.3** | 0.04 | 0.03 | 0.36 |
| *TATP73* | **18.9** | **7.20** | 1.21 | 0 | 0.03 |
| *TFF3* | **14.1** | **14.7** | 0.40 | 0.64 | **4.79** |
| *KRT18* | **12.4** | **5.93** | 0.76 | 1.12 | **12.9** |
| *CA12* | **10.6** | **24.2** | 0.82 | 0 | **16.3** |
| *ERBB4* | **10.6** | **8.11** | 0.08 | 0.19 | 0.74 |
| *TNFRSF10C/TRAILR3* | **9.79** | **2.28** | 0.26 | 1.39 | 0.66 |
| *SULT2B1* | **8.88** | **3.31** | 0 | 0 | 0.76 |
| *AR* | **7.96** | **6.67** | 0.11 | 1.17 | 1.53 |
| *STC2* | **7.34** | **7.37** | 0 | 0.39 | 0.03 |
| *CGA* | **3.61** | **5.53** | 0 | 0 | **322** |
| *Genes upregulated in only the REα-positive breast tumor cell line MCF7* | | | | | |
| *SLC7A2* | **88.8** | 0.11 | 1.02 | 2.59 | 0.25 |
| *SEMA3B* | **15.6** | 0.88 | 2.62 | 1.33 | **18.7** |
| *RET* | **15.2** | 1.21 | 0 | 0.04 | 0.05 |
| *CLU* | **4.67** | 1.17 | 0.15 | 0 | **11.2** |
| *DNTP73* | **3.98** | 1.30 | 0.31 | 0 | 0 |
| *CCND1* | **3.90** | 0.37 | 0.64 | 0.50 | 0.64 |
| *NAT1* | **3.60** | 1.29 | 0.69 | 1.01 | 2.80 |
| *Genes upregulated in only the REα-positive breast tumor cell line T-47D* | | | | | |
| *CYP2B6* | 1.40 | **355** | 0.23 | 1.05 | **14.0** |
| *RERG* | 0.73 | **10.0** | 0 | 0.07 | 0.36 |
| *ZMYND10/BLU* | 1.17 | **8.77** | 0.23 | 0.97 | 0.53 |
| *EGF* | 0.52 | **3.09** | 0.04 | 0.75 | **14.9** |
| *Genes upregulated in the REa-negative breast tumor cell lines: MDA-231. MDA-435 or SK-BR-3* | | | | | |
| *BDNF* | 2.39 | 0.42 | **42.8** | 1.87 | 1.21 |
| *PMAIP1/NOXA* | 1.67 | 0.36 | **3.39** | 0.59 | 0.74 |
| *PAX3* | 0.46 | 0.57 | 0.02 | **778** | 0.17 |
| *DNAJC12* | 2.37 | 0.14 | 0.08 | **6.53** | **6.40** |
| *ACADSB* | 0.76 | 0.95 | 0.29 | **5.88** | **3.15** |
| *FGF2* | 0 | 0 | 0 | **3.86** | 0 |
| *BBC3/PUMA* | 0.74 | 0.49 | 0.32 | 1.82 | **4.58** |
| *XBP1* | 0.85 | 1.35 | 0.02 | 0.07 | **3.60** |
| *Genes with normal expression in the five breast tumor cell lines* | | | | | |
| *SLC39A6* | 1.74 | 0.72 | 0.09 | 1.27 | 2.16 |
| *P27/CDKN1B* | 1.41 | 0.33 | 0.29 | 0.44 | 1.48 |
| *LRBA* | 0.89 | 0.53 | 0.11 | 0.45 | 1.27 |
| *EMS-1* | 0.85 | 0.98 | 0.37 | 4.22 | 0.65 |
| *PTPRT* | 0.68 | 0.40 | 0.15 | 0.38 | 0.45 |
| *RABEP1* | 0.65 | 1.19 | 0.65 | 1.54 | 1.17 |
| *LOC255743* | 0.65 | 0.62 | 0 | 0.01 | 0.19 |
| *IL6ST* | 0.43 | 2.49 | 1.06 | 2.31 | 1.09 |
| *TIM14* | 0.38 | 1.67 | 0.30 | 0.83 | 1.11 |
| *HPN* | 0.35 | 0.94 | 0 | 0 | 0.27 |
| *BCL2* | 0.18 | 0.25 | 0.40 | 1.97 | 0.02 |
| *FGFR2* | 0.14 | 0.44 | 0 | 0.01 | 0.61 |
| *MYB* | 0.10 | 1.31 | 0.01 | 0.30 | 0.31 |
| *IGFBP4* | 0.07 | 1.71 | 0.47 | 0 | 0 |
| *IGFBP5* | 0.04 | 0.44 | 0 | 0 | 0.45 |
| *GJA1* | 0 | 0.05 | 0.24 | 0.05 | 0.01 |
| *VEGFR1* | 0 | 0.18 | 0 | 0 | 0 |
| *RARRES3* | 0 | 0.17 | 0.01 | 0.01 | 0.68 |
| *Estrogen receptor genes* | | | | | |
| *ESR2/ERβ* | 0.61 | 0.15 | 0 | 0 | 2.40 |
| *ESR1/ERα* | **66.5** | **26.9** | 0 | 0.16 | 0.17 |

[a]bold type indicates the mRNA level values 3-fold higher than the median of the ERα-negative breast tumor values.
[b]'0' indicate very low levels of target gene mRNA that were not reliably quantifiable by means of real-time quantitative RT-PCR assays ($C_t > 35$).

**Table 3** mRNA expression of 51 identified genes in MCF7 cell line treated with E2

| Genes | MCF7 − E2 | MCF7 + E2 | *N*-fold variation |
|---|---|---|---|
| *Selected genes* | | | |
| *PGR* | 75.3 | 1995 | **+26.5**[a] |
| *TFF1/PS2* | 3066 | 37 406 | **+12.2** |
| *MYB* | 0.10 | 1.17 | **+11.7** |
| *IGFBP4* | 0.07 | 0.63 | **+9.0** |
| *RET* | 15.2 | 103 | **+6.8** |
| *PMAIP1/NOXA* | 1.67 | 10.1 | **+5.3** |
| *SEMA3B* | 15.6 | 57.7 | **+3.7** |
| *CA12* | 10.6 | 33.9 | **+3.2** |
| *TFF3* | 14.1 | 35.3 | +2.5 |
| *BCL2* | 0.18 | 0.37 | +2.1 |
| *IGFBP5* | 0.04 | 0.07 | +1.7 |
| *CCND1* | 3.90 | 6.10 | +1.6 |
| *STC2* | 7.34 | 8.81 | +1.2 |
| *TATP73* | 18.9 | 22.1 | +1.2 |
| *ZMYND10/BLU* | 1.17 | 1.30 | +1.1 |
| *DNAJC12* | 2.37 | 2.26 | +1 |
| *EMS-1* | 0.85 | 0.85 | +1 |
| *FGFR2* | 0.14 | 0.14 | +1 |
| *TIM14* | 0.38 | 0.38 | +1 |
| *XBP1* | 0.85 | 0.83 | +1 |
| *NAT1* | 3.60 | 3.17 | −1.1 |
| *RERG* | 0.73 | 0.61 | −1.2 |
| *TNFRSF10C/TRAILR3* | 9.79 | 8.00 | −1.2 |
| *AR* | 7.96 | 6.29 | −1.3 |
| *DNTP73* | 3.98 | 3.20 | −1.3 |
| *BDNF* | 2.39 | 1.70 | −1.4 |
| *EGF* | 0.52 | 0.37 | −1.4 |
| *RABEP1* | 0.65 | 0.46 | −1.4 |
| *ACADSB* | 0.76 | 0.51 | −1.5 |
| *CDKN1B* | 1.41 | 0.94 | −1.5 |
| *SLC39A6* | 1.74 | 1.19 | −1.5 |
| *SULT2B1* | 8.88 | 5.30 | −1.6 |
| *KRT18* | 12.4 | 7.32 | −1.7 |
| *IL6ST* | 0.43 | 0.23 | −1.8 |
| *PTPRT* | 0.68 | 0.38 | −1.8 |
| *CYP2B6* | 1.40 | 0.73 | −1.9 |
| *LRBA* | 0.89 | 0.42 | −2.1 |
| *HPN* | 0.35 | 0.16 | −2.2 |
| *FOXA1* | 27.2 | 8.42 | −3.2 |
| *GATA3* | 22.9 | 6.65 | −3.5 |
| *SLC7A2* | 88.8 | 21.4 | −4.1 |
| *BBC3/PUMA* | 0.74 | 0.15 | −4.9 |
| *CLU* | 4.67 | 0.75 | −6.2 |
| *ERBB4* | 10.6 | 1.30 | −8.1 |
| *LOC255743* | 0.65 | 0.07 | −9.3 |
| *PAX3* | 0.46 | 0.02 | −23.0 |
| *CGA* | 3.61 | 0.11 | −32.8 |
| *FGF2* | 0[b] | 0 | – |
| *RARRES3* | 0 | 0 | – |
| *VEGFR1* | 0 | 0 | – |
| *GJA1* | 0 | 0 | – |
| *Estrogen receptor genes* | | | |
| *ESR2/ERβ* | 0.61 | 0.40 | −1.5 |
| *ESR1/ERα* | 66.5 | 28.6 | −2.3 |

[a]bold type indicates the mRNA level increase (>3-fold higher) in the MCF7 treated with E2, relative to those in the MCF7 not treated.
[b]'0' indicates very low levels of target gene mRNA, that were not reliably quantifiable by means of real-time quantitative RT-PCR assays ($C_t$>35).

with ERα-negative cell lines) and/or E2-regulated (i.e., regulated by E2 in MCF7 cells). There are several possible explanations for these findings. (a) The existence of other ER-signaling pathways, independent of estrogen has been postulated and (Zwijsen *et al.* 1998, Ding *et al.* 2003). For example, Sabbah *et al.* (1999) described a mechanism by which ERα regulates *CCND1* gene transcription through a cyclic AMP response element (CRE); (b) Expression of genes in ERα-positive breast tumors can also reflect the presence of different types of epithelial cells in the mammary gland, independently of the presence of estrogen and ERα. In this regard, ERα-positive breast tumors have been suggested to exhibit the phenotype of luminal epithelial cells, whereas ERα-negative tumors resemble myoepithelial (basal) cells (Perou *et al.* 2000); (c) Downregulation of genes in ERα-negative tumors may also simply reflect dedifferentiation of epithelial cells during malignant progression of ERα-negative breast tumors evolving from ERα-positive precursors; (d) Finally, cultured cell lines (*in vitro* models) have lost many features that characterize tumor specimens *in vivo* (Welsh *et al.* 2001, Dangles *et al.* 2002). The mechanism that leads to *in vivo* gene overexpression in ERα-positive breast tumors involves several factors, including ERα and several known or unknown transcriptional coactivators, not all of, which present in classical *in vitro* models. We were particularly surprised to identify genes that were tightly linked to ERα-positive status in breast tumor biopsies but were downregulated in MCF7 cells after E2 treatment (Table 1). It is also noteworthy that we cannot rule out the possibility that we identified some genes by chance, which can happen when large numbers of variables (gene expressions) are analyzed, in particular the genes showing a weak link to the ERα status.

Our results provide further evidence that gene expression databases based on breast tumor cell lines, used to identify new ERα status markers or new candidate markers of the response to endocrine therapy, must be carefully interpreted (Soulez & Parker 2001, Ngwenya & Safe 2003, Frasor *et al.* 2003, Vendrell *et al.* 2004).

A large proportion of the 51 genes identified in this study have previously been reported to be related to ERα status. *PGR*, *TFF1/PS2*, *BCL2*, *ERBB4* and *CCND1* are well-known ERα-induced genes in breast cancer. Several new genes, such as *GATA3*, *TFF3*, *MYB*, *IGFBP4*, *IGFBP5*, *STC2*, *KRT18*, *HPN/HEPSIN*, *FOXA1*, *XBP1*, *SLC39A6/LIV-1* and *CA12* M, were recently identified by microarray studies (Gruvberger *et al.* 2001, Bertucci *et al.* 2002, van't Veer *et al.* 2002). For our part, we have previously identified *CGA*, *NAT1*

and *CYP2B6* as candidate ERα-responsive genes in human breast cancer (Bieche *et al.* 2001*b*, 2004*b*).

In addition, to known ERα-induced genes, we identified an appreciable number of novel genes, and particularly weakly expressed genes, validating our large-scale real-time RT-PCR approach as a method complementary to cDNA microarrays for molecular tumor profiling. These new genes mainly encode secreted proteins and growth factors (*BDNF*, *FGF2*, *EGF*, *SEMA3B* and *CLU*), growth factor receptors (*IL6ST*, *PTPRT*, *RET*, *VEGFR1* and *FGFR2*) and metabolic enzymes (*CYP2B6*, *CA12*, *ACADSB*, *NAT1*, *LRBA*, *SLC7A2* and *SULT2B1*). Interestingly, in addition to *BCL2*, we identified a large number of genes encoding proteins involved in apoptosis (*TNFRSF10C*/*TRAILR3*, *PUMA*, *NOXA*, *DNTP73* and *TATP73*).

*DNTP73* and *TATP73*, produced by alternative splicing of the same gene (*TP73*), are expressed under the control of two independent promoters and have opposite activities. TAp73 is the transcriptionally active full-length protein, while ΔNp73 is the amino-terminally truncated dominant-negative protein (Melino *et al.* 2002). Unlike *TP53*, the genes *DNTP73* and *TATP73* are mainly regulated at the transcriptional level. TAp73 induces cell-cycle arrest and apoptosis, whereas ΔNp73 inhibits both TAp73-induced and p53-induced apoptosis. Furthermore, ΔNp73 is induced by TAp73 and p53, in a dominant-negative feedback loop that regulates p53 and p73 functions (Melino *et al.* 2002). NOXA and PUMA are recently identified BH3-only Bcl-2 family proteins, and are key mediators of p53-mediated apoptosis (Fridman & Lowe 2003). PUMA was shown to be downregulated by estradiol and to be associated with OH-Tam resistance in MCF-7-derived cell lines (Vendrell *et al.* 2005). Finally, *TRAILR3* encode a TNF-related apoptosis-inducing-ligand receptor that acts as a decoy receptor for TRAIL, a member of the tumor necrosis factor family (Ashkenazi 2002). In several cell types, decoy receptors inhibit TRAIL-induced apoptosis by binding TRAIL and thereby preventing its binding to pro-apoptotic TRAIL receptors. Surprisingly, we observed upregulation of both pro-apoptotic genes (*PUMA*, *NOXA* and *TATP73*) and anti-apoptotic genes (*BCL2*, *DNTP73* and *TRAILR3*) in the ERα-positive tumors. Further, studies are needed to determine the respective roles of these apoptotic genes in ERα-positive tumorigenesis.

Identification of genes that co-cluster with ERα status is a first step towards identifying reliable markers with which to predict ERα status or the response to endocrine therapy. In addition, to *CYP2B6* and *CA12* that are already known to be ERα-related in breast cancer (Gruvberger *et al.* 2001, Bieche *et al.* 2004*b*), we identified a third gene (*IL6ST*) that perfectly predicted ERα status in our breast tumor series (AUC-ROC, 1.000). *IL6ST* encodes gp130, the subunit shared by the different receptors of IL-6 family cytokines, including interleukin-6, interleukin-11, leukemia inhibitory factor, oncostatin M, ciliary neurotrophic factor, and cardiotrophin-1 (Kishimoto *et al.* 1994). Interestingly, Grant *et al.* (2002) have reported a functional interaction between gp130 and the EGF receptor family in breast cancer cells. However, while these three genes are potentially valuable predictive markers of ERα status, they would be less useful for predicting the response to endocrine therapy, being too strongly linked to ERα. About one-half of all patients with ERα-positive breast tumors fail to respond favorably to antiestrogen treatment, and thus there is a need for new molecular markers with which to identify them. This study identifies new candidate markers of endocrine responsiveness because they are upregulated in only a subgroup of ERα-positive tumors (for example, the genes with AUC-ROC $<0.900$ in Table 1). Interesting, some of these genes (*IGFBP5*, *FGF2*, *CGA*, etc.) encode secreted proteins that could serve as serum-based predictive biomarkers. We tested the 51 genes as candidate prognostic molecular markers in our small series of 24 postmenopausal ERα-positive breast cancer patients, who were treated with primary surgery, followed by adjuvant tamoxifen alone, and 12 of them relapsed. The only gene showing significantly different expression ($P=0.024$) between patients, who relapsed ($n=12$) and those, who did not relapse ($n=12$) was *NAT1*. It is noteworthy that, in a previous study of 125 ERα-positive postmenopausal breast cancer patients, we identified *NAT1* and *CGA* (also identified in the present study) as independent predictors of the response to tamoxifen (Bieche *et al.* 2001*b*, 2004*b*).

Some results of this study – identification of new ERα-induced genes, the three genes (*CYP2B6*, *CA12*, *IL6ST*) that highly predicted ERα status and new candidate markers of endocrine responsiveness must now be confirmed in larger series of breast tumors.

In conclusion, by using a large-scale real-time quantitative RT-PCR approach, we identified 51 genes that co-cluster with ERα status. Many of these genes were identified for the first time as being linked to ERα status and several are involved in apoptosis (*TNFRSF10C*/*TRAILR3*, *PUMA*, *NOXA*, *DNTP73* and *TATP73*). These 51 genes should help to delineate the estrogen receptor pathway and function, and some of the genes may prove useful for developing diagnostic tests or new markers of responsiveness to the different

available strategies of endocrine therapy (aromatase inhibitor, tamoxifen or pure antiestrogen).

## Acknowledgements

We thank the staff of Centre René Huguenin for assistance in specimen collection and patient care. JV was supported by a scholarship from the Ligue Nationale contre le Cancer (France). This work was supported by the Comité Régional des Hauts-de-Seine de la Ligue Nationale Contre le Cancer. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

## References

Ashkenazi A 2000 Targeting death and decoy receptors of the tumour-necrosis factor superfamily. *Nature Reviews Cancer* **2** 420–430.

Bertucci F, Nasser V, Granjeaud S, Eisinger F, Adelaïde J, Tagett R, Loriod B, Giaconia A, Benziane A, Devilard E *et al.* 2002 Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. *Human Molecular Genetics* **11** 863–872.

Bieche I, Onody P, Laurendeau I, Olivi M, Vidaud D, Lidereau R & Vidaud M 1999 Real-time reverse transcription-PCR assay for future management of ERBB2-based clinical applications. *Clinical Chemistry* **45** 1148–1156.

Bieche I, Parfait B, Le Doussal V, Olivi M, Rio MC, Lidereau R & Vidaud M 2001*a* Identification of CGA as a novel estrogen receptor-responsive gene in breast cancer: an outstanding candidate marker to predict the response to endocrine therapy. *Cancer Research* **61** 1652–1658.

Bieche I, Parfait B, Nogues C, Andrieu C, Vidaud D, Spyratos F, Lidereau R & Vidaud M 2001*b* The *CGA* gene as new predictor of the response to endocrine therapy in ERalpha-positive postmenopausal breast cancer patients. *Oncogene* **20** 6955–6959.

Bieche I, Parfait B, Laurendeau I, Girault I, Girault M & Lidereau R 2001*c* Quantification of estrogen receptor alpha and beta expression in sporadic breast cancer. *Oncogene* **20** 8109–8115.

Bieche I, Lerebours F, Tozlu S, Espie M, Marty M & Lidereau R 2004*a* Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. *Clinical Cancer Research* **10** 6789–6795.

Bieche I, Girault I, Urbain E, Tozlu S & Lidereau R 2004*b* Relationship between intratumoral expression of genes coding for xenobiotic-metabolizing enzymes and benefit from adjuvant tamoxifen in estrogen receptor alpha-positive postmenopausal breast carcinoma. *Breast Cancer Research* **6** R252–R263.

Dangles V, Lazar V, Validire P, Richon S, Wertheimer M, Laville V, Janneau JL, Barrois M, Bovin C, Poynard T *et al.* 2002 Gene expression profiles of bladder cancers: evidence for a striking effect of *in vitro* cell models on gene patterns. *British Journal of Cancer* **86** 1283–1289.

DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA & Trent JM 1996 Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics* **14** 457–460.

Ding L, Yan J, Zhu J, Zhong H, Lu Q, Wang Z, Huang C & Ye Q 2003 Ligand-independent activation of estrogen receptor alpha by XBP-1. *Nucleic Acids Research* **31** 5266–5274.

Frasor J, Danes JM, Komm B, Chang KCN, 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.

Fridman JS & Lowe SW 2003 Control of apoptosis by p53. *Oncogene* **22** 9030–9040.

Grant SL, Hammacher A, Douglas AM, Goss GA, Mansfield RK, Heath JK & Begley CG 2002 An unexpected biochemical and functional interaction between gp130 and the EGF receptor family in breast cancer cells. *Oncogene* **21** 460–474.

Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Fernö M, Peterson C & Meltzer PS 2001 Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Research* **61** 5979–5984.

Hanley J & McNeil B 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143** 29–36.

Kishimoto T, Taga T & Akira S 1994 Cytokine signal transduction. *Cell* **76** 253–262.

Latil A, Bieche I, Chene L, Laurendeau I, Berthon P, Cussenot O & Vidaud M 2003 A four gene expression model predicts recurrence in patients with clinically localized prostate cancer. *Clinical Cancer Research* **9** 5477–5485.

Mann H & Whitney D 1947 On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18** 50–60.

McDonnell DP & Norris JD 2002 Connections and regulation of the human estrogen receptor. *Science* **296** 1642–1644.

Melino G, Laurenzi V & Vousden KH 2002 p73: Friend or foe in tumorigenesis. *Nature Reviews Cancer* **2** 605–615.

Ngwenya S & Safe S 2003 Cell context-dependent differences in the induction of *E2F-1* gene expression by 17 Beta-estradiol in MCF-7 and ZR-75 cells. *Endocrinology* **144** 1675–1685.

Paradis V, Bieche I, Dargere D, Laurendeau I, Bioulac-Sage P, Degott C, Belghiti J, Vidaud M & Bedossa P 2003 Molecular profiling of hepatocellular carcinomas (HCC) using a large scale real-time RT-PCR approach: determination of a molecular diagnostic index. *American Journal of Pathology* **163** 733–741.

Perou CM, Sorlie T, Eisen MB, Van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al.* 2000 Molecular portraits of human breast tumours. *Nature* **406** 747–752.

Pike M, Spicer D, Dahmoush L & Press M 1993 Estrogens, progesterones, normal breast cell proliferation, and breast cancer risk. *Epidemiologic Reviews* **15** 17–35.

Prall O, Sarcevic B, Musgrove E, Watts C & Sutherland R 1997 Estrogen-induced activation of Cdk4 and Cdk2 during G1-S phase progression is accompanied by increased cyclin D1 expression and decreased cyclin-dependent kinase inhibitor association with cyclin E-Cdk2. *Journal of Biological Chemistry* **272** 10882–10894.

Sabbah M, Courilleau D, Mester J & Redeuilh G 1999 Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element. *PNAS* **96** 11217–11222.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Van De Rijn M, Jeffrey SS *et al.* 2001 Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* **98** 10869–10874.

Soulez M & Parker MG 2001 Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. *Journal of Molecular Endocrinology* **27** 259–274.

van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT *et al.* 2002 Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415** 530–536.

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.

Vendrell JA, Bieche I, Desmetz C, Badia E, Tozlu S, Nguyen C, Nicolas JC, Lidereau R & Cohen PA 2005 Molecular changes associated with the agonist activity of hydroxytamoxifen and the hyper-response to estradiol in hydroxytamoxifen-resistant breast cancer cell lines. *Endocrine-Related Cancer* **12** 75–92.

Welsh JB, Sapinoso LM, Su AI, Kern SG, Wang-Rodriguez J, Moskaluk CA, Frierson HF & Hampton GM 2001 Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Research* **61** 5974–5978.

West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA, Marks JR & Nevins JR 2001 Predicting the clinical status of human breast cancer by using gene expression profiles. *PNAS* **98** 11462–11467.

Zwijsen RM, Buckle RS, Hijmans EM, Loomans CJ & Bernards R 1998 Ligand-independent recruitment of steroid receptor coactivators to estrogen receptor by cyclin D1. *Genes & Development* **12** 3488–3498.