

Methodology article

Open Access

## Evaluation and validation of candidate endogenous control genes for real-time quantitative PCR studies of breast cancer

Roisin E McNeill\*, Nicola Miller and Michael J Kerin

Address: Department of Surgery, Clinical Science Institute, National University of Ireland, Galway, Ireland

Email: Roisin E McNeill\* - roisin.mcneill@nuigalway.ie; Nicola Miller - nicola.miller@nuigalway.ie;

Michael J Kerin - michael.kerin@nuigalway.ie

\* Corresponding author

Published: 27 November 2007

Received: 15 June 2007

BMC Molecular Biology 2007, 8:107 doi:10.1186/1471-2199-8-107

Accepted: 27 November 2007

This article is available from: <http://www.biomedcentral.com/1471-2199/8/107>

© 2007 McNeill et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Real-time quantitative PCR (RQ-PCR) forms the basis of many breast cancer biomarker studies and novel prognostic assays, paving the way towards personalised cancer treatments. Normalisation of relative RQ-PCR data is required to control for non-biological variation introduced during sample preparation. Endogenous control (EC) genes, used in this context, should ideally be expressed constitutively and uniformly across treatments in all test samples. Despite widespread recognition that the accuracy of the normalised data is largely dependent on the reliability of the EC, there are no reports of the systematic validation of genes commonly used for this purpose in the analysis of gene expression by RQ-PCR in primary breast cancer tissues. The aim of this study was to identify the most suitable endogenous control genes for RQ-PCR analysis of primary breast tissue from a panel of eleven candidates in current use. Oestrogen receptor alpha (*ESR1*) was used a target gene to compare the effect of choice of EC on the estimate of gene quantity.

**Results:** The expression and validity of candidate ECs (*GAPDH*, *TFRC*, *ABL*, *PPIA*, *HPRT1*, *RPLP0*, *B2M*, *GUSB*, *MRPL19*, *PUM1* and *PSMC4*) was determined in 6 benign and 21 malignant primary breast cancer tissues. Gene expression data was analysed using two different statistical models. *MRPL19* and *PPIA* were identified as the most stable and reliable EC genes, while *GUSB*, *RPLP0* and *ABL* were least stable. There was a highly significant difference in variance between ECs. *ESR1* expression was appreciably higher in malignant compared to benign tissues and there was a significant effect of EC on the magnitude of the error associated with the relative quantity of *ESR1*.

**Conclusion:** We have validated two endogenous control genes, *MRPL19* and *PPIA*, for RQ-PCR analysis of gene expression in primary breast tissue. Of the genes in current use in this field, the above combination offers increased accuracy and resolution in the quantitation of gene expression data, facilitating the detection of smaller changes in gene expression than otherwise possible. The combination identified here is a good candidate for use as a two-gene endogenous control in a broad spectrum of future research and diagnostic applications in breast cancer.

## Background

Breast cancer is the most common form of malignancy among women in almost all of Europe and in North America. Each year over one million women worldwide are diagnosed with the disease and it causes over 400,000 deaths annually. Much of the current translational research in this area is based on "transcriptomics", the elucidation of the transcriptional programs underlying disease initiation, promotion and progression, through tumour gene expression profiling. Real-time quantitative PCR (RQ-PCR) [1,2] is one of the most sensitive and specific quantitation methods for gene expression analysis and is firmly established as a mainstream research tool [3-5]. With the development of high throughput and reliable instrumentation, improved detection chemistries, more efficient protocols and appropriate analysis software, RQ-PCR has become the basis of many breast cancer biomarker studies as well as several novel diagnostic and prognostic assays [6-12]. In addition, RQ-PCR is used to validate microarray expression profiles and quantify genes of interest identified from those analyses.

The most commonly used method to quantify gene expression involves the analysis of target gene expression relative to a control gene. As in other relative gene expression analysis techniques such as Northern blotting and ribonuclease protection assays (RPAs), normalisation of RQ-PCR data is required to control for variation introduced during the steps from RNA extraction to quantitation, especially to control for differences in the quantity and quality of RNA used in reactions [13]. The use of endogenous control (EC) genes, known variously as housekeepers, reference or simply control genes is based on the principle that these genes are expressed constitutively and uniformly in all test samples, so that expression of the target gene can be normalised against them to control for systematic variation in sample handling. Results are then expressed as the ratio of target gene expression relative to the EC gene. In many cases, control genes are inherited from earlier studies using less sensitive forerunner techniques such as Northern blotting and little if any consideration has been paid to validating these genes as controls for specific experiments.

Two of the most commonly used endogenous control genes for breast cancer gene expression studies are glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and  $\beta$ -actin (*ACTB*) but their reliability in this context has not been demonstrated. In some studies the use of either gene may be inappropriate, as these genes have been implicated in aspects of disease aetiology [14-20]. For example, bisphosphonates; used to inhibit bone resorption in diseases including osteoporosis, Paget's disease and metastatic breast cancer, target *GAPDH*, decreasing its expression in both breast and prostate cell lines [21],

while the actin filament protein family, of which *ACTB* is a member, may be modulated in malignancy [22] particularly during processes involving reorganisation of the cytoskeleton such as invasion and migration.

The precision of the estimate of change in target gene expression is dependent on the stability of the endogenous control, the variability associated with the target gene and any covariance between the two. Thus the use of non-validated endogenous control genes results in, at best, unreliable data. It is now recognised that a universal, invariably expressed gene is unlikely to exist [23] and may not exist even within individual tissue or cell types. The goal therefore is to identify the most reliable gene or set of genes as endogenous controls for a particular experiment. As a result of this, several groups have developed statistical models and software programs for the analysis of candidate gene stability. The aim of this study was to identify the most stable endogenous control genes from a panel of eleven candidates commonly used as endogenous controls in the context of, but not limited to, breast cancer: *GAPDH*, *TFRC*, *ABL*, *PPIA*, *HPRT1*, *RPLP0*, *B2M*, *GUSB*, *MRPL19*, *PUM1*, *PSMC4*, for the quantification of gene expression by relative comparative RQ-PCR in primary breast cancer tissues. The oestrogen receptor alpha (*ESR1*) transcript, a gene of special significance in breast cancer, was used as a target gene to compare the effect of choice of EC on the estimate of gene quantity.

## Results

To identify suitable EC genes for breast cancer gene expression studies in fresh-frozen primary tissue, a panel of 11 genes commonly used as ECs was selected from the literature for analysis of stability: *GAPDH*, *TFRC*, *ABL*, *PPIA*, *HPRT1*, *RPLP0*, *B2M*, *GUSB*, *MRPL19*, *PUM1* and *PSMC4*. Genes were analysed in tumours recovered from patients with benign or malignant breast disease using RQ-PCR. Stability of candidate EC genes was analysed using two statistical analysis tools, geNorm and NormFinder, which employ different statistical models to define the most reliable EC genes for normalisation. The effect of choice of EC gene on target gene expression was analysed using *ESR1* as target.

### Range of expression of candidate EC genes and *ESR1*

The candidate ECs displayed a range of  $C_t$  values. Mean  $C_t$  values per gene and the range of  $C_t$  values for each gene are shown in Table 1. Mean  $C_t$  values ranged from 19.13 ( $\pm$  0.21 s.e.m.) for *B2M* to 26.48 ( $\pm$  0.15 s.e.m.) for *MRPL19*. *MRPL19* showed the narrowest range followed by *PPIA*. The genes broadly fell into two categories, those highly expressed with mean  $C_t$  values of 19–20 (*B2M*, *RPLP0*, *GAPDH*, *PPIA*) and moderate abundance genes with mean  $C_t$  values of 23–26 (*PSMC4*, *ABL*, *GUSB*, *TFRC*, *PUM1*, *HPRT1* and *MRPL19*). The target gene *ESR1*

**Table 1: Cycle threshold ( $C_t$ ) values of candidate EC genes and *ESRI*. Among the candidate ECs, *MRPL19* and *PPIA* showed the smallest range in  $C_t$  values while *GAPDH*, *HPRT1*, and *RPLP0* showed the greatest. Candidates fell into two groups in terms of abundance, high ( $C_t$  19–20; *B2M*, *RPLP0*, *GAPDH* and *PPIA*) and moderate abundance ( $C_t$  23–26; *PSMC4*, *ABL*, *GUSB*, *TFRC*, *PUM1*, *HPRT1* and *MRPL19*). *ESRI* ranged over 10.35  $C_t$  values**

Gene symbol	$C_t$ Range	$C_t$ Min.	$C_t$ Max.	Mean $C_t \pm$ s.e.m.
<i>MRPL19</i>	4.06	24.79	28.84	26.48 $\pm$ 0.15
<i>PPIA</i>	4.4	18.01	22.41	19.74 $\pm$ 0.18
<i>B2M</i>	4.67	17.22	21.89	19.13 $\pm$ 0.21
<i>PUM1</i>	5.06	22.38	27.44	24.78 $\pm$ 0.21
<i>GUS</i>	5.19	21.37	26.55	23.94 $\pm$ 0.20
<i>PSMC4</i>	5.69	20.84	26.52	23.32 $\pm$ 0.22
<i>ABL</i>	5.89	21.27	27.15	23.75 $\pm$ 0.30
<i>TFRC</i>	6.03	21.4	27.42	24.40 $\pm$ 0.29
<i>GAPDH</i>	6.88	16.79	23.67	19.71 $\pm$ 0.25
<i>HPRT1</i>	7.05	21.57	28.61	24.84 $\pm$ 0.27
<i>RPLP0</i>	8.8	16.1	24.9	19.46 $\pm$ 0.33
<i>ESRI</i>	10.35	17.10	27.45	21.96 $\pm$ 0.34

showed the broadest range of  $C_t$  values (10.35) from 17.10–27.45.

With conversion of  $C_t$  values to relative quantity values ( $Q_{Rel.}$ ), there was no difference in candidate EC gene quantities between benign and malignant tissues ( $P > 0.05$ ; Fig. 1a). There was however, a significant difference in variance between genes ( $P = 0.001$ ; Fig. 1b) indicative of differing stabilities of the candidates.

#### Analysis of EC gene stability

The significant difference in EC variability demonstrated the necessity to validate their use in this context. Expression stability was analysed using the two softwares geNorm [24,25] and NormFinder [26].

GeNorm uses a pair-wise comparison-based model to select from a panel of candidate EC genes, the gene-pair showing least variation in expression ratio across samples. It calculates a measure of gene stability ( $M$ ) of each gene based on the average pairwise variation between all tested genes. Genes with the lowest  $M$  values are those demonstrating most stable expression. Table 2 shows the  $M$  values for all tested genes. Eight of the eleven genes analysed; *PPIA*, *MRPL19*, *GAPDH*, *PUM1*, *B2M*, *HPRT1*, *PSMC4* and *TFRC*, showed  $M$  values less than the geNorm default threshold of 1.5, while the three remaining genes; *RPLP0*, *GUSB* and *ABL*, showed  $M$  values greater than that threshold. In a stepwise progression, geNorm excludes the least stable gene, recalculating  $M$  for the remaining genes, resulting in the characterisation of the stability of each gene on a ranked scale and ultimately the identification of the two most stably expressed genes (Fig. 2a). As shown, *ABL* and *GUSB* were the first and second genes respectively excluded from the analysis on the basis of

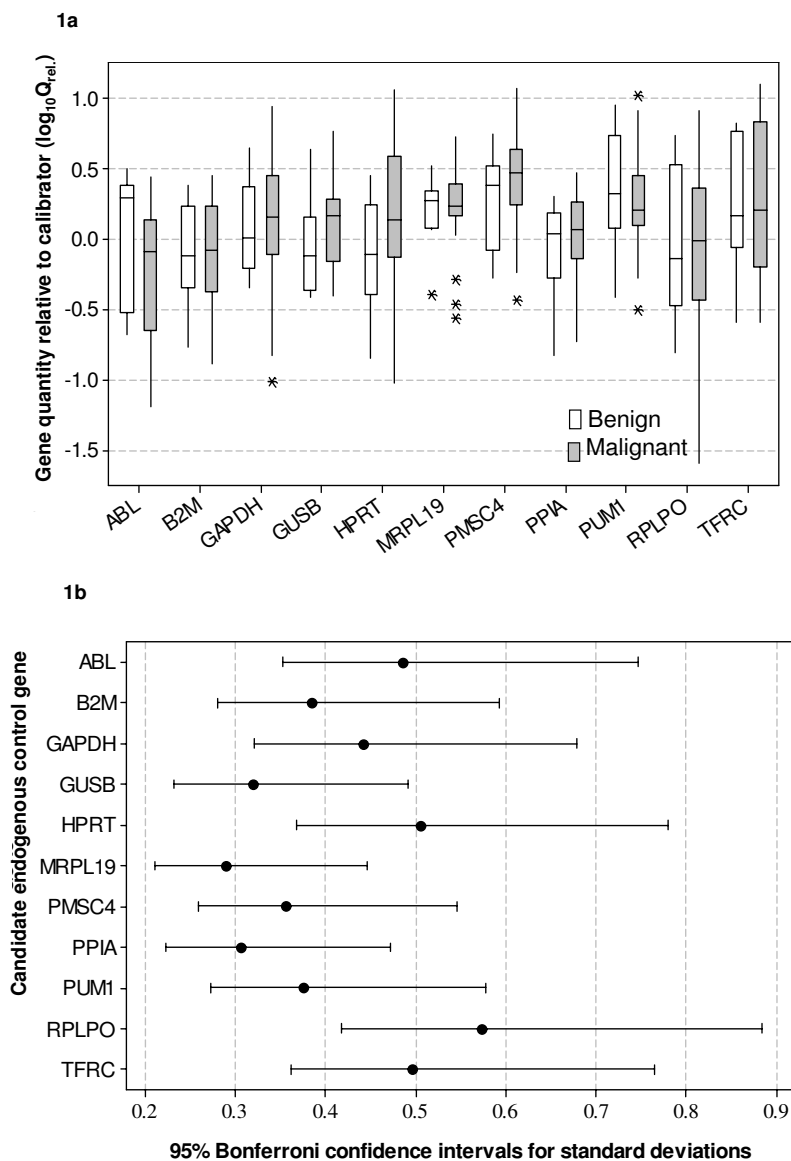
instability and *MRPL19* and *PPIA* were identified as most stable gene-pair.

GeNorm also calculates a normalisation factor (NF) required to determine the optimal number of EC genes required for accurate normalisation. This factor is calculated using the variable  $V$  as the pairwise variation ( $V_n/V_{n+1}$ ) between two sequential NFs ( $NF_n$  and  $NF_{n+1}$ ). To meet the recommended cut off  $V$ -value of 0.15, the point at which it is unnecessary to include additional genes in a normalisation strategy [24], the programme indicated the use of 5 of the six most stable genes *i.e.*, *MRPL19*, *PPIA*, *GAPDH*, *PUM1* and *B2M* (Fig. 2b). However, there was no significant effect on relative quantity of *ESRI* expression using the 5 gene panel of *MRPL19*, *PPIA*, *B2M*, *GAPDH* and *PUM1*, compared to the two-gene combination of *MRPL19* and *PPIA* ( $P > 0.05$ ).

Stability of gene expression was also analysed using NormFinder [26]. This programme uses a combined estimate of the intra- and inter-group variation to determine the most stably expressed candidate EC gene and gene-pair. Table 2 shows the ranking of the candidates. As for geNorm, NormFinder identified *MRPL19* and *PPIA* as the most stable pair of genes and *MRPL19* as the single most stable gene.

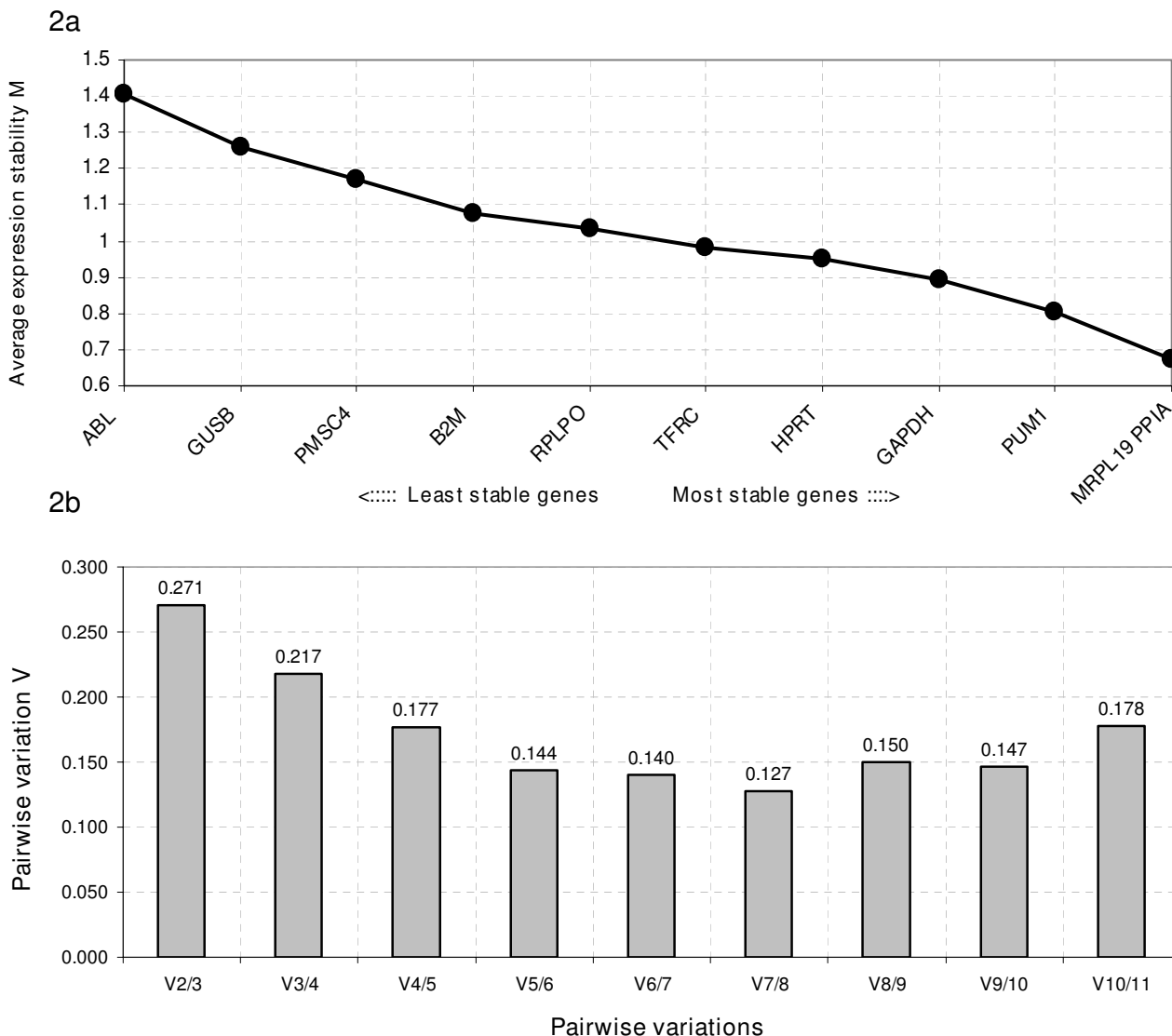
#### Associations between candidates EC genes and *ESRI*

The geNorm programme assumes no co-regulation of candidate ECs as obviously this would lead to an erroneous choice of optimum normaliser pair. As stated above, to our knowledge the candidate ECs tested in this study are functionally independent. In addition co-variance between target gene and EC would affect results. Regression analysis demonstrated significant, negative, linear



**Figure 1**

**Relative quantity and variation associated with each candidate EC.** a) Quantity of candidate endogenous control genes *GAPDH*, *TFRC*, *ABL*, *PPIA*, *HPRTI*, *RPLPO*, *B2M*, *GUSB*, *MRPL19*, *PUM1* and *PSMC4* in benign and malignant breast tumour tissue relative to calibrator ( $Q_{rel} = E^{-\Delta C_t}$ ). Boxplot shows median value, interquartile-range box and outliers (\*). Within gene there was no difference in gene quantities between benign and malignant tissues ( $P > 0.05$ ). b) Variation associated with candidate endogenous control genes *GAPDH*, *TFRC*, *ABL*, *PPIA*, *HPRTI*, *RPLPO*, *B2M*, *GUSB*, *MRPL19*, *PUM1* and *PSMC4* in all breast tumours relative to calibrator. Relative gene expression was calculated using the  $\Delta C_t$  method and corrected for efficiency of amplification ( $Q_{rel} = E^{-\Delta C_t}$ ). There was a significant difference in variance associated with relative gene expression ( $P = 0.001$ ) with genes such as *RPLPO*, *TRFC*, *HPRTI* and *GAPDH* showing greater variance than genes such as *MRPL19* and *PPIA*.



**Figure 2**  
**GeNorm analysis of the candidate EC genes.** Results are presented as per the output file of the geNorm programme [24]. (a) Stepwise exclusion of the least stable genes. The gene stability value *M* is based on the average pairwise variation between all tested genes. Low *M* values characterise genes with greater stability, thus the x-axis from left to right indicates the ranking of the EC genes according to expression stability and the y-axis indicates the stability measure, *M*. (b) Determination of the optimal number of ECs for normalisation. The recommended upper limit of the pairwise variation value *V* is set at 0.15 but in meeting this criterion, sample availability, the practicality of using multiple EC genes and the degree of required resolution must be considered.

associations between the relative quantities ( $Q_{Rel.}$ ) of the target gene, *ESR1* and two commonly used ECs; *TFRC* ( $TFRC = 0.079 - 0.272 ESR1$ ;  $R^2 = 0.18$ ;  $P < 0.05$ ) and *HPRT1* ( $HPRT1 = - 0.202 - 0.386 ESR1$ ;  $R^2 = 0.35$ ;  $P = 0.001$ ). In addition there was a significant linear association between  $Q_{Rel.}$  of *ESR1* and *ABL* described by the equation:  $ABL = 0.044 + 0.290 ESR1$ ;  $R^2 = 0.21$ ;  $P = 0.012$ .

There was no association between *ESR1* and either *MRPL19* or *PPIA* ( $P > 0.05$ ).

**Effect of EC on *ESR1* relative gene expression**

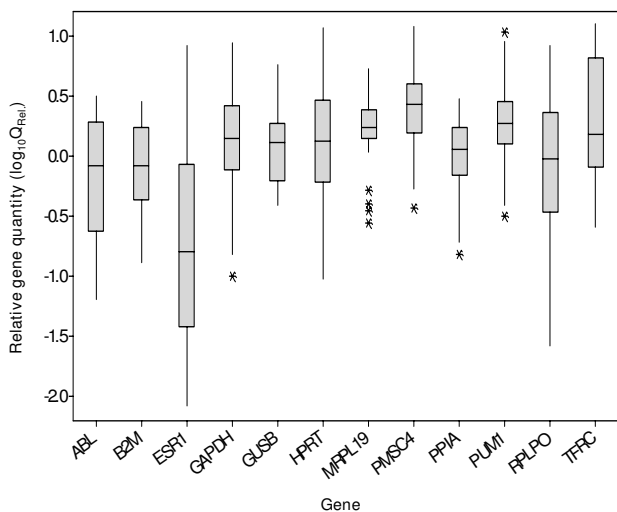
There was no effect of choice of EC on the relative quantity of *ESR1* ( $P > 0.05$ ) probably due to the large variation associated with the expression of *ESR1* compared to any

**Table 2: Expression stability values of EC genes calculated by geNorm and NormFinder programmes**

Rank	Gene	NormFinder		geNorm	
		Gene	Stability <sup>b</sup>	Gene	Stability
1	MRPL19		0.105	PPIA	1.116
2	PPIA		0.119	MRPL19	1.134
3	B2M		0.169	GAPDH	1.240
4	GAPDH		0.176	PUM1	1.275
5	PUM1		0.222	B2M	1.305
6	HPRT1		0.233	HPRT1	1.357
7	TFRC		0.236	PSMC4	1.419
8	PSMC4		0.247	TFRC	1.437
9	GUSB		0.268	RPLPO	1.549
10	RPLPO		0.295	GUSB	1.568
11	ABL		0.413	ABL	2.070

<sup>b</sup>High expression stability is indicated by a low stability value. For geNorm, stability is based on an estimate of the pairwise variation (M) relative expression while for NormFinder, stability is based on combined inter- and intra-group variation.

of the EC genes as shown in Fig. 3. However, there was a significant effect of EC gene on the estimate of the error associated with relative gene expression ( $P < 0.05$ ). The magnitude of the error was significantly reduced using *MRL19* and *PPIA* as a combined endogenous control compared to the use of *ABL*, *B2M*, *GAPDH*, *GUSB*, *HPRT1*, *PSMC4*, *PPIA*, *PUM1*, *RPLPO* and *TFRC*. While there was no difference in error using *MRPL19* and *PPIA* in combination compared to *MRPL19* on its own ( $P > 0.05$ ), there was a trend towards a reduction in the estimate of the error using the combined normalising factor. There was a significant increase in relative quantity of *ESR1* in malignant versus benign breast tissues compared to normal breast tissue ( $P < 0.01$ ; Fig. 4).



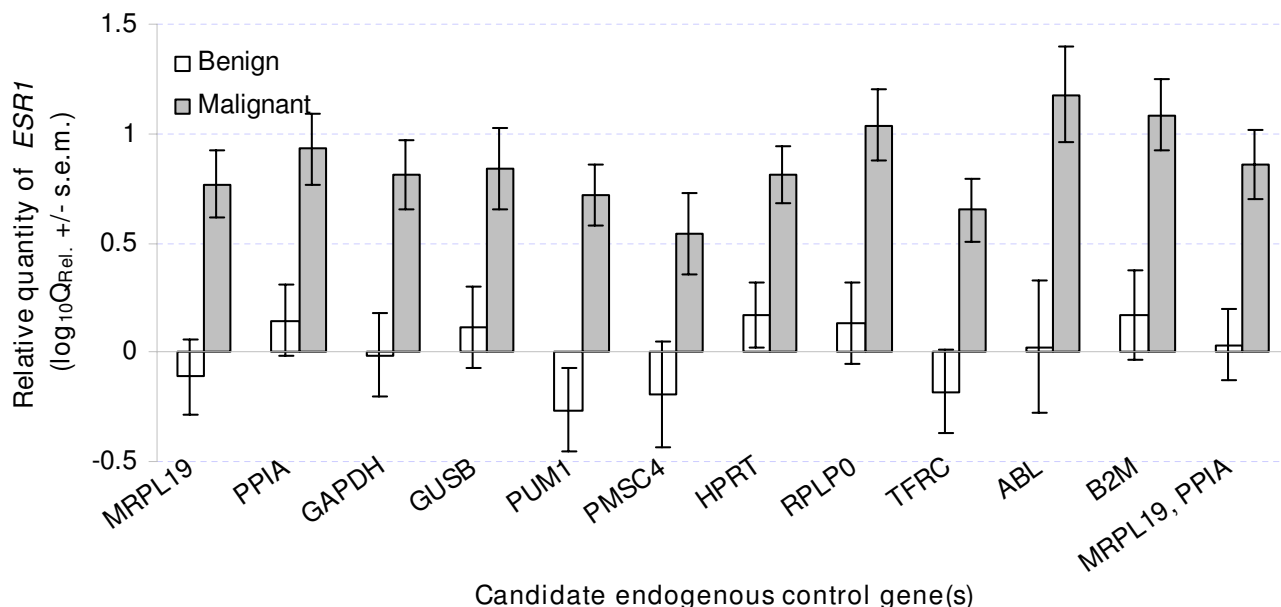
**Figure 3**  
**Quantity of all genes expressed in breast tissues relative to calibrator ( $Q_{Rel} = E^{-\Delta Ct}$ ).** A pool of cDNA from two normal tissues was used as calibrator.

**Discussion**

To our knowledge this is the first systematic evaluation of the reliability of a large number of genes used as endogenous controls for RQ-PCR analysis in breast cancer studies. The literature cites just two articles in relation to the evaluation of EC genes for breast cancer [27,28]. However, one study characterised ECs not in primary breast tissues but in cell lines [27] and the other, while employing primary breast tissues, compared just two commonly used EC genes with other genes selected from a microarray dataset of breast cancer tissues and cell lines, as well as cell lines of different origins [28].

Quantitative PCR is the basis of most nucleic acid-based breast cancer biomarker studies and its potential clinical utility is foretold by the development of the *Oncotype Dx* assay (Genomic Health). This 21-gene assay can predict metastatic recurrence [11] and magnitude of response to chemotherapy [29] in Tamoxifen-treated ER-positive early breast cancer patients. RQ-PCR will undoubtedly feature prominently in the move toward personalised medicine so the necessity of validating ECs in clinical samples as opposed to cell lines is clear. The diversity of the tissues used in this study in terms of histological and clinical parameters (Table 3) makes the results of interest to a broad spectrum of the breast cancer research community. With the exception of *ABL*, used as an EC in other settings [30], genes were selected for evaluation based their prior use in breast cancer studies, to determine the most reliable EC of those used in this field. Certain genes were excluded based on evidence that their use in this context is inappropriate [20,22,31-33].

Validation of EC genes raises the circular issue of how to normalise normalising genes. This issue governs the validity of the conclusions of such studies so at each stage of this experiment sources of non-biological variation were minimised and data were scaled relative to a calibrator.

**Figure 4****Relative quantity of oestrogen receptor alpha mRNA (*ESR1*) in benign and malignant breast tumour tissue.**

Quantity of gene expression was calculated relative to each candidate endogenous control gene and to the geometric mean of *MRPL19* and *PPIA* ( $Q_{rel.} = E^{-\Delta\Delta C_t} \pm \text{s.e.m.}$ ). A pool of cDNA from two normal tissues was used as calibrator. There was no effect of EC on the relative quantity of *ESR1* in either group ( $P > 0.05$ ), however, there was a significant effect of EC gene on the estimate of the error associated with relative gene expression ( $P < 0.05$ ). The error was significantly reduced using the combination of *MRPL19* and *PPIA* compared to the use of all EC genes individually with the exception of *MRPL19*.

For example, RNA integrity, quality and purity were stringently analysed. A threshold RIN value of 7 was applied, below which samples were excluded from analysis. This aspect is of importance given the relationship between RNA integrity and expression quantitation [34-36]. Duplicate cDNA reactions were performed and genes were amplified in triplicate using more stringent cut-offs for replicate variability than recommended elsewhere [37]. In addition, the efficiency of amplification of each assay was determined (Table 4) and data were corrected appropriately. Determination of assay efficiency is critical in comparing gene expression [38] but has not been addressed in similar studies [39]. Cycle threshold ( $C_t$ ) data were scaled relative a pooled normal tissue calibrator. Similar studies describe the comparison of genes based on raw  $C_t$  values [40,41], an inappropriate approach as discussed below and elsewhere [36].

There was no effect of tissue type on EC expression, validating comparison of their stability. This is an essential but often overlooked precursor analysis when using geNorm and NormFinder [42] since these methodologies assume the candidates are not differentially expressed between experimental groups. There was however a signif-

icant difference in variance between candidates ( $P = 0.001$ ; Fig. 1), with genes such as *RPLP0*, *TFRC*, *HPRT1* and *GAPDH* showing greater variance than others *e.g.*, *MRPL19* and *PPIA*. Since the resolution of RQ-PCR is defined by the variance associated with the EC [13] these results emphasise the necessity to evaluate and validate EC genes.

A single universal EC is unlikely to exist [43] and since the function of most genes is largely unknown it is impossible to predict their expression under different experimental conditions. The use of more than one EC hedges the bet and increases the accuracy of quantitation compared to the use of a single EC [13,24,26,36,44]. Studies show substantial errors, up to 6.5-fold, in expression quantitation using single as opposed to multiple EC genes [24]. In this study, stability of expression was analysed using two distinct statistical models, a pairwise comparison model, geNorm, and an ANOVA-based model, NormFinder. The geNorm applet selects from a panel of genes, the pair showing least variation in expression ratio across samples and estimates the minimum number of genes required for optimal normalisation. NormFinder estimates stability values for ECs considering combined intra- and inter-

**Table 3: Clinical and histological data relating to the benign (Ben.) and malignant (Mal.) breast tissues. Data includes patient menopausal status and histological type, and tumour size, T, N, M, UICC stage, grade, ER, PR and HER2/neu status and intrinsic subtype of malignant tissues where available**

Tissue type	Size (mm)	T	N	M	UICC	Grade	Menopausal status	Histological type	Subtype	ER	PR	HER2/neu
Mal.	35	2	1	0	2B	3	pre	ductal				
Mal.	22	2	1	0	2B	2	post	ductal	luminal A	8	4	negative
Mal.	22	2	0	0	2B	3	pre	ductal	basal	0	0	negative
Mal.	25	2	0	0	2A		pre	ductal	unknown			
Mal.	37	2	1	0	2B	3	pre	ductal	luminal A	8	8	negative
Mal.	35	2	1	0	2B	1	pre	lobular	luminal A	7	8	negative
Mal.	45	2	1	0	2B	3	pre	ductal	basal	0	0	negative
Ben.							pre	fibrocystic				
Mal.	20	1	1	0	2A	1	post	ductal	luminal A	8	8	negative
Mal.	50	2	1	0	2B	3	post	ductal	luminal B	8	3	positive
Mal.	15	1	1	0	2A	2	post	ductal	luminal A	8	3	negative
Mal.	20	1	0	0	1	2	post	ductal	luminal A	8	8	negative
Mal.	25	4	1	0	3B	3	post	ductal	luminal B	4	4	positive
Mal.	10	1	0	0	1		post	ductal, some tubular	luminal A	8	8	negative
Mal.	33	2	1	0	2B	1	post	lobular	luminal A	8	8	negative
Mal.	30	2	1	0	2B	3	pre	ductal	luminal A	7	8	negative
Mal.	30	2	1	0	2B	3	pre	ductal		0	0	
Mal.	20	1				1	post	colloid/mucinous				
Mal.	40	2	1	0	2B	2	post	lobular	luminal B	8	0	positive
Ben.												
Ben.							pre	fibroadenoma				
Ben.							pre	parenchymal inflammation				
Mal.	35	4	2	0	3B	3	post	ductal	luminal A	8	8	negative
Mal.	35	2	1	0	2B	3	post	ductal	luminal A	8	6	negative
Mal.							pre	ductal	luminal A	7	0	negative
Mal.	25	2	0	0	2A	2	pre	ductal	Her2	0	0	positive
Mal.	60	4	2	1	4	2						
Ben.							pre					
Ben.							pre	fibroadenoma				

Abbreviations: T: size or extent of primary tumour; N: spread to regional lymph nodes; M: distant metastasis; UICC, tumour stage according to the International Union Against Cancer TNM classification; ER: oestrogen receptor status; PR: progesterone receptor status; HER2/neu: v-erb-b2 erythroblastic leukaemia viral oncogene status.

group variation and identifies the most stable gene and gene-pair, where the stability of the pair exceeds that of the single gene. Despite their differences both models identified *MRPL19* and *PPIA* as the most reliable ECs while *RPLP0*, *GUSB* and *ABL* were least reliable. This result reflects those of the equality of variance analysis and, broadly, ranking by range of  $C_t$  values (Table 2). However, the ranking of genes by  $C_t$  range and by the model-based methods differed for some genes *e.g.*, *GAPDH*, demonstrating the necessity to scale and correct raw  $C_t$ s for amplification efficiency before analysis. GeNorm indicated that optimal normalisation could be achieved using the five most stable genes but there was no difference in *ESR1* gene expression using this approach as opposed to the two-gene combination. While it is not known whether this would hold for other less variable target genes, cost and sample availability are limiting factors

for most studies so the two-gene combination may be more practical for most applications.

The effect of using less stable ECs was assessed using *ESR1* as a target gene. Due to the high variability of the expression of this gene (Fig. 4) there was no effect of EC on quantitation. Without further discussion of the relevance of the differential expression of *ESR1* in benign and malignant tissues; a gene whose role in breast cancer is widely appreciated, it is likely that had the target gene shown a more discreet change in gene expression, an effect of EC on quantitation would have been apparent. There was however a clear tendency for *ESR1* expression to change depending on EC (Fig. 4). In the benign samples *ESR1* could be made to appear up- or down-regulated depending on EC, while its expression in the malignant samples could be numerically altered by one order of magnitude –



**Table 4: Details of gene-specific RQ-PCR assays**

Gene symbol	Gene name	Molecular function	Applied Biosystems assay identifier	Amplicon size (bp)	Slope of inhibition curve	PCR Amplification efficiency (%)
<i>ABL</i>	Abelson murine leukaemia viral I	non-receptor tyrosine protein kinase	Hs00245443_m1	54	-3.47	93.9
<i>B2M</i>	Beta-2-microglobulin	defense/immunity protein	4333766	75	-3.48	93.6
<i>GAPDH</i>	Glyceraldehyde-3-phosphate dehydrogenase	oxidoreductase, dehydrogenase	4333764	168	-3.52	92.3
<i>GUSB</i>	Glucuronidase, beta	galactosidase	4333767	63	-3.38	97.3
<i>HPRT1</i>	Hypoxanthine guanine phosphoribosyl transferase I	glycosyltransferase	4333768	100	-3.37	97.7
<i>MRPL19</i>	Mitochondrial ribosomal protein L19	protein biosynthesis	Hs00608519_m1	72	-3.14	107.7
<i>PPIA</i>	Peptidylprolyl isomerase A	isomerase	Hs99999904_m1	98	-3.38	97.3
<i>PSMC4</i>	Proteasome 26S subunit, ATPase, 4	protease, hydrolase	Hs00197826_m1	83	-3.38	97.6
<i>PUM1</i>	Pumilio, Drosophila, homolog of, I	RNA binding, translation regulation	Hs00982776_m1	62	-3.30	100.7
<i>RPLP0</i>	Ribosomal protein, large, P0	protein biosynthesis	4333761	154	-3.51	92.7
<i>TFRC</i>	Transferrin receptor	ion receptor	4333770	130	-3.56	90.9
<i>ESR1</i>	Oestrogen receptor alpha	nuclear steroid receptor	Hs00174860_m1	62	-3.45	94.5

artificial results due simply to the choice of EC. Clearly this type of error is unacceptable, especially in the analysis of markers for potential clinical application.

Furthermore, there was a significant effect of EC on the magnitude of the error associated with the estimate of *ESR1* expression. The use of the *MRPL19*, *PPIA* combination minimised the error compared to all other ECs with the exception of *MRPL19*. Apart from the fact that the use of single EC genes can compromise data as already described, the use of either gene alone is not recommended for two further reasons. Firstly, geNorm identified them based on a pairwise comparison model so their individual use is inconsistent with that analysis method. As shown in Fig. 4, *ESR1* expression can be made numerically increase or decrease in the benign groups if these genes are used singly. Secondly, stability, as assessed by NormFinder, increased roughly 30 percentage points, from 0.105 using *MRPL19* alone to 0.072 using it in combination with *PPIA* indicating improved reliability of the two-gene combination [26].

One previous study analysed EC stability in primary breast tissue [28]. That study compared two traditionally used ECs, *GAPDH* and *ACTB* and four genes identified in microarray studies [45,46]: *MRPL19*, *PUM1*, *PSMC4* and *SF3A1*. The authors recommended *MRPL19* be used with *PSMC4* and *PUM1*. However, in the present study, *PPIA*, which was not assessed by Szabo and colleagues, showed greater stability than *PSMC4* and *PUM1*. Of the genes analysed in this study, *PPIA* ranked first and second by Normfinder and geNorm respectively compared to fifth and seventh for *PSMC4* and eighth and fourth for *PUM1*.

The reason(s) *PPIA* did not emerge as a candidate in Szabo and colleagues' microarray study is unclear. As part of their selection procedure, data was filtered to remove near background signals from low abundance genes yet in this study *PPIA* showed the third highest mean  $C_t$  value. Cohort-specific effects are also unlikely since the tissues are broadly similar in terms of their clinical and histological parameters. Apart from the obvious differences in the quantitative capacities of microarray and RQ-PCR technology, a possible explanation is that the cohort from which these authors selected their candidates was not breast cancer-specific and included tissue from metastatic breast cancer lymph nodes and cell lines from an assortment of origins including dermal, leukemic, umbilical and melanoma samples [46].

This study also analysed associations between genes. An assumption of the geNorm model is that candidate ECs are not co-regulated yet the analysis of such genes by that method would lead to an erroneous choice of best gene-pair. To our knowledge the EC genes evaluated here are functionally independent as shown in Table 4. In addition, covariance of target and EC is clearly unacceptable. However, regression analysis showed significant linear relationships between *ESR1* and three of the candidates: *TFRC*, *HPRT1* and *ABL*. Approximately two thirds of breast tumours are oestrogen-dependent and the number of genes whose expression is known to be, or likely to be mediated through the receptor is sizeable [47]. This makes analysis of associations between oestrogen-responsive target and control genes of particular importance in breast cancer studies. Although there is little evidence that these genes are regulated by oestrogen, the results suggest that

aside from their poor stability it would be wise not to use them as ECs in the analysis of oestrogen-responsive breast cancer.

Despite the clear increase in accuracy afforded by the use of more than one validated EC gene, a recent survey of working practices indicated that over half of those polled continue to use one reference gene and that two thirds of these do not validate that gene [48]. Obviously levels of awareness and/or willingness to address this issue must be improved.

## Conclusion

The current emphasis on personalised cancer treatment has resulted in the development of prognostic and predictive multi-gene RQ-PCR assays. However, with such developments comes the demand for greater accuracy and resolution of gene expression quantitation. In this study we have validated two genes, *MRPL19* and *PPIA* as EC candidates for RQ-PCR analysis of primary breast tissue using two different statistical models and demonstrate that of the genes in current use in this field, the above gene combination offers increased accuracy and resolution in the relative quantitation of gene expression data. The genes identified should be of use in a broad spectrum of translational research and diagnostic applications in breast cancer.

## Methods

### Breast tissue samples

Primary breast tumour tissues (n = 27) were obtained from patients during primary curative resection, at Galway University Hospital, Galway, Ireland. Samples were categorised into benign (n = 6) or malignant groups (n = 21) according to analysis of standard histopathological parameters. Clinical data relating to the tumour tissues used in this study are shown in Table 3. RNA from normal tissues, recovered from patients undergoing reduction mastopexy were used as calibrator samples for comparative relative RQ-PCR (n = 2). After excision, tissue samples were immediately snap-frozen in liquid nitrogen and stored at -80°C until RNA extraction. Prior written and informed consent was obtained from each patient and the study was approved by the ethics review board of Galway University Hospital. Clinical data were obtained from the Breast Cancer Database at the Department of Surgery, Galway University Hospital.

### Candidate endogenous control genes

Eleven commonly used candidate endogenous control genes were selected for analysis (Table 4). To our knowledge, all genes are constitutively expressed in breast cancer tissues and all have independent cellular functions and are assumed not to be co-regulated. Only *RPLP0* and

*MRPL19* share a molecular function, *i.e.*, protein biosynthesis.

### Minimisation of non-biological variation

While target gene expression is normalised using EC genes to correct for variation introduced during sample processing using, this is obviously not possible in EC validation studies. Since this critical issue governs the reliability of the data generated and the validity of the conclusions it was addressed as described below.

Firstly, while it was not possible to control for variation in the acquisition of clinical samples collected over a number of years, every effort was made to minimise systematic variation downstream of sample acquisition. All equipment and instruments were calibrated before use. Benign and malignant samples were homogenised separately but on the same day. All RNA was extracted using the same protocol and reagent lot by one person to avoid batch-to-batch variation. Where possible, two extractions from separate areas of the each tissue sample were pooled to control for tissue heterogeneity. RNA integrity and purity were stringently analysed as described below. Duplicate cDNA reactions were performed to minimise variation from the reverse transcription step. No-RT controls were included with each batch of cDNA synthesised. All PCR reactions were performed on the same pool of aliquotted cDNA and no-template controls were included in each run for each gene. Appropriate inter-assay controls were included in each run. In addition, the efficiency of amplification was calculated for each assay and expression results were corrected for the small differences in efficiency observed between genes (Table 4). All cycle threshold (C<sub>t</sub>) data was scaled relative to a calibrator sample amplified using the same gene.

### Total RNA Isolation

Tissue (50–100 mg) was homogenised in 1 ml of QIAzol Lysis Reagent (Qiagen, Crawley, UK), using a bench-top homogeniser (Polytron PT1600E, Kinematica AG, Littau-Luzern, Switzerland). Total RNA was isolated from homogenised breast tissue using the RNeasy® Tissue Mini Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. RNA was eluted in 60 µl nuclease-free water and stored at -80°C. In addition to the on-column DNase treatment performed during the RNA extraction procedure, RNA was DNase-treated after extraction using the DNA-free™ DNase Treatment and Removal Reagents (Ambion, Cambridgeshire, UK). RNA concentration and purity was assessed in duplicate samples using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, DE, USA). RNA integrity was assessed using the RNA 6000 Nano LabChip Series II Assay with the 2100 Bioanalyzer System (Agilent Technologies, Palo Alto, CA, USA). Electropherograms and gel-like images were evaluated

using the Agilent 2100 Expert software (Version B.02.03) which generated the RNA integrity number (RIN) enabling estimation of RNA integrity. The RIN value describes a graded scale of RNA integrity ranging from 1 (completely degraded total RNA) to 10 (intact total RNA). Based on this tool, total RNA integrity is determined not only by the ratio of the ribosomal bands but by the entire electrophoretic trace of the sample including presence or absence of degradation products [49].

In agreement with recent reports [34,35], there was a significant negative linear relationship between RNA integrity, as determined by analysis of RIN and  $C_t$  values ( $P < 0.05$ ), with increased  $C_t$  values associated with RIN less than 5 (data not shown). Therefore the threshold RIN value for inclusion of RNA samples in analysis was  $\geq 7$ . RNA purity was verified by an average  $A_{260}/A_{280}$  ratio of 1.98, ranging from 1.97 to 2.01.  $A_{260}/A_{230}$  ratios averaged 1.7, ranging from 1.5 to 1.83.

#### First strand cDNA synthesis by reverse transcription

First strand cDNA was synthesised in duplicate reactions for each RNA sample (2 by 1  $\mu$ g each) using Superscript III reverse transcriptase (Invitrogen Life Technologies, Paisley, UK) and random primers ( $N_9$ ; 1  $\mu$ g; MWG Biotech, AG, Ebersberg, Germany). Negative controls consisting of non-reverse transcribed samples were included in each set of reactions. The reactions were incubated at 25°C for 5 min followed by 50°C for 1 h and finally 72°C for 15 min. Duplicate cDNA reactions were pooled, diluted to 120  $\mu$ l in nuclease-free water (Invitrogen Life Technologies), aliquotted and stored at -20°C till further use.

#### Real-time Quantitative PCR

The expression of the 11 candidate EC genes was analysed by RQ-PCR using TaqMan® Endogenous Control Assays or TaqMan® Gene Expression Assays and the ABI Prism® 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Each gene was tested in triplicate within the same PCR run for the majority of samples, with the remaining samples tested on an additional 96-well plate. TaqMan® Endogenous Control Assay and Gene Expression Assay IDs are listed in Table 4. Samples with standard deviations  $>0.3$  from the mean  $C_t$  of the triplicates were excluded from analysis. *HPRT1*, amplified from pooled normal cDNA, was run on each plate to assess inter-assay variation. cDNA (2  $\mu$ l) from each tumour sample was added to a PCR reaction mix containing 1 $\times$  TaqMan® Universal PCR Master Mix, No AmpErase® UNG and 1  $\mu$ l Endogenous Control Assay or Gene Expression Assay (Applied Biosystems) in a 20  $\mu$ l reaction volume. Standard cycling conditions were used [95°C for 10 minutes, (95°C for 15 seconds, 60°C for 60 seconds)  $\times$  40 cycles]. The inter-assay percent coefficient of variation (%CV) for

samples with a mean  $C_t$  of  $25.81 \pm 0.07$  (mean  $\pm$  s.e.m.) was 0.81% ( $n = 15$ ).

#### PCR amplification efficiencies

Percent PCR amplification efficiencies (E) for each assay were calculated as  $E = (10^{-1/\text{slope}} - 1) \times 100$ , using the slope of the semi-log regression plot of  $C_t$  versus log input of cDNA (10-fold dilution series of five points) as shown in Table 4. A threshold of 10% above or below 100% efficiency was applied. Amplification efficiencies ranged from 90.9% for *TFRC* to 107.7% for *MRPL19*, indicative of approximate exponential efficiencies for these assays.

#### Conversion of cycle threshold ( $C_t$ ) to quantity relative to calibrator

The baseline (3–15 cycles) and average threshold cycle ( $C_t$ ) were automatically calculated using the ABI Prism SDS Software (version 1.2.3). The  $C_t$  value is defined as the PCR cycle number at which the fluorescence generated from amplification of the target gene within a sample increases to a threshold value of 10 times the standard deviation of the baseline emission and is inversely proportional to the starting amount of target cDNA.  $C_t$  results were converted into quantities relative to normal ( $Q_{\text{Rel.}}$ ), and corrected for PCR amplification efficiency (E), using the following formula:  $Q_{\text{Rel.}} = E^{-\Delta C_t}$ , where  $\Delta C_t = C_t$  test sample -  $C_t$  calibrator sample.

#### Comparative quantitation of target gene *ESR1* relative to endogenous control

To calculate the expression of *ESR1*, relative to an EC gene(s), the  $\Delta\Delta C_t$  method was used where  $\Delta\Delta C_t = (C_t \text{ target gene, test sample} - C_t \text{ endogenous control, test sample}) - (C_t \text{ target gene, calibrator sample} - C_t \text{ endogenous control, calibrator sample})$ . Fold change in gene expression between groups was calculated as  $E^{-\Delta\Delta C_t} \pm$  s.e.m. Where target gene expression was normalised using more than one endogenous control, fold change estimates were calculated using the geometric mean of EC quantities relative to the calibrator sample and the errors were calculated following the rules of error propagation described previously [25].

#### Analysis of EC stability

Candidate EC gene stability was evaluated using two statistical models for the analysis candidate EC genes, geNorm [[24], Ver. 3.4] and NormFinder [26].  $C_t$  values were converted into relative quantities considering the PCR amplification efficiencies as detailed above. GeNorm is a Visual Basic application tool for Microsoft® Excel and is freely available by request from the authors [50]. NormFinder is a Microsoft® Excel add-in, also freely available [51]. For NormFinder analysis tissues samples were categorised into benign ( $n = 6$ ) or malignant groups ( $n =$

21) according to analysis of standard histopathological parameters as described above.

### Statistical analyses

Statistical analyses were performed with Minitab® 15 Statistical Software for Windows® (Minitab Ltd., Coventry, UK). *P* values <0.05 were considered statistically significant. The Anderson-Darling normality test was applied. Two-sample *t* tests were used to compare calibrator-scaled EC gene quantities between benign and malignant tissue groups. Equality of variance between scaled EC  $Q_{Rel}$  values and effect of EC on the estimate of the error associated with gene expression was analysed using Bartlett's test. Associations between scaled EC and target gene quantities were determined by regression analysis which examined linear, quadratic and cubic relationships. One-way ANOVA was used to compare *ESR1* expression normalised using different EC genes.

### Abbreviations

RQ-PCR, Real-time quantitative polymerase chain reaction

EC, endogenous control

$C_t$ , cycle threshold

RIN, RNA integrity number

$Q_{Rel}$ , quantity of gene expression relative to normal

NF, geNorm normalisation factor

*V*, geNorm pairwise variation

*M*, geNorm stability measure

T, size or extent of primary tumour

N, spread to regional lymph nodes

M, distant metastasis

UICC, tumour stage according to the International Union Against Cancer TNM classification

ER, oestrogen receptor

PR, progesterone receptor

HER2/*neu*, v-erb-b2 erythroblastic leukaemia viral oncogene

### Authors' contributions

REM conceived, designed and performed the experiments, was responsible for data analyses and for writing the manuscript. NM contributed throughout the experiment and critically reviewed the manuscript. MJK participated throughout as Head of Department, particularly in sample acquisition during surgery and in critically reviewing the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

This work was funded by the National Breast Cancer Research Institute (NBCRI) who we wish to thank for their continued generous support. We gratefully acknowledge the assistance of Ms. Catherine Curran in collating the clinical and histological data and Ms. Emer Hennessy, curator of the Department of Surgery, NUI, Galway BioBank. We also thank Ms. Jill Sales, Biomathematics & Statistics Scotland, for her advice and help with the data analysis.

### References

- Higuchi R, Dollinger G, Walsh PS, Griffith R: **Simultaneous amplification and detection of specific DNA sequences.** *Biotechnology (N Y)* 1992, **10(4)**:413-417.
- Higuchi R, Fockler C, Dollinger G, Watson R: **Kinetic PCR analysis: real-time monitoring of DNA amplification reactions.** *Biotechnology (N Y)* 1993, **11(9)**:1026-1030.
- Ginzinger DG: **Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream.** *Exp Hematol* 2002, **30(6)**:503-512.
- Valasek MA, Repa JJ: **The power of real-time PCR.** *Adv Physiol Educ* 2005, **29(3)**:151-159.
- Walker NJ: **Tech.Sight. A technique whose time has come.** *Science* 2002, **296(5567)**:557-559.
- Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, Mone M, Hansen H, Buys SS, Rasmussen K, Orrico AR, Dreher D, Walters R, Parker J, Hu Z, He X, Palazzo JP, Olopade OI, Szabo A, Perou CM, Bernard PS: **Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay.** *Breast Cancer Res* 2006, **8(2)**:R23.
- Leissner P, Verjat T, Bachelot T, Paye M, Krause A, Puisieux A, Mougin B: **Prognostic significance of urokinase plasminogen activator and plasminogen activator inhibitor-1 mRNA expression in lymph node- and hormone receptor-positive breast cancer.** *BMC Cancer* 2006, **6**:216.
- Vinatzer U, Dampier B, Streubel B, Pacher M, Seewald MJ, Stratowa C, Kaserer K, Schreiber M: **Expression of HER2 and the coamplified genes GRB7 and MLN64 in human breast cancer: quantitative real-time reverse transcription-PCR as a diagnostic alternative to immunohistochemistry and fluorescence in situ hybridization.** *Clin Cancer Res* 2005, **11(23)**:8348-8357.
- Span PN, Lindberg RL, Manders P, Tjan-Heijnen VC, Heuvel JJ, Beex LV, Sweep CG: **Tissue inhibitors of metalloproteinase expression in human breast cancer: TIMP-3 is associated with adjuvant endocrine therapy success.** *J Pathol* 2004, **202(4)**:395-402.
- Esteva FJ, Sahin AA, Cristofanilli M, Coombes K, Lee SJ, Baker J, Cronin M, Walker M, Watson D, Shak S, Hortobagyi GN: **Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy.** *Clin Cancer Res* 2005, **11(9)**:3315-3319.
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351(27)**:2817-2826.
- Zhang Z, Yamashita H, Toyama T, Sugiura H, Ando Y, Mita K, Hamaguchi M, Hara Y, Kobayashi S, Iwase H: **NCOR1 mRNA is an independent prognostic factor for breast cancer.** *Cancer Lett* 2006, **237(1)**:123-129.

13. Huggett J, Dheda K, Bustin S, Zumla A: **Real-time RT-PCR normalisation; strategies and considerations.** *Genes and immunity* 2005, **6(4)**:279-284.
14. Chang TJ, Juan CC, Yin PH, Chi CW, Tsay HJ: **Up-regulation of beta-actin, cyclophilin and GAPDH in N1S1 rat hepatoma.** *Oncol Rep* 1998, **5(2)**:469-471.
15. Rondinelli RH, Epner DE, Tricoli JV: **Increased glyceraldehyde-3-phosphate dehydrogenase gene expression in late pathological stage human prostate cancer.** *Prostate Cancer Prostatic Dis* 1997, **1(2)**:66-72.
16. Weisman AS, Tixier-Vidal A, Gourdj D: **Thyrotropin-releasing hormone increases the levels of c-fos and beta-actin mRNA in GH3/B6 pituitary tumor cells.** *In Vitro Cell Dev Biol* 1987, **23(8)**:585-590.
17. Lu S, Gu X, Hoestje S, Epner DE: **Identification of an additional hypoxia responsive element in the glyceraldehyde-3-phosphate dehydrogenase gene promoter.** *Biochim Biophys Acta* 2002, **1574(2)**:152-156.
18. Tokunaga K, Nakamura Y, Sakata K, Fujimori K, Ohkubo M, Sawada K, Sakiyama S: **Enhanced expression of a glyceraldehyde-3-phosphate dehydrogenase gene in human lung cancers.** *Cancer Res* 1987, **47(21)**:5616-5619.
19. Goidin D, Mamessier A, Staquet MJ, Schmitt D, Berthier-Vergnes O: **Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations.** *Anal Biochem* 2001, **295(1)**:17-21.
20. Naora H, Naora H: **Differential expression patterns of beta-actin mRNA in cells undergoing apoptosis.** *Biochem Biophys Res Commun* 1995, **211(2)**:491-496.
21. Valenti MT, Bertoldo F, Dalle Carbonare L, Azzarello G, Zenari S, Zanatta M, Balducci E, Vinante O, Lo Cascio V: **The effect of bisphosphonates on gene expression: GAPDH as a house-keeping or a new target gene?** *BMC Cancer* 2006, **6**:49.
22. Naylor MS, Stamp GW, Balkwill FR: **Beta actin expression and organization of actin filaments in colorectal neoplasia.** *Epithelial Cell Biol* 1992, **1(3)**:99-104.
23. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, Sindelka R, Sjoberg R, Sjogreen B, Strombom L, Stahlberg A, Zoric N: **The real-time polymerase chain reaction.** *Mol Aspects Med* 2006, **27(2-3)**:95-125.
24. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** *Genome Biol* 2002, **3(7)**:RESEARCH0034.
25. Vandesompele J, De Peter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **GeNorm software manual, update 13th March 2007.** [<http://medgen.ugent.be/~jvdesomp/genorm/>].
26. Andersen CL, Jensen JL, Orntoft TF: **Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets.** *Cancer Res* 2004, **64(15)**:5245-5250.
27. Morse DL, Carroll D, Weberg L, Borgstrom MC, Ranger-Moore J, Gillies RJ: **Determining suitable internal standards for mRNA quantification of increasing cancer progression in human breast cells by real-time reverse transcriptase polymerase chain reaction.** *Anal Biochem* 2005, **342(1)**:69-77.
28. Szabo A, Perou CM, Karaca M, Perreard L, Quackenbush JF, Bernard PS: **Statistical modeling for selecting housekeeper genes.** *Genome Biol* 2004, **5(8)**:R59.
29. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, Costantino JP, Geyer CE Jr., Wickerham DL, Wolmark N: **Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer.** *J Clin Oncol* 2006, **24(23)**:3726-3734.
30. Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, van der Schoot E, Delabesse E, Macintyre E, Gottardi E, Saglio G, Watzinger F, Lion T, van Dongen JJ, Hokland P, Gabert J: **Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program.** *Leukemia* 2003, **17(12)**:2474-2486.
31. Chan MW, Wei SH, Wen P, Wang Z, Matei DE, Liu JC, Liyanarachchi S, Brown R, Nephew KP, Yan PS, Huang TH: **Hypermethylation of 18S and 28S ribosomal DNAs predicts progression-free survival in patients with ovarian cancer.** *Clin Cancer Res* 2005, **11(20)**:7376-7383.
32. Nadano D, Sato TA: **Caspase-3-dependent and -independent degradation of 28 S ribosomal RNA may be involved in the inhibition of protein synthesis during apoptosis initiated by death receptor engagement.** *J Biol Chem* 2000, **275(18)**:13967-13973.
33. White RJ: **RNA polymerases I and III, growth control and cancer.** *Nature reviews* 2005, **6(1)**:69-78.
34. Fleige S, Pfaffl MW: **RNA integrity and the effect on the real-time qRT-PCR performance.** *Mol Aspects Med* 2006, **27(2-3)**:126-139.
35. Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW: **Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR.** *Biotechnol Lett* 2006, **28(19)**:1601-1613.
36. Bustin SA, Nolan T: **Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction.** *J Biomol Tech* 2004, **15(3)**:155-166.
37. Nolan T, Hands RE, Bustin SA: **Quantification of mRNA using real-time RT-PCR.** *Nature protocols* 2006, **1(3)**:1559-1582.
38. Yuan JS, Reed A, Chen F, Stewart CN Jr.: **Statistical analysis of real-time PCR data.** *BMC bioinformatics* 2006, **7**:85.
39. Aerts JL, Gonzales MI, Topalian SL: **Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR.** *Biotechniques* 2004, **36(1)**:84-6, 88, 90-1.
40. de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN: **Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes.** *Laboratory investigation; a journal of technical methods and pathology* 2005, **85(1)**:154-159.
41. Liu DW, Chen ST, Liu HP: **Choice of endogenous control for gene expression in nonsmall cell lung cancer.** *Eur Respir J* 2005, **26(6)**:1002-1008.
42. Ohl F, Jung M, Radonic A, Sachs M, Loening SA, Jung K: **Identification and validation of suitable endogenous reference genes for gene expression studies of human bladder cancer.** *The Journal of urology* 2006, **175(5)**:1915-1920.
43. Bustin SA: **Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays.** *J Mol Endocrinol* 2000, **25(2)**:169-193.
44. Pfaffl MW, Horgan GW, Dempfle L: **Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR.** *Nucleic Acids Res* 2002, **30(9)**:e36.
45. Perou CM, Brown PO, Botstein D: **Tumor classification using gene expression patterns from DNA microarrays.** *New technologies for life sciences: A trends guide* 2000, **Dec**:67-76.
46. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406(6797)**:747-752.
47. Tang S, Tan SL, Ramadoss SK, Kumar AP, Tang MH, Bajic VB: **Computational method for discovery of estrogen responsive genes.** *Nucleic Acids Res* 2004, **32(21)**:6212-6217.
48. Bustin SA: **Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences.** *Expert review of molecular diagnostics* 2005, **5(4)**:493-498.
49. Mueller O, Lightfoot S, Schroeder A: **RNA integrity number (RIN) - standardization of RNA quality control.** *Agilent Technologies, Palo Alto* 2004.
50. geNorm [<http://medgen.ugent.be/~jvdesomp/genorm/>]
51. NormFinder [<http://www.mdl.dk/publicationsnormfinder.htm>]