



**HAL**  
open science

## **RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: a marker of ER-positive luminal-like subtype**

Hany Onsy Habashy, Desmond G. Powe, Enrico Glaab, Graham Ball, Inmaculada Spiteri, Natalio Krasnogor, Jonathan M. Garibaldi, Emad A. Rakha, Andrew R. Green, C. Caldas, et al.

### ► To cite this version:

Hany Onsy Habashy, Desmond G. Powe, Enrico Glaab, Graham Ball, Inmaculada Spiteri, et al.. RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: a marker of ER-positive luminal-like subtype. *Breast Cancer Research and Treatment*, 2010, 128 (2), pp.315-326. 10.1007/s10549-010-1073-y . hal-00615369

**HAL Id: hal-00615369**

**<https://hal.science/hal-00615369v1>**

Submitted on 19 Aug 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: A marker of ER-positive luminal-like subtype**

**Hany Onsy Habashy<sup>1,2</sup>, Desmond G Powe<sup>3</sup>, Enrico Glaab<sup>4</sup>, Graham Ball<sup>5</sup>,  
Inmaculada Spiteri<sup>6</sup>, Natalio Krasnogor<sup>4</sup>, Jonathan M Garibaldi<sup>4</sup>, Emad A  
Rakha<sup>1</sup>, Andrew R Green<sup>1</sup>, C Caldas<sup>6</sup>, Ian O Ellis<sup>1</sup>**

<sup>1</sup>Department of Pathology, School of Molecular Medical Sciences, Nottingham  
University Hospitals NHS Trust and University of Nottingham, Nottingham, UK

<sup>2</sup>Department of Histopathology, Faculty of Medicine, Mansoura University,  
Mansoura, Egypt

<sup>3</sup>Department of Cellular Pathology, Queen's Medical Centre, Nottingham University  
Hospitals NHS Trust, Derby Road, Nottingham NG7 2UH, UK

<sup>4</sup>School of Computer Science, University of Nottingham, Nottingham, UK

<sup>5</sup> John Van Geest Cancer Research Centre, School of Science and Technology,  
Nottingham Trent University, Nottingham, NG11 8NS, UK

<sup>6</sup>Department of Oncology, University of Cambridge and Cancer Research UK  
Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK

## **Correspondence:**

**Prof. Ian O Ellis**

Molecular Medical Sciences, University of Nottingham  
Department of Histopathology, Nottingham City Hospital NHS Trust, Hucknall Road,  
Nottingham, NG5 1PB, UK

Tel: (44) 0115-9691169

Fax: (44) 0115- 9627768

Email: Ian.Ellis@nottingham.ac.uk

## **Key Words:**

Breast carcinoma, RERG, Gene expression, Luminal, Oestrogen receptor,  
Immunohistochemistry

## **Abstract**

Global gene expression profiling studies have classified breast cancer into a number of distinct biological and molecular classes with clinical relevance. The heterogeneous luminal group, which is largely characterised by oestrogen receptor (ER) expression, appears to contain distinct subgroups with differing behaviour.

In this study, we analysed 47,293 gene transcripts in 128 invasive breast carcinomas (BC) using Artificial Neural Networks and a cross-validation analysis used in combination with an ensemble sample classification to identify genes that can be used to subclassify ER+ luminal tumours. The results were validated using immunohistochemistry on TMAs containing 1,140 invasive breast cancers.

Our results showed that the RERG-gene is one of the highest ranked genes to differentiate between ER+ luminal tumours and ER- non-luminal cancers based on a 10-fold external cross-validation analysis with an average classification accuracy of 89%. This was confirmed in our protein expression studies that showed RERG positive associations with markers of luminal differentiation including ER, luminal cytokeratins (CK19, CK18, and CK7/8) and FOXA1 ( $p=0.004$ ) and other markers of good prognosis in BC including small size, lower histologic grade and positive expression of androgen receptor, nuclear BRCA1, FHIT and cell cycle inhibitors p27 and p21. RERG expression was inversely associated with the proliferation marker MIB1 ( $p=0.005$ ) and p53.

Strong RERG expression showed an association with longer breast cancer specific survival and distant metastasis free interval in the whole series as well as in the ER+ luminal group and these associations were independent of other prognostic variables

In conclusion, we used novel bioinformatics methods to identify candidate genes to characterise ER+ luminal-like breast cancer. RERG gene is a key marker of the luminal BC class and can be used to separate distinct prognostic subgroups.

## **Introduction**

Gene expression profiling (GEP) studies have classified breast cancer (BC) into a number of distinct biological classes which show relationships with clinical outcome [1-3]. The luminal group is the largest class and is fundamentally characterised by oestrogen receptor (ER) expression, but it has increasingly been acknowledged that several subgroups exist within this class, each with different molecular features and clinical behaviour. The number and characterization of these subgroups however remains controversial. In one of the first GEP studies, Perou et al[1] reported four molecular subtypes including the luminal subtype. In a subsequent study, three luminal subtypes were described (Luminal A, B and C)[2]. Subsequent studies have settled for the existence of only two luminal classes (A and B)[4]. Because luminal tumours constitute approximately 70% of all BC, there is a clear need to improve understanding of their heterogeneity through improved sub-classification and with emphasis on clinical behaviour.

Previously, we have reported on biomarkers selected for investigation on the basis of their biological function and their potential ability to distinguish different prognostic subclasses within the luminal group [5-7]. However, it is thought that novel approaches may be more appropriate for biomarker discovery in breast cancer. Among the machine learning based methods, Artificial Neural Networks (ANNs) are emerging as valuable tools for this purpose. ANNs are a form of artificial intelligence inspired by learning in human neuronal systems and have been shown to be capable of modelling complex systems with high predictive accuracies on several large scale datasets [8-10]. We have used a combination of analytical computational approaches to identify genes associated with the luminal subtype of breast cancer and from the top ranked genes; RERG was validated as a prognostic luminal biomarker at the

protein level in a well-characterised and annotated consecutive patient series of invasive breast cancer using immunohistochemistry and TMA analysis. This approach allowed us to assess the prognostic and biological importance of RERG protein in invasive BC and its role in subclassification of luminal-like cancer subtype.

Previous studies have shown that RERG mRNA expression was found to be induced by estradiol and repressed by ER-antagonist in ER-responsive MCF-7 cells [11].

## **Material and Methods**

This study was approved by the Nottingham Research Ethics Committee 2 under the title “Development of a molecular genetics classification of breast cancer”.

### **Gene expression study**

The study population used was derived from the Nottingham Tenovus Primary Breast Carcinoma Series of women aged 70 years or less, who presented with stage I and II primary operable invasive breast carcinomas.

Gene expression profiling has been previously described [12-14]. Briefly, total RNA was extracted from a series of frozen breast cancers retrieved from Nottingham Hospitals NHS Trust Tumour Bank between 1986 and 1992. RNA integrity and DNA contamination were analysed using Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Total RNA was biotin-labelled using the Illumina TotalPrep RNA Amplification kit (Ambion, Austin, TX, USA) according to manufacturer’s instructions. Biotin-labelled cRNA (1.5  $\mu$ g) was used for each hybridisation on Sentrix Human-6 BeadChips (Illumina, San Diego, CA, USA) in accordance with the manufacturer’s protocol. Illumina gene expression data containing 47,293 transcripts were analysed and summarised in the Illumina Bead Studio software. Analyses of the probe level data were done using the beadarray Bioconductor package. The expression data are available at the EBI website (<http://www.ebi.ac.uk/miamexpress/>) with the accession number E-TABM-576.

### **Bioinformatics analysis (1): Artificial neural network model**

We identified the ER+ cases (n=84) identified by immunohistochemistry [7, 15] in our patient cohort (n=128) and applied an artificial neural network (ANN) model for sample classification to the gene expression data comprising 47,293 inputs for each sample. The output node was coded as 0 if a case was low ER expression (the median

H-score<140; n = 42), and 1 if high ER expression (H score> 140; n = 42). Each input (gene) was considered singly as an input to the model. More precisely, the data was analysed using multi-layer perception architecture with a sigmoidal transfer function, where weights were updated by a back propagation algorithm as previously described [10, 14]. Inputs were ranked in ascending order based on predictive error. The detailed methods are described previously [10, 14].

### **Bioinformatics analysis (2): Ensemble classification and cross-validation analysis**

In a second bioinformatics analysis step, we sought to obtain a robust ranking of genes that are differentially expressed between the ER+ (n=84) cases and the ER negative non-luminal cases (all other cases) and have high predictive power, by applying an ensemble sample classification method (see description below) within a leave-one-out cross-validation scheme. For this purpose, the 128 patient samples were first grouped into 128 different training/test set partitions, using 127 samples for the training sets and the remaining sample as the test set. For each of the 128 training sets differentially expressed genes were selected independently with the "Empirical Bayes moderated t-statistic" (Smyth, 2004)[16] and used to train a machine learning model, which was evaluated based on the left-out sample (a procedure known as "external cross-validation")[28]. To classify the left-out sample, the prediction results of four algorithms (Support Vector Machine, Random Forest, kNN and Prediction Analysis for Microarrays, with all parameters being optimised by using a grid search within a nested cross-validation)[17] were combined to a majority-vote ensemble classifier as to compensate for the inevitable inherent biases and variances that exists amongst each of these machine learning algorithms. In order to rank the genes based on the cross-validation results, their frequency of occurrence in the list of significantly differentially expressed genes (p-value < 0.05) across different cross-validation cycles



was recorded, and genes received higher scores the more often they had been selected. RERG belonged to the top-ranked genes which were selected in each of the 128 cross-validation cycles. All steps of the analysis were conducted using an in-house web-application for microarray analysis, available at [www.arraymining.net](http://www.arraymining.net)[18].

### **Patient selection and tissue microarray construction**

To validate our main conclusion from the bioinformatics analysis of the gene expression data, tissue microarrays (TMAs) prepared from the whole series cases were used as described previously[15]. Briefly, Formalin fixed paraffin embedded (FFPE) TMAs were prepared from a series cases of primary operable (stage I and II) breast carcinoma cases from patients age <70 presenting consecutively to the Nottingham Breast Unit with tumours of less than 5 cm diameter between 1988 and 1998. This is a well-characterized resource that contains patients' clinical and pathological data including patients' age, histologic tumour type[19], primary tumour size, lymph node status, mitotic count and histologic grade [20], Nottingham prognostic index (NPI) [21], vascular invasion (VI), development of recurrence, and distant metastases (DM). Patient management was based on Nottingham Prognostic Index (NPI) score and ER status as previously described [22]. The NPI, which is widely used in the UK, was developed for the prognostic management of breast cancer by using multivariate analysis to determine the most important prognostic factors. It includes 3 factors; tumour histologic grade (1-3 using the Nottingham Grading System[23]), lymph node (LN) stage (1-3; 1=LN negative, 2=1-3 positive nodes and 3 $\geq$  4 positive nodes) and primary tumour size (0.2x size in cm). NPI is then categorized into 3 groups: the good prognostic group (score  $\leq$ 3.4), the moderate group (>3.4 - 5.4) and the poor prognostic group (score > 5.4) [24].

Survival data including survival time and disease-free interval (DFI) were maintained on a prospective basis. Breast cancer specific survival (BCSS) was defined as the time (in months) from the date of the primary surgical treatment to the time of death from (or with active) breast cancer. DFI was defined as the interval (in months) from the date of the primary surgical treatment to the first locoregional or distant metastasis (DMFI).

Mean follow-up time of the studied cohort was 124 months. The data on other biomarkers with strong relevance to breast cancer including ER, progesterone receptor (PgR), androgen receptor (AR), BRCA1, p53, FHIT, HER2, E-cadherin, P-cadherin, basal and luminal cytokeratins (CKs) (CK5/6, CK14, CK18, CK19 and CK7/8), cell cycle inhibitors (p21 and p27), smooth muscle actin (SMA), MIB1, BCL2, and FOXA1 protein expression were available [7, 15, 25]. Source, dilution, pretreatment and cut-offs of antibodies used are summarised in Table 1. MIB1 staining was done on full face tumour sections [26].

### **Immunohistochemistry**

The RERG specific rabbit polyclonal antibody (Purified rabbit anti-human RERG polyclonal antibody, 10687-1-AP, Proteintech Group, Chicago, IL, USA) was optimized at a working dilution of 1:20 using randomly selected full-face sections of breast cancer tissue to assess the staining distribution in cancer and normal tissue. After microwave antigen retrieval in citrate buffer pH 6 and protein blocking, the TMAs sections and control sections (normal breast tissue and a known positive carcinoma) were put in a humidity chamber followed by blocking of endogenous peroxidase by applying hydrogen peroxide in methanol for 10 min. The TMA slides were incubated in primary antibody at 4°C overnight. The immunohistochemical detection of RERG was carried out using a labelled streptavidin biotin technique

LSAB<sup>®</sup> in accordance with the manufacturer's instructions (DakoCytomation, Cambridge, UK). The slides were incubated with the biotinylated secondary antibody for 30 min, followed by HRP-Streptavidin for another 30 min. Tris buffer saline (TBS) were used for washing between steps (3x2 with stirrer), for visualisation of the reaction, the slides were incubation in freshly prepared peroxidase substrate solution (DAB) diluted 1:50 for 10 min. After application of DAB, the slides were washed in running tap water, counterstained in haematoxylin, dehydrated in alcohols, cleared in xylene and coverslipped using DPX mounting medium. Negative controls were performed by omitting the primary antibody while positive control BC sections were used in each run.

### **Statistical analysis**

Statistical analysis was performed using SPSS 15.0 statistical software (SPSS Inc., Chicago, IL, USA). Association between RERG immunoreactivity and different clinicopathological parameters was evaluated using Chi-square test. Standard cutoff values for the different biomarkers, needed to determine categorical scores before statistical analysis, were the same as those published in previous studies [15, 25, 27] (Table 1). Survival curves were estimated by the Kaplan-Meier method with a log rank test to assess significance. Patients that died due to causes other than breast cancer were censored during survival analysis. We used REMARK guidelines [28] for reporting on prognostic biomarkers and therefore studied the outcome in patients in the whole series and a subset of ER+ luminal like patients and according to the systemic therapy groups. Multivariate Cox proportional hazard regression models were used to evaluate any independent prognostic effect of the variables with 95% confidence interval. A *p*-value of <0.05 was considered to indicate statistical significance.

## **Results**

### **Gene expression studies**

#### **Novel genes associated with ER+ status (high expression vs. low expression) using Artificial Neural Network**

High ER expression is associated with better prognosis when compared to low ER expression. Our aim was to find; using a novel prediction method (ANN), a set of genes that can associate with high ER expression and to validated the genes using protein expression. To study this, ER+ cases (84 tumours) were categorized according to the level of ER expression into high and low expression using the median of the H-score value (H-score 140).

The ranking order of ANN results was based on predictive error for the unseen cohort in the Monte Carlo Cross validation with the lowest being higher in the ranking order. Table 2 illustrates top transcripts according to ER status ranked by selection error. RERG transcript was associated with high ER status that characterise luminal A molecular subtype.

In this study ESR1, the gene for ER, was ranked as the most important gene for ER membership and this was used a proof for the validation of the model.

#### **Novel genes associated with ER+ luminal phenotype using the ensemble cross-validation analysis**

The RERG-gene was selected among the significantly differentially expressed genes in every cycle of a leave-one-out external cross-validation analysis. The prediction models obtained from this procedure distinguished the luminal from the non-luminal samples with an average accuracy of 88.3% (sensitivity: 95.2%, specificity: 75.0%).

Very similar results were obtained in a 10-fold cross-validation analysis, which was conducted for further verification (average accuracy: 89%, sensitivity: 95.2%,

specificity: 77.3%). Table 3 lists the 10 top-ranked genes, which were identified as being differentially expressed in ER+ luminal and non-luminal samples. These were identified using both in a leave-one-out and a 10-fold cross-validation analysis, i.e. using different subsets of samples, and they were always selected as significantly differentially expressed in each cycle of the analysis (see details on the method in section "Bioinformatics analysis (2)").

Fig 1a shows a box plot of RERG gene (mRNA) expression (normalised expression value) in ER+ versus non-luminal samples with higher expression in ER+ cohort, while Fig 1b shows a box plot of RERG gene (mRNA) expression (normalised expression value) in different tumour grades which shows that the expression of RERG mRNA is higher in low grade tumours. Figure 2 shows a heat map visualization displaying the microarray expression values of 30 genes (rows) in 128 breast cancer samples (columns) using different colour codes (red = high expression, green = low expression). The 30 rows correspond to the 30 top-ranked genes from the cross-validation analysis (see method details in section "Bioinformatics analysis (2)"), grouped according to the results of an average linkage hierarchical clustering using the Euclidean distance metric (see the dendrogram on the left in Figure 2). The 128 columns in this figure represent the microarray samples, grouped into non-luminal samples (left) and ER+ samples (right).

RERG was therefore selected for further study using a protein expression assay to assess the biological and prognostic significance of its protein expression in large breast cancer patient cohort as well as in the ER+ subgroup.

## **Protein expression study**

### **Expression of RERG protein in breast tumours**

Evaluation of RERG protein expression in breast cancer showed that the immunoreactivity was localised to the cytoplasm of invasive tumour cells with homogenous distribution (Fig 3a and 3b) and was strongly expressed in the cytoplasm of luminal cells of normal acini.

Of the whole series, 1,140 informative cases for RERG expression were studied. RERG cytoplasmic expression was scored as negative (no staining) in 28% of cases, low (weak staining hardly visible at low magnification) in 45.6% of cases or high (strong staining easily visible at low magnification) in 26.4%. When studying the expression in relation to BCSS we found no difference in outcome in patients between those with negative and low expression of RERG cases and therefore we combined them into one group of negative/low RERG expression.

### **Association of RERG with clinico-pathological parameters**

Table 4 summarises the associations between RERG expression and clinico-pathological variables in the whole series while Table 5 shows the associations in the ER+ luminal cohort. In the whole series, high RERG expression was positively associated with postmenopausal status ( $p=0.016$ ), tumours of small size, ( $p=0.032$ ), low grade ( $p=0.002$ ), and low mitotic counts ( $p=0.006$ ) with good NPI ( $p=0.006$ ). RERG expression also showed association with tumours that were less likely to develop DM ( $p=0.001$ ) or tumour recurrence ( $p=0.003$ ). No associations were found between RERG and other clinicopathological variables.

In the ER+ luminal cohort, RERG expression showed similar significant associations with regards to tumour size ( $p=0.022$ ), NPI ( $p=0.026$ ), DM ( $p=0.002$ ) and recurrence. ( $p=0.009$ ).

## **Association of RERG with other biomarkers**

In the whole patient series, RERG expression was positively associated with markers of luminal differentiation including ER ( $p=0.001$ ), luminal CKs (CK19 ( $p=0.028$ ), CK18 ( $p=0.001$ ), CK7/8 ( $p=0.013$ )), E-cadherin ( $p=0.001$ ), and the ER-related gene FOXA1 ( $p=0.004$ ), androgen receptor (AR) ( $p<0.001$ ), nuclear BRCA1 ( $p=0.023$ ), FHIT ( $p<0.001$ ) and the cell cycle inhibitors p27 ( $p=0.005$ ) and p21 ( $p=0.023$ ). In contrast, RERG expression was inversely associated with the negative expression of the proliferation marker Ki67 (MIB1) ( $p=0.005$ ) and p53 ( $p=0.050$ ) (Table 6). In the ER+ luminal cohort RERG expression retained similar associations.

## **Patient Outcome**

### **Univariate analysis**

Breast cancer patients with strong RERG expression showed a significantly longer BCSS (LR=12.267,  $p<0.001$ ; Fig 4a) and longer DMFI (LR=7.472,  $p=0.006$ ) (Fig 4b). Similar associations were found in the ER+ luminal group in terms of BCSS (LR=9.887,  $p=0.002$ ; Fig 4c) and DMFI (LR=7.205,  $p=0.007$ ; Fig 4d).

### **According to systemic therapy groups**

The association between RERG protein expression and longer BCSS was also confirmed in the group of patients that did not receive systemic therapy ( $n=397$ ) (LR=6.467,  $p=0.01$ ) (Fig 4e). In ER+ tamoxifen-only treated patients, high RERG expression indicated better response to tamoxifen monotherapy (LR=4.871,  $p=0.027$ ) (Fig 4f).

### **Multivariate analysis**

#### **BCSS**

Multivariate analyses including well-established prognostic variables showed that RERG expression was an independent prognostic marker for longer BCSS in the

whole series (Hazard ratio (HR) =0.573,  $p$  =0.001, 95% CI =0.411-0.799) and in ER+ luminal-like cohort (HR =0.555,  $p$  =0.006, 95% CI =0.364-0.846) (Table 7).

#### DMFI

Multivariate analyses including the same well-established prognostic variables as above showed that RERG expression was not an independent prognostic marker for DMFI in the whole series (HR=0.752,  $p$  =0.053, 95% CI =0.563-1.003) and in ER+ luminal-like cohort (HR =0.748,  $p$  =112, 95% CI =0.522-1.070) (data not shown).



## Discussion

Gene expression studies have classified breast cancer into a number of distinct biological classes which show an association with clinical outcome. The largest “luminal” group is mainly characterised by ER expression [2-4, 29]. However, this class is comprised of tumours showing wide variation at the morphological level, immunophenotypic features and a recognised heterogeneity in terms of response to therapy and clinical outcome indicating that the current molecular classification requires further refinement.

It is thought that novel approaches may be more appropriate for biomarkers discovery in breast cancer. Among the machine learning based methods, Artificial Neural Networks (ANNs) are emerging as valuable tools for this purpose. In the present study, we used a transcript expression profiling of 128 frozen breast cancer cases and analysed the normalised expression values using ANNs. Additionally, a cross-validation analysis in combination with a majority-vote ensemble sample classification was applied in order to obtain a more robust selection and ranking of genes to characterise luminal breast cancer samples. The ensemble classifier combines the prediction results of four algorithms (Support Vector Machine, Random Forest, k-Nearest Neighbour, and Prediction Analysis for Microarrays ) [17]. We identified Ras-related, oestrogen-regulated growth inhibitor (RERG) as a candidate marker for differentiating between luminal and non-luminal BC classes among other genes including GATA3, CA12 and ESR1 which have been reported previously for characterising luminal class membership.

RERG is a GTP-binding protein with intrinsic GTPase activity [11] and was initially identified as one of the genes that characterise luminal A tumours using gene

expression arrays. The expression of RERG has been reported to be decreased in the aggressive ER negative subtypes [2, 4].

In our analysis of BC gene expression data, RERG shown to correlate with high ER expression status using an ANN model and clearly belonged to the best-ranked genes for differentiating between luminal and non-luminal BC, being selected in each cycle a leave-one-out external cross-validation analysis that provided an average classification accuracy of 88.3%. Although RERG gene expression has been reported to be associated with ER+ breast cancer in our analysis and previous studies [4], its protein expression has not been studied in BC. To validate the gene expression findings we studied RERG protein expression in invasive BC using TMAs and immunohistochemistry. We found good agreement between protein and gene expression results, highlighting the importance of RERG as a candidate luminal marker. RERG protein expression showed a highly significant association with luminal CKs, FOXA1, E-cadherin, steroid receptors, BRCA1, cell cycle inhibitors (p21 and p27) and FHIT which are all markers associated with good prognosis and luminal phenotype. *In vitro* studies also lend support to its association with good-prognostic phenotypes because RERG mRNA expression is induced rapidly in ER-responsive MCF-7 cells stimulated by estradiol and repressed by ER-antagonist tamoxifen treatment[11].

In contrast, we found an inverse relation between RERG protein expression and indicators of cell proliferation such as tumour grade, mitosis and MIB1 expression and this observation of altered cellular proliferation has been proposed as an explanation for the difference in prognosis seen within luminal tumours. **The association of RERG with high tumour grade was also seen on mRNA level as shown in Fig 1a.** In agreement with this observation, our protein expression results confirmed

that RERG expression is associated with tumours displaying low MIB1 expression supporting the growth inhibitory function of RERG. For these reasons, we propose that within the luminal classes, expression of ER, RERG and MIB1 could be used to define biological subgroups with different prognoses. ER<sup>+</sup> RERG<sup>+</sup> MIB<sup>low</sup> could represent a luminal subgroup with good prognosis while those with an ER<sup>+</sup> RERG<sup>-</sup> MIB1<sup>high</sup> phenotype could define a luminal subclass with poor prognosis. Further work is currently underway to better characterise the luminal subclasses with emphasis on predicting prognostic signatures.

A key aim of this study was to assess the prognostic ability of RERG protein expression in ER-positive luminal-like BC patients. In this important group of patients, we found that RERG expression was significantly associated with longer BCSS and longer DMFI which implies its role in subclassification of ER-positive groups into prognostic subgroups. To overcome the confounding effects of systemic therapy, these findings were also confirmed in a cohort of patients that did not receive adjuvant treatment.

In conclusion, a novel bioinformatics approach was used to analyse high dimensional datasets and we were able to identify candidate genes to characterise the ER<sup>+</sup> luminal like BC. Subsequently, these genes were translated into protein biomarkers that were shown to subclassify BC in terms of biology and prognosis. In particular, we propose that RERG can differentiate between patients with luminal-like tumours with good and poor clinical outcome and as such it could be a useful marker for a more precise definition of the luminal phenotype.

## **Acknowledgement**

We thank the ministry of higher education (Egypt) for funding HO Habashy and E Rakha. Funding (salaries and infrastructure) was provided by the University of Nottingham and Nottingham University Hospitals Trust. N Krasnogor and E Glaab would like to acknowledge funding by Marie Curie Early Stage Training Programme (MEST-CT-2004-007597).

All authors declare no conflict of interest.

## References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**:747 - 752.
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *PNAS* 2001, **98**(19):10869-10874.
3. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: **Breast cancer classification and prognosis based on gene expression profiles from a population-based study**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(18):10393-10398.
4. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: **Repeated observation of breast tumor subtypes in independent gene expression data sets**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(14):8418-8423.
5. Habashy H, Powe D, Rakha E, Ball G, Macmillan R, Green A, Ellis I: **The prognostic significance of PELP1 expression in invasive breast cancer with emphasis on the ER-positive luminal-like subtype**. *Breast Cancer Research and Treatment*, **120**(3):603-612.
6. Habashy H, Powe D, Staka C, Rakha E, Ball G, Green A, Aleskandarany M, Paish E, Douglas Macmillan R, Nicholson R *et al*: **Transferrin receptor (CD71) is a marker of poor prognosis in breast cancer and can predict response to tamoxifen**. *Breast Cancer Research and Treatment*, **119**(2):283-293.
7. Habashy HO, Powe DG, Rakha EA, Ball G, Paish C, Gee J, Nicholson RI, Ellis IO: **Forkhead-box A1 (FOXA1) expression in breast cancer and its prognostic significance**. *European Journal of Cancer* 2008, **44**(11):1541-1551.
8. Lancashire LJ, Rees RC, Ball GR: **Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach**. *Artificial Intelligence in Medicine* 2008, **43**(2):99-111.
9. Rumelhart DE, Hinton GE, Williams RJ: **LEARNING REPRESENTATIONS BY BACK-PROPAGATING ERRORS**. *Nature* 1986, **323**(6088):533-536.
10. Lancashire L, Powe D, Reis-Filho J, Rakha E, Lemetre C, Weigelt B, Abdel-Fatah T, Green A, Mukta R, Blamey R *et al*: **A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks**. *Breast Cancer Research and Treatment*, **120**(1):83-93.
11. Finlin BS, Gau CL, Murphy GA, Shao HP, Kimel T, Seitz RS, Chiu YF, Botstein D, Brown PO, Tamanoi F *et al*: **RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer**. *Journal of Biological Chemistry* 2001, **276**(45):42259-42267.
12. Chin S, Teschendorff A, Marioni J, Wang Y, Barbosa-Morais N, Thorne N, Costa J, Pinder S, van de Wiel M, Green A *et al*: **High-resolution aCGH and**

- expression profiling identifies a novel genomic subtype of ER negative breast cancer.** *Genome Biology* 2007, **8**(10):R215.
13. Blenkiron C, Goldstein L, Thorne N, Spiteri I, Chin S-F, Dunning M, Barbosa-Morais N, Teschendorff A, Green A, Ellis I *et al*: **MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype.** *Genome Biology* 2007, **8**(10):R214.
  14. Zhang H, Rakha E, Ball G, Spiteri I, Aleskandarany M, Paish E, Powe D, Macmillan R, Caldas C, Ellis I *et al*: **The proteins FABP7 and OATP2 are associated with the basal phenotype and patient outcome in human breast cancer.** *Breast Cancer Research and Treatment*, **121**(1):41-51.
  15. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JFR, Macmillan D, Blamey RW, Ellis IO: **High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses.** *International Journal of Cancer* 2005, **116**(3):340-350.
  16. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
  17. Tibshirani R, Hastie T, Narasimhan B, Chu G: **Diagnosis of multiple cancer types by shrunken centroids of gene expression.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(10):6567-6572.
  18. Glaab E, Garibaldi J, Krasnogor N: **ArrayMining: a modular web-application for microarray analysis combining ensemble and consensus methods with cross-study normalization.** *BMC Bioinformatics* 2009, **10**(1):358.
  19. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW: **PATHOLOGICAL PROGNOSTIC FACTORS IN BREAST-CANCER .2. HISTOLOGICAL TYPE - RELATIONSHIP WITH SURVIVAL IN A LARGE STUDY WITH LONG-TERM FOLLOW-UP.** *Histopathology* 1992, **20**(6):479-489.
  20. Elston CW, Ellis IO: **Pathological Prognostic Factors in Breast-Cancer .1. the Value of Histological Grade in Breast-Cancer - Experience from a Large Study with Long-Term Follow-Up.** *Histopathology* 1991, **19**(5):403-410.
  21. Galea MH, Blamey RW, Elston CE, Ellis IO: **The Nottingham Prognostic Index in Primary Breast-Cancer.** *Breast Cancer Research and Treatment* 1992, **22**(3):207-219.
  22. Madjd Z, Pinder SE, Paish C, Ellis IO, Carmichael J, Durrant LG: **Loss of CD59 expression in breast tumours correlates with poor survival.** *J Pathol* 2003, **200**(5):633-639.
  23. Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, Blamey RW, Ellis IO: **Prognostic significance of Nottingham histologic grade in invasive breast carcinoma.** *J Clin Oncol* 2008, **26**(19):3153-3158.
  24. Galea MH, Blamey RW, Elston CE, Ellis IO: **The Nottingham Prognostic Index in primary breast cancer.** *Breast Cancer Res Treat* 1992, **22**(3):207-219.
  25. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA *et al*: **Triple-**

- Negative Breast Cancer: Distinguishing between Basal and Nonbasal Subtypes.** *Clinical Cancer Research* 2009, **15**(7):2302-2310.
26. Aleskandarany MA, Green AR, Rakha EA, Mohammed RA, Elsheikh SE, Powe DG, Paish EC, Macmillan RD, Chan S, Ahmed SI *et al*: **Growth fraction as a predictor of response to chemotherapy in node-negative breast cancer.** *International Journal of Cancer*, **126**(7):1761-1769.
27. Abd El-Rehim DM, Pinder SE, Paish CE, Bell JA, Rampaul RS, Blamey RW, Robertson JFR, Nicholson RI, Ellis IO: **Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma.** *British Journal of Cancer* 2004, **91**(8):1532-1542.
28. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM: **REporting recommendations for tumor MARKer prognostic studies (REMARK).** *Breast Cancer Research and Treatment* 2006, **100**(2):229-235.
29. Sorlie T, Perou CM, Fan C, Geisler S, Aas T, Nobel A, Anker G, Akslen LA, Botstein D, Borresen-Dale AL *et al*: **Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer.** *Molecular Cancer Therapeutics* 2006, **5**(11):2914-2918.

Table 1: Source, dilution and pretreatment of antibodies used

Antibody (clone)	Dilution	Source	Pretreatment	Cut-off
<b>Hormone receptors and ER-related proteins</b>				
ER (clone 1D5)	1:80	Dako Cytomation	Microwave	20%
PR (clone PgR 636)	1:100	Dako Cytomation	Microwave	20%
AR (clone F39.4.1)	1:30	Biogenex	Microwave	20%
FOXA1(clone2F83)	1:2000	ABCAM, UK	Microwave	10*
<b>Luminal and myoepithelial/basal associated markers and adhesion molecules</b>				
<b><u>Luminal associated CKs</u></b>				
CK18 (clone DC10)	1:50	Dako Cytomation	Microwave	50*
CK19 (clone BCK 108)	1:100	Dako Cytomation		50*
<b><u>Basal CKs</u></b>				
CK5/6(cloneD5/16134)	1:100	Boehringer	Microwave	10%
CK 14 (clone LL002)	1:100	Biochemica Novocastra		10%
<b><u>Other basal/ME associated markers</u></b>				
SMA (clone 1A4)	1:2000	Dako Cytomation	No	10%
<b><u>Cell adhesion molecules</u></b>				
Anti E-cadherin (clone HECD-1)	1:100	Zymed Laboratories	Microwave	100*
Anti P-cadherin (clone 56)	1:200	BD Biosciences		5%
<b>Oncogenes and Tumour suppressor genes</b>				
HER-2 (cerbB-2)	1:250	Dako Cytomation	Microwave	#
p53 (clone DO7)	1:50	Novocastra		10%
BRCA1 Ab-1 (clone MS110)	1:150	Oncogene		5%
Anti-FHIT (clone ZR44)	1:600	Research Zymed Laboratories		5%
<b>Cell cycle associated, proliferation and apoptosis-related proteins</b>				
p21	1:100	Dako, UK	Microwave	5%
p27	1:25	Dako, Denmark		50%
BCL2 (clone124)	1:100	Dako, UK		30%
MIB1	1:100	Dako		10%

# =scored according to Hercept test guidelines

\*=H-score

^=Scored as (0, absent; 1, weak; 2, moderate; 3, strong cytoplasmic expression)



Table 2: A gene signature of ER expression status (summary of step 1 of the ANN approach- 10 genes shown)

<b>Gene</b>	<b>Selection Error</b>
ESR1	0.403422327
RERG	0.438626499
AMN1	0.441492448
ZNF271	0.445580899
PCDHA5	0.446326207
PRKAR2B	0.447933195
TCEAL1	0.448787999
CTBP2	0.449052205
LDB3	0.449750785
DDIT4	0.451080472

Table 3: A gene signature to differentiate between luminal (ER+) and non-luminal cases (ER-) using cross validation analysis

<b>Gene identifier</b>	<b>Gene</b>
GI_4503602-S	ESR1
GI_14249703-S	RERG
GI_9951924-S	CA12
GI_37551139-S	C6orf115
GI_34452698-S	ACTR3
GI_22779933-S	WDR19
GI_38455428-S	AGR3
GI_38146007-A	TTC8
GI_40788002-S	PSME4
GI_4503928-S	GATA3

Table 4: Relation of RERG immunostaining with various clinicopathological parameters in whole series of breast cancer patients

Variable	Negative	Positive	$\chi^2$	P value
<b>Age</b>			4.870	0.182
<40	70(79.5)	18(20.5)		
40-50	251(76.1)	79(23.9)		
51-60	277(73.3)	101(26.7)		
>60	241(70.1)	103(29.9)		
<b>Size</b>			4.604	0.032
≤2 cm	406(70.9)	167(29.1)		
>2 cm	432(76.5)	133(23.5)		
<b>Lymph Node (LN) Stage</b>			1.334	0.513
1(Negative)	497(72.6)	188(27.4)		
2(1-3 LN)	261(74.6)	89(25.4)		
3(>3 LN)	79(77.5)	23(22.5)		
<b>Grade</b>			12.419	0.002
1	127(66.8)	63(33.2)		
2	260(70.3)	110(29.7)		
3	451(78)	127(22)		
<b>NPI(Nottingham Prognostic Index)</b>			10.330	0.006
<b>Good</b>	212(67.3)	103(32.7)		
<b>Moderate</b>	470(75)	157(25)		
<b>Poor</b>	157(79.3)	41(20.7)		
<b>DM (Distant Metastasis)</b>			11.948	0.001
No	565(70.9)	232(29.1)		
Positive	266(80.9)	63(19.1)		
<b>Recurrence</b>			8.642	0.003
No	474(70.6)	197(29.4)		
Positive	348(78.6)	95(21.4)		
<b>Tumour type</b>			4.783	0.443
Ductal/NST	498(75)	166(25)		
Lobular	94(72.9)	35(27.1)		
Tubular and Tubular mixed	163(70.6)	68(29.4)		
Medullary	25(83.3)	5(16.7)		
Other special types*	12(75)	4(25)		
Mixed**	47(67.1)	23(32.9)		
<b>Mitotic Frequency</b>			10.274	0.006
1	268(68.7)	122(31.3)		
2	144(73.5)	52(26.5)		
3	407(78.1)	114(21.9)		
<b>Menopause</b>			6.116	0.016
Premenopausal	341(77.7)	98(22.3)		
Postmenopausal	498(71)	203(29)		

\*Includes Mucoïd, invasive cribriform and invasive papillary carcinoma, \*\* Include ductal/NST mixed with lobular or special types

Table 5: Relation of RERG immunostaining with various clinicopathological parameters in ER+ cohort of breast cancer patients

Variable	Negative	Positive	$\chi^2$	P value
<b>Age</b>			1.875	0.599
<40	31(70.5)	13(29.5)		
40-50	160(73.4)	58(26.6)		
51-60	187(72.2)	72(27.8)		
>60	177(68.1)	83(31.9)		
<b>Size</b>			5.335	0.021
≤2 cm	288(67.6)	138(32.4)		
>2 cm	266(75.1)	88(24.9)		
<b>LN Stage</b>			2.044	0.360
1(Negative)	325(69.1)	145(30.5)		
2(1-3 LN)	182(73.4)	66(26.6)		
3(>3 LN)	46(75.4)	15(24.6)		
<b>Grade</b>			5.379	0.068
1	111(65.7)	58(34.3)		
2	227(69.8)	98(30.2)		
3	216(75.5)	70(24.5)		
<b>NPI</b>			7.313	0.026
<b>Good</b>	185(65.8)	96(34.2)		
<b>Moderate</b>	285(72.7)	107(27.3)		
<b>Poor</b>	85(78.7)	23(21.3)		
<b>DM</b>			9.537	0.002
No	382(68.2)	178(31.8)		
Positive	170(79.4)	44(20.6)		
<b>Recurrence</b>			7.189	0.007
No	321(67.9)	152(32.1)		
Positive	226(76.9)	68(23.1)		
<b>Tumour type</b>			3.301	0.654
Ductal/NST	272(70.6)	113(29.4)		
Lobular	90(76.9)	27(23.1)		
Tubular and Tubular mixed	142(70)	61(30)		
Medullary	3(75)	1(25)		
Other special types*	9(75)	3(25)		
Mixed**	39(65)	21(35)		
<b>Mitosis</b>			3.223	0.200
1	238(68.6)	109(31.4)		
2	116(72)	45(28)		
3	186(75.3)	61(24.7)		
<b>Menopause</b>			2.984	0.084
Premenopausal	203(74.9)	68(25.1)		
Postmenopausal	352(69)	158(31)		

\*Includes Mucoid, invasive cribriform and invasive papillary carcinoma, \*\* Include ductal/NST mixed with lobular or special types

Table 6: Relation of RERG immunostaining with various biomarkers parameters in whole cohort of breast cancer patients

Variable	Negative	Positive	$\chi^2$	p-value
<b>CK 5/6</b>				
Negative	688(73.7)	245(26.3)	0.003	0.954
Positive	125(73.5)	45(26.5)		
<b>CK 14</b>				
Negative	711(73.8)	253(26.2)	0.290	0.590
Positive	92(76)	29(24)		
<b>CK18</b>				
Negative	118(84.9)	21(15.1)	10.823	0.001
Positive	647(71.7)	256(28.3)		
<b>CK19</b>				
Negative	90(82.6)	19(17.4)	4.829	0.028
Positive	726(72.8)	271(27.2)		
<b>CK7/8</b>				
Negative	17(100)	0(0)	6.128	0.013
Positive	800(73.4)	290(26.6)		
<b>ER</b>				
Negative	249(80.8)	59(19.2)	10.938	0.001
Positive	555(71.1)	226(28.9)		
<b>PgR</b>				
Negative	384(74)	122(26)	0.026	0.872
Positive	449(73.6)	161(26.4)		
<b>AR</b>				
Negative	300(82.9)	62(17.1)	23.614	<0.001
Positive	469(69)	211(31)		
<b>P53</b>				
Negative	571(71.7)	225(28.3)	3.839	0.050
Positive	223(77.7)	64(22.3)		
<b>FHIT</b>				
Negative	158(84.9)	28(15.1)	14.235	<0.001
Positive	574(71.5)	229(28.5)		
<b>BRCA1</b>				
Negative	114(82)	25(18)	5.163	0.023
Positive	586(72.9)	218(27.1)		
<b>Bcl2</b>				
Negative	268(77.5)	78(22.5)	2.280	0.131
Weak	385(72.9)	143(27.1)		
<b>MIB1</b>				
low	172(67.7)	82(32.3)	7.915	0.005
High	488(76.9)	147(23.1)		
<b>P-Cad</b>				
Negative	321(71.7)	127(28.3)	2.601	0.107
Positive	382(76.2)	119(23.8)		
<b>E-Cad</b>				
Negative	332(79.6)	85(20.4)	11.370	0.001
Positive	473(70.4)	199(29.6)		
<b>FOXA1</b>				
Negative	344(78.9)	92(21.1)	8.082	0.004
Positive	272(70.3)	115(29.7)		
<b>HER2</b>				
Negative	711(73.8)	253(26.2)	0.243	0.622
Positive	103(75.7)	33(24.3)		
<b>Actin</b>				
Negative	678(73)	251(27)	2.631	0.105
Positive	125(79.1)	33(29.1)		
<b>p21</b>				
Negative	366(77.9)	104(22.1)	5.162	0.023
Positive	281(71.1)	114(28.9)		
<b>p27</b>				
Negative	317(79.1)	84(20.9)	7.711	0.005
Positive	264(70.4)	111(29.6)		

Table 7: COX model for predictors of BCSS in the whole patient series and in the ER+ luminal subgroup

Variable	Whole series				ER-positive cohort			
	p value	HR	95 % CI		p value	HR	95 % CI	
			Lower	Upper			Lower	Upper
RERG expression	0.001	0.573	0.411	0.799	0.006	0.555	0.364	0.846
Endocrine therapy	0.020	0.670	0.478	0.938	0.126	0.713	0.463	1.100
Chemotherapy	0.001	0.508	0.338	0.762	0.028	0.521	0.292	0.931
Tumour size (>2cm)	<0.001	1.951	1.472	2.585	<0.001	2.083	1.464	2.963
Tumour stage 1*	<0.001	1			<0.001	1		
Tumour stage 2 vs. 1	<0.001	1.814	1.341	2.454	0.004	1.769	1.201	2.606
Tumour stage 3 vs. 1	<0.001	4.604	3.204	6.617	<0.001	3.948	2.388	6.526
Tumour grade 1*	<0.001	1			<0.001	1		
Tumour grade 2 vs.1	0.011	1.982	1.174	3.348	0.021	1.955	1.107	3.454
Tumour grade 3 vs. 1	<0.001	4.175	2.495	6.985	<0.001	4.487	2.483	8.107

\*fitted as categorical variable

## **Legends**

### **Tables**

Table 1: Source, dilution and pretreatment of antibodies us

Table 2: A gene signature of ER expression status (summary of step 1 of the ANN approach)

Table 3: A gene signature to differentiate between luminal (ER+) and non-luminal cases (ER-) using cross validation analysis

Table 4: Relation of RERG immunostaining with various clinicopathological parameters in whole series of breast cancer patients

Table 5: Relation of RERG immunostaining with various clinicopathological parameters in ER+ cohort of breast cancer patients

Table 6: Relation of RERG immunostaining with various biomarkers parameters in whole cohort of breast cancer patients

Table 7: COX model for predictors of BCSS

### **Figures legends**

Figure 1:

- (a) Box plot of RERG gene (mRNA) expression (normalised expression value) in ER+ Luminal versus non-luminal samples
- (b) Box plot of RERG gene (mRNA) expression (normalised expression value) in different tumour grades.

Figure 2:

A heatmap created to visualise the differential expression of the 30 top-ranked genes, grouping the samples into ER+ and non-luminal samples and applying an average

linkage hierarchical clustering using the Euclidean distance metric to the 30 gene expression vectors.

Figure 3:

Grade II invasive ductal carcinoma with high RERG expression (a) lower magnification (b) higher magnification

Figure 4:

Figure 4a: Kaplan Meier plot of RERG protein expression with respect to BCSS. The expression was significantly associated with improved survival (LR=12.267,  $p<0.001$ ).

Figure 4b: Kaplan Meier plot of RERG expression with respect of DMFI shows a significant longer DMFI in patients with high RERG expression (LR=7.472,  $p=0.006$ )

Figure 4c: Kaplan Meier plot of RERG expression with respect to BCSS in ER+ luminal like cohort (LR=9.887,  $p=0.002$ ).

Figure 4d: Kaplan Meier plot of RERG expression with respect of DMFI in ER+ luminal like cohort (LR=7.205,  $p=0.007$ ).

Figure 4e: Kaplan Meier plot of RERG expression patterns with respect to BCSS in patient cohort that didn't receive adjuvant systemic therapy (LR=6.467,  $p=0.011$ )

Figure 4f: Kaplan Meier plot of RERG expression patterns with respect to BCSS in tamoxifen only treated patients (LR=4.871,  $p=0.027$ )









