



Article

Rho-GTPase activating-protein 18: a biomarker associated with good prognosis in invasive breast cancer

Aleskandarany, Mohammed A, Sonbul, Sultan, SurrIDGE, Rachel, Mukherjee, Abhik, Caldas, Carlos, Diez-Rodriguez, Maria, Ashankyty, Ibraheem, Albrahim, Khalil I, Elmouna, Ahmed M, Aneja, Ritu, Martin, Stewart G, Ellis, Ian O, Green, Andrew R and Rakha, Emad A

Available at <http://clock.uclan.ac.uk/34541/>

Aleskandarany, Mohammed A ORCID: 0000-0001-5538-5946, Sonbul, Sultan, SurrIDGE, Rachel, Mukherjee, Abhik, Caldas, Carlos, Diez-Rodriguez, Maria, Ashankyty, Ibraheem, Albrahim, Khalil I, Elmouna, Ahmed M et al (2017) Rho-GTPase activating-protein 18: a biomarker associated with good prognosis in invasive breast cancer. British Journal of Cancer, 117 . pp. 1176-1184. ISSN 0007-0920

It is advisable to refer to the publisher's version if you intend to cite from the work.
<http://dx.doi.org/10.1038/bjc.2017.261>

For more information about UCLan's research in this area go to <http://www.uclan.ac.uk/researchgroups/> and search for <name of research Group>.

For information about Research generally at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the [policies](#) page.

Keywords: ARHGAP18; lymphovascular invasion; immunohistochemistry; breast cancer; prognosis

Rho-GTPase activating-protein 18: a biomarker associated with good prognosis in invasive breast cancer

Mohammed A Aleskandarany^{*,1,2}, Sultan Sonbul¹, Rachel Surridge¹, Abhik Mukherjee¹, Carlos Caldas³, Maria Diez-Rodriguez¹, Ibraheem Ashankyty⁴, Khalil I Albrahim⁴, Ahmed M Elmouna⁴, Ritu Aneja⁵, Stewart G Martin¹, Ian O Ellis¹, Andrew R Green¹ and Emad A Rakha^{1,2}

¹Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham NG5 1PB, UK; ²Pathology Department, Faculty of Medicine, Menoufyia University, Menoufyia 110532, Egypt; ³Cancer Research UK, Cambridge Research Institute, Cambridge CB 0RE, UK; ⁴Molecular Diagnostics and Personalised Therapeutics Unit, University of Ha'il, Ha'il 2440, Saudi Arabia and ⁵Department of Biology, Georgia State University, Atlanta, GA 30303, USA

Background: The prognostic value of lymphovascular invasion (LVI) in breast cancer (BC) has been demonstrated in several independent studies. However, identification of driver molecules for LVI remains a challenging task. Large-scale transcriptomic profiling of histologically validated LVI can potentially identify genes that regulate LVI.

Methods: Integrative bio-informatics analyses of the METABRIC study were performed utilising a subset of strictly defined LVI using histological and immunohistochemical (IHC) criteria. *ARHGAP18* was among the top differentially expressed genes between LVI+ and LVI- BC with a 1.8-fold change. The prognostic impact of *ARHGAP18* gene expression was assessed in the METABRIC data set ($n = 1980$) and externally validated using the online BC gene expression data sets utilising bc-GenExMiner v4.0 ($n = 2016$). Subsequently, *ARHGAP18* protein expression was assessed on a large cohort of invasive BC ($n = 959$) with long-term follow-up using IHC.

Results: Pooled analysis of *ARHGAP18* mRNA expression showed that overexpression was associated with better outcome ($P < 0.001$, hazard ratio (HR) = 0.82, 95% CI 0.75–0.90). *ARHGAP18* protein was expressed in the cytoplasm and nuclei of the tumour cells and its expression was positively associated with good prognostic variables. Lack of cytoplasmic expression showed associations with LVI ($P = 0.006$), epithelial-mesenchymal transition and the HER+ subtype ($P = 0.01$). Loss of nuclear expression was associated with higher grade, HER2+ and high Ki67LI ($P = 0.001$). Cytoplasmic and nuclear expression showed a positive association with improved survival independent of other variables ($P = 0.01$, HR = 0.74, 95% CI 0.60–87).

Conclusions: *ARHGAP18* expression at transcriptomic and protein levels is associated with improved patients' outcomes whose deregulation may play a role in tumour progression and the development of LVI in BC. Further assessment of its potential therapeutic value in BC is warranted.

Lymphovascular invasion (LVI) is an indicator of metastatic potential and poor outcome in breast cancer (BC) (Lee *et al*, 1990; Quinlan, 1993; Altman and Bland, 1994; Lauria *et al*, 1995; Choi *et al*, 2003; Truong *et al*, 2005; Debled *et al*, 2010; Song *et al*, 2011; Ugras *et al*, 2014). In a previous study (Quinlan, 1993) we have demonstrated that in a lymph node-negative BC cohort, LVI could

*Correspondence: Dr MA Aleskandarany; E-mail: mohammed.aleskandarany@nuh.nhs.uk or m.aleskandarany@gmail.com

Received 7 April 2017; revised 3 July 2017; accepted 14 July 2017; published online 22 August 2017

© 2017 Cancer Research UK. All rights reserved 0007–0920/17

be used as a high-risk criterion conferring survival disadvantage equivalent to that provided by involvement of one or two lymph nodes and to that provided by one higher size category (pT1 to pT2) (Quinlan, 1993). Despite its recognised prognostic roles, the molecular mechanisms underlying the development of LVI in BC and genes driving this process that could be used as potential therapeutic targets remain largely unknown (reviewed in Aleskandarany *et al*, 2015). This is related not only to the complexity of the mechanisms involved in the development of LVI as part of the invasion-metastasis process and the complex interaction between tumour cells and their microenvironment but also due to the difficulty in studies' design and the subjectivity of identification of true LVI-negative and LVI-positive BC. Here we hypothesised that large-scale transcriptomic and genomic profiling of a cohort with strictly defined LVI status could potentially yield candidate key/driver genes. This approach is likely to narrow down the number of potential genes that can be further investigated using *in vitro* and *in vivo* assays models with the aim of identifying novel therapeutic targets.

In this study, the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) data were used for exploring genes associated with LVI in invasive BC through extensive genome-wide analysis of gene expression and copy number aberrations (CNAs) (Curtis *et al*, 2012). LVI was defined using strict criteria with the aid of immunohistochemistry (IHC). Genes with significant differential expression were ranked according to the strength of their association with LVI. The CNAs of these genes were determined. One of the top genes that is currently gaining interest as a therapeutic target (Rao *et al*, 2015a) showed differential expression together with CNAs was the Rho-GTPase activating-protein 18 (*ARHGAP18*). The prognostic impact of *ARHGAP18* gene expression was externally validated using online BC gene expression data sets. Protein expression of *ARHGAP18* was assessed on a large clinically annotated cohort of invasive BC with long-term follow-up using IHC. The association between *ARHGAP18* and LVI as well as clinicopathologic criteria and patients' outcome were explored.

MATERIALS AND METHODS

Gene identification using METABRIC cohort. To identify putative LVI driver genes, the METABRIC cohort data were explored as regards the gene expression and CNA between cases with definite LVI and cases with no evidence of LVI. Details of tissue processing, analysis and interpretation of the findings of the METABRIC study were previously described (Curtis *et al*, 2012). Briefly, the extracted and purified DNA probes were hybridised to Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA, USA) at AROS Applied Biotechnology (Aarhus, Denmark). The Illumina Totalprep RNA amplification kit (Ambion, Warrington, UK) was used, and then hybridised onto Illumina Human HT-12 v3 Expression Beadchips from the same manufacturer. For this purpose, the discovery set formed of cases from Nottingham primary series of invasive BC included in the METABRIC study ($n = 328$). Tumours in this discovery set were treated using standardised methods of specimen fixation, sampling and processing as previously described (Rakha *et al*, 2012). The clinicopathological data including LVI status, axillary nodal status and other variables were available. The LVI status of these cases was determined using a set of criteria as follows: cases were defined as LVI positive (LVI+) when LVI was reported in routine practice as positive based on H&E sections of the whole tumours and LVI was confirmed using IHC using CD34 and/or D2-40 (Rakha *et al*, 2012) on one full-face tumour section from the surgical specimens. LVI-negative (LVI-) cases were defined by negativity of LVI on

both H&E sections and IHC. Importantly, cases in the discovery set with positive lymph nodes in the LVI- subgroup were excluded to avoid bias caused by potentially undetectable foci of LVI in the primary tumour. This discovery set was used to identify genes differentially expressed between the two LVI subgroups that are likely to be related to LVI. The validation set (Addenbrookes cases; $n = 914$ cases) set was used to validate the expression of the differentially expressed genes, to ensure that genes were differentially expressed in both the Nottingham and Addenbrookes cohorts. To test for genes/transcripts significantly differentially expressed between LVI+ and LVI- cases, a supervised differential gene expression analysis approach of gene expression data was followed. This was performed using the Linear Model for Microarray and RNA-seq data (LIMMA) software package that is compatible with the Affymetrix data. The Affymetrix SNP 6.0 array data were further analysed for CNAs of each transcript in the list of top associated genes with LVI using Affymetrix SNP6 Copy Number Inference Pipeline (Cancer Genomics Computation Analysis group of the Broad Institute, USA). The top differentially expressed genes were ranked based on their *P*-value of association with LVI. Subsequently the CNAs of the top differentially expressed genes were determined. Of the top five differentially expressed genes, *ARHGAP18* was identified.

ARHGAP18 gene expression. *ARHGAP18* gene expression was evaluated in the METABRIC cohort of 1980 BC samples (Curtis *et al*, 2012). In this cohort, patients with ER-positive and/or lymph node-negative tumours did not receive adjuvant chemotherapy, while those with ER-negative and/or lymph node-positive tumours received adjuvant chemotherapy. Dichotomisation of *ARHGAP18* mRNA expression was performed at the median mRNA expression of the studied cohort.

External validation. To further assess the prognostic significance of *ARHGAP18* mRNA expression, bc-GenExMiner v4.0 (Breast Cancer Gene-Expression Miner v3.0) online data set (<http://bcgenex.centregauducheau.fr>) was used. This is composed of two statistical mining modules; the 'prognostic module', offering the possibility to evaluate the *in vivo* prognostic impact of candidate genes in BC, and the 'correlation module', to compute correlation coefficients between gene expressions or to find lists of correlated genes in BC. The prognostic module was in this external validation, where Cox model, Kaplan-Meier and forest plots were performed (Jezequel *et al*, 2012). The prognostic significance was used as LVI status was not available for these cohorts.

ARHGAP18 protein expression. *ARHGAP18* protein expression was assessed in formalin-fixed paraffin-embedded tissue samples from BC cases enrolled into the Nottingham Tenovus Breast Carcinoma Series: a well-characterised series of primary operable invasive BC presenting between 1987 and 1998 at Nottingham City Hospital ($n = 959$). Clinical and pathological data had been recorded and patients' outcome data are regularly updated and prospectively maintained. Outcome data recorded included survival status, mean survival in months, recurrence of disease (including distant metastases) and cause of death. Disease-free interval (DFI) was defined as months from date of primary surgery to appearance of local, regional, or distant recurrence, breast cancer-specific survival (BCSS) was defined as months from primary surgery until BC-specific death, and time to distant metastasis (distant metastasis-free interval, DMFI) was defined as months from primary surgery to occurrence of first distant metastasis/recurrence. Adjuvant therapy was based upon tumour prognostic and predictive factors including the Nottingham Prognostic Index (NPI), lymph node (LN) status, menopausal status and ER positivity/negativity and included hormonal therapy and CMF (cyclophosphamide, methotrexate and 5-Fluorouracil).

Expression of a large number of molecular biomarkers relevant to BC tumourigenesis and progression including oestrogen receptor (ER), progesterone receptors (PR), human epidermal growth factor receptor 2 (HER2), p53, Ki67, and others have been previously studied (Abd El-Rehim *et al*, 2005). The tumours were subtyped into luminal (ER + /PR + /HER2 -), HER2 +, and triple negative (PR - /ER - /HER2 -) subtypes (Rakha *et al*, 2009). Table 1 summarises the clinicopathological data of this cohort. The median and mean age of the study population was 54 years (range: 24–70). The median overall survival in months was 168 months while the median DFI was 109 months. This work was approved by Nottingham Research Ethics Committee 2 under the title: Development of a molecular genetic classification of breast cancer. All tissue samples included in this study were from patients who were consented prior to inclusion in the study cohort.

Validation of primary antibody specificity using western blotting. Prior to IHC, the specificity of the anti-ARHGAP18 antibody (ab175970, Abcam, UK) was validated using western blotting performed on whole cell lysates of high expressor cells of ARHGAP18 mRNA according to the manufacturer's protocol. For this purpose, lysates of human cervical cancer (HeLa CCL-2, American Type Culture Collection (ATCC), USA) cell lines and human breast adenocarcinoma (MDA-MB-231, ATTC HTB-26, USA) cell lines were used. This was performed using 1:1000 dilution of the primary antibody (ab175970, Abcam, UK), and 1:15000 of the horseradish peroxidase-labelled secondary anti-rabbit antibody, with β -actin (diluted 1:2000) used as a loading control. Bovine serum albumin was used for blocking. Enhanced chemiluminescence was used to visualise the membrane as previously described (Mruk and Cheng, 2011). The images were developed via chemiluminescence using an Odyssey Fc (Li-cor Biosciences, Lincoln, NE, USA).

Immunohistochemistry. Heat-assisted antigen retrieval was performed in accordance to the manufacturer's recommendations (citrate buffer pH 6.0 at 1000 W for 20 min using microwave). The Novolink Max Polymer Detection System was used to visualise the reaction (RE7280-k, Leica, Newcastle, UK). Endogenous peroxidase activity was blocked by the addition of a peroxidase block (Novolink Peroxidase Block) for 5 min. Non-specific binding of the primary antibody, post primary block (Novolink Protein block) was used. The slides were washed and incubated with the anti-ARHGAP18 primary antibody (ab175970, Abcam, UK, diluted 1:75 in Bond Primary Antibody Diluent, Leica, Germany), for 30 min at room temperature. This working dilution was chosen based on repeated attempts of optimisation using different dilutions, guided by the manufacturer's recommendation, in order to achieve specific staining with minimal background. The positive control, β_2 microglobulin (Polyclonal Rabbit Anti-Human Beta-2-Microglobulin, Dako, Glostrup, Denmark) was incubated at 1:2000 dilution. A DAB (3,3'-diaminobenzadine) in a buffered stabiliser solution (Novolink DAB Chromogen; 1:20 dilution) was added to the slides for 5 min. The slides were counterstained with haematoxylin for 6 min (Novolink Haematoxylin). Dehydration, clearing, mounting and cover-slipping were performed as previously described. Full-face tissue sections from 25 BC excision specimens were also stained for ARHGAP18 to assess for the distribution of staining to decide for TMAs suitability. Negative and positive controls (by omission of the primary antibody, and human liver tissues, respectively) were included in each staining run.

Scoring of TMA cores. Stained TMA slides were digitally scanned (NanoZoomer, Hamamatsu Photonics, Tokyo, Japan) to high resolution images and viewed by Aperio ImageScope (Aperio Technologies, Milton Keynes, UK) at $\times 20$ magnification. Cores/

Table 1. Summary of patient demographics of the Nottingham primary BC series used in this study

Parameter	Number (%)
Age	
≤ 50 years	367 (38.3)
> 50 years	591 (61.6)
Median/mean	54
Menopausal status	
Premenopausal	381 (39.7)
Postmenopausal	577 (60.2)
Tumour grade	
1	129 (13.6)
2	300 (31.6)
3	521 (54.8)
Pleomorphisms	
1	13 (1.4)
2	314 (33.9)
3	599 (64.7)
Tubule formation	
1	45 (4.8)
2	308 (33.3)
3	574 (61.9)
Mitotic figures	
1	267 (28.8)
2	196 (21.2)
3	463 (50.0)
Tumour size	
≤ 2 cm	543 (56.6)
> 2 cm	404 (42.1)
Axillary nodal stage	
1 (Node negative)	571 (59.5)
2 (1–3 positive nodes)	304 (31.7)
3 (≥ 4 positive nodes)	78 (8.1)
Nottingham Prognostic Index (NPI)	
Good NPI (<3.4)	247 (26.0)
Moderate NPI (3.41–5.4)	530 (55.7)
Poor NPI (≥ 5.4)	174 (18.3)
Histological tumour type	
Ductal no special type (NST)	578 (62.3)
Tubular mixed	158 (17.0)
Medullary (typical and atypical)	23 (2.5)
Lobular*	73 (7.9)
Excellent Prognosis Special Groups**	38 (4.1)
Mixed NST and lobular	39 (4.2)
Mixed NST and other special type	19 (2.0)
LVI Negative	612 (65.2)
Definite	327 (34.8)
Distant metastasis	
Negative	567 (60.1)
Positive	376 (39.2)
Survival (month)	
Overall survival: Median/mean (range)	168/149.6 (1–308)
DFI: Median/mean (range)	109/100 (2–239)
ER status	
ER positive	271 (28.5)
ER negative	681 (71.5)
PR status	
PR positive	407 (43.8)
PR negative	522 (56.2)
HER2 status	
HER2 positive	786 (85.4)
HER2 negative	131 (13.2)
Molecular subtype	
Luminal	614 (65.9)
HER2 positive	131 (14.1)
Triple negative	187 (20.0)
Abbreviations: ER = oestrogen receptor; HER2 = human epidermal growth factor receptor 2; LVI = lymphovascular invasion; PR = progesterone receptors.	
*Invasive lobular includes: classic lobular, tubulo-lobular, alveolar lobular, solid lobular, pleomorphic lobular, and lobular mixed carcinomas.	
**Excellent prognostic special types comprise: mucinous carcinoma, cribriform carcinoma, tubular carcinoma and invasive papillary carcinoma.	

images were individually evaluated and assigned an 'H-Score' by identification of the percentage of the malignant cells that were stained (0–100) and the staining intensity (0 – negative, 1 – weak, 2 – moderate and 3 – strong), and the final H-Score was calculated as previously described (McCarty *et al*, 1985). Scoring was performed by a single observer (RS) and a subset of cases (20%) was scored by an independent observer (MA) to assess inter-rater reliability. All identifying patient information and clinical variables were completely anonymised during scoring.

Statistical analysis. The H-Scores were linked to anonymised patient codes to match all other variables with biomarker expression. Analysis was performed in the statistical software package SPSS (IBM SPSS statistics, Version 22). The data were exported to X-Tile (X-Tile Bioinformatics Software, Yale University, version 3.6.1) to determine an unbiased optimally significant cut off point based on patients' outcomes. Kappa statistic was used to test scoring reproducibility between observers (inter-observer agreement). Analyses were performed for nuclear and cytoplasmic H scores independently. Univariate analysis was performed using the chi-squared test to evaluate the significance of the association between expression of the biomarkers and the clinicopathological parameters of the data, as well as other previously investigated biomarkers. Kaplan–Meier analysis was performed to assess BCSS and DFI survival differences. Multivariate Cox Regression analysis with adjustment of co-variables was fitted to test independence from standard prognostic factors. A *P*-value of <0.05 (two tailed) was considered significant.

RESULTS

Gene identification using the METABRIC study. Analysis of the METABRIC transcriptomic data in the discovery and validation sets revealed 514 genes differentially expressed with regard to LVI status. These genes were ranked in order based on the *P*-value of expression. Biological functions of the top differentially expressed genes were also assessed using published literature. In addition, the CNAs of these genes were analysed to identify those genes whose expression was associated with their gene CNAs (*cis* acting). Of the top five genes that showed differential expression associated with CNAs with regard to LVI was *ARHGAP18* and this gene was subject to further analysis in this study.

Using the median expression as a cut-off point, high *ARHGAP18* mRNA expression was significantly associated with variables of good prognosis including smaller size ($P < 0.04$), early stage ($P < 0.001$), tumours of tubular subtype ($P < 0.001$) and HER2 – tumours ($P = 0.029$). When comparing the levels of *ARHGAP18* mRNA expression in the intrinsic (PAM50) subtypes, significant differences were observed ($P = 0.006$), with the HER2 + subtype showing the least expression levels. Similarly, significant difference in the expression levels were observed within the different METABRIC Integrative Clusters ($P = 0.021$), with clusters 5 (ERBB2 amplified) and cluster 9 (Luminal B subgroup) showed the least *ARHGAP18* mRNA expression. A trend towards significantly improved patients' survival ($P = 0.082$) in *ARHGAP18* mRNA over-expression was also identified.

ARHGAP18 mRNA expression levels in the external validation cohorts. The prognostic impact of *ARHGAP18* mRNA expression was subsequently assessed using bc-GenExMiner v4. (Breast Cancer Gene-Expression Miner v4) online data set as external validation cohorts ($n = 21$ data sets, 4177 patients). Of these, 12 data sets provided information on DMFI for *ARHGAP18* mRNA expression. As shown in the Forest plot (Supplementary Figure 1), 4 out of 12 studies showed high *ARHGAP18* mRNA expression was significantly associated with improved DMFI; however, the remaining eight studies did not show significant associations.

When the data were pooled together ($n = 2016$), high *ARHGAP18* mRNA expression was significantly associated with longer DMFI ($P < 0.001$, hazard ratio (HR) = 0.82, 95%CI 0.75–0.90), Supplementary Figure 2. Interestingly, exhaustive prognostic analysis, which permits to screen the prognostic impact of *ARHGAP18* mRNA on all possible combinations of population, showed significantly improved outcome of patients in 9 out of 18 studies using all possible combinations, Supplementary Table 1.

Immunohistochemical expression of ARHGAP18. Validation of the antibody specificity using western blot showed a single specific band at the predicted size (75kDa) of ARHGAP18 protein, confirming the specificity of the antibody, Figure 1A. Full-face sections stained with ARHGAP18 antibody showed homogenous staining distribution throughout stained sections; therefore, validating the use of TMA. IHC expression showed ARHGAP18 protein was localised in the nuclei and cytoplasm of tumour cells with varying intensities, Figure 1B–D. In addition, positivity was observed in the cells of ductal carcinoma *in situ* and in the normal ductal epithelial cells entrapped in some cores. Overall, the expression was reduced in some invasive tumours compared to the cells of the *in situ* component or the normal ductal cells.

Cases re-scored by an independent observer showed a very good (Kappa = 0.83) agreement between the two scorers. The H-Scores of both nuclear and cytoplasmic expression did not follow normal distribution. X-Tile determined optimal cut-off values as H-Score 140 for cytoplasmic staining, and 65 for nuclear staining. At these cut-offs, 848 out of 959 (88.4%) cases showed negative/low nuclear expression and 111 out of 959 (12.6%) showed positive/high nuclear expression. Cases with cytoplasmic staining were dichotomised as 784 out of 959 (81.8%) cores negative/low expression and 175 out of 959 (18.2%) with positive/high expression.

ARHGAP18 protein expression and clinicopathological variables. Cytoplasmic expression of ARHGAP18 showed an inverse association with the LVI ($P = 0.006$) and Nottingham Prognostic Index (NPI, $P = 0.010$; Table 2). Loss of its expression was associated with epithelial mesenchymal transition with loss of E-cadherin and overexpression of N-cadherin, and with HER2 overexpression. Consistent with mRNA expression, ARHGAP18 cytoplasmic protein expression showed an association with IHC-defined molecular subtypes where it was less expressed in the HER2 + and TNBC classes ($P = 0.035$; Table 3). Nuclear expression of ARHGAP18 was associated with lower grade ($P = 0.022$), with less mitotic counts and nuclear pleomorphism, smaller tumour size ($P = 0.002$), the better prognostic NPI subgroup ($P = 0.001$), histological types of excellent/good prognosis ($P = 0.002$) but not with LVI (Table 4). Positive nuclear expression showed significant associations with HER2 negativity ($P = 0.003$), negative/low Ki67 ($P = 0.001$) luminal-A subtype ($P = 0.001$), and the negative/low expression of N-cadherin, transforming growth factor β (TGF- β 1), PIK3CA and C-terminal tensin-like (CTEN) ($P < 0.05$; Table 5).

Similar statistical associations were observed when the expression of ARHGAP18 was taken into account irrespective of the localisation (i.e. negative *versus* positive cytoplasmic or nuclear expression).

Expression of ARHGAP18 patient outcome. Increased cytoplasmic expression of ARHGAP18 was significantly associated with improved outcome in terms of longer BCSS ($P = 0.004$) and DMFI ($P = 0.012$; Figure 2). Using multivariate cox regression analysis, this association was independent of tumour size, tumour grade, nodal stage and BC molecular subtype ($P = 0.017$, HR = 0.68, 95% CI 0.49–0.93, and $P = 0.031$, HR = 0.72, 95% CI = 0.54–0.97, for BCSS and DMFI, respectively). High nuclear expression was also significantly associated with longer BCSS ($P = 0.001$) and DMFI ($P = 0.003$; Figure 3). Again, similar statistical associations with

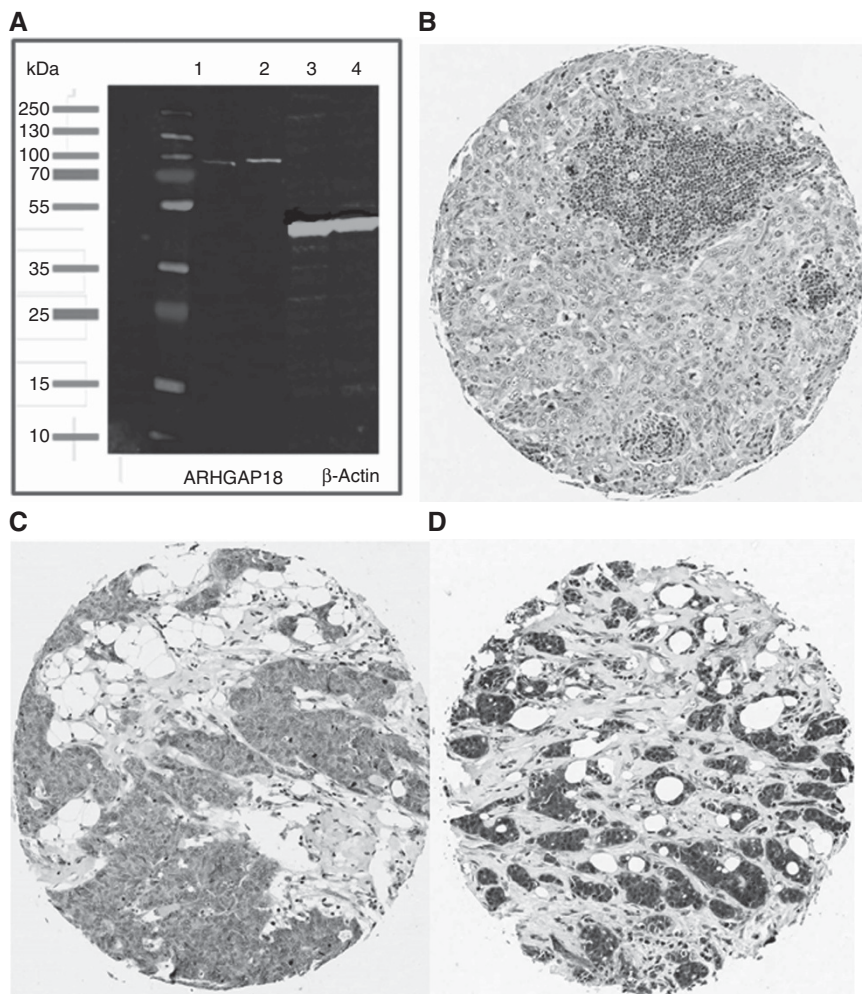


Figure 1. Western blot and immunohistochemical expression of ARHGAP18 in BC. **(A)** Western blot of ARHGAP18 and the housekeeping Beta-actin. Western blotting performed on whole cell lysates of HeLa CCL-2 and MDA-MB-231 (lanes 1 and 2, respectively). Primary antibody (Anti-ARHGAP18, 1 : 1000 dilution), and of the HRP-labelled secondary anti-rabbit antibody (1 : 15 000), with β -actin (1 : 2000, lanes 3 and 4) used as a loading control. The images were developed via chemiluminescence using an Odyssey Fc (Li-cor Biosciences, USA). **(B–D)** Immunohistochemical expression of ARHGAP18 in invasive BC: **(B)** Negative, **(C)** Invasive BC case showing positive cytoplasmic expression, and **(D)** a case of invasive BC case showing both nuclear and cytoplasmic expression.

patients' outcome were observed when the expression of ARHGAP18 was taken into account irrespective of the localisation. These associations were independent of the standard prognostic factors including tumour size, nodal stage and tumour grade using Cox proportional multivariate analysis ($P = 0.023$, HR = 0.62, 95% CI = 0.41–0.94, and $P = 0.047$, HR = 0.69, 95% CI = 0.47–0.10, for BCSS and DMFI, respectively).

DISCUSSION

The ARHGAP genes encode a family of at least 32 members of RhoGAP proteins with variable functions (Katoh and Katoh, 2004). The downstream target of the ARHGAPs is the Rho GTPase protein family; therefore, they are logical candidates to investigate their roles in BC. Members of the Rho GTPase family have numerous cellular functions including maintenance and control of the actin cytoskeleton, angiogenesis, cell proliferation, differentiation, and survival (Peck *et al*, 2002). Consequently, deregulation of these proteins has been implicated in multiple types of tumours. In this study and using an integrative bio-informatics analysis of gene expression and CNA data associated with LVI coupled with strict

defining criteria for definition of LVI in BC we have identified ARHGAP18 as one of the top differentially expressed genes and its expression was associated with its gene CNA. ARHGAP18 is located on chromosome 6q22.33 and is a member of the ARHGAP family (Potkin *et al*, 2008). After confirmation of its prognostic value in the METABRIC cohort and in external validation series, we sought to assess the protein expression of ARHGAP18 in a large unselected annotated series of BC with long-term follow-up.

Studies exploring the impact of tissue-based ARHGAP18 protein expression in cancers including BC are lacking and most of the studies so far were performed using *in vitro* tissue culture platform. In the current study, ARHGAP18 IHC staining was detected in the cytoplasm and in the nuclei of invasive BC cells with variable percent and intensity, in the cells of ductal carcinoma *in situ* and in the normal ductal epithelial cells. An overall tendency of reduced expression was observed in invasive tumours compared to the non-invasive and normal ductal cells, with the latter showing the highest expression; observations which may be denoting a tumour suppressor role.

Cytoplasmic expression of ARHGAP18 showed significant negative association with LVI. Recently, Chang *et al* reported on roles of ARHGAP18 in limiting proangiogenic signalling and promoting vascular stability through limiting endothelial cell

Table 2. Statistical association of cytoplasmic H score expression of ARHGAP18 and the clinicopathological parameters of the studied series

Parameter	ARHGAP18 cytoplasmic expression		χ^2	P-value
	Low no. (%)	High no. (%)		
Patients' age				
≤ 50 yrs	290 (79.0)	77 (21.0)	3.178	0.075
> 50 yrs	494 (83.6)	97 (16.4)		
Menopausal status				
Premenopausal	303 (79.5)	78 (20.5)	2.270	0.132
Postmenopausal	481 (83.4)	96 (916.6)		
Grade				
1	98 (76.0)	31 (24.0)	3.439	0.179
2	247 (82.3)	53 (17.7)		
3	432 (82.9)	89 (17.1)		
Mitotic figures				
1	221 (82.5)	47 (17.5)	1.713	0.425
2	154 (78.6)	42 (21.4)		
3	383 (82.7)	80 (17.3)		
Pleomorphisms				
1	8 (61.5)	5 (38.5)	6.015	0.049
2	249 (79.3)	65 (20.7)		
3	500 (83.5)	99 (16.5)		
Tubule formation				
1	34 (75.6)	11 (24.4)	5.147	0.075
2	242 (78.6)	66 (21.4)		
3	482 (84.0)	92 (16.0)		
Axillary nodal stage				
1	458 (80.2)	113 (19.8)	2.900	0.235
2	258 (84.9)	46 (15.1)		
3	64 (82.1)	14 (17.9)		
Tumour size				
≤ 2 cm	437 (80.5)	106 (19.5)	1.581	0.209
> 2 cm	338 (83.7)	66 (16.3)		
NPI				
Good (<3.4)	195 (78.9)	52 (21.2)	17.898	0.010
Moderate (3.41–5.4)	431 (81.3)	99 (18.7)		
Poor (≥ 5.41)	152 (87.4)	22 (12.6)		
Vascular invasion				
Negative	485 (79.2)	127 (20.8)	7.617	0.006
Definite	283 (86.5)	44 (13.5)		
Tumour type				
Ductal carcinoma No Special Type (NST)	476 (82.4)	102 (17.6)	7.914	0.244
Tubular mixed	125(79.1)	33(20.9)		
Medullary-like	21(91.3)	2 (8.7)		
Invasive lobular	63 (86.3)	10 (13.7)		
Excellent Prognosis Special Type**	26 (68.4)	12 (31.6)		
Mixed NST and lobular	32 (82.1)	7 (17.9)		
Mixed NST and other special type	16 (84.2)	169 (18.2)		

Abbreviation: NPI = Nottingham Prognostic Index.
 **Excellent prognostic special types comprise: mucinous carcinoma, cribriform carcinoma, tubular carcinoma and invasive papillary carcinoma.

sprouting and stabilising junctional integrity (Rao *et al*, 2015b). Moreover, knockdown of ARHGAP18 has been recently reported to promote endothelial cells migration in scratch wound assay with a protrusive and irregular migratory front and disrupted cell junctions at the proximal edge of the leading cell (Chang *et al*, 2014). Sustained angiogenesis is one of the hallmarks of cancer and is essential for tumour progression. Moreover, in the current study, reduced expression was associated with high Ki67, epithelial

Table 3. Statistical association of cytoplasmic H score expression of ARHGAP18 and the expression of other biomarkers

Parameter	ARHGAP18 cytoplasmic expression		Significance	
	Low no. (%)	High no. (%)	χ^2	P-value
ER				
Negative	227 (83.8)	44 (16.2)	0.955	0.328
Positive	552 (81.1)	129 (18.9)		
PR				
Negative	343 (84.3)	64 (15.7)	2.962	0.085
Positive	417 (79.9)	105 (20.1)		
HER2				
Negative	636 (80.9)	150 (19.1)	5.391	0.019
Positive	117 (89.3)	14 (10.7)		
Molecular subtypes				
Luminal	492 (80.1)	122 (19.9)	6.719	0.035
HER2 +	117 (89.3)	14 (10.7)		
Triple negative	157 (84.0)	30 (16.0)		
p53				
Negative	528 (82.2)	114 (17.8)	0.081	0.776
Positive	224 (81.5)	51 (18.5)		
Ki67				
Negative	179 (82.1)	39 (17.9)	0.004	0.951
Positive	435 (81.9)	96 (18.1)		
E-Cadherin				
Negative	284 (86.1)	46 (13.9)	5.223	0.023
Positive	448 (80.0)	112 (20.0)		
N-Cadherin				
Negative	147 (78.6)	40 (21.4)	5.752	0.019
Positive	440(86.1)	71 (13.9)		
TGF-β1				
Negative	274 (79.7)	70 (20.3)	1.465	0.262
Positive	246 (83.4)	49 (16.6)		
PIK3CA				
Negative	129 (79.1)	34 (20.9)	1.334	0.247
Positive	471 (83.1)	96 (16.9)		
CTEN				
Negative	31 (79.5)	8 (20.5)	0.363	0.514
Positive	575 (83.2)	116 (16.8)		

Abbreviations: CTEN = C-terminal tensin-like; ER = oestrogen receptor; HER2 = human epidermal growth factor receptor 2; PR = progesterone receptors; TGF-β1 = Transforming Growth Factor β1.

mesenchymal transition with expression of N-cadherin and TGFβ1 and reduced expression of E-cadherin. Similar significance was also identified with nuclear expression, showing positive association with other prognostic variables and outcome. A large proportion of cases strongly expressing ARHGAP18 were of the excellent prognosis histological subtypes including tubular, invasive cribriform and mucinous carcinomas. Nuclear expression was also significantly associated with lower tumour grade, lower mitotic scores, less nuclear pleomorphism and smaller tumour size. Such associations suggest ARHGAP18 may be a marker of related to tumour differentiation.

Investigations specifically pertaining to the subcellular localisation of ARHGAP18 are lacking. However, similar to findings of the current study with ARHGAP18 in BC it has been reported that ARHGAP21 is localised both in the cytoplasm and nuclei of prostatic carcinoma (Barcellos *et al*, 2013) and nuclear and perinuclear localisation seen in glioblastoma-derived cell lines (Bigarella *et al*, 2009). We suggest ARHGAP18 might have a similar trafficking from the cytoplasm, where synthesis occurs, to the nucleus, possibly upon occasion of tumourigenic changes.

Table 4. Associations of nuclear H scores for ARHGAP18 in relation to the clinicopathological parameters

Parameter	ARHGAP18 Nuclear Expression		χ^2	P-value
	Low expression no. (%)	High expression no. (%)		
Patients' age				
≤ 50 yrs	320 (87.2)	47 (12.8)	0.585	0.471
> 50 yrs	525 (88.8)	66 (11.2)		
Menopausal status				
Premenopausal	335 (87.9)	46 (12.1)	0.047	0.838
Postmenopausal	510 (88.4)	67 (11.6)		
Grade				
1	103 (79.8)	26 (20.2)	12.832	0.002
2	262 (87.3)	38 (12.7)		
3	474 (91.0)	47 (9.0)		
Mitotic figures				
1	222 (82.8)	46 (17.2)	11.914	0.003
2	176 (89.8)	20 (10.2)		
3	422 (91.1)	41 (8.9)		
Pleomorphisms				
1	10 (76.9)	3 (23.1)	7.705	0.021
2	267 (85.0)	47 (15.0)		
3	542 (90.5)	57 (9.5)		
Tubule formation				
1	40 (88.9)	5 (11.1)	3.441	0.179
2	264 (85.7)	44 (14.3)		
3	516 (89.9)	58 (10.1)		
Axillary nodal stage				
1	496 (86.9)	75 (13.1)	2.656	0.265
2	275 (90.5)	29 (9.5)		
3	70 (89.7)	8 (10.3)		
Tumour size				
≤ 2 cm	464 (85.5)	79 (14.5)	9.835	0.002
> 2 cm	372 (92.1)	32 (7.9)		
NPI				
Good (<3.4)	203 (82.2)	44 (17.8)	13.748	0.001
Moderate (3.41–5.4)	475 (89.6)	55 (10.4)		
Poor (≥5.4)	162 (93.1)	12 (6.9)		
Vascular invasion				
Negative	531 (86.8)	81 (13.2)	3.372	0.066
Definite	297 (90.8)	30 (9.2)		
Tumour type				
Ductal carcinoma No Special Type (NST)	529 (91.5)	49 (8.5)	21.170	0.002
Tubular mixed	136 (86.1)	22 (13.9)		
Medullary (typical and atypical)	20 (87.0)	3 (13.0)		
Lobular*	59 (80.8)	14 (19.2)		
Excellent Prognosis Special Type**	29 (76.3)	9 (23.7)		
Mixed NST and lobular	32 (82.1)	7 (17.9)		
Mixed NST and other special type	14 (73.7)	5 (26.3)		

Abbreviation: NPI = Nottingham Prognostic Index.
 *Invasive lobular includes: classic lobular, tubulo-lobular, alveolar lobular, solid lobular, pleomorphic lobular, and lobular mixed carcinomas.
 **Excellent prognostic special types comprise: mucinous carcinoma, cribriform carcinoma, tubular carcinoma and invasive papillary carcinoma.

The associations of both nuclear and cytoplasmic ARGAP18 with the cadherin switch in the current data; E-cadherin loss and N-cadherin gain, as well as the negative association with the EMT triggers TGFβ1 and PIK3CA (Zhao *et al*, 2006), as well as the migration-associated CTEN (Albarsi *et al*, 2009), reflects its potential role in controlling cellular cytoskeletal dynamics and

Table 5. Association of nuclear ARHGAP18 expression and the expression of other biomarkers

Parameter	ARHGAP18 nuclear expression		Significance	
	Low expression no. (%)	High expression no. (%)	χ^2	P-value
ER				
Negative	248 (91.5)	23 (8.5)	3.702	0.054
Positive	593 (87.1)	88 (12.9)		
PR				
Negative	369 (90.7)	38 (9.3)	3.381	0.066
Positive	453 (86.8)	69 (13.2)		
HER2				
Negative	683 (86.9)	103 (13.1)	11.433	0.003
Positive	127 (96.9)	4 (3.1)		
Molecular subtypes				
Luminal	528 (86.0)	86 (14.0)	13.522	0.001
HER2+	127 (96.9)	4 (3.1)		
Triple negative	169 (90.4)	18 (9.6)		
p53				
Negative	567 (88.3)	75 (11.7)	0.032	0.859
Positive	244 (88.7)	31 (11.3)		
Ki67				
Low	179 (82.1)	39 (17.9)	10.599	0.001
High	481 (90.6)	50 (9.4)		
E-cadherin				
Negative	299 (90.6)	31 (9.4)	1.784	0.182
Positive	491 (87.7)	69 (12.3)		
N-cadherin				
Negative	156 (83.4)	31 (16.6)	10.828	0.002
Positive	470 (92.0)	41 (8.0)		
TGF-β1				
Negative	296 (86.0)	48 (14.0)	6.870	0.011
Positive	273(92.5)	22 (7.5)		
PIK3CA				
Negative	126 (77.3)	37 (22.7)	24.932	< 0.001
Positive	519 (91.5)	48 (8.5)		
CTEN				
Negative	29 (74.4)	10 (25.6)	8.836	0.007
Positive	620 (89.7)	71 (10.3)		

Abbreviations: ER = oestrogen receptor; HER2 = human epidermal growth factor receptor 2; PR = progesterone receptors.

migration. The latter is attributed to the recognised roles of Rho GTPases in controlling the actin cytoskeleton. ARHGAP18 has been postulated to suppress the function of RhoA and cause a disruption in the production of stress fibres; contractile bundles of actin found in non-muscle cells (Maeda *et al*, 2011). These functions may suggest a mechanism through which ARHGAP18 is involved in limiting spread and metastasis of cancer cells therefore may provide some explanation for its good prognostic value.

High ARHGAP18 mRNA expression within the METABRIC data set was significantly associated with variables of good prognosis including the PAM50 and Integrative Clusters of good prognostic classes. Although analysis of the METABRIC study revealed a trend towards improved patients' outcome in cases overexpressing ARHGAP18, pooled analysis of the publicly available data sets, using the bc-GenExMiner v4, showed significant association between ARHGAP18 expression and better outcome. Although gene expression data reflect the overall expression levels, they show concordance with our protein expression data where increased ARHGAP18 expression in both the nucleus and cytoplasm was significantly associated with better outcome.

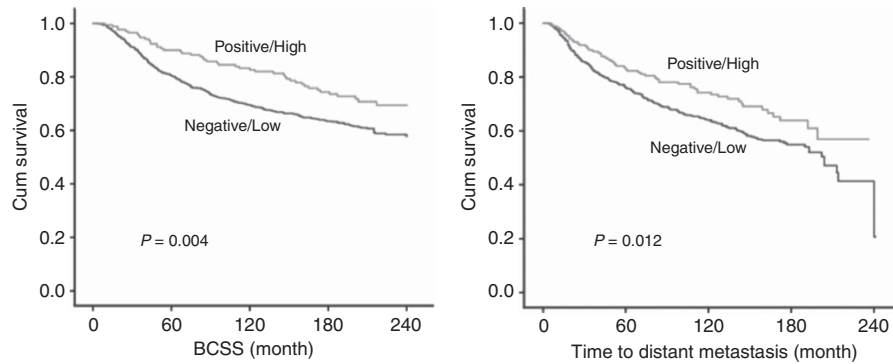


Figure 2. Kaplan–Meier plots of the association between cytoplasmic ARHGAP18 expression and BCSS and time to distant metastasis during the follow-up period.

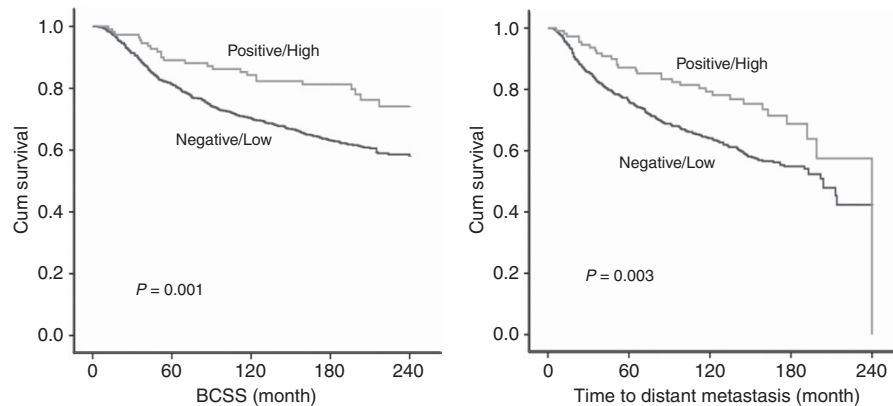


Figure 3. Kaplan–Meier plots of the association between nuclear ARHGAP18 expression and BCSS and time to distant metastasis during the follow-up period.

In conclusion, results presented herein suggest ARHGAP18 may be acting, directly or indirectly, as a putative metastasis suppressor gene, based on its association with favourable prognostic features including negative association with LVI and longer survival at both protein and mRNA expression levels. The general tumour suppressor effect probably indicates that the relationship to LVI is perhaps within a coordinated set of events related to multiple events in BC development and progression. They also indicate the validity of our approach to identify novel biomarkers associated with complex biological processes that are related not only to the investigative techniques but also on robust histological characterisation. Further investigation of this biomarker as well as other genes differentially expressed with regard to LVI is warranted to decipher the mechanism underlying its development and to identify potential therapeutic targets.

ACKNOWLEDGEMENTS

We thank the Nottingham Health Science Biobank and Breast Cancer Now Tissue Bank for the provision of tissue samples.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JFR, Macmillan D, Blamey RW, Ellis IO (2005) 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. *Int J Cancer* **116**(3): 340–350.
- Albasri A, Seth R, Jackson D, Benhasouna A, Crook S, Nateri AS, Chapman R, Ilyas M (2009) C-terminal Tensin-like (CTEN) is an oncogene which alters cell motility possibly through repression of E-cadherin in colorectal cancer. *J Pathol* **218**(1): 57–65.
- Aleskandarany MA, Sonbul SN, Mukherjee A, Rakha EA (2015) Molecular mechanisms underlying lymphovascular invasion in invasive breast cancer. *Pathobiology* **82**(3–4): 113–123.
- Altman DG, Bland JM (1994) Statistics notes: diagnostic tests 2: predictive values. *BMJ* **309**(6947): 102.
- Barcellos KS, Bigarella CL, Wagner MV, Vieira KP, Lazarini M, Langford PR, Machado-Neto JA, Call SG, Staley DM, Chung JY, Hansen MD, Saad ST (2013) ARHGAP21 protein, a new partner of alpha-tubulin involved in cell-cell adhesion formation and essential for epithelial-mesenchymal transition. *J Biol Chem* **288**(4): 2179–2189.
- Bigarella CL, Borges L, Costa FF, Saad STO (2009) ARHGAP21 modulates FAK activity and impairs glioblastoma cell migration. *Biochim Biophys Acta* **1793**(5): 806–816.
- Chang GH, Lay AJ, Ting KK, Zhao Y, Coleman PR, Powter EE, Formaz-Preston A, Jolly CJ, Bower NI, Hogan BM, Rinkwitz S, Becker TS, Vadas MA, Gamble JR (2014) ARHGAP18: an endogenous inhibitor of angiogenesis, limiting tip formation and stabilizing junctions. *Small GTPases* **5**(3): 1–15.
- Choi SH, Barsky SH, Chang HR (2003) Clinicopathologic analysis of sentinel lymph node mapping in early breast cancer. *Breast J* **9**(3): 153–162.
- Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale A-L, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* **486**(7403): 346–352.

- Debled M, de Mascarel I, Brouste V, Mauriac L, MacGrogan G (2010) Re: Population-based study of peritumoral lymphovascular invasion and outcome among patients with operable breast cancer. *J Natl Cancer Inst* **102**(4): 275–276author reply 276–277.
- Jezequel P, Campone M, Gouraud W, Guerin-Charbonnel C, Leux C, Ricolleau G, Campion L (2012) bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Res Treat* **131**(3): 765–775.
- Katoh M, Katoh M (2004) Characterization of human ARHGAP10 gene in silico. *Int J Oncol* **25**(4): 1201–1206.
- Lauria R, Perrone F, Carlomagno C, De Laurentiis M, Morabito A, Gallo C, Varriale E, Pettinato G, Panico L, Petrella G, Raffaele Bianco A, De Placido S (1995) The prognostic value of lymphatic and blood vessel invasion in operable breast cancer. *Cancer* **76**(10): 1772–1778.
- Lee AK, DeLellis RA, Silverman ML, Heatley GJ, Wolfe HJ (1990) Prognostic significance of peritumoral lymphatic and blood vessel invasion in node-negative carcinoma of the breast. *J Clin Oncol* **8**(9): 1457–1465.
- Maeda M, Hasegawa H, Hyodo T, Ito S, Asano E, Yuang H, Funasaka K, Shimokata K, Hasegawa Y, Hamaguchi M, Senga T (2011) ARHGAP18, a GTPase-activating protein for RhoA, controls cell shape, spreading, and motility. *Mol Biol Cell* **22**(20): 3840–3852.
- McCarty Jr KS, Miller LS, Cox EB, Konrath J, McCarty Sr KS (1985) Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch Pathol Lab Med* **109**(8): 716–721.
- Mruk DD, Cheng CY (2011) Enhanced chemiluminescence (ECL) for routine immunoblotting: an inexpensive alternative to commercially available kits. *Spermatogenesis* **1**(2): 121–122.
- Peck J, Douglas G, Wu CH, Burbelo PD (2002) Human RhoGAP domain-containing proteins: structure, function and evolutionary relationships. *FEBS Lett* **528**(1–3): 27–34.
- Potkin SG, Turner JA, Fallon JA, Lakatos A, Keator DB, Guffanti G, Macciardi F (2008) Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia. *Mol Psychiatry* **14**(4): 416–428.
- Quinlan JR (1993) *C4.5: Programs for Machine Learning* Los Altos, CA, USA.
- Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA, Evans AJ, Blamey R, Reis-Filho JS, Foulkes WD, Ellis IO (2009) Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clin Cancer Res* **15**(7): 2302–2310.
- Rakha EA, Martin S, Lee AHS, Morgan D, Pharoah PDP, Hodi Z, MacMillan D, Ellis IO (2012) The prognostic significance of lymphovascular invasion in invasive breast carcinoma. *Cancer* **118**(15): 3670–3680.
- Rao N, Lee YF, Ge R (2015a) Novel endogenous angiogenesis inhibitors and their therapeutic potential. *Acta Pharmacol Sin* **36**(10): 1177–1190.
- Rao N, Lee YF, Ge R (2015b) Novel endogenous angiogenesis inhibitors and their therapeutic potential. *Acta Pharmacol Sin* **36**(10): 1177–1190.
- Song YJ, Shin SH, Cho JS, Park MH, Yoon JH, Jegal YJ (2011) The role of lymphovascular invasion as a prognostic factor in patients with lymph node-positive operable invasive breast cancer. *J breast cancer* **14**(3): 198–203.
- Truong PT, Yong CM, Abnoui F, Lee J, Kader HA, Hayashi A, Olivetto IA (2005) Lymphovascular invasion is associated with reduced locoregional control and survival in women with node-negative breast cancer treated with mastectomy and systemic therapy. *J Am Coll Surg* **200**(6): 912–921.
- Ugras S, Stempel M, Patil S, Morrow M (2014) Estrogen receptor, progesterone receptor, and HER2 status predict lymphovascular invasion and lymph node involvement. *Ann Surg Oncol* **21**(12): 3780–3786.
- Zhao JJ, Cheng H, Jia S, Wang L, Gjoerup OV, Mikami A, Roberts TM (2006) The p110alpha isoform of PI3K is essential for proper growth factor signaling and oncogenic transformation. *Proc Natl Acad Sci USA* **103**(44): 16296–16300.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)