Clinicopathological and prognostic significance of Ras Association and Pleckstrin Homology domains 1 (RAPH1) in breast cancer

Sasagu Kurozumi[1,2], Chitra Joseph[1], Sultan Sonbul[1], Mohammed A Aleskandarany[1,3], Marian Pigera[1], Mansour Alsaleem[1], Sami Alsaeed[1], Yousif Kariri[1], Christopher C Nolan[1], Maria Diez-Rodriguez[1,] Simon Johnston[1], Nigel P. Mongan[4,5], Takaaki Fujii[2], Ken Shirabe[2], Stewart G Martin[1], Ian O Ellis[1], Andrew R Green[1] and Emad A Rakha[1,3]

[1]Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham, UK; [2]Department of General Surgical Science, Gunma University Graduate School of Medicine, Gunma, Japan; [3]Faculty of Medicine, Menoufyia University, Shebin al Kawm, Egypt; [4]Cancer Biology and Translational Research, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, UK; [5]Department of Pharmacology, Weill Cornell Medicine, New York, 10065, USA.

Corresponding author:

Professor Emad Rakha

Department of Histopathology, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, NG5 1PB, UK

Email: Emad.Rakha@nottingham.ac.uk

RUNNING TITLE: Prognostic value of RAPH1 in breast cancer

ABSTRACT

BACKGROUND: Ras association and pleckstrin homology domains 1 (RAPH1) is involved in cytoskeleton regulation and re-epithelialisation in invasive carcinoma and therefore may play a key role in carcinogenesis and metastasis. We herein investigated the biological and clinical significance of RAPH1 in breast cancer using large annotated cohorts.

METHODS: The clinicopathological and prognostic significance of *RAPH1* was assessed at the genomic and transcriptomic levels using The Cancer Genome Atlas (TCGA) dataset (n=1039) and the results were validated using the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (n=1980). RAPH1 protein expression was evaluated by immunohistochemistry in a large, well-characterised cohort of early-stage breast cancer (n=1040).

RESULTS: In both the TCGA-BRCA and METABRIC cohorts, *RAPH1* mRNA expression and *RAPH1* copy number alteration were strongly correlated. *RAPH1* mRNA overexpression was significantly correlated with high expression of adhesion and EMT markers including *CDH1, TGFβ1* and *CD44*. *RAPH1* mRNA overexpression was a significant predictor of a poor prognosis (Hazard ratio: 3.88; $p = 0.049$). High RAPH1 protein expression was associated with higher grade tumours with high proliferation index, triple negative phenotype and high E-cadherin expression. High RAPH1 protein expression was an independent predictor of shorter survival (Hazard ratio: 4.37; $p = 0.037$).

CONCLUSIONS: High RAPH1 expression is correlated with aggressive breast cancer phenotypes and provides independent prognostic value in invasive breast cancer.

List of abbreviations
BC, Breast cancer; BCSS, BC-specific survival; CI, Confidence intervals; CAN, Copy number alteration; EMT, Epithelial-mesenchymal transition; HR, Hazard ratio; MET, Mesenchymal-epithelial transition; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; TCGA, The Cancer Genome Atlas

INTRODUCTION

Recent developments in diagnostic accuracy and drug therapies have improved outcomes in early-stage breast cancer (BC). However, more than 20% of BC patients have poor clinical outcomes related to recurrence and metastasis [1, 2]. Various studies have been conducted to identify factors associated with tumour invasion and metastasis to uncover novel candidates for targeted treatment.

Ras association and pleckstrin homology domains 1 (RAPH1), also referred to as lamellipodin (LPD) [3], acts as a downstream effector of the Ras pathway [4], which plays an important role in epithelial-mesenchymal transition (EMT) [5, 6]. EMT is thought to play an important role in the development of distant metastasis in BC [7]. Previous studies showed changes in intracellular tight junction characteristics and weakened cell adhesive properties during EMT as underlying mechanisms inducing the loss of epithelial features and the gain of mesenchymal morphology [8, 9]. RAPH1 has a role in regulating cell migration and re-epithelialisation [10]. Aberrant expression of RAPH1 may also promote cancer cell adhesion at distant, metastatic sites [10]. However, the clinical significance and prognostic role of RAPH1 in BC remains to be defined.

In this study, RAPH1 was investigated in BC at the gene and protein phases using large well-characterised and annotated cohorts of early stage BC with emphasis on the relation between RAPH1 expression and clinicopathological factors including the biomarkers associated with EMT and patients' outcomes.

MATERIALS AND METHODS

Genomic and transcriptomic analysis of *RAPH1* gene

In this study, the Cancer Genome Atlas (TCGA) [11] dataset (n = 1039) provided by cBioPortal [12] was used for copy number and mRNA analysis of *RAPH1* gene. Briefly, *RAPH1* copy number alteration (CNA) datasets from Affymetrix SNP 6.0 arrays and *RAPH1* mRNA expression data from RNASeqV2 were accessed, and information on several clinicopathological factors and outcomes were also collected.

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset [13, 14] including 1980 patients was used for validation of the findings based on the TCGA dataset. In this cohort, patients with oestrogen receptor (ER)-positive tumours with negative lymph node metastasis did not receive adjuvant chemotherapy, whereas those with ER-negative or lymph node metastasis-positive tumours received adjuvant chemotherapy. No patients with HER2-positive BC were treated with trastuzumab.

RAPH1 protein analysis expression

A total of 1040 BC patients who underwent surgery at Nottingham City Hospital in the UK between 1987 and 1998 were included in this study. Characteristics of the 1040 cases are shown in Supplementary Table 1. All patients underwent breast-conserving surgery or modified radical mastectomy without receiving neoadjuvant treatment. A total of 405 (38.9%) cases had endocrine therapy and 212 (20.4%) patients were treated with chemotherapy. Haematoxylin & eosin staining and immunostaining with antibodies against CD34 and D2-40 according to our previously published methods [15] were used for assessment of lymphovascular invasion (LVI) status. Availability and assessment of oestrogen receptor (ER), progesterone receptor (PR), HER2, the proliferation marker Ki67, EMT related genes

(E-cadherin and N-cadherin), basal markers (CK5/6 and EGFR), and the stem cell markers (CD24 and CD44) were described in previous studies [16-22]. This study was approved by the Nottingham Research Ethics Committee 2 (Reference title: Development of a molecular genetic classification of breast cancer). This study was conducted in accordance with the Declaration of Helsinki, and all patients participated in the study after informed consent.

Specificity of the RAPH1 antibody (HPA016744, Merck, Germany) was confirmed by Western blotting using Jurkat and MCF7 cell lines (The American Type Culture Collection; Rockville, MD, USA), which indicated a specific band at approximately 130 kDa (Supplementary Figure 2). RAPH1 protein expression was assessed in 15 full-face slides of BC with aggressive phenotype for evaluation of its association with morphological characteristics. Tumour samples, available as 0.6-mm cores, were arrayed as previously described [23]. For IHC, the Novolink Max polymer detection system (RE7280-k, Leica, Newcastle, UK) was used to visualise the reaction. . Following optimisations, the primary antibody for anti-RAPH1 was diluted 1:300 in Bond primary antibody diluent (Leica, Germany). For the staining control, a polyclonal rabbit anti-human beta-2 microglobulin antibody (1:2000; Dako, Glostrup, Denmark) was used. 3-3' diaminobenzidine tetrahydrochloride (Novolink DAB substrate buffer plus) was treated as the chromogen. Counter-staining was performed using Meyer's haematoxylin for 6 minutes. Immunostained tissue microarray sections were digitally scanned using a Nano Zoomer (Hamamatsu Photonics, Tokyo, Japan). RAPH1 IHC expression was classified into the following groups according to the cytoplasmic staining intensity of the cancer cells: no staining, weak staining, moderate staining and strong staining (Figure 1). The H-score system, as previously reported [24, 25], was used based on intensity scoring (0, negative; 1, weak; 2, moderate and 3, strong) and proportion scoring (0–100) based on the cytoplasmic RAPH1 expression pattern.

Statistical analysis

Statistical analyses were conducted using the SPSS statistical software v24.0 (IBM, Armonk, NY, USA). The relationship of *RAPH1* CNA with *RAPH1* mRNA expression was examined using analysis of variance test with Bonferroni correction. Pearson correlation test was used to evaluate the relationship of *RAPH1* mRNA expression with the mRNA expression of other genes. The chi-squared test was used to assess differences in protein expression level stratified by several clinicopathological factors including tumour size, lymph node status and ER and HER2 expression. To assess the association of RAPH1 with prognosis, Kaplan–Meier survival curves for BC-specific survival (BCSS) were determined, and the log-rank test was conducted. The BCSS in this study was determined as the time from the day of surgery to the day of death due to breast cancer. For univariate and multivariate survival assessment of clinicopathological factors including RAPH1 expression, 95% confidence intervals (CIs) were assessed using the Cox proportional hazards regression model. In these analyses based on mRNA expression, median mRNA value was determined as a cut-off point. For RAPH1 protein, significant cut-off point was calculated as H-score 90 using X-Tile (X-Tile Bioinformatics Software, Yale University, version 3.6.1) based on patient outcome, and samples were divided into high and low expression groups.

RESULTS

RAPH1 copy number alteration and mRNA expression

In the TCGA cohort of 1039 patients, *RAPH1* copy number gain and *RAPH1* loss were observed in 99 (9.5%) and 219 (21.1%) cases, respectively. In 1980 patients of the METABIRIC cohort, *RAPH1* copy number gain and *RAPH1* loss were observed in 27 (1.4%) and 33 (1.7%) cases, respectively. In both cohorts, *RAPH1* mRNA expression was higher in the tumours with copy number gain than in the neutral cases ($p < 0.0001$) and was lower in the tumours with copy number loss than in the neutral cases *(both $p < 0.0001$).*

High RAPH1 mRNA overexpression was associated with higher *CDH1* (TCGA: $p < 0.0001$; METABRIC: $p = 0.0096$), lower *TGFβ1* (TCGA: $p < 0.0001$; METABRIC: $p = 0.0011$), lower *KRT6C* (TCGA: $p = 0.0059$; METABRIC: $p = 0.026$) and higher *CD44* (TCGA: $p < 0.0001$; METABRIC: $p < 0.0001$) expression s in both the TCGA and METABRIC cohorts (Table 1). Patients with high *RAPH1* mRNA expression had worse prognosis than those with low *RAPH1* mRNA expression in TCGA cohort (hazard ratio [HR] 3.88; $p = 0.049$; Figure 2a), although this was not observed in the METABRIC validation cohort (Supplementary Figure 3).

Immunohistochemical expression of RAPH1 protein

Among full-face tissue sections, the RAPH1 expression grades and patterns in carcinoma cells were apparently different from those of normal mammary glands, as shown in Figure 3. RAPH1 expression in normal glandular epithelium was uniformly weak. The RAPH1 immunoreactivity of myoepithelial cells was lower than those of glandular cells. In contrast, RAPH1 immunopositivity was observed in the cytoplasm of invasive cancer cells, which was stronger compared to normal epithelial cells. In intraductal carcinoma cells, the

IHC expression pattern of RAPH1 was similar to that of invasive cancer cells, whereas the RAPH1 immunoreactivity was stronger than that of normal epithelial cells. The RAPH1 expression of myoepithelial cells around ductal components was weak, similar to that observed in normal myoepithelial cells. In both the invasive and intraductal cancer cells, the nuclei of a few cancer cells showed weak to moderate immunoreactivity concomitantly with cytoplasmic immunostaining.

RAPH1 protein expression

There were 393 (37.8%) patients with low RAPH1 expression (H-score < 90) and 647 (62.2%) patients with high RAPH1 expression group (H-score ≥ 90). In univariate analysis, the RAPH1 expression was significantly associated with histological grade ($p < 0.0001$), ER status ($p = 0.034$), intrinsic subtype ($p = 0.0088$), Ki67 labelling index ($p < 0.0001$) and E-cadherin expression ($p = 0.0046$; Table 2). No correlation between RAPH1 expression and other genes, such as N-cadherin, CK5/6, EGFR, CD44 and CD24 were identified (Table 2).

The BCSS in BC patients with high RAPH1 expression was significantly different from that of those with low RAPH1 expression (HR 4.37, $p = 0.037$; Figure 2b). Univariate analysis using the Cox proportional hazards regression model identified high RAPH1 expression (HR 1.44, $p = 0.038$), low E-cadherin expression (HR 0.68, $p = 0.0023$), positive lymphovascular invasion (HR 2.89, $p < 0.0001$), large tumour size (HR 2.99, $p < 0.0001$), positive nodal status (HR 3.20, $p < 0.0001$) and HER2-positive (HR 4.69, $p < 0.0001$) and triple negative BC (HR 3.45, $p < 0.0001$) as poor prognostic factors. RAPH1 protein expression was an independent poor prognostic factor in multivariate analysis (HR 1.43, $p = 0.046$; Table 3).

DISCUSSION

Cancer cell migration and cancer cell adhesion at distant metastatic sites have key roles in BC metastasis. EMT is involved in initial transformation of the primary invasive carcinoma to a metastatic phenotype, whereas cell adhesion has its functional effect during the establishment of cancer metastasis at distant sites [4, 26, 27]. In present study, although RAPH1 protein expression was not associated with the biomarkers related to EMT, RAPH1 overexpression was significantly associated with high tumour proliferation and E-cadherin positivity. Moreover, high *RAPH1* mRNA expression was associated with the high mRNA expression levels of *CDH1*, which encodes E-cadherin. Interestingly, there is evidence that RAPH1 interacts with E-cadherin [28]. Analysis of RAPH1 protein expression by IHC also revealed that its immunoreactivity in normal epithelium, albeit uniformly weak, was higher than that in myoepithelial cells. E-cadherin, a member of a transmembrane glycoprotein family, is involved in cell-to-cell adhesion [29, 30]. The extracellular domain of E-cadherin is involved in cell-cell interactions, whereas the intracellular domain regulates the assembly of the actin cytoskeleton via β-catenin, which has an important role in cell adhesion and signal transduction [31-33]. Following invasion, migrating BC cells keep or recover epithelial characteristic to adhere metastatic sites [34, 35]. Previous studies reported that RAPH1, also referred to as Lamellipodin/LPD, regulates formation of lamellipodia on cells and regulated cell migration and adhesion [10, 36, 37] by inducing the activity of Ena/VASP and Scar/WAVE, which are involved in the construction of actin cytoskeleton at the cellular membrane to promote invasion and metastasis of BC [3, 36-39]. High expression and membrane localisation of the actin regulator RAPH1 was shown to be associated with poor outcomes in BC patients [3]. This result suggested that RAPH1 might be associated with cancer cell adhesion at distant metastatic sites after LVI. In the present study, a strong RAPH1

10

expression was observed in the cytoplasm of invasive BC cells; therefore, further investigation is necessary to elucidate the role of intracellular localisation of RAPH1 on its protein activity.

Previous studies indicated that *RAPH1* gene abnormalities were not associated with outcomes in BC [38]. Although the number of cases with *RAPH1* CNA was limited in the current study, with less than 10% of the cases in the TCGA cohort exhibiting copy number gain, *RAPH1* mRNA expression was strongly associated with *RAPH1* CNA. Additionally, the association between *RAPH1* mRNA expression and outcome was significant. RAPH1 mRNA expression was significantly associated with the biomarkers related to cancer stem cells and basal type BC. However, there was no association between RAPH1, cancer stem cells and basal type BC at the protein expression level. Recent studies suggest that microRNA-203, which is considered to be associated with tumour invasion and transformation [40, 41], regulated *RAPH1* gene expression at the transcriptional level to contribute to re-epithelialisation [42]. Additionally, microRNA-203 was proposed to inhibit tumour growth and metastasis through podoplanin [43]. Further functional studies of aberrant *RAPH1* regulation by microRNA-203 are a reasonable approach to explore the mechanisms of metastasis in invasive BC.

CONCLUSIONS

RAPH1 expression is associated with aggressive BC phenotypes and E-cadherin positivity. High RAPH1 expression is an independent prognostic factor. RAPH1 association with proteins involved in cell adhesion at the plasma membrane might promote BC metastasis.

Acknowledgements

Compliance with Ethical Standards

Disclosure of potential conflicts of interest

Takaaki Fujii has received research funding from Eisai Co, Ltd.

Any of the other authors declare that they have no conflict of interest.

Research involving human participants

This study was approved by the Nottingham Research Ethics Committee 2 (Reference title: Development of a molecular genetic classification of breast cancer). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

REFERENCES

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2015) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365: 1687-1717.

2. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA et al (2008) Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 26: 1275-1281.

3. Carmona G, Perera U, Gillett C, Naba A, Law AL, Sharma VP et al (2016) Lamellipodin promotes invasive 3D cancer cell migration via regulated interactions with Ena/VASP and SCAR/WAVE. Oncogene 35: 5155-5169.

4. Stephen AG, Esposito D, Bagni RK, McCormick F (2014) Dragging ras back in the ring. Cancer Cell 25: 272-281.

5. Siewertsz van Reesema LL, Lee MP, Zheleva V, Winston JS, O'Connor CF, et al (2016) RAS pathway biomarkers for breast cancer prognosis. Clin Lab Int 40: 18-23.

6. Okada T, Sinha S, Esposito I, Schiavon G, López-Lago MA, Su W et al (2015) The Rho GTPase Rnd1 suppresses mammary tumorigenesis and EMT by restraining Ras-MAPK signalling. Nat Cell Biol 17: 81-94.

7. Wu Y, Sarkissyan M, Vadgama JV (2016) Epithelial-Mesenchymal Transition and Breast Cancer. J Clin Med 5.

8. Kalluri R, Neilson EG (2003) Epithelial-mesenchymal transition and its implications for fibrosis. J Clin Invest 112: 1776-1784.

9. Kalluri R, Weinberg RA (2009) The basics of epithelial-mesenchymal transition. J Clin Invest 119: 1420-1428.

10. Krause M, Leslie JD, Stewart M, Lafuente EM, Valderrama F, Jagannathan R et al (2004) Lamellipodin, an EnaNASP ligand, is implicated in the regulation of lamellipodial dynamics. Dev Cell 7: 571-583.

11. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. Nature 490: 61-70.

12. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA et al (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2: 401-404.

13. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486: 346-352.

14. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA et al (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun 10: 11479.

15. Mohammed RA, Martin SG, Mahmmod AM, Macmillan RD, Green AR, Paish EC et al (2016) Objective assessment of lymphatic and blood vascular invasion in lymph node-negative breast carcinoma: findings from a large case series with long-term follow-up. J Pathol 223: 358-365.

16. Rakha EA, Agarwal D, Green AR, Ashankyty I, Ellis IO, Ball G et al (2017) Prognostic stratification of oestrogen receptor-positive HER2-negative lymph node-negative class of breast cancer. Histopathology 70: 622-631.

17. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG et al (2009) Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clin Cancer Res 15: 2302-2310.

18. Ahmed MA, Aleskandarany MA, Rakha EA, Moustafa RZ, Benhasouna A, Nolan C et al (2012) A $CD44^-/CD24^+$ phenotype is a poor prognostic marker in early invasive breast cancer. Breast Cancer Res Treat 133: 979-995.

19. Kurozumi S, Joseph C, Sonbul S, Gorringe KL, Pigera M, Aleskandarany MA et al (2018) Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis. Breast Cancer Res Treat 170: 525-533.

20. Green AR, Powe DG, Rakha EA, Soria D, Lemetre C, Nolan CC et al (2013) Identification of key clinical phenotypes of breast cancer using a reduced panel of protein biomarkers. Br J Cancer 109: 1886-1894.

21. Rakha EA, Soria D, Green AR, Lemetre C, Powe DG, Nolan CC et al (2014) Nottingham Prognostic Index Plus (NPI+): a modern clinical decision making tool in breast cancer. Br J Cancer 110: 1688-1697.

22. Joseph C, Macnamara O, Craze M, Russell R, Provenzano E, Nolan CC et al (2018) Mediator complex (MED) 7: a biomarker associated with good prognosis in invasive breast cancer, especially ER+ luminal subtypes. Br J Cancer 118: 1142-1151.

23. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF et al (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: 340-350.

24. McCarty KS Jr, Miller LS, Cox EB, Konrath J, McCarty KS Sr (1985) Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Arch Pathol Lab Med 109: 716-721.

25. Detre S, Saclani Jotti G, Dowsett MA (1995) "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas. J Clin Pathol 48: 876-878.

26. Wang Y, Zhou BP (2011) Epithelial-mesenchymal transition in breast cancer progression and metastasis. Chinese Journal of Cancer 30: 603-611.

27. Yao D, Dai C, Peng S (2011) Mechanism of the mesenchymal-epithelial transition and its relationship with metastatic tumor formation. Mol Cancer Res 9: 1608-1620.

28. Guo Z, Neilson LJ, Zhong H, Murray PS, Zanivan S, Zaidel-Bar R (2014) E-cadherin interactome complexity and robustness resolved by quantitative proteomics. Sci Signal 7: rs7.

29. Yoshida-Noro C, Takeichi M (1992) Teratocarcinoma cell-adhesion: identification of cell-surface protein involved in calcium dependent cell-aggregation. Cell 28: 217-224.

30. Nose A, Takeichi M (1986) A novel cadherin cell-adhesion surface molecule. Its expression patterns associated implantation and organogenesis of mouse embryos. J Cell Biol 103: 2649-2658.

31. Takeichi M (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. Science 51: 1451-1455.

32. Aberle H, Schwartz H, Kemler R (1996) Cadherin-catenin complex: protein interactions and their implications for cadherin functions. J Cell Biol 61: 514-523.

33. Chintamani, Rekhi B, Bansal A, Bhatnagar D, Saxena S (2010) Expression of E-Cadherin in breast carcinomas and its association with other biological markers - a prospective study. Indian J Surg Oncol 1: 40-46.

34. Foroni C1, Broggini M, Generali D, Damia G (2012) Epithelial-mesenchymal transition and breast cancer: role, molecular mechanisms and clinical impact. Cancer Treat Rev 38: 689-697.

35. Mallini P, Lennard T, Kirby J, Meeson A (2014) Epithelial-to-mesenchymal transition: what is the impact on breast cancer stem cells and drug resistance. Cancer Treat Rev 40: 341-348.

36. Bae YH, Ding ZJ, Das T, Wells A, Gertler F, Roy P (2010) Profilin1 regulates PI(3,4)P(2) and lamellipodin accumulation at the leading edge thus influencing motility of MDA-MB-231 cells. P Natl Acad Sci USA 107: 21547-21552.

37. Lafuente EM, van Puijenbroek AA, Krause M, Carman CV, Freeman GJ, Berezovskaya A et al (2004) RIAM, an Ena/VASP and Profilin ligand, interacts with Rap1-GTP and mediates Rap1-induced adhesion. Dev Cell 7: 585-595.

38. Batistela MS, Boberg DR, Andrade FA, Pecharki M, de S F Ribeiro EM, Cavalli IJ et al (2013) Amplification and deletion of the RAPH1 gene in breast cancer patients. Mol Biol Rep 40: 6613-6637.

39. Liu YL, Saraf A, Lee SM, Zhong X, Hibshoosh H, Kalinsky K et al (2016) Lymphovascular invasion is an independent predictor of survival in breast cancer after neoadjuvant chemotherapy. Breast Cancer Res Treat 157: 555-564.

40. He S, Zhang G, Dong H, Ma M, Sun Q (2016) miR-203 facilitates tumor growth and metastasis by targeting fibroblast growth factor 2 in breast cancer. Onco Targets Ther 9: 6203-6210.

41. Wang C, Zheng X, Shen C, Shi Y (2012) MicroRNA-203 suppresses cell proliferation and migration by targeting BIRC5 and LASP1 in human triple-negative breast cancer cells. J Exp Clin Cancer Res 31: 58.

42. Viticchiè G, Lena AM, Cianfarani F, Odorisio T, Annicchiarico-Petruzzelli M, Melino G, Candi E (2012) MicroRNA-203 contributes to skin re-epithelialization. Cell Death Dis 3: e435.

43. Wang L, Mou Y, Meng D, Sun Y, Chen X, Yang X et al (2017) MicroRNA-203 inhibits tumour growth and metastasis through PDPN. Clin Otolaryngol 42: 620-628.

FIGURE LEGENDS

Figure 1. Representative tissue microarray images of breast cancer tissue samples immunohistochemically stained for RAPH1.

In the cytoplasm of cancer cells, RAPH1 shows (a) no staining, (b) weak staining, (c) moderate staining and (d) strong staining (magnification: x200 for all images).

Figure 2. Survival curves of breast cancer patients stratified by RAPH1 expression.

(a) In TCGA cohort, breast cancer-specific survival is significantly worse in the RAPH1-high group than in the RAPH1-low group ($p = 0.049$). (b) Significant difference is observed in the breast cancer-specific survival between RAPH1-high and RAPH1-low tumours at protein phase ($p = 0.037$).

Figure 3. Morphological characteristics of RAPH1 immunohistochemistry

(a) Normal mammary gland cells show absent or weak RAPH1 staining (magnification: x200). (b) RAPH1 immunoreactivity is uniformly weak in normal epithelial cells and lower in myoepithelial cells than in glandular cells (magnification: x400). (c) RAPH1

immunoreactivity is significantly different between intraductal cancer cells and the normal

mammary gland adjacent to the tumour (white arrow: normal mammary gland; back arrow:

intraductal cancer cells; magnification: x100)). (d) Invasive cancer cells showing strong

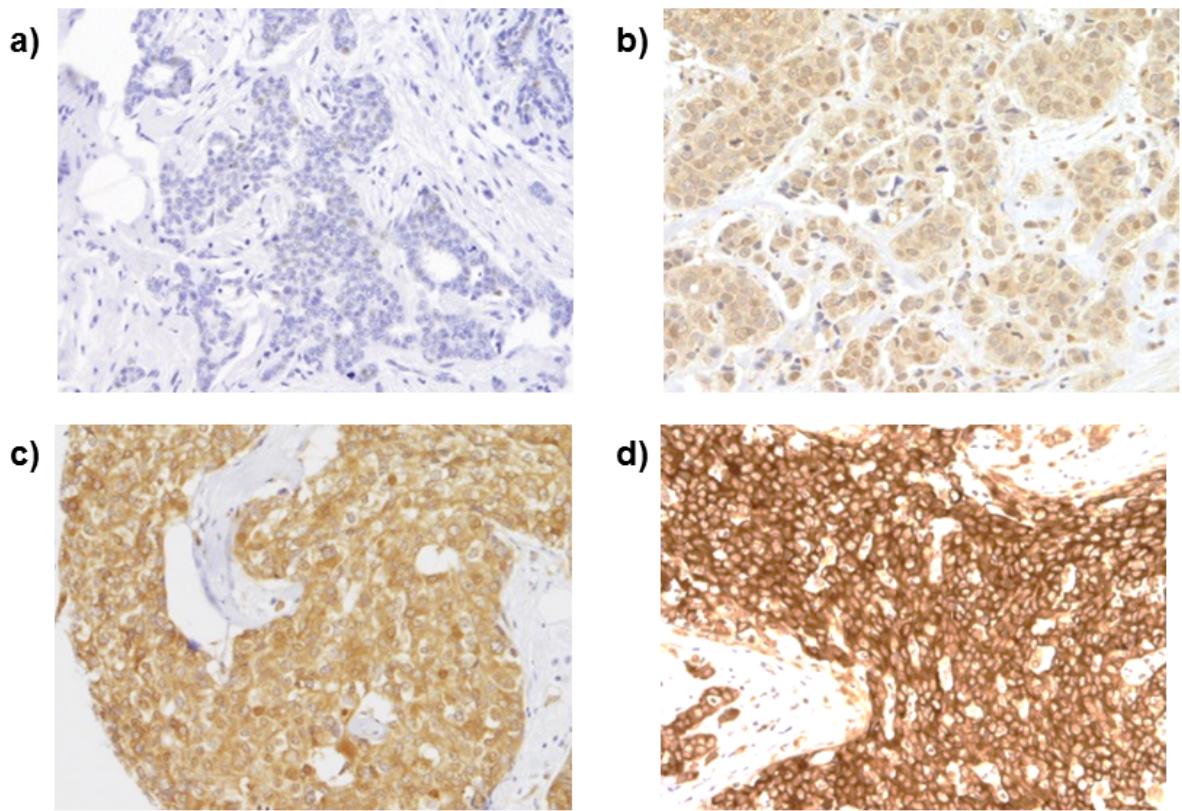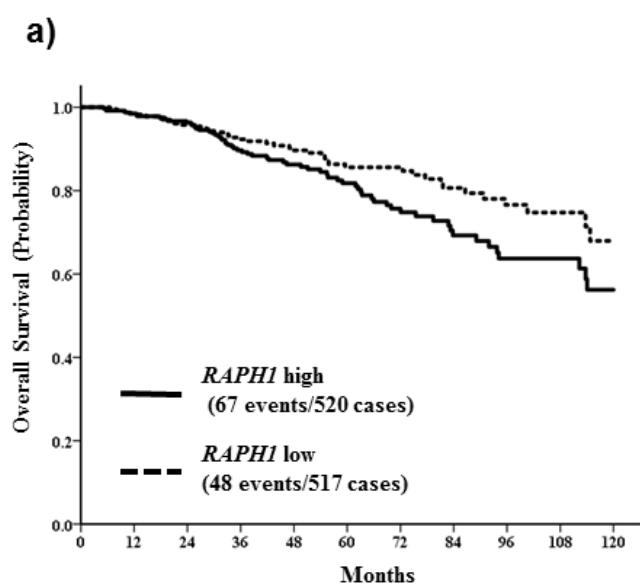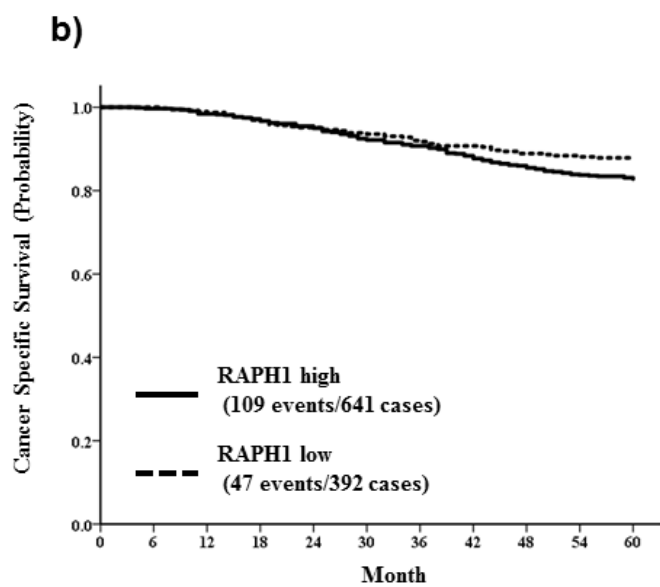RAPH1 immunoreactivity mainly recognised in the cytoplasm (magnification: x200).

Figure 1

a)

Overall Survival (Probability)

RAPH1 high
(67 events/520 cases)

RAPH1 low
(48 events/517 cases)

Months

High vs. low *RAPH1* mRNA expression
Hazard ratio 3.88, $p = 0.049$

b)

Cancer Specific Survival (Probability)

RAPH1 high
(109 events/641 cases)

RAPH1 low
(47 events/392 cases)

Month

High vs. low RAPH1 protein expression
Hazard ratio 4.37, $p = 0.037$

Figure 2

Figure 3

130 kDa ➡️

95 kDa ➡️

70 kDa ➡️

⬅️ *RAPH1*
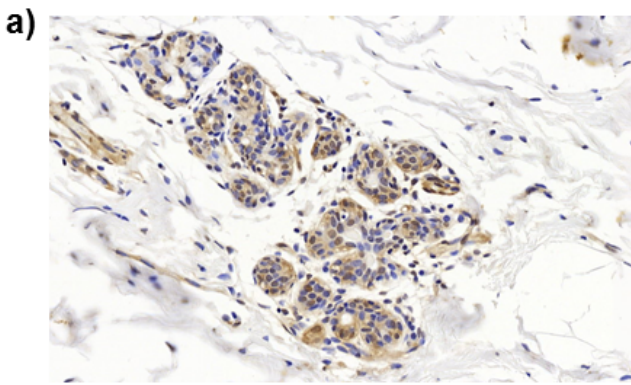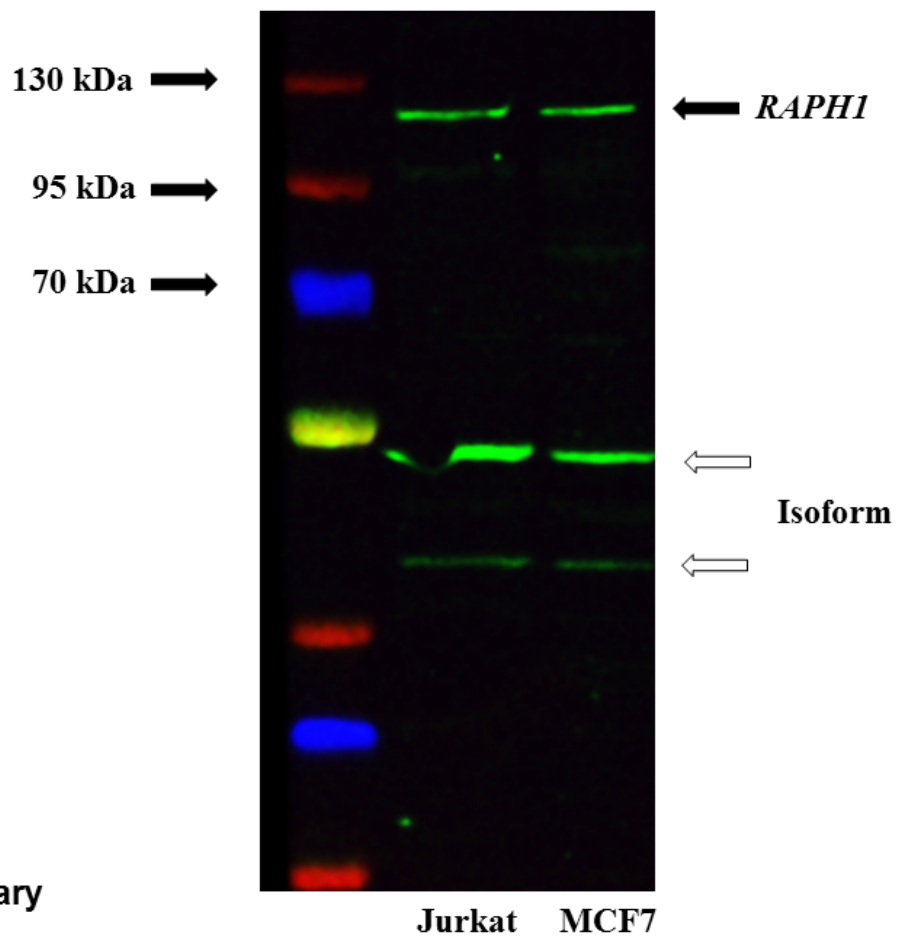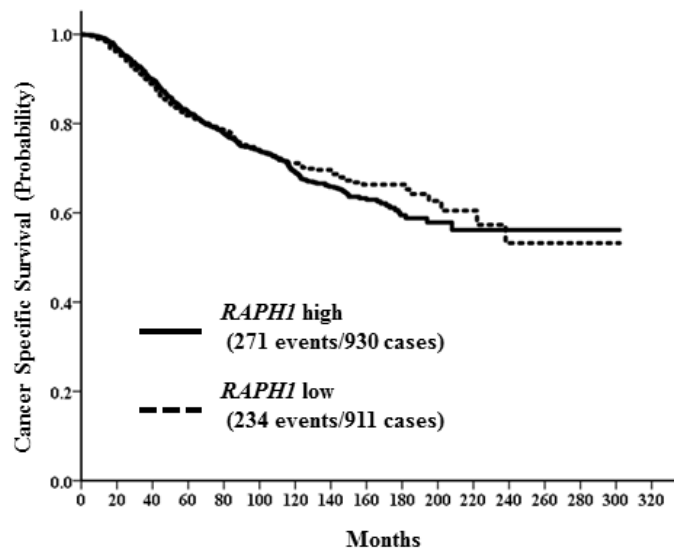
Isoform

Supplementary Figure 2

Jurkat   MCF7
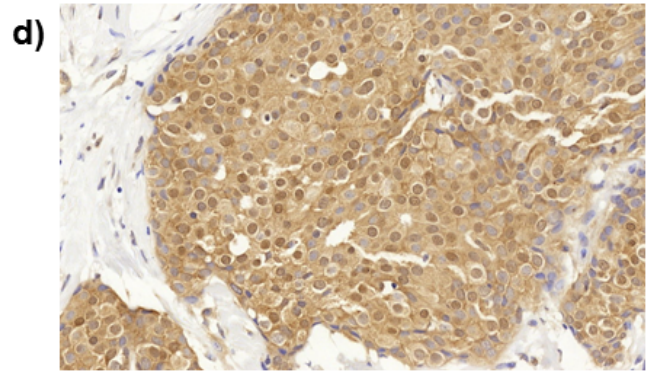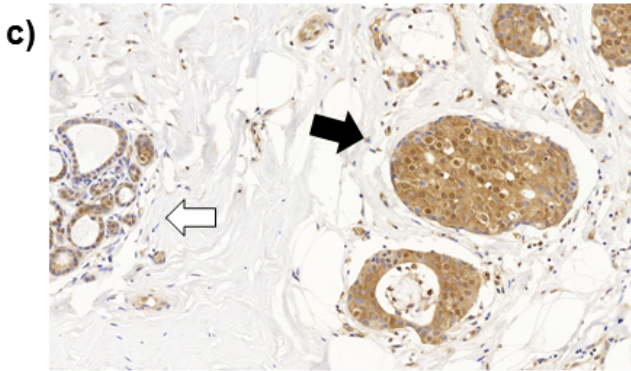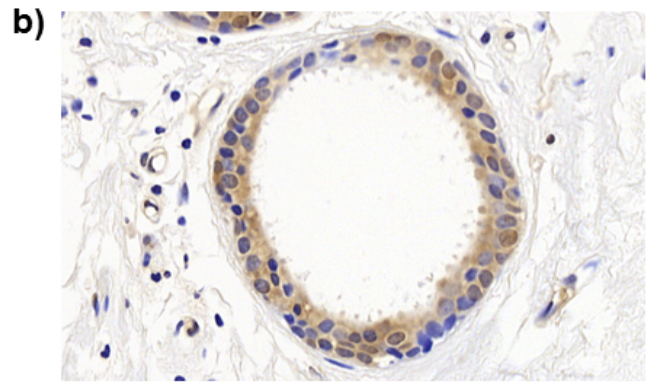
High vs. low *RAPH1* mRNA expression
Hazard ratio 0.39, *p* = 0.53

**Supplementary
Figure 3**

Table 1. Association of *RAPH1* mRNA expression with mRNA expression of other genes associated with EMT related genes, cancer stem cell related genes and basal markers

Gene names	TCGA cohort		METABRIC cohort	
	Correlation value	*p*-value	Correlation value	*p*-value
CDH1	0.20	< 0.0001*	0.058	0.0096*
CDH2	0.012	0.69	-0.056	0.012*
TGFB1	-0.17	< 0.0001*	-0.073	0.0011*
KRT5	-0.09	0.0035*	-0.041	0.071
KRT6A	-0.053	0.087	-0.047	0.035*
KRT6B	-0.058	0.063	-0.055	0.014*
KRT6C	-0.085	0.0059*	-0.050	0.026*
EGFR	-0.024	0.44	-0.017	0.46
CD24	0.025	0.41	0.063	0.0048*
CD44	0.19	< 0.0001*	0.27	< 0.0001*
ALDH1A3	0.086	0.0057*	0.0081	0.72

Abbreviations: TCGA, The Cancer Genome Atlas; METABRIC, The Molecular Taxonomy of Breast Cancer International Consortium.
* Significant difference *p*<0.05.

Table 2. Correlation between RAPH1 protein expression and clinicopathological characteristics

Factors		Expression of RAPH1		Total	p-value
		Low	High		
Tumor size	≥ 2cm	219 (38.8%)	345 (61.2%)	564	0.45
	< 2cm	174 (36.6%)	302 (63.5%)	476	
Nodal status	Positive	140 (34.7%)	264 (65.4%)	404	0.097
	Negative	253 (39.8%)	383 (60.2%)	636	
Histological grade	Grade 3	164 (31.0%)	365 (69.0%)	529	< 0.0001*
	Grade 1, 2	229 (44.8%)	282 (55.2%)	511	
Lymphovascular invasion	Positive	170 (40.0%)	255 (60.0%)	425	0.220
	Negative	223 (36.3%)	392 (63.7%)	615	
ER	Positive	308 (39.6%)	469 (60.4%)	777	0.034
	Negative	85 (32.3%)	178 (67.7%)	263	
PgR	Positive	242 (39.4%)	372 (60.6%)	614	0.19
	Negative	151 (35.5%)	275 (64.6%)	426	
HER2	Positive	53 (41.1%)	76 (58.9%)	129	0.41
	Negative	340 (37.3%)	571 (62.7%)	911	
Subtypes	HR positive/HER2 negative	284 (39.8%)	429 (60.2%)	713	0.0088*
	HER2 positive	53 (41.1%)	76 (58.9%)	129	
	Triple negative	56 (28.3%)	142 (71.7%)	198	
Ki67	High (≥ 10%)	145 (27.9%)	374 (72.1%)	519	< 0.0001*
	Low (< 10%)	145 (47.1%)	163 (52.9%)	308	

E-cadherin	**High**	146 (32.7%)	300 (67.3%)	446	**0.0046***
	Low	236 (41.4%)	334 (58.6%)	570	
N-cadherin	**High**	194 (34.7%)	365 (65.3%)	559	**0.45**
	Low	83 (37.6%)	138 (62.4%)	221	
CK 5/6	**High**	67 (39.2%)	104 (60.8%)	171	**0.67**
	Low	321 (37.5%)	536 (62.5%)	857	
EGFR	**High**	74 (35.2%)	136 (64.8%)	210	**0.43**
	Low	311 (38.2%)	503 (61.8%)	814	
CD24	**High**	194 (37.2%)	327 (62.8%)	521	**0.30**
	Low	46 (42.6%)	62 (57.4%)	108	
CD44	**High**	79 (33.1%)	160 (67.0%)	239	**0.23**
	Low	107 (38.1%)	174 (61.9%)	281	

Abbreviations: ER, oestrogen receptor; PR, progesterone receptor; HR, hormonal receptor.

Some variables do not add up to 1040 for all patients due to missing data.

* Significant difference $p<0.05$.

Table 3. Survival analysis based on clinicopathological characteristics, including RAPH1 protein expression in lymphovascular invasion positive cases

Factors		Univariate analysis			Multivariate analysis		
		Hazard Ratio	95% CI	*p*-value	Hazard Ratio	95% CI	*p*-value
RAPH1 expression	Low	Reference			Reference		
	High	1.44	1.02-2.02	0.038*	1.43	1.01-2.03	0.046*
Tumor size	< 2cm	Reference			Reference		
	≥ 2cm	2.99	2.07-4.33	< 0.0001*	1.78	1.21-2.62	0.0033*
Nodal status	Negative	Reference			Reference		
	Positive	3.20	2.30-4.45	<0.0001*	2.49	1.76-3.53	< 0.0001*
Lymphovascular invasion	Negative	Reference			Reference		
	Positive	2.89	2.08-4.01	< 0.0001*	2.16	1.52-3.05	< 0.0001*
Subtypes	HR positive/HER2 negative	Reference			Reference		
	HER2 positive	4.69	3.18-6.92	< 0.0001*	3.78	2.55-5.63	< 0.0001*
	Triple negative	3.45	2.38-4.99	< 0.0001*	3.12	2.14-4.57	< 0.0001*
E-cadherin	Low	Reference			Reference		
	High	0.68	0.49-0.95	0.023*	0.67	0.48-0.94	0.019*

Abbreviations: ER, oestrogen receptor; PR, progesterone receptor; HR, hormonal receptor.
* Significant difference *p*<0.05.

Supplementary Table 1. Characteristics of the Nottingham primary cohort patients

Age range (years)	Patients (n)
24–40	114
41–59	595
60 and over	331
Menopausal status	
Pre	426
Post	614
Tumour size	
<2.0 cm	476
≥2.0 cm	564
Nodal status	
Negative	636
Positive	404
Lymphovascular invasion	
Negative	615
Positive	425
Type of breast surgery	
Breast-conserving surgery	450
Mastectomy	590
Axillary surgery	
Sampling alone	631
Axillary lymph node dissection	404
No surgery	5
Chemotherapy	
Yes	212
No	789
Unknown	39
Endocrine therapy	
Yes	405
No	597
Unknown	38