



Identification and validation of DOCK4 as a potential biomarker for risk of bone metastasis development in patients with early breast cancer.

Journal:	<i>The Journal of Pathology</i>
Manuscript ID	18-325.R1
Wiley - Manuscript type:	Original Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Westbrook, Jules; University of Sheffield, Academic Unit of Clinical Oncology, Department of Oncology and Metabolism; University of Leeds, Clinical and Biomedical Proteomics Group, Leeds Institute of Cancer and Pathology</p> <p>Wood, Steven; University of Sheffield, Academic Unit of Clinical Oncology, Oncology and Metabolism; University of Leeds, Clinical and Biomedical Proteomics Group, Leeds Institute of Cancer and Pathology</p> <p>Cairns, David; University of Leeds, Clinical and Biomedical Proteomics Group, Leeds Institute of Cancer and Pathology</p> <p>McMahon, Kathryn; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Gahlaut, Renu; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Thygesen, Helene; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Shires, Mike; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Roberts, Stephanie; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Marshall, Helen; University of Leeds, Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research</p> <p>Oliva, Maria; University of Sheffield, Academic Unit of Clinical Oncology, Department of Oncology and Metabolism</p> <p>Dunning, Mark; University of Sheffield, Neurosciences</p> <p>Hanby, Andrew; University of Leeds, Leeds Institute of Cancer and Pathology (LICAP)</p> <p>Selby, Peter; University of Leeds, Clinical and Biomedical Proteomics Group, Leeds Institute of Cancer and Pathology</p> <p>Speirs, Valerie; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Mavria, Georgia; University of Leeds, Leeds Institute of Cancer and Pathology</p> <p>Coleman, Robert; University of Sheffield, Academic Unit of Clinical Oncology, Department of Oncology and Metabolism</p> <p>Brown, Janet; University of Sheffield Department of Oncology and Metabolism, Oncology and Metabolism</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tissue:	
Pathology:	
Technique:	

SCHOLARONE™
Manuscripts

For Peer Review

1
2
3 **Identification and validation of DOCK4 as a potential biomarker for risk of bone**
4 **metastasis development in patients with early breast cancer**
5
6
7
8

9 Running Title: DOCK4 in breast cancer bone metastasis
10
11

12
13 Jules A. Westbrook^{1,2}, Steven L. Wood^{1,2}, David A. Cairns^{2,4}, , Kathryn McMahon³,
14 Renu Gahlaut³, Helene Thygesen³, Mike Shires³, Stephanie Roberts², Helen
15 Marshall⁴, [Maria R. Oliva¹](#), [Mark J. Dunning⁵](#), Andrew M. Hanby³, Peter J. Selby^{2,3},
16 Valerie Speirs^{3,6}, Georgia Mavria³, Robert E. Coleman¹, Janet E. Brown^{1,2}.
17
18
19
20
21
22
23

24 ¹Academic Unit of Clinical Oncology, Department of Oncology and Metabolism,
25 University of Sheffield, UK; ²Clinical and Biomedical Proteomics Group, Leeds
26 Institute of Cancer and Pathology, University of Leeds, UK; ³Leeds Institute of
27 Cancer and Pathology, University of Leeds, UK; ⁴Clinical Trials Research Unit,
28 Leeds Institute of Clinical Trials Research, University of Leeds, UK; ⁵[Institute for](#)
29 [Translational Neuroscience, Department of Neuroscience, University of Sheffield,](#)
30 [UK;](#) ⁶[Institute of Medical Sciences, University of Aberdeen, UK.](#)
31
32
33
34
35
36
37
38
39
40
41
42
43

44 Conflicts: Prof Janet Brown received fees from Novartis and Amgen for Advisory
45 Boards and Speakers Bureaux; Prof Rob Coleman received consultancy fees from
46 [Amgen, Astellas and Boehringer Ingelheim](#), and lecture fees from Amgen. All other
47 authors declared no conflicts.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Corresponding Author: Janet E Brown (j.e.brown@sheffield.ac.uk), Academic Unit
of Clinical Oncology, Department of Oncology and Metabolism, University of
Sheffield, Weston Park Hospital, Whitham Road, Sheffield S10 2SJ, UK.

For Peer Review

ABSTRACT

Skeletal metastasis occurs in around 75% of advanced breast cancers, with the disease incurable once cancer cells disseminate to bone, but there remains an unmet need for biomarkers to identify patients at high risk of bone recurrence. This study aimed to identify such a biomarker and to assess its utility in predicting response to adjuvant zoledronic acid.

We used quantitative proteomics (SILAC-MS), to compare protein expression in a bone-homed variant (BM1) of the human breast cancer cell line MDA-MB-231 with parental non-bone-homing cells to identify novel biomarkers for risk of subsequent bone metastasis in early breast cancer. SILAC-MS showed that Dedicator of cytokinesis protein4 (DOCK4) was upregulated in bone-homing BM1 cells, confirmed by Western blotting. BM1 cells also had enhanced invasive ability compared with parental cells which could be reduced by DOCK4-shRNA.

In a training Tissue Microarray (TMA) comprising 345 patients with early breast cancer, immunohistochemistry followed by Cox regression revealed that high DOCK4 expression correlated with histological grade ($p=0.004$) but not oestrogen receptor status ($p=0.19$) or lymph node involvement ($p=0.15$). A clinical validation TMA used tissue samples and the clinical database from the large AZURE adjuvant study ($n=689$). Adjusted Cox regression analyses showed that high DOCK4 expression in the control arm (no zoledronic acid) was significantly prognostic for first recurrence in bone (HR 2.13, 95%CI 1.06-4.30, $p=0.034$). No corresponding association was found in patients who received zoledronic acid (HR 0.812, 95%CI 0.176-3.76, $p=0.790$), suggesting that treatment with zoledronic acid may counteract the higher risk for bone relapse from high DOCK4-expressing tumours.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

High DOCK4 expression was not associated with metastasis to non-skeletal sites when these were assessed collectively. In conclusion, high DOCK4 in early breast cancer is significantly associated with aggressive disease and with future bone metastasis and is a potentially useful biomarker for subsequent bone metastasis risk.

Keywords: DOCK4; Bone metastasis; breast cancer; biomarker; proteomics

For Peer Review

INTRODUCTION

Despite substantial progress in early detection and treatment, breast cancer still accounts for 15% of female cancer-related deaths, with skeletal metastasis occurring in approximately 75% of patients with advanced disease [1]. Relapse in bone typically occurs years after apparently successful treatment of early breast cancer and a period of tumour dormancy. Bone-targeted agents such as bisphosphonates [2] and denosumab [3] are widely used to treat the skeletal complications of established bone metastases, but have also recently been the focus of several large adjuvant studies in early breast cancer to assess their potential to reduce the frequency of relapse in bone and subsequent breast cancer mortality. The phase III AZURE trial (BIG01/04-ISRCTN79831382) recruited 3360 patients with stage II/III breast cancer randomised (1:1) to 5 years of standard adjuvant therapy alone (control) or standard therapy with zoledronic acid (zoledronate[4]. Although there was no significant difference in invasive disease-free survival in the overall population, zoledronate improved disease outcomes for women who were >5 years postmenopausal at diagnosis and a meta-analysis of 26 randomised trials (N=18,766), demonstrated that bone recurrences (HR=0.72; 95%CI 0.60, 0.86, 2p=0.0002) and breast cancer deaths (HR 0.82; 95%CI 0.73, 0.93, 2p=0.002) were reduced by adjuvant bisphosphonates in post-menopausal women [5]. Breast cancer practice has changed as a result of these studies, but they also highlight the unmet need for biomarkers to identify patients with early breast cancer who are most at risk of developing bone recurrence, thus permitting tailoring of treatment to patients most likely to benefit and sparing patients who would not benefit, from potential complications [6]. ~~Tissue samples from AZURE trial patients and the associated clinical database, prospectively designed to support biomarker~~

1
2
3 studies, offer a major opportunity for clinical validation of novel potential biomarkers
4
5 for bone metastasis and other disease related outcomes.
6
7
8

9 Proteomic studies are yielding key information about breast cancer metastasis to bone
10 [7-9] and, in a recent proteomics-based study, validated in 571 patients, we showed
11 that the proteins CAPG and GIPC1 had both prognostic and predictive potential as
12 biomarkers of bone metastasis [10]. In the current study, we hypothesized that
13 proteins, identified by proteomics and up-regulated in breast cancer cells which have
14 a propensity to home to bone, would be potential biomarkers for metastasis and could
15 play key mechanistic roles in the process of metastatic dissemination to bone.
16
17
18
19
20
21
22
23
24
25

26 **MATERIALS AND METHODS**

27
28
29
30 **Proteomic studies** (Further details to those below are given in [the text](#) and in [Figure](#)
31 [S1 of Supporting Information](#))
32
33
34

35 ***Cell culture and SILAC (Stable Isotope Labelling by Amino acids in Cell culture)***

36 The human breast cancer cell line MDA-MB-231 (“PCC”, parental control cells,
37 obtained originally from ATCC) and a bone-homing variant (“BM1”, bone metastatic
38 cells) [11] (the latter supplied by Prof. Joan Massague, Sloan-Kettering Institute, New
39 York), were used in a ‘classical’ SILAC experiment [12,13]. PCC and BM1 cells
40 were cultured in ‘heavy’ SILAC medium and ‘light’ medium for ~10 doublings to
41 ensure >95% isotope label incorporation (Heavy media), and all cells were used
42 within 10 doublings. ‘Heavy’ SILAC media consisted of DMEM containing L-
43 arginine ($^{13}\text{C}_6$, $^{15}\text{N}_4$) and L-lysine ($^{13}\text{C}_6$, $^{15}\text{N}_2$) (R10K8, DMEM-15, Dundee Cell
44 Products, Dundee, UK) supplemented with dialysed bovine serum (D-FBS100,
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Dundee Cell Products, Dundee, UK). 'Light' media was DMEM but without the
4
5 heavy isotopes (R0K0).
6
7

8 9 ***LC-MS/MS***

10
11 Equal amounts of protein (40 µg) from light and heavy labelled samples were
12
13 combined, reduced, alkylated, and separated on a 1-D SDS-PAGE gel. LC-MS-MS
14
15 was performed by Dundee Cell products, Dundee, Scotland, UK. Ten slices of the
16
17 gel-resolved proteins were cut and proteins were digested to peptides using trypsin.
18
19 Tryptic peptides were separated using a nanoflow LC-System coupled to an LTQ-
20
21 Orbitrap mass spectrometer (ThermoFisher Scientific).
22
23

24 ***Quantitation and bioinformatics analysis***

25
26 Quantitation was performed using the software Max Quant
27
28 (<http://www.maxquant.org/downloads.htm>), with peptide ratios calculated for each
29
30 arginine- and/or lysine-containing peptide as the peak area of labelled arginine/lysine
31
32 divided by the peak area of non-labelled arginine/lysine for each single-scan mass
33
34 spectrum. Peptide ratios for all arginine- and lysine-containing peptides sequenced
35
36 for each protein were averaged. Data output from Max Quant were analysed further
37
38 using Excel and R (v. 3.2, <http://www.r-project.org/>) to select differentially expressed
39
40 proteins.
41
42
43

44 **Western blotting**

45
46 The differential expression of DOCK4, and confirmation of DOCK4 knockdown, was
47
48 assessed in cell lysates using Western blotting and an infra-red (IR) immunodetection
49
50 system (LI-COR Biosciences), as well as by Enhanced ChemiLuminescence (ECL).
51
52 Primary antibodies used were: DOCK4 (Abcam ab56743, mouse
53
54 monoclonal, 0.1µg/mL); beta-tubulin (Abcam ab6046, rabbit polyclonal, 1/5000
55
56
57
58
59
60

1
2
3 dilution). Secondary antibodies were: #925-68070 IRDye 680RD goat anti-mouse
4 IgG (H+L), LI-COR Biosciences, 1/5000 dilution, for detection of DOCK4 in the
5 700nm channel (red); #925-32211 IRDye 800CW goat anti-rabbit IgG (H+L), LI-
6 COR Biosciences, 1/5000 dilution, for detection of beta-tubulin in the 800nm channel
7 (green). For ECL secondary antibodies were Goat-anti-mouse-HRP (Abcam, ab 6789)
8 and goat anti-rabbit-HRP (Abcam, Ab6721, both 1:2,500). Normalised densitometric
9 data from six replicate runs of the fluorescent-antibody immunoprobed samples, (and
10 three blots from the ECL-visualized samples), were tested for significance using
11 Student's t-test.
12
13
14
15
16
17
18
19
20
21
22

23 **Generation of cell lines with stable DOCK4 knockdown and 3D invasion assay**

24
25
26 DOCK4 expression was knocked down in the BM1 and PCC cell lines by lentiviral
27 delivery of a validated shRNA targeting DOCK4 (shR), with the use of an empty
28 vector (cv) as control [14]. Untransduced, wild-type (wt) cells were also studied as
29 control. Inclusion of a GFP marker protein within the lentiviral vectors enabled
30 FACS separation of vector-bearing clones and subsequent culture of pure cell
31 populations. Infected cells were selected by FACS sorting 48 hrs after lentiviral
32 infection. Confirmation of knockdown of DOCK4 protein expression was carried out
33 by Western blotting as described above. Quantification of the blots was performed by
34 densitometric scanning.
35
36
37
38
39
40
41
42
43
44
45

46 To study the effects of reduced expression of DOCK4 protein in PCC and BM1, an
47 invasion assay was carried out using the IncuCyte® platform (Essen BioScience)
48 following the manufacturer's instructions for a 'scratch-wound' procedure. Briefly,
49 triplicate wells of an ImageLock 96-well plate (Essen BioScience 4379) were over-
50 laid with thin Matrigel onto which wt, cv, and shR cells for PCC and BM1 were
51
52
53
54
55
56
57
58
59
60

1
2
3 seeded (50k cells/well and 100k cells/well, respectively), and a homogenous scratch
4
5 was created using the Incucyte® WoundMaker tool when the cell monolayer was
6
7 confluent. Cell invasion was assessed by measuring the closure of the scratch
8
9 introduced on the confluent cell monolayer. Experiments were performed in triplicate
10
11 using independent cultures of cells. Data were analysed using IncuCyte® software
12
13 and Excel.
14
15
16
17

18 **Patients, samples and immunohistochemistry:**

19
20 All analyses on patient samples were performed with Ethics approval (Leeds training
21
22 set, 06/Q1206/180; AZURE validation set 55/03/182).
23

24 Initial studies of DOCK4 expression were performed in a training TMA comprising
25
26 345 specimens with available tumour type, grade, ER, lymph node (LN) and overall
27
28 survival (OS) data, from breast tumours diagnosed at the Leeds Teaching Hospitals
29
30 NHS Trust (1987 – 2005). Samples were stained with a rabbit polyclonal DOCK4
31
32 antibody (1:100, Bethyl Laboratories Inc;A302-263A) and corroborated in a smaller
33
34 cohort with a mouse monoclonal antibody (1:50, Abcam;Ab56743).
35
36

37 The main analyses for correlations with risk of bone metastasis were performed on
38
39 TMAs constructed from primary tumours from [a sub-set of patients within the overall](#)
40
41 AZURE trial (n = 689). Due to the relatively high prevalence of bone metastatic
42
43 outcomes and the long follow-up (median 84 months [interquartile range 66-93]),
44
45 these TMAs provide an excellent resource for validation of protein biomarkers
46
47 emerging from our proteomics studies. Protein expression was assessed using the
48
49 Bethyl DOCK4 antibody (A302-263A). [DOCK4 specificity for this antibody was](#)
50
51 [confirmed by immunohistochemistry using FFPE cell pellets from BM1 cells with](#)
52
53 [and without DOCK4 knockdown \(Figure S2, Supporting Information\).](#) Full gel
54
55
56
57
58
59
60

1
2
3 images for the antibodies using ECL as the visualization method with use of positive
4 and negative control cell-lines are included (Figure S3, Supporting Information).
5
6
7

8
9 Digital images were created using a digital scanner (Aperio Scan Scope XT).
10
11 Cytoplasmic staining assessment in the invasive margins of primary breast tumours
12 was carried out independently by two trained operators, blinded to outcome data,
13 under the supervision of an experienced breast histopathologist (AMH) who also
14 adjudicated discrepant scores and the level of agreement of the two scores was
15 measured using Cohen's kappa coefficient. All scores were based upon intensity of
16 staining and not the number of positive cells. Staining intensity was ranked based on
17 a three tier ordinal categorical system used to rank the tumours based on intensity of
18 cytoplasmic staining [15,16] where 1 = weak staining; 2 = moderate, easily perceived
19 staining; 3 = strong/intense staining, i.e. the scoring was based on staining intensity
20 only.
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 **Statistical analyses**

36
37
38 All immunohistochemical analyses followed REMARK guidelines [17]. Statistical
39 analyses evaluated the associations between protein expression and relevant clinical
40 and pathological variables (e.g. ER/PR/HER2 status) using Fisher's Exact test
41 (categorical variables) and the Kruskal-Wallis test (continuous variables), before
42 assessing prognostic and predictive associations with time-to-event data (time to first
43 distant recurrence, time to first skeletal recurrence, time to first non-skeletal
44 recurrence) using Cox proportional hazards regression, the Kaplan-Meier estimate of
45 the survival function and the log-rank test. Time to first distant recurrence was
46 defined as the time from the date of randomization to the date of the distant
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 recurrence. In analyses, other types of events were censored, e.g. if a local recurrence
4 occurred prior to any distant recurrence, the patient would be censored at the date of
5 the local recurrence. Time to first skeletal recurrence and first non-skeletal recurrence
6 were defined similarly. Time to first skeletal recurrence irrespective of all other
7 previous recurrences was also investigated. Time to event analysis was first
8 performed within treatment arms to identify prognostic associations with the
9 biomarkers. The predictive heterogeneity of effect between treatment arms for time
10 to distant events was assessed in multivariable analysis by including an interaction
11 term in the Cox proportional hazard regressions for treatment arm and biomarker
12 (while adjusting for systemic therapy plan, ER status and lymph node involvement).
13 All significance tests were two-sided and were designated significant at the 5% level.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 **RESULTS**

30 31 **Proteomic identification of proteins specifically associated with breast cancer** 32 **bone metastasis** 33 34

35 Proteins identified in the 'forward' and 'reciprocal' labelling experiments (Figure S1)
36 were aligned and the most robust, highest quality, data were extracted for analysis
37 using the following stringent selection criteria: protein identified in both datasets; ≥ 2
38 so-called razor+unique peptides assigned to identifications (clarifies assignment of
39 protein identification); equal numbers of peptides and razor+unique peptides assigned
40 to identifications. This resulted in 2006 proteins taken forward for analysis out of a
41 total of 2999 identified in the complete data set (Full proteomic data are available on
42 the publicly accessible database ORDA (<https://orda.shef.ac.uk/>)).
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The aligned dataset was filtered further by using a 1.75-fold cut-off to distinguish
4 change (up or down) in protein expression, resulting in 48 proteins up-regulated in
5 BM1 relative to PCC cells (Table S1, Supporting Information). These were
6 prioritized for further study using literature evidence of relevance to bone metastasis,
7 the magnitude of the differential expression fold change and evidence (where
8 available) of non-association with lung metastases, based on correlation with our
9 other proteomic datasets that included a MDA-MB-231 variant which specifically
10 homes to lung (10, 11). This allowed us to identify proteins likely to be involved in
11 breast cancer metastasis to bone. Consequently, four proteins had potential for further
12 consideration: Dedicator of cytokinesis protein 4 (DOCK4, fold change 2.7),
13 SerpinB2 (fold change 15.6), cell-division cycle protein 20 homolog (CDC20, fold
14 change 3.7), and pericentrin (fold change 2.5).
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 While all four have published evidence linking them to breast cancer we focused on
32 DOCK4 for further investigation and clinical validation based on its published role in
33 cell migration including breast cancer cell migration, metastasis [18-21] and tumour
34 angiogenesis (14), processes known to be integral to dissemination of tumour cells
35 and development of bone metastases. Moreover, DOCK4 functions as a guanine
36 nucleotide exchange factor for the GTPase Rac1, a key regulator of motility
37 [14,19,20] and localises at actin-rich protrusions in migrating breast cancer cells [19],
38 whilst SNPs within the promoter region of DOCK4 have been detected in breast
39 cancer [22].
40
41
42
43
44
45
46
47
48
49
50
51

52 **Confirmation of DOCK4 upregulation in BM1 cells**

53
54
55
56
57
58
59
60

1
2
3 Analysis of DOCK4 expression levels by Western blotting in the BM1 and PCC cell-
4 lines showed a 2-fold increase in DOCK4 expression within the bone homing BM1
5 cell-line compared to parental PCC (Figure 1A,C). Increased expression of DOCK4
6 within BM1 cells was also confirmed using ECL-based visualization.
7
8
9
10

11 12 13 **DOCK4 is upregulated during breast cancer cell-motility**

14
15 There was significant knockdown of DOCK4 protein expression in the BM cell type
16 following lentiviral delivery of DOCK4 shRNA (shR) compared with the empty
17 vector (cv) control (Figure 1B,D). In the wound healing assay, BM1 cells (control
18 vector) had enhanced invasive ability compared with the PCC cells, with 13% and
19 39% wound closure at 6 and 12h respectively, compared to 9 and 20% for PCC at 6
20 and 12hrs (Figure 1E,F). Invasion through Matrigel was reduced in cells with
21 DOCK4 knockdown (shR) compared to control, with significant differences in wound
22 closure observed for both cell types at 6 hours and 12 hours, and with a greater effect
23 seen at 6 hrs.
24
25
26
27
28
29
30
31
32
33
34
35
36

37 **Association between DOCK4 expression and tumour grade in local breast** 38 **tumour array**

39
40 DOCK4 expression was initially assessed in a local breast tumour array of 345
41 unselected breast tumours (88% ductal, 9% lobular, 3% other) with patient data
42 available on tumour grade (18% Grade 1, 44% Grade 2, 38% Grade 3) oestrogen
43 receptor (ER), and axillary lymph node (LN) involvement. Examples of typical
44 staining patterns are shown in Figure 2. Analysis revealed a significant association
45 between DOCK4 expression and histological tumour type ($p = 0.002$) and tumour
46 grade ($p=0.004$) with 86.4% Grade 3 and 77.3% grade 2 ductal carcinomas expressing
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 moderate/high DOCK4 (as opposed to 62.5% grade 1 carcinomas), but no association
4
5 with ER status (positive versus negative, $p=0.185$) or LN status (involved versus not
6
7 involved, $p=0.15$). These data suggested that DOCK4 expression was associated with
8
9 tumour aggressiveness and further supported our selection of DOCK4 for clinical
10
11 validation in the AZURE patient cohort.
12
13
14

15 16 **Clinical validation of DOCK4 expression in breast cancer patients in the AZURE** 17 **study.**

18 ***Patients.***

19
20
21
22
23 Independent scoring of stained TMAs by two trained operators (JW and SR) under the
24
25 supervision of AMH yielded a Cohen's kappa coefficient value of 0.86, signifying
26
27 excellent agreement. Possible associations between clinical outcomes and
28
29 immunohistochemistry scores for DOCK4 were tested for 689 patients in the AZURE
30
31 study (330 control arm, 359 zoledronate arm). Demographic data for these patients
32
33 are similar to those of the whole AZURE population (Table 1). DOCK4 had no
34
35 significant associations with age, lymph node involvement, ER status, tumour grade
36
37 (though this approached significance, $p = 0.062$), menopausal status, systemic
38
39 therapy, chemotherapy and statin use. HER2 status was not a mandated assessment
40
41 but was available for 307 participants. DOCK4 was significantly associated with
42
43 HER2 status with a smaller proportion of patients with low DOCK4 being HER2
44
45 positive ($P<0.001$). However, adjustment for HER2 status had no impact on
46
47 subsequent statistical analyses in either control or zoledronate arms.
48
49
50
51
52

53 ***AZURE patients: Association of DOCK4 expression with distant event recurrence.***

54
55
56
57
58
59
60

1
2
3 **Control arm.** Initial analyses considered possible associations between DOCK4
4 expression (scored as 1, 2 or 3) and disease-free survival (DFS) for distant recurrence
5 components of DFS. Whilst we found no statistically significant association between
6 DOCK4 and first event in non-skeletal sites, taken as a group ($p = 0.08$, [Figure 3a](#)) or
7 in any distant site ($p = 0.475$, [Figure 3d](#)) higher DOCK4 was significantly associated
8 with increased risk of developing first event as skeletal recurrence whether only in
9 bone ($p = 0.043$, [Figure 3c](#)) or in bone and other distant sites concurrently ($p = 0.033$,
10 [Figure 3b](#)).

11
12
13
14
15
16
17
18
19
20
21
22 Subsequently, dichotomised TMA expression scores were used where high DOCK4
23 expression (score of 3) was compared with low DOCK4 expression (score of 1 or 2).
24 Using these categories, Kaplan-Meier estimates of the survival function for time to
25 distant recurrence confirmed that high DOCK4 is significantly prognostic for first
26 distant recurrence involving bone only (HR 2.1; 95% CI 1.09, 4.15, $p = 0.024$, [Figure](#)
27 [4a](#)). For first distant event involving both skeletal and other site(s) concurrently,
28 although the same trend was observed, the association did not reach significance (HR
29 1.6; 95%CI 0.88, 3.05, $p = 0.113$, [Figure 4c](#)).

30
31
32
33
34
35
36
37
38
39
40
41 Where the first distant event was in non-skeletal sites (taken as a group), the
42 corresponding dichotomised analyses ([Figure 4e](#)), suggested a **non-significant** reduced
43 risk of non-skeletal events with high DOCK4 (HR = 0.2; 95%CI 0.03, 1.51, $p = 0.12$).
44 In the dichotomised analyses, we found no association between high DOCK4 and first
45 skeletal event, whether or not other distant events had occurred first, (HR = 1.3;
46 95%CI 0.73, 2.33, $p = 0.373$), suggesting that the inclusion of metastasis in other
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 sites, before eventual spread to bone, counters the significant association seen with the
4
5 bone only analyses.
6
7

9 ***Zoledronate arm.***

11 In the zoledronate arm, dichotomised analyses revealed no association between
12
13 DOCK4 and bone only distant events (HR 0.6; 95%CI 0.14, 2.62, p = 0.488, [Figure](#)
14
15 [4b](#)), ie the increased risk of bone-only first event posed by high DOCK4 in the control
16
17 arm was abolished by zoledronate. This suggests that DOCK4 may act as a predictive
18
19 biomarker for prevention of bone metastases by zoledronate. This was confirmed by
20
21 examining Kaplan-Meier plots for DOCK4 high and DOCK4 low comparing control
22
23 and zoledronate patients for the various events (Supporting Information, [Figure S5](#)).
24
25 [Figure S5\(b\)](#) clearly shows the reduction in bone metastasis-only risk as a first event
26
27 (HR 0.1; 95% CI 0.03, 0.5, p=0.003) in patients with high DOCK4 treated with
28
29 zoledronate.
30
31
32
33
34

35 For the first event involving bone and other sites concurrently, there was no
36
37 significant association seen in the zoledronate group ([Figure 4d](#)).
38
39
40

41 ***Cox proportional hazards regressions adjusted for systemic therapy plan, ER status,*** 42 43 ***HER2 status and lymph node involvement***

44
45 *Control arm.* Adjusted analyses (Table 2) confirmed that high DOCK4 expression
46
47 was significantly prognostic for skeletal only events (HR 2.13, 95% CI 1.06, 4.30, p =
48
49 0.034). Similar analyses for first recurrence events involving the skeleton and other
50
51 site(s) concurrently showed no significant association (HR 1.63, 95%CI [0.86, 3.12],
52
53 p = 0.137; Table 2). No significant association was found in these adjusted analyses
54
55
56
57
58
59
60

1
2
3 for non-skeletal distant events when assessed collectively. Analysis of relationships to
4
5 other individual metastatic sites was not possible with this data-set.
6
7

8
9 *Zoledronate arm.* The significant prognostic effect of DOCK4 for skeletal only
10
11 metastases seen in the control arm patients was not observed in the zoledronate arm
12
13 (HR 0.812, 95%CI [0.18, 3.76], $p = 0.790$) (Table 2), suggesting again that the
14
15 increased risk for skeletal metastasis in patients with high DOCK4 levels at baseline,
16
17 may be counteracted by treatment with zoledronate. This was tested formally by
18
19 including an interaction term in the Cox proportional hazards regressions for
20
21 treatment arm and DOCK4 level for time to first event, skeletal only. These analyses
22
23 suggested a predictive effect for treatment with zoledronate (HR 0.12 95%CI 0.03,
24
25 0.56; likelihood ratio test $p=0.063$, [hazard ratio <1 indicates improvement with
26
27 zoledronate]), though this did not reach significance.
28
29
30
31
32

33 ***Overall survival and menopausal status***

34
35 As shown in Figure 4 (i) and (j) and Supplementary Figure S5 (i) and (j), although
36
37 there appeared to be a trend towards high DOCK4 producing worse outcome,
38
39 DOCK4, expression level did not impact significantly on overall survival in either the
40
41 control or the zoledronate arms. When post-menopausal and pre-menopausal patients
42
43 were analysed separately, although similar associations with bone metastasis to the
44
45 full group were observed, there was a loss of statistical significance due to the smaller
46
47 numbers involved.
48
49
50
51

52 **DISCUSSION**

53
54
55
56
57
58
59
60

1
2
3 Breast cancer bone metastasis causes significant morbidity and biomarkers which can
4 predict the development of bone metastases are badly needed. In this translational
5 study, SILAC-MS-based comparison of bone homing and non-homing cell variants
6 and functional *in vitro* work, coupled with rigorous clinical validation enabled us to
7 identify DOCK4 as a potential biomarker for this purpose.
8
9
10
11
12

13
14
15 ~~Mechanistically, there have been few studies identifying differentially expressed~~
16 ~~proteins within breast cancer bone metastasis; as such the current study provides a~~
17 ~~novel perspective on homing of breast cancer to bone sites of distant relapse. SILAC~~
18 ~~is an established proteomics approach, which gives quantitative information on the~~
19 ~~relative expression of proteins when comparing two or more cell types. This~~
20 ~~approach was successful in terms of the overall numbers of proteins identified with~~
21 ~~high confidence in the total data set (n = 2999).~~
22
23
24
25
26
27
28
29
30
31
32

33 Our discovery science used MDA-MB-231 and its bone homing variant BM1. The
34 MDA-MB-231 cell line is regarded as ‘gold standard’ for this type of research as it
35 was derived from a pleural effusion of a breast cancer patient with widespread
36 metastasis, many years after resection of the primary tumour [23]. Furthermore, the
37 validity of this approach has already been proven in our previous work in which the
38 importance of the proteins CAPG and GIPC1 as biomarkers was discovered and
39 subsequently clinically validated (10). Here, we focused on DOCK4, a key guanine
40 nucleotide exchange factor (GEF) regulating the activation of the small GTPase Rac1
41 (19,20,24,25). DOCK4-mediated activation of Rac1 has been demonstrated to
42 promote actin reorganization and the formation of lamellipodia at the leading edge of
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 breast cancer cells [19] as well as the formation of lateral filopodia, and blood vessel
4
5 lumen morphogenesis within tumour angiogenesis [14].
6
7

8
9 DOCK4 was expressed in parental MDA-MB-231 cells (PCC), but was more
10
11 abundant in the bone-homing variant (BM1). DOCK4 knockdown inhibited the
12
13 migration of both PCC and BM cell-lines with inhibition being greater at 6 hours than
14
15 12 hours post assay-initiation, suggesting that DOCK4-mediated cell invasion may be
16
17 important in the earlier stages of breast cancer cell-migration.
18

19
20 However, enhanced cell-migration and invasiveness resulting from high DOCK4
21
22 expression is only one aspect of the bone-homing cells which makes them bone
23
24 homing. In particular, c-MAF targets and other proteins elevated in the bone homing
25
26 cells might drive the bone-homing phenotype as well as DOCK4. Our observation
27
28 that, in patients, high DOCK4 is specifically associated with first distant metastasis in
29
30 bone, may be linked to these factors, including those in the bone micro-environment.
31
32 Notably in this regard, the specific association of high DOCK4 with bone metastasis
33
34 at any time, is lost once metastasis has occurred elsewhere. This presumably indicates
35
36 the substantially altered metastatic environment influencing bone metastasis once
37
38 non-bone metastases have occurred. Association with bone as first metastatic site is
39
40 lost in zoledronate-treated patients suggesting that zoledronate treatment reduces the
41
42 risk of bone metastasis to a level similar to that in non-high DOCK4 patients.
43
44
45
46
47

48 This work has shown that DOCK4 has a similar prognostic and predictive profile to
49
50 CAPG and GIPC1 for the prognosis of skeletal-only relapse within control arm
51
52 patients.
53
54
55
56
57
58
59
60

1
2
3 DOCK4 was also similar to the previously discovered markers in not being predictive
4 of non-skeletal recurrence events within control arm patients. Treatment with
5 zoledronate abolished the association of high levels of all three of these proteins in
6 development of skeletal-only metastasis.
7
8
9

10
11
12
13 DOCK4 expression is induced by the cytokine transforming growth factor- β (TGF β)
14 acting via the Smad pathway, and this is a key step in TGF β 's pro-metastatic effect
15 [21]. GIPC1 is also a key scaffolding protein which functions to transmit signals from
16 the TGF β -receptor and its co-receptor endoglin to downstream Smad-phosphorylation
17 [26]. TGF β is a regulator of numerous steps within metastasis including intravasation,
18 extravasation and cancer-cell survival at distant organ sites [27]. The transcription
19 factor c-MAF has recently been identified as a key regulator of breast cancer bone
20 metastasis [28,29]. c-MAF expression is induced by TGF β and a recent patent
21 application showed that DOCK4-expression correlates with MAF-expression within
22 primary tumours [30]. Examination of our quantitative proteomic data set identified
23 36 proteins also present in the c-MAF gene set (a panel of 109 genes in total). Within
24 the 36 proteins quantified by proteomics, 15 proteins displayed the same change
25 (increase or decrease of expression within bone-homing cells compared to parental
26 cells) as the relevant gene transcripts in response to c-MAF expression. DOCK4 may
27 therefore be a component of a protein panel which responds to elevated c-MAF
28 expression within bone homing breast cancer cells.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50 As well as being a prognostic biomarker, our data also provide evidence that DOCK4
51 is a potential predictive biomarker in terms of the treatment effect of zoledronate for
52 bone as first metastatic site, since the addition of zoledronate appears to reduce the
53
54
55
56
57
58
59
60

1
2
3 risk of patients with high DOCK4 levels to that of patients with lower DOCK4 levels.
4
5 There was also a substantial HR in favour of a treatment effect when an interaction
6
7 term was included in the Cox proportional hazards regression, though this fell short of
8
9 statistical significance and additional testing in a further patient cohort is needed
10
11 before DOCK4 can be confirmed as a predictive biomarker.
12
13

14
15
16 Our current studies are supported further through analysis of DOCK4 gene expression
17
18 within a publicly available database where high DOCK4 transcript levels predicted
19
20 distant bone metastatic spread of breast cancer (Figure S4, Supporting Information)
21
22

23
24
25
26 Interestingly, DOCK4 was not prognostic for DFS in terms of metastasis at non-
27
28 skeletal sites, taken as a group (N=28). Indeed, data for the control arm in Figure 4
29
30 and Table 2 (and in Figure S5) suggest that high DOCK4 may be associated with a
31
32 reduction in the occurrence of non-skeletal metastasis, an effect which is not seen in
33
34 the zoledronate arm. Whilst zoledronate is advantageous in terms of preventing
35
36 skeletal metastasis in DOCK4 high patients, it may also remove a potentially
37
38 advantageous effect for non-skeletal metastases. Whilst our data do not point to any
39
40 particular mechanistic explanation for this effect, it is possible that zoledronate may
41
42 have effects other than on the skeleton and, in this respect, we note the negative
43
44 impact of zoledronate on overall survival in pre-menopausal women recently
45
46 reported, where in non-postmenopausal patients with MAF-positive tumours,
47
48 zoledronate was associated with worse invasive-disease-free survival and overall
49
50 survival [28]. Clearly, these factors need to be borne in mind in the consideration of
51
52 DOCK4 as a predictive biomarker for zoledronate response in bone metastasis
53
54
55
56
57
58
59
60

1
2
3 ~~prevention. DOCK4 may play a functional role in the process of breast cancer~~
4 ~~metastasis to the bone, as it was also previously shown to play a role in the process of~~
5 ~~extravasation of lung adenocarcinoma cells to the liver [21]. However, DOCK4 may~~
6 ~~not be necessary for metastasis to all secondary sites, so that other metastatic sites~~
7 ~~such as the lung, may not require elevated DOCK4 expression for metastasis. This~~
8 ~~may be related to the role of DOCK4 for cancer cell extravasation and vascular~~
9 ~~differences at different metastatic sites, or differential requirement for metastatic~~
10 ~~growth in different organ microenvironments. Because our analyses included non-~~
11 ~~skeletal sites as a single group, they do not exclude the possibility of DOCK4~~
12 ~~association with a less common site for metastatic spread than the skeleton.~~

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27 There are some limitations to the current study. Although the number of patients
28 available for analysis from the AZURE study considerably exceeds that required for
29 statistical powering, there is currently no equivalent independent sample set available
30 for further validation. Also, breast cancer metastasis to bone involves numerous
31 autocrine and paracrine signalling events [31], and these and the directionality and
32 recruitment of key signalling pathways that migration involves, is clearly not
33 replicated within the invasion assay used in the current study. More complex tools to
34 assess this are not currently available.

35
36
37
38
39
40
41
42
43
44
45
46 Further studies of the mechanistic role of DOCK4 in breast cancer bone metastasis
47 ~~and implications for pharmacological inhibition~~ is justified by our work. ~~In addition,~~
48 ~~the potential for the DOCK4 mediated signalling pathway to function as a target for~~
49 ~~pharmacological inhibition within breast cancer bone metastasis is an exciting~~
50 ~~possibility. This study suggests that DOCK4 may have clinical utility as a potential~~
51
52
53
54
55
56
57
58
59
60

1
2
3 prognostic biomarker for assessment of risk of cancer progression to the skeleton,
4 possibly in combination with other biomarkers, since the specificity and sensitivity of
5 a biomarker test can be improved by the incorporation of multiple markers into a
6 diagnostic/prognostic panel of markers [27]. DOCK4 may therefore form part of a
7 future panel based test (possibly including CAPG, GIPC1 and the MAF gene [28])
8 used for informing patient treatment options. Future validation studies of DOCK4 as
9 a biomarker of bone metastasis in other datasets with denosumab treatment or other
10 use of bisphosphonates would be useful, if samples from such datasets become
11 available in the future.
12
13
14
15
16
17
18
19
20
21
22

23 **Acknowledgements:**

24 We thank the patients in the AZURE study. We thank Cancer Research UK for the
25 award of a Clinician Scientist Fellowship (to JEB, C18605/A 10048) and for support
26 to Sheffield ECMC, Yorkshire Cancer Research for a project award (S315 to JEB,
27 REC, VS), Breast Cancer Now for pilot and project grant awards ([2012MaySP047](#)
28 [and 2016MayPR74](#) to GM and VS) and the UK Medical Research Council for a
29 Career Development Fellowship [G0802416] to DAC. We also thank Adam Davison
30 and Liz Straszynsky for assistance with cell sorting, and Paul Morris for assistance
31 with immunohistochemistry.
32
33
34
35
36
37
38
39
40
41
42

43 **Contributions:**

44 JW, SW, KM, RG, HT, [MO](#), [MD](#), SR, AH, PS, VS GM, RC and JB conceived and/or
45 carried out the experiments. JW, DC, SW, HM, AH, [MD](#), PS, VS, GM, RC and JB
46 analysed the data, all authors were involved in writing the manuscript and had final
47 approval of the submitted version
48
49
50
51
52

53 **Supporting Information**

1
2
3 A file has been uploaded for on-line access only containing Supporting Information.
4
5 This file contains further details on the methods used in this study, and further
6
7 relevant data in table and figure format. This table (Table S1) and the figures (Figures
8
9 S1-S5) have been cited in the main article.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

REFERENCES

1. UK CR. [Available from: <http://www.cancerresearchuk.org/cancerinfo/cancerstats/types/breast/mortality/uk-breast-cancer-mortality-statistics>. 2014.
2. Singh T, Kaur V, Kumar M, Kaur P, Murthy RS, Rawal RK. The critical role of bisphosphonates to target bone cancer metastasis: an overview. *J Drug Target*. 2015;23(1):1-15.
3. Casas A, Llombart A, Martin M. Denosumab for the treatment of bone metastases in advanced breast cancer. *Breast*. 2013;22(5):585-92.
4. Coleman RE, Marshall H, Cameron D, Dodwell D, Burkinshaw R, Keane M, et al. Breast-cancer adjuvant therapy with zoledronic acid. *N Engl J Med*. 2011;365(15):1396-405.
5. Early Breast Cancer Trialists' Collaborative G. Adjuvant bisphosphonate treatment in early breast cancer: meta-analyses of individual patient data from randomised trials. *Lancet*. 2015;386(10001):1353-61.
6. Rathbone EJ, Brown JE, Marshall HC, Collinson M, Liversedge V, Murden GA, et al. Osteonecrosis of the jaw and oral health-related quality of life after adjuvant zoledronic acid: an adjuvant zoledronic acid to reduce recurrence trial subprotocol (BIG01/04). *J Clin Oncol*. 2013;31(21):2685-91.
7. Santana-Codina N, Carretero R, Sanz-Pamplona R, Cabrera T, Guney E, Oliva B, et al. A transcriptome-proteome integrated network identifies endoplasmic reticulum thiol oxidoreductase (ERp57) as a hub that mediates bone metastasis. *MolCell Proteomics*. 2013;12(8):2111-25.
8. Lam SW, Jimenez CR, Boven E. Breast cancer classification by proteomic technologies: current state of knowledge. *Cancer Treat Rev*. 2014;40(1):129-38.
9. Dumont B, Castronovo V, Peulen O, Bletard N, Clezardin P, Delvenne P, et al. Differential proteomic analysis of a human breast tumor and its matched bone metastasis identifies cell membrane and extracellular proteins associated with bone metastasis. *J Proteome Res*. 2012;11(4):2247-60.
10. Westbrook JA, Cairns DA, Peng J, Speirs V, Hanby AM, Holen I, et al. CAPG and GIPC1: Breast Cancer Biomarkers for Bone Metastasis Development and Treatment. *J Natl Cancer Inst*. 2016;108(4).
11. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell*. 2003;3(6):537-49.

12. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002;1(5):376-86.
13. Ong SE, Mann M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc*. 2006;1(6):2650-60.
14. Abraham S, Scarcia M, Bagshaw RD, McMahon K, Grant G, Harvey T, et al. A Rac/Cdc42 exchange factor complex promotes formation of lateral filopodia and blood vessel lumen morphogenesis. *Nat Commun*. 2015;6:7286.
15. Simpkins SA, Hanby AM, Holliday DL, Speirs V. Clinical and functional significance of loss of caveolin-1 expression in breast cancer-associated fibroblasts. *J Pathol*. 2012;227(4):490-8.
16. Hornigold N, Craven RA, Keen JN, Johnson T, Banks RE, Mooney AF. Upregulation of Hic-5 in glomerulosclerosis and its regulation of mesangial cell apoptosis. *Kidney Int*. 2010;77(4):329-38.
17. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Urol*. 2005;2(8):416-22.
18. Kawada K, Upadhyay G, Ferandon S, Janarthanan S, Hall M, Vilardaga JP, et al. Cell migration is regulated by platelet-derived growth factor receptor endocytosis. *Mol Cell Biol*. 2009;29(16):4508-18.
19. Hiramoto-Yamaki N, Takeuchi S, Ueda S, Harada K, Fujimoto S, Negishi M, et al. Ephexin4 and EphA2 mediate cell migration through a RhoG-dependent mechanism. *J Cell Biol*. 2010;190(3):461-77.
20. Kobayashi M, Harada K, Negishi M, Katoh H. Dock4 forms a complex with SH3YL1 and regulates cancer cell migration. *Cell Signal*. 2014;26(5):1082-8.
21. Yu JR, Tai Y, Jin Y, Hammell MC, Wilkinson JE, Roe JS, et al. TGF-beta/Smad signaling through DOCK4 facilitates lung adenocarcinoma metastasis. *Genes Dev*. 2015;29(3):250-61.
22. Jalali A, Ebrahimi H, Ohadi M, Karimloo M, Shemirani AI, Mohajer B, et al. New Variations in the Promoter Regions of Human DOCK4 and RAP1A Genes, and Coding Regions of RAP1A in Sporadic Breast Tumors. *Avicenna J Med Biotechnol*. 2009;1(2):117-23.
23. Cailleau R, Olive M, Cruciger QV. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro*. 1978;14(11):911-5.
24. Brugnera E, Haney L, Grimsley C, Lu M, Walk SF, Tosello-Tramont AC, et al. Unconventional Rac-GEF activity is mediated through the Dock180-ELMO complex. *Nat Cell Biol*. 2002;4(8):574-82.

- 1
2
3
4 25. Cote JF, Vuori K. Identification of an evolutionarily conserved superfamily of
5 DOCK180-related proteins with guanine nucleotide exchange activity. *J Cell Sci.*
6 2002;115(Pt 24):4901-13.
7
- 8 26. Lee NY, Ray B, How T, Blobel GC. Endoglin promotes transforming growth
9 factor beta-mediated Smad 1/5/8 signaling and inhibits endothelial cell migration
10 through its association with GIPC. *J Biol Chem.* 2008;283(47):32527-33.
11
- 12 27. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and
13 evolving paradigms. *Cell.* 2011;147(2):275-92.
14
- 15 28. Coleman R, Hall A, Albanell J, Hanby A, Bell R, Cameron D, et al. Effect of
16 MAF amplification on treatment outcomes with adjuvant zoledronic acid in early
17 breast cancer: a secondary analysis of the international, open-label, randomised,
18 controlled, phase 3 AZURE (BIG 01/04) trial. *Lancet Oncol.* 2017;18(11):1543-52.
19
- 20 29. Pavlovic M, Arnal-Estape A, Rojo F, Bellmunt A, Tarragona M, Guiu M, et
21 al. Enhanced MAF Oncogene Expression and Breast Cancer Bone Metastasis. *J Natl*
22 *Cancer Inst.* 2015;107(12):djv256.
23
- 24 30. Gomis RR, inventor Method for the diagnosis, prognosis and treatment of
25 metastatic cancer 2018.
26
- 27 31. Kang Y. Dissecting Tumor-Stromal Interactions in Breast Cancer Bone
28 Metastasis. *Endocrinol Metab (Seoul).* 2016;31(2):206-12.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Legends to Figures

Figure 1. Schematic showing the key steps in the SILAC proteomic approach used.

MDA-MB-231 (MDA231) and bone homed variant (BM) cell lines were incubated with media containing 'heavy' and 'light' isotopically labelled amino acids arginine (R) and lysine (K). The heavy medium contained [$^{13}\text{C}_6, ^{15}\text{N}_4$]-L-Arginine and [$^{13}\text{C}_6, ^{15}\text{N}_2$]-L-Lysine, i.e. stable isotopes of [$^{12}\text{C}_6, ^{14}\text{N}_4$]-L-Arginine and [$^{12}\text{C}_6, ^{14}\text{N}_2$]-L-Lysine, respectively. Two experiments were performed involving 'forward' and 'reverse' labelling (reciprocal labelling) in which the cell lines were incubated in both media types and then combined in heavy:light pairs in a 1:1 ratio (based on extracted protein assay) prior to separation by molecular weight using 1-DE. The entire lane of gel-separated proteins was cut into equal slices (n=10). The proteins in each gel slice were reduced to peptides by enzymatic digestion (using trypsin) and were further separated using high performance liquid chromatography (HPLC) and analysed using high resolution mass spectrometry (MS).

Figure 1: A and C. Higher DOCK4 expression in BM cell type; B and D

Knock-down of DOCK4 expression by shRNA. DOCK4 was confirmed to be of higher expression in BM1 cell type by Western blotting (2.0-fold induction, $p = 0.027$), and 81% and 88% knock-down of DOCK4 protein expression was achieved in PCC and BM1 cells, respectively. Representative whole-gel lane images are shown. wt = wild type (no vector involvement); cv = control vector; sh = shRNA.

E and F: Invasion assay. The ability of PCC and BM1 cells to move through MatrigelTM matrix was assessed using a scratch-wound assay. Significant differences

1
2
3 between the control vector and DOCK4 knock-down cells were seen at the 6 and 12hr
4
5 time points.
6
7
8

9
10 **Figure 2. Examples of staining profiles for DOCK 4 in patients from the**
11 **AZURE study.**
12

13 Composite of differential protein expression intensities for DOCK4 protein
14
15 expression as revealed by IHC and visualised at magnification x 20. The scoring is
16
17 based on the intensity of staining in the cytoplasmic compartment in the tumour cells
18
19 only. Scale bar = 200µm.
20
21
22
23

24 **Figure 3. Association of DOCK4 with DFS events.** Kaplan-Meier estimates of
25
26 the relationship between expression of DOCK4 and (a) non-skeletal DFS events (b)
27
28 skeletal DFS events (where other distant events may have been recorded at the same
29
30 time); (c) solely skeletal DFS events (where no other distant event was recorded at
31
32 the same time) and (d) any DFS events (where first event was recorded at any distant
33
34 site) in patients in the control arm of the AZURE trial (n = 434). P-value is from the
35
36 logrank test for testing equality of survival functions.
37
38
39
40
41

42 **Figure 4. Univariate associations of distant recurrence outcomes with biomarker**
43 **expression in control and zoledronate arms.** (Estimates are from Cox proportional
44
45 hazards regressions) Kaplan-Meier estimates of the survival function for time to
46
47 distant recurrence (DR) and overall survival for control and zoledronate arms.
48
49 Numbers 1 to 3 refer to the DOCK4 staining intensity scores. These were
50
51 dichotomised, ie DOCK4 low (1 and 2); DOCK4 high (3). Comparisons shown to be
52
53 significant are also significant in analyses adjusting for the effect of systemic therapy
54
55
56
57
58
59
60

1
2
3 plan, ER status and lymph node involvement. (a) and (b): Skeletal only; (c) and (d):
4 skeletal and other; (e) and (f): Non-skeletal; (g) and (h): First skeletal irrespective of
5 whether other distant events have occurred previously (ie bone metastasis-free
6 survival). (i) and (j): Overall Survival (OS). P-values refer to the logrank test. For
7 definitions of non-skeletal, skeletal and other and skeletal only see legend to Table 1.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Table 1: Characteristics of the patients whose tissue was assessed on TMAs in this study (as at baseline on the AZURE study) and first disease-free survival (DFS) events. Non-skeletal (first distant recurrence event does not include any skeletal component); Skeletal and other (first distant recurrence event reported includes both skeletal and other sites of metastasis, as well as skeletal only); Skeletal only (first distant recurrence event only skeletal - this group is a subset of those classified as skeletal and other).

Characteristic	DOCK4 dataset		Full AZURE trial population	
	zoledronate (n=359)	Control (n=330)	zoledronate (N=1681)	Control (n=1678)
Age (years) Median (range)	50 (26, 75)	51 (32, 79)	51 (20-89)	51 (21, 89)
Axillary lymph nodes - no. (%)				
0	4 (1.2)	4 (1.2)	29 (1.7)	32 (1.9)
1-3	223 (67.6)	223 (67.6)	1041 (61.9)	1032 (61.5)
≥4	103 (31.2)	103 (31.2)	604 (35.9)	608 (36.2)
Tumour stage - no. (%)				
T1	115 (32.0)	116 (35.2)	542 (32.2)	523 (31.2)
T2	196 (54.6)	163 (49.4)	851 (50.6)	867 (51.7)
T3	37 (10.3)	43 (13)	227 (13.5)	228 (13.6)
T4	11 (3.1)	8 (2.4)	58 (3.5)	59 (3.5)
Histological grade - no. (%)				
1	26 (7.2)	23 (7.0)	145 (8.6)	140 (8.3)
2	144 (40.1)	130 (39.4)	731 (43.5)	708 (42.2)
3	187 (52.1)	174 (52.7)	765 (45.5)	787 (46.9)
Not specified		3 (0.9)		
Missing	2 (0.6)			
ER status - no. (%)				
ER positive	255 (77.3)	255 (77.3)	1319 (78.5)	1316 (78.4)
ER negative	80 (22.3)	72 (21.8)	349 (20.8)	355 (21.2)
ER unknown	1 (0.3)	3 (0.9)	13 (0.8)	7 (0.4)
PR status - no. (%)				
PR positive	126 (35.1)	102 (30.9)	725 (43.1)	698 (41.6)
PR negative	63 (17.5)	73 (22.1)	382 (22.7)	424 (25.3)
PR unknown	169 (47.1)	153 (46.4)	571 (34.0)	548 (32.7)
Missing	1 (0.3)	2 (0.6)		
HER2 status - no. (%)				
HER2 positive	38 (10.6)	48 (12.7)	192 (11.4)	223 (13.3)
HER2 negative	106 (29.5)	84 (26.0)	648 (38.5)	603 (35.9)
HER2 unknown/not measured	209 (58.3)	197 (61.2)	831 (49.5)	843 (50.1)
Missing	6 (1.7)	1 (0.3)		
Menopausal status - no. (%)				
Pre-menopausal	167 (46.5)	151 (45.8)	751 (44.7)	752 (44.8)
≤ 5 years since menopause	53 (14.8)	54 (16.4)	247 (14.7)	244 (14.5)
> 5 years since menopause	112 (31.2)	97 (29.4)	519 (30.9)	522 (31.1)
Menopausal status unknown	27 (7.5)	28 (8.5)	164 (9.8)	160 (9.5)
Planned systemic therapy - no. (%)				
Endocrine therapy alone	24 (6.7)	17 (5.2)	76 (4.5)	74 (4.5)
Chemotherapy alone	79 (22.0)	72 (21.8)	362 (21.5)	360 (21.5)
Endocrine therapy plus chemotherapy	256 (71.3)	241 (73.0)	1243 (73.9)	1243 (74.1)
Type of chemotherapy				
Anthracyclins - no. (%)	328 (91.4)	307 (93.0)	1567 (97.6)	1564 (97.6)
Taxanes - no. (%)	48 (13.4)	41 (12.4)	390 (24.3)	385 (24.0)
Timing of chemotherapy				
Neo-adjuvant	14 (3.3)	11 (3.3)	104 (6.5)	104 (6.5)
Post-operative	345 (96.7)	319 (96.7)	1501 (93.5)	1499 (93.5)
Statin use - no. (%)	19 (5.3)	15 (4.5)	97 (5.8)	101 (6.0)
Type of first disease-free survival event - no. (%)				
Loco-regional recurrence	26 (7.2)	17 (5.2)	79 (4.7)	78 (4.7)
Distant recurrence	68 (18.9)	74 (22.4)	332 (19.8)	341 (20.3)
Distant and loco-regional recurrence	5 (1.4)	3 (0.9)	18 (1.1)	21 (1.3)
Death without prior recurrence	11 (3.1)	13 (3.9)	53 (3.2)	44 (2.6)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

First distant recurrence is nonskeletal - no. (%)	46 (12.8)	28 (8.5)	194 (11.5)	165 (9.8)
First distant recurrence includes skeletal and other - no. (%)	27(7.5)	49 (14.8)	156 (9.3)	197 (11.7)
First distant recurrence is skeletal only - no. (%)	16 (4.5)	38 (11.5)	97 (5.8)	140 (8.3)

For Peer Review

Table 2 Cox regression analysis for associations between protein IHC and distant recurrence events by AZURE trial arm.

Reference category in each multivariable model is DOCK4 low (1 or 2) Comparisons shown to be significant are also significant in analyses adjusting for the effect of systemic therapy plan, ER status, HER2 status and lymph node involvement. Time to first bone metastasis is defined as time to the first skeletal distant recurrence, irrespective of whether other distant recurrence has occurred earlier (ie bone metastasis-free survival). For definitions of non-skeletal, skeletal plus other and skeletal only, see legend to Table 1. NB: n = number of events and N = number at risk.

		Standard Treatment				Standard Treatment + zoledronic Acid			
		N (at risk)	n (events)	HR (95% CI)	p-value	N (at risk)	n (events)	HR (95% CI)	p-value
Nonskeletal distant recurrence	Unadj	330	28	0.154 [0.021,1.131]	0.066	359	46	1.162 [0.577,2.342]	0.674
	Adj	329	28	0.201 [0.027,1.506]	0.118	353	46	0.999 [0.478,2.086]	0.997
Distant recurrence including skeletal	Unadj	330	49	1.642 [0.883,3.052]	0.117	359	27	0.96 [0.363,2.535]	0.934
	Adj	329	49	1.634 [0.855,3.121]	0.137	353	26	0.999 [0.336,2.967]	0.998
Skeletal only distant recurrence	Unadj	330	38	2.121 [1.085,4.148]	0.028	359	16	0.595 [0.135,2.619]	0.493
	Adj	329	38	2.133 [1.058,4.304]	0.034	353	16	0.812 [0.176,3.756]	0.79
Bone metastasis at any time	Unadj	330	62	1.302 [0.728,2.328]	0.374	359	49	0.81 [0.38,1.728]	0.586
	Adj	329	62	1.344 [0.734,2.46]	0.338	353	47	0.882 [0.387,2.012]	0.765

Supporting Information.

Further experimental details of proteomics methods

Cell culture and SILAC (Stable Isotope Labelling by Amino acids in Cell culture)

The human breast cancer cell line MDA-MB-231 (“PCC”, parental control cells, obtained originally from ATCC) and a bone-homing variant (“BM1”, bone metastatic cells) (1) (the latter supplied by Prof. Joan Massague, Sloan-Kettering Institute, New York), were used in a ‘classical’ SILAC experiment (2, 3). Quarterly Mycoplasma checks and annual STR profiling confirmed cell provenance. PCC and BM1 cells were cultured in ‘heavy’ SILAC medium and ‘light’ medium for ~10 doublings to ensure >95% isotope label incorporation (Heavy media), and all cells were used within 10 doublings. ‘Heavy’ SILAC media consisted of DMEM containing L-arginine ($^{13}\text{C}_6$, $^{15}\text{N}_4$) and L-lysine ($^{13}\text{C}_6$, $^{15}\text{N}_2$) (R10K8, DMEM-15, Dundee Cell Products, Dundee, UK) supplemented with dialysed bovine serum (D-FBS100, Dundee Cell Products, Dundee, UK). ‘Light’ media was DMEM but without the heavy isotopes (R0K0). Essentially, two experiments were performed involving ‘forward’ and ‘reverse’ labelling (i.e. reciprocal labelling) in which the cell lines were incubated in both media types and then combined in heavy:light pairs in a 1:1 ratio (based on extracted protein amount) prior to separation by molecular weight using 1-DE. For proteomic analysis, cells were extracted for total protein content using Laemmli buffer (S-3401 Sigma-Aldrich).

LC-MS/MS

Equal amounts of protein (40 μg) from light and heavy labelled samples were combined, reduced, alkylated, and separated on a 1-D gel. LC-MS-MS was performed by Dundee Cell products, Dundee, Scotland, UK. Ten slices of the gel-resolved proteins were cut and proteins were digested to peptides using trypsin. Tryptic peptides were separated using a nanoflow LC-System coupled to an LTQ-Orbitrap mass spectrometer (ThermoFisher

1
2
3 Scientific). Survey full scan MS spectra (m/z 300-1700) were acquired in the Orbitrap with a
4 resolution of 60,000 at m/z 400 after accumulation of 1,000,000 ions. The five most intense
5 ions from the preview survey scan delivered by the Orbitrap were sequenced by collision
6 induced dissociation (normalized collision energy = 40%) after accumulation of 5,000 ions
7 concurrently to full scan acquisition in the Orbitrap. Data were acquired using the Xcalibur
8 software.

16 ***Quantitation and bioinformatics analysis***

19 Quantitation was performed using the software Max Quant
20 (<http://www.maxquant.org/downloads.htm>), with peptide ratios calculated for each arginine-
21 and/or lysine-containing peptide as the peak area of labelled arginine/lysine divided by the
22 peak area of non-labelled arginine/lysine for each single-scan mass spectrum. Peptide ratios
23 for all arginine- and lysine-containing peptides sequenced for each protein were averaged.
24 Data output from Max Quant were analysed further using Excel and R (v. 3.2, [http://www.r-
25 project.org/](http://www.r-project.org/)) to select differentially expressed proteins for consideration as candidates for
26 further verification and validation.

37 **Western blotting**

40 The differential expression of DOCK4, and confirmation of DOCK4 knockdown, was
41 assessed in cell lysates using Western blotting and an infra-red (IR) immunodetection system
42 (LI-COR Biosciences) as well as by Enhanced Chemiluminescence (ECL). The primary
43 antibodies used were: DOCK4 (Abcam ab56743, mouse monoclonal, 0.1 μ g/mL; beta-tubulin
44 loading control (Abcam ab6046, rabbit polyclonal, 1/5000 dilution. Secondary antibodies
45 were: #925-68070 IRDye 680RD goat anti-mouse IgG (H+L), LI-COR Biosciences, 1/5000
46 dilution, for detection of DOCK4 in the 700nm channel (red); #925-32211 IRDye 800CW
47 goat anti-rabbit IgG (H+L), LI-COR Biosciences, 1/5000 dilution, for detection of beta-

1
2
3 tubulin in the 800nm channel (green). Normalised densitometric data from six replicate runs
4
5 of the immunoprobed samples were tested for significance using Student's t-test. For ECL the
6
7 secondary antibodies used were: goat-anti-mouse HRP (Abcam, Ab6789) and goat anti-
8
9 rabbit-HRP (Abcam, Ab 6721) both used a 1 in 2500 dilution followed by visualization with
10
11 ECL-Reagent (Promega).
12

13 14 **Further experimental details of immunohistochemistry** 15

16
17 Bone homing MDA-MB-231 cells (BM1) were subjected to lentiviral transfection of either
18
19 control plasmid (control vector) or plasmid expressing anti-DOCK4 miRNA (DOCK4
20
21 miRNA) as described in the main manuscript, Materials and Methods. Cells were cultured in
22
23 DMEM + 10% (v/v) FBS until confluent and then harvested by trypsinization, cells pelleted
24
25 at 1000 x g and the cell pellet washed twice with PBS. Cell pellets were then resuspended in
26
27 10% formalin in Neutral Buffered Saline and fixed for 48 hours at 4°C. In preparation for
28
29 sectioning, 100µl of 10% (w/v) molten agarose was allowed to set in eppendorf tubes
30
31 forming an agarose plug. Following fixation of the cells in 10% (v/v) formalin in neutral
32
33 buffered saline they were pelleted at 1000 x g and then resuspended in 300µl (10% w/v)
34
35 molten agarose. Cells in molten agarose were introduced into the pre-prepared Eppendorf
36
37 tubes with 100µl agarose plugs and allowed to set overnight at 4°C. The next day cell-pellets
38
39 were progressively dehydrated in graded alcohol followed by xylene and then finally molten
40
41 wax treatment. Cell-pellets were embedded in wax blocks and sectioned at 5µm onto
42
43 Superfrost Slides (Manufacturer: Menzel-Gläser), then dried for approximately 48 hours at
44
45 37°C. Sections were dewaxed in xylene, and rehydrated in graded alcohol and then
46
47 endogenous peroxidase-blocked with 10% (v/v) hydrogen peroxide. Antigen retrieval was
48
49 done using citrate buffer (pH 6.0) using a microwave for 20 minutes. Sections were blocked
50
51 in 10% (v/v) goat serum and then primary antibody (Bethyl anti-DOCK4 antibody, Cat No:
52
53 A302-263A) added overnight at a dilution of 1 in 100. Sections were developed using
54
55
56
57
58
59
60

1
2
3 biotinylated secondary antibody (Vector labs) at a dilution of 1 in 200 for 1 hour. Sections
4
5 were rinsed and developed using the Vector Labs ABC kit (Cat No: PK-6100, standard) for
6
7 30 minutes. DAB was applied for 1 minute and then sections rinsed and counter stained with
8
9 Gills Haematoxylin. Sections were then rehydrated with graded alcohols and mounted using
10
11 DPX. Slides were scanned using a Panoramic 250 slide scanner (3DHitech) and images
12
13 analysed for DAB staining intensity within the QuPath software program.
14
15
16
17

18 **Testing antibody specificity: Western blotting:**

19
20
21 Total cell lysates from bone-homing MDA-MB-231 cells (BM1) and parental controls (PCC)
22
23 were prepared in 1x Laemmli Sample Buffer. In addition cell lysates were also prepared from
24
25 bone homing (BM1) cells following lentiviral transfection of either control vector or DOCK4
26
27 miRNA). All cell-lysates were assayed for protein content using the Bio-Rad RC-DC protein
28
29 assay kit. 50µg of total cell protein was loaded onto each lane of a 10-lane Bio-Rad Mini-
30
31 Protean 4-20% TGX SDS-PAGE gel and blotted onto nitrocellulose membrane. Following
32
33 blocking of the membranes overnight at 4°C with 5% (w/v) milk powder in phosphate-
34
35 buffered saline (PBS) they were probed with either: Abcam anti-DOCK4 (ab56743) at a
36
37 dilution of 1 in 3400 or the Bethyl anti-DOCK4 antibody (A302-263A) at a 1 in 2000 dilution
38
39 for 1 hour. Blots were washed and secondary antibody applied (for Abcam ab56743, mouse
40
41 secondary was applied and for Bethyl anti-DOCK4 a rabbit secondary, both secondaries were
42
43 used at a 1 in 2500 dilution) for one hour. Blots were then washed with PBS-tween and
44
45 developed with ECL-reagent (Promega, Cat. No: W1001) prior to exposure to x-ray film.
46
47
48
49

50 **Gene-expression analysis for DOCK4:**

51
52
53 Gene expression data for Wang et al (4) were obtained from the breastCancerVDX (5)
54
55 Bioconductor package, and sample metadata from the GEO submission (GSE2034) was used
56
57
58
59
60

1
2
3 to determine which samples had metastasis to the bone. A recursive partitioning (RP) analysis
4
5 (6) was performed to determine if the samples can be split into two groups based on the
6
7 expression data of DOCK4. Difference in the time to metastasis in the two groups was then
8
9 assessed using the “survival” R package (7).
10

11 **References for Supporting Information**

- 14 1. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A
15 multigenic program mediating breast cancer metastasis to bone. *Cancer Cell*. 2003;3(6):537-
16 49.
17
- 21 2. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable
22 isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to
23 expression proteomics. *Mol Cell Proteomics*. 2002;1(5):376-86.
24
- 28 3. Ong SE, Mann M. A practical recipe for stable isotope labeling by amino acids in cell
29 culture (SILAC). *Nat Protoc*. 2006;1(6):2650-60.
30
- 32 4. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-
33 expression profiles to predict distant metastasis of lymph-node-negative primary breast
34 cancer. *Lancet*. 2005;365(9460):671-9.
35
- 38 5. Schroeder M H-K-B. breastCancerVDX: Gene expression datasets published by
39 Wang et al. [2005] and Minn et al. [2007] (VDX). R package version 1.16.0.2011.
40
- 43 6. Hothorn T, Hornik, K and Zeileis, A. Unbiased Recursive Partitioning: A Conditional
44 Inference Framework. *Journal of Computational and Graphical Statistics*.2006.
45
- 47 7. Therneau TMaG, P.M. Modeling Survival Data: Extensing the Cox Model. Springer.,
48 editor. New York2000.
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary Table S1: Proteomic Results for up-regulated proteins within the SILAC comparison of parental (PCC) and bone-homing (BM1) MDA-MB-231 cells:

Protein Name	Gene	UniProt	BM1 / PCC Fold change	Number of Peptides	Mol. Weight [kDa]	Sequence Length
Interstitial collagenase	MMP1	P03956	22.56	12	54.006	469
Sperm protein associated with the nucleus on the X chromosome B1	SPANXB1	Q9NS25	15.58	6	11.826	103
Plasminogen activator inhibitor 2	SERPINB2	P05120	15.56	12	46.596	415
Protein EVI2B	EVI2B	P34910	7.43	4	50.583	463
Creatine kinase B-type	CKB	P12277	6.92	6	42.644	381
Regulator of G-protein signaling 10	RGS10	O43665	4.70	7	21.21	181
Neutral cholesterol ester hydrolase 1	NCEH1	Q6PIU2	4.69	9	49.881	448
Protein-glutamine gamma-glutamyltransferase 2	TGM2	P21980	4.40	24	77.328	687
Annexin A3	ANXA3	P12429	4.08	14	36.417	323
Cell division cycle protein 20 homolog	CDC20	Q12834	3.67	4	54.722	499
CD44 antigen	CD44	P16070	3.22	10	81.553	742
Formin-like protein 1	FMNL1	O95466	3.19	14	122.41	1104
Alpha-taxilin	TXLNA	P40222	2.76	6	61.89	546
Dedicator of cytokinesis protein 4	DOCK4	Q8N110	2.69	6	225.15	1966
Niemann-Pick C1 protein	NPC1	O15118	2.68	9	142.17	1278
Membrane-associated progesterone receptor component 2	PGRMC2	O15173	2.53	7	26.17	247
Pericentrin	PCNT	O95613	2.52	6	378.08	3336
Four and a half LIM domains protein 2	FHL2	Q14192	2.42	3	44.205	389
Ferritin light chain	FTL	P02792	2.38	6	20.019	175
Vimentin	VIM	P08670	2.37	46	53.651	466
Serine/threonine-protein phosphatase CPPED1	CPPED1	Q9BRF8	2.36	4	35.576	314
CUB domain-containing protein 1	CDCP1	Q9H5V8	2.33	4	92.901	836
Tyrosine-protein phosphatase non-receptor type 1	PTPN1	P18031	2.31	12	49.966	435
Nicotinamide phosphoribosyltransferase	NAMPT	P43490	2.29	26	55.52	491
Xaa-Pro aminopeptidase 1	XPNPEP1	Q9NQW7	2.27	8	74.797	666
Integrin alpha-6	ITGA6	P23229	2.25	22	126.63	1130
Cytosolic non-specific dipeptidase	CNDP2	Q96KP4	2.21	20	52.878	475
Fascin	FSCN1	Q16658	2.21	21	55.135	500
Enoyl-CoA delta isomerase 2, mitochondrial	ECI2	O75521	2.16	2	43.585	394
Coiled-coil domain-containing protein 124	CCDC124	Q96CT7	2.11	6	25.835	223
Glutathione S-transferase Mu 1	GSTM1	P09488	2.04	3	25.712	218
Cap-specific mRNA	CMTR1	Q8N1G2	2.01	5	95.32	835
Nucleobindin-1	NUCB1	Q02818	2.00	5	53.879	461
Adenosine kinase	ADK	P55263	1.97	13	40.545	362
Unconventional myosin-IXb	MYO9B	Q13459	1.96	13	243.56	2158

1	Stathmin	STMN1	P16949	1.96	7	17.302	149
2	Non-syndromic hearing impairment protein 5	DFNA5	O60443	1.96	6	54.554	496
3	C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	P11586	1.92	49	101.56	935
4	Tryptophan--tRNA ligase, cytoplasmic	WARS	P23381	1.92	15	53.165	471
5	Protein CutA	CUTA	O60888	1.90	2	20.925	198
6	Epidermal growth factor receptor substrate 15-like 1	EPS15L1	Q9UBC2	1.89	9	99.606	910
7	Cdc42-interacting protein 4	TRIP10	Q15642	1.88	8	68.351	601
8	Mitochondrial import inner membrane translocase subunit Tim9	TIMM9	Q9Y5J7	1.88	3	10.378	89
9	Neutral amino acid transporter B	SLC1A5	Q15758	1.81	12	56.598	541
10	Chloride intracellular channel protein 4	CLIC4	Q9Y696	1.80	17	28.772	253
11	Copine-1	CPNE1	B0QZ18	1.79	12	59.717	542
12	E3 ubiquitin-protein ligase Itchy homolog	ITCH	Q96J02	1.76	4	102.8	903
13	Nucleosome assembly protein 1-like 1	NAP1L1	P55209	1.75	14	45.374	391

Legends to Figures in Supporting Information

Figure S1. Schematic showing the key steps in the SILAC proteomic approach used.

MDA-MB-231 (MDA231) and bone-homed variant (BM) cell lines were incubated with media containing 'heavy' and 'light' isotopically labelled amino acids arginine (R) and lysine (K). The heavy medium contained [$^{13}\text{C}_6$, $^{15}\text{N}_4$]-L-Arginine and [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-L-Lysine, i.e. stable isotopes of [$^{12}\text{C}_6$, $^{14}\text{N}_4$]-L-Arginine and [$^{12}\text{C}_6$, $^{14}\text{N}_2$]-L-Lysine, respectively. Two experiments were performed involving 'forward' and 'reverse' labelling (reciprocal labelling) in which the cell lines were incubated in both media types and then combined in heavy:light pairs in a 1:1 ratio (based on extracted protein assay) prior to separation by molecular weight using 1-DE. The entire lane of gel-separated proteins was cut into equal slices (n=10). The proteins in each gel slice were reduced to peptides by enzymatic digestion (using trypsin) and were further separated using high-performance liquid chromatography (HPLC) and analysed using high-resolution mass spectrometry (MS)

Figure S2: Testing antibody specificity by immunohistochemistry within FFPE-cells:

- (A) Western blotting confirmation of DOCK4 knockdown in DOCK4 miRNA transfected MDA-MB-231 cells,
- (B) Representative microscopy fields from FFPE-embedded bone-homing MDA-MB-231 cells (BM1) stably transfected with either control lentivirus (Control vector) or DOCK4-miRNA-expressing lentivirus (DOCK4-miRNA) and subjected to immunohistochemical staining using the Bethyl anti-DOCK4-antibody with Haematoxylin counterstaining. Two representative fields are shown for each cell-type.

1
2
3 (C) Quantification of mean cellular DAB staining level across five replicates of each
4
5 cell-type (mean \pm SEM, $n = 5$).
6
7

8 **Figure S3: Testing antibody specificity by western blotting of cell-lysates:**
9

10 (A) Full length ECL images of Western blots of 50 μ g total cell lysate from PCC-cells
11
12 (lane 1) and BM1 cells (lane 2). Western blotting was performed using the Abcam
13
14 anti-DOCK4 antibody (ab56743) as described. Tubulin loading control is also
15
16 shown. Quantification of the normalised DOCK4 band intensity is depicted in the
17
18 attached histogram ($n = 3$ replicate gels, mean band intensity \pm SEM).
19
20

21 (B) Full length ECL images of Western blots of 50 μ g total cell lysate from PCC (lane
22
23 1) and BM1-cells (lane 2) probed with the Bethyl anti-DOCK4 antibody. Tubulin
24
25 loading control included. Histogram depicts the quantification of the normalised
26
27 DOCK4 band intensity (mean \pm SEM, $n = 3$, replicate gels).
28
29
30
31
32
33
34
35

36 **Figure S4: Gene expression analysis of DOCK4 expression and time to bone metastasis:**
37

38
39 Time to bone metastasis analysis of DOCK4 expression level within breast cancer patients
40
41 from Wang et al. with high (> 8.53) and low (< 8.53) levels of DOCK4-gene expression.
42
43
44
45

46
47 **Figure S5 Univariate associations of distant recurrence outcomes with biomarker**
48
49 **expression for DOCK4 low and DOCK4 high.** (Estimates are from Cox proportional
50
51 hazards regressions).
52

53
54 Kaplan-Meier estimates of the survival function for time to distant recurrence (DR) and
55
56 overall survival for control and zoledronate arms for dichotomised DOCK4 low (1 and 2) and
57
58
59
60

1
2
3 high (3). Numbers 1 to 3 refer to the DOCK4 staining intensity scores. Comparisons shown
4
5 to be significant are also significant in analyses adjusting for the effect of systemic therapy
6
7 plan, ER status and lymph node involvement. (a) and (b): Skeletal only; (c) and (d): Skeletal
8
9 and other; (e) and (f): Non-skeletal; (g) and (h): First skeletal irrespective of whether other
10
11 distant events have occurred previously (ie bone metastasis-free survival). (i) and (j): Overall
12
13 Survival (OS). P-values refer to the logrank test. For definitions of non-skeletal, skeletal and
14
15 other and skeletal only see legend to Table 1.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Figure S1.

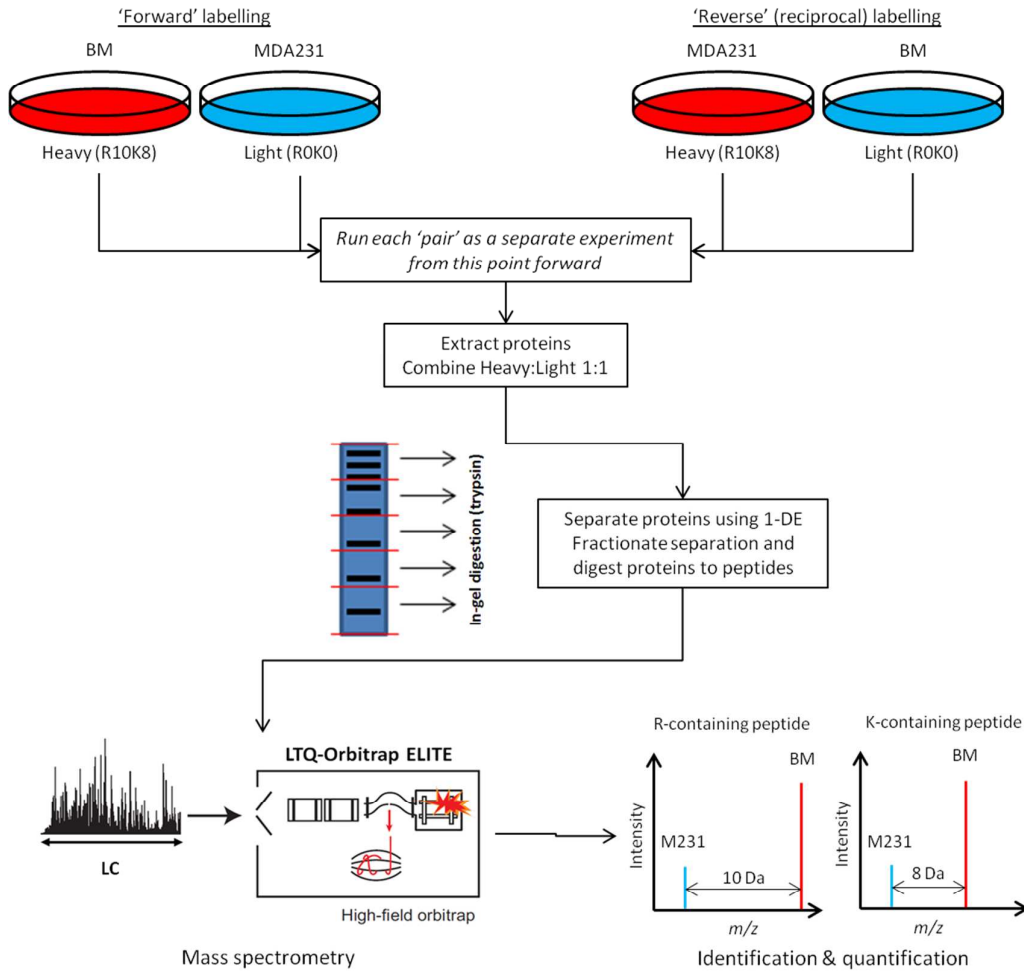
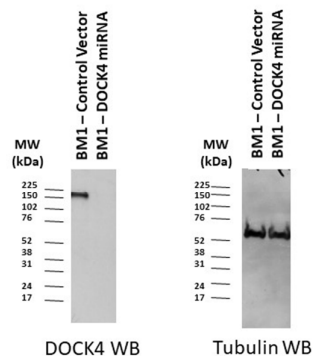


Figure S2:

(A)

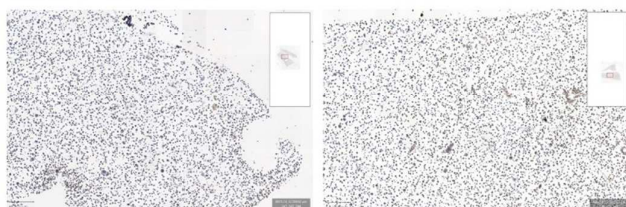


(B)

BM1 – Cells (Control Vector):



BM1-Cells – DOCK4 miRNA



(C)

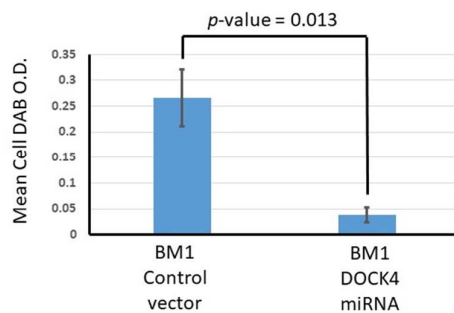
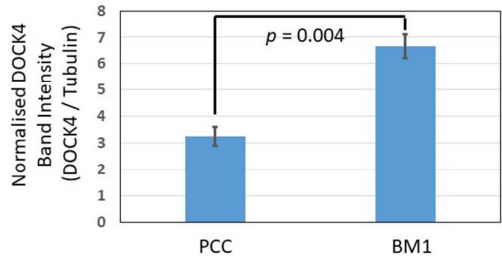
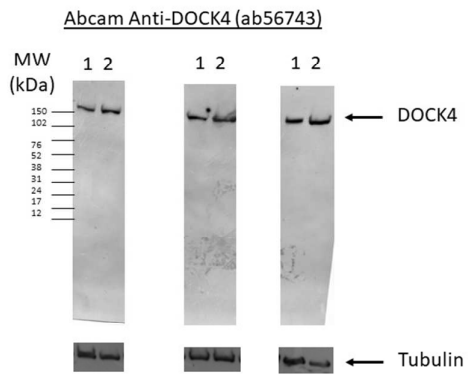


Figure S3:

(A)



(B)

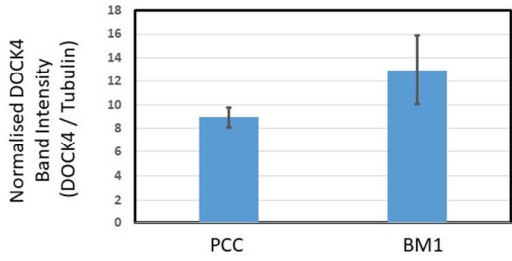
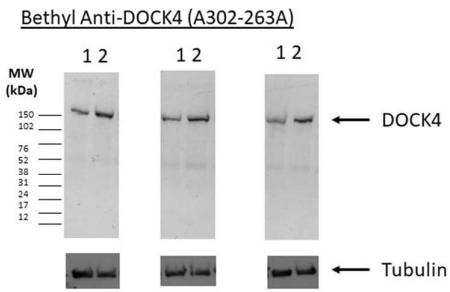


Figure S4

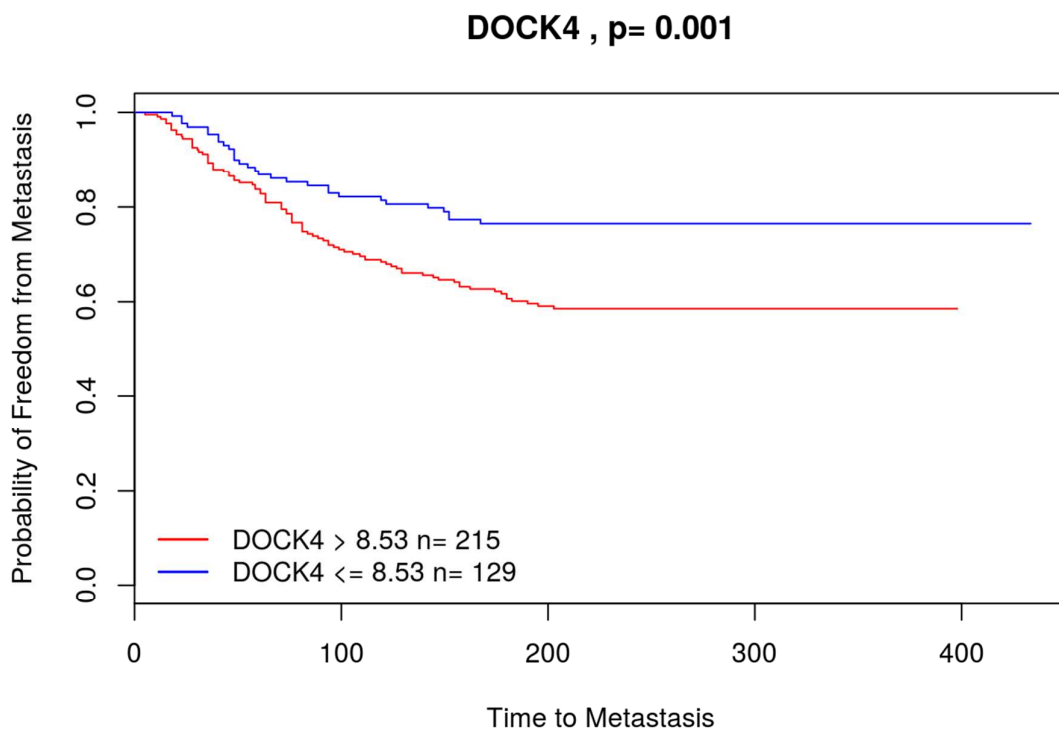
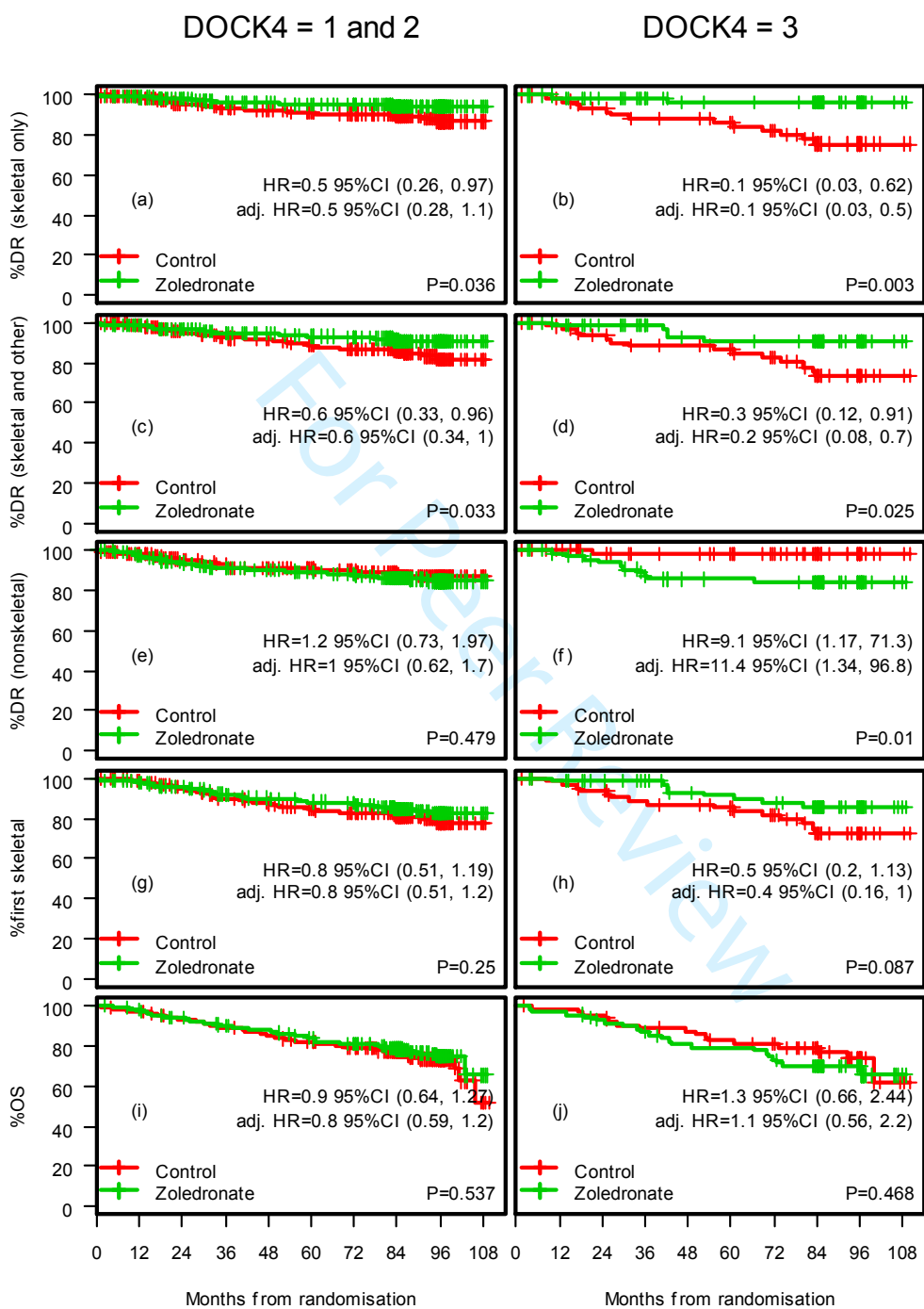


Figure S5



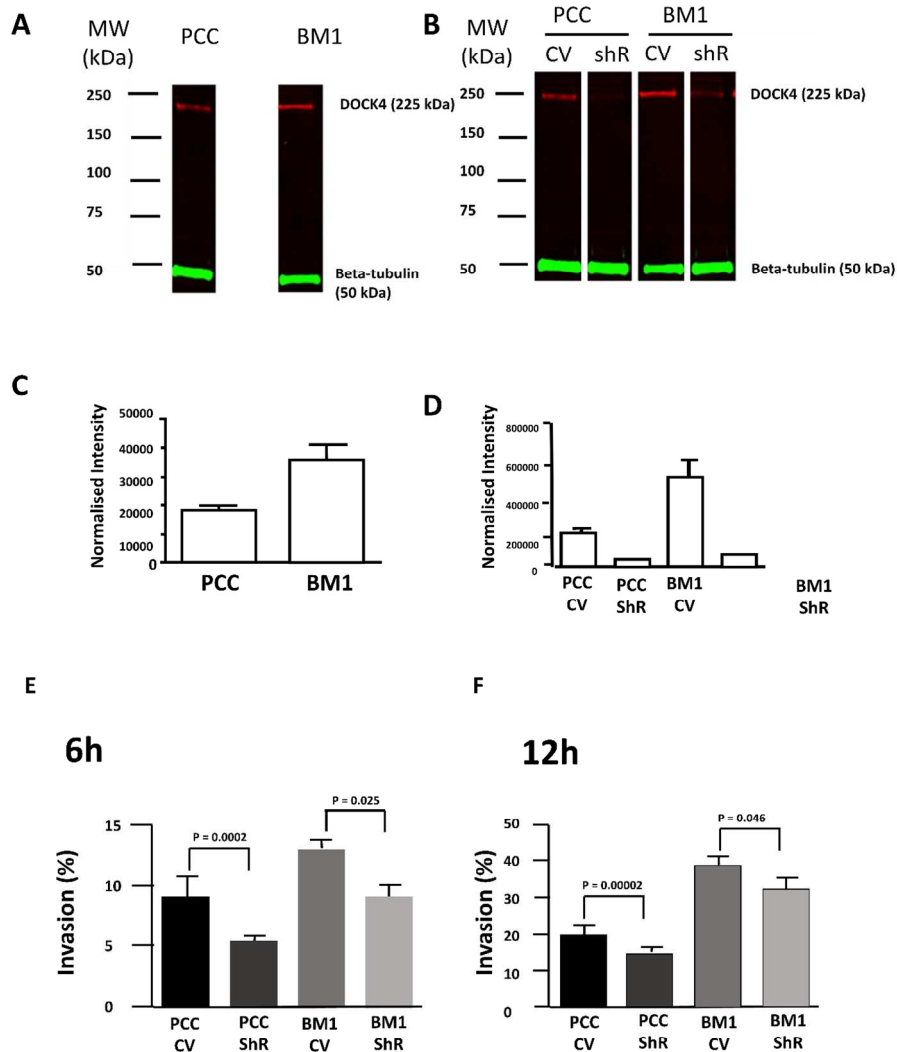


Figure 1: A and C. Higher DOCK4 expression in BM cell type; B and D Knock-down of DOCK4 expression by shRNA. DOCK4 was confirmed to be of higher expression in BM1 cell type by Western blotting (2.0-fold induction, $p = 0.027$), and 81% and 88% knock-down of DOCK4 protein expression was achieved in PCC and BM1 cells, respectively. Representative whole-gel lane images are shown. wt = wild type (no vector involvement); cv = control vector; sh = shRNA. !! + E and F: Invasion assay. The ability of PCC and BM1 cells to move through Matrigel™ matrix was assessed using a scratch-wound assay. Significant differences between the control vector and DOCK4 knock-down cells were seen at the 6 and 12hr time points. !! +

209x259mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

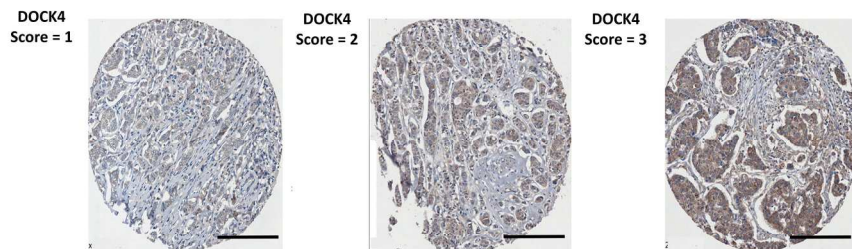


Figure 2. Examples of staining profiles for DOCK 4 in patients from the AZURE study. Composite of differential protein expression intensities for DOCK4 protein expression as revealed by IHC and visualised at magnification x 20. The scoring is based on the intensity of staining in the cytoplasmic compartment in the tumour cells only. Scale bar = 200µm.

190x107mm (300 x 300 DPI)

review

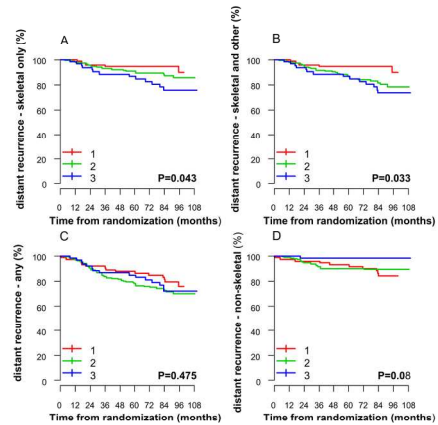


Figure 3. Association of DOCK4 with DFS events. Kaplan-Meier estimates of the relationship between expression of DOCK4 and (a) non-skeletal DFS events (b) skeletal DFS events (where other distant events may have been recorded at the same time); (c) solely skeletal DFS events (where no other distant event was recorded at the same time) and (d) any DFS events (where first event was recorded at any distant site) in patients in the control arm of the AZURE trial ($n = 434$). P-value is from the logrank test for testing equality of survival functions.

190x107mm (300 x 300 DPI)

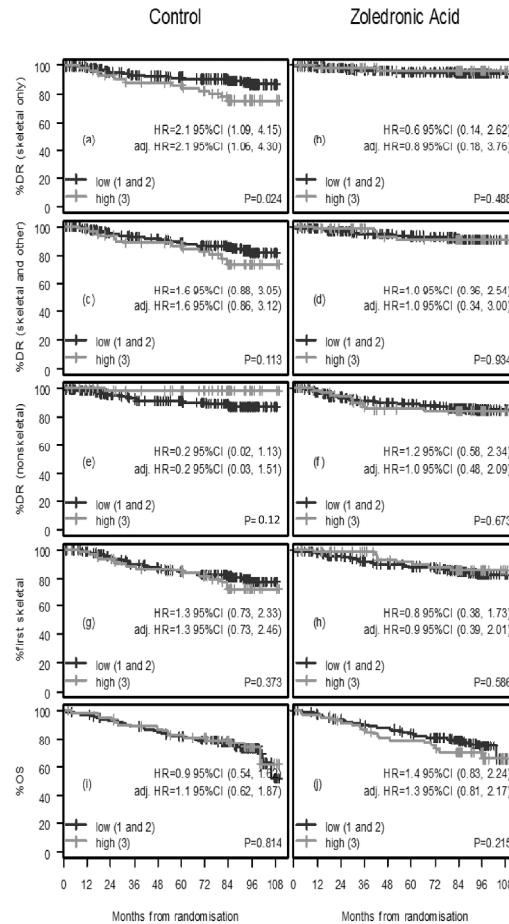


Figure 4. Univariate associations of distant recurrence outcomes with biomarker expression in control and zoledronate arms. (Estimates are from Cox proportional hazards regressions) Kaplan-Meier estimates of the survival function for time to distant recurrence (DR) and overall survival for control and zoledronate arms. Numbers 1 to 3 refer to the DOCK4 staining intensity scores. These were dichotomised, ie DOCK4 low (1 and 2); DOCK4 high (3). Comparisons shown to be significant are also significant in analyses adjusting for the effect of systemic therapy plan, ER status and lymph node involvement. (a) and (b): Skeletal only; (c) and (d): skeletal and other; (e) and (f): Non-skeletal; (g) and (h): First skeletal irrespective of whether other distant events have occurred previously (ie bone metastasis-free survival). (i) and (j): Overall Survival (OS). P-values refer to the logrank test. For definitions of non-skeletal, skeletal and other and skeletal only see legend to Table 1.

338x602mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review